Construction of Image Retrieval Module for Cultural and Creative Products Based on DF-CNN

Meng Jiang
School of Art and Design, Henan University of Engineering, Zhengzhou, 451191, China

***Abstract*—With the growth of the cultural and creative product industry, more and more cultural and creative products have been designed and published in different channels. A method based on image retrieval module is proposed to address the search problem of Chinese creative products in online channels. During the process, a cascaded forest is proposed to achieve layer by layer processing, with class vectors as the main transfer content in the entire forest system. An image attribute feature extraction process that introduces extreme gradient enhancement is designed, and the aggregation of multi-scale and multi-region features is utilized to improve image retrieval performance. The experimental results showed that in the similarity test of extracting image feature information when the image contained three composite cultural and creative objects and the total pixel amount of the image reached 7M, the similarity of image feature information was 97.6%. In the analysis of running time, the research method only took 7.4ms to generate search results in seven fields. In the analysis of the proportion of false search content, the research method maintained a false search proportion of within 6.0% when searching for a single cultural and creative product object. This indicates that the research method has higher accuracy and efficiency in image retrieval of cultural and creative products. Research methods can provide certain technical support for the development of the cultural and creative industry.**

Keywords—Image retrieval; class vector; extreme gradient enhancement; Chinese creative products; layer by layer processing

I. Introduction

In today's rapidly developing era of digitization and informatization, the cultural and creative industry has become an important force driving economic growth and cultural inheritance. Cultural and creative products (CCPs), with their unique design and cultural connotations, not only satisfy people's pursuit of a better life but also become an important medium for spreading culture and shaping brand image [1]. However, with the increasing variety of CCPs, efficiently retrieving and identifying these products has become a problem that needs to be solved [2]. Image retrieval is an important retrieval method, and some scholars have conducted relevant research on image retrieval technology. Zhang et al. proposed a research method based on generative adversarial learning for image retrieval in text writing. The process enhanced retrieval efficiency by enhancing the semantics of generated images, and the resolution of generated images was determined by local discriminators. The results indicated that the designed method was efficient and accurate for image retrieval [3]. Humenberger et al. proposed a visual localization-based research method for image retrieval in autonomous driving. The method prioritized the judgment of image types by setting scene road conditions and then used fuzzy matching in dynamic scenes for image localization. The results indicated that the proposed method had good fitness for image retrieval in dynamic scenes [4]. Salih et al. proposed a research method based on double-layer feature judgment for multimedia image retrieval problems. The method filtered out images with deviant types by setting a coarse search layer, and then locked in the required images through subdivision class retrieval. The outcomes indicated that the designed method had high accuracy for image retrieval on multimedia [5]. Wang et al. proposed a research method based on spatial and exchange domains for texture image retrieval. The process used dual-tree complex wavelet transform to decompose the image and re-model it, achieving a complementary global retrieval structure. The outcomes indicated that the designed method improved the retrieval efficiency in the database [6]. Li et al. proposed a research algorithm based on a hash method for image retrieval on the Internet. The algorithm retrieved data patterns by collecting image data in a stationary environment, and simulating image transformations in non-stationary environments to adapt to its retrieval environment. The results indicated that the proposed method was efficient for image retrieval in large datasets [7].

Traditional image retrieval techniques, although achieving certain results in certain scenarios, often fail to meet the high demands of users for retrieval accuracy and efficiency. With the advancement of computer technology, deep learning technology has emerged and made revolutionary progress in the field of image recognition and retrieval [8]. Convolutional neural network (CNN) has become an important technology in the field of image processing due to its powerful feature extraction ability and ability to learn complex patterns. Some scholars have conducted relevant research on CNN. Xin's team proposed a CNN-based feature extraction method for the diagnosis of ocean turbine attachments. This method consists of three steps: data preprocessing, feature extraction, and fault diagnosis. The results indicated that the method could work smoothly in harsh environments [9]. Kumar et al. put forward a training algorithm using Dolphin-SCA to address the compression problem of CNN models applied in tumor classification. During the process, a fuzzy deformable fusion model was used for image segmentation. The experiment findings showed that the proposed method had good diagnostic accuracy [10]. Thirusangu et al. proposed a deep CNN based on the U-Net architecture to address the issue of low resolution in transcranial ultrasound images. During the process, filters were used for preprocessing and the effects of other architectures were compared. The results indicated that the U-Net architecture had better accuracy in identifying SN [11]. Ma proposed a method combining CNN to address the design issue of facial feature tracking systems. During the process, heterogeneous convolution was used to reduce the parameters of the convolution kernel. The search box mechanism was inserted into the network acceptance domain adjustment module, and the attention dispersal mechanism was used to standardize the arrangement of heterogeneous convolutions. The experiment outcomes indicated that the proposed method had good tracking accuracy and interpretability [12]. Ashtiani et al. proposed a method combining CNN technology to address the issue of medical image classification. During the process, the light waves incident on the pixel array were directly processed to obtain image classification results, and linear calculations were performed optically to reduce time consumption. The experiment outcomes indicated that the proposed method had good classification accuracy [13].

In summary, there have been many techniques using CNN for image retrieval and processing, but existing image retrieval techniques often struggle to accurately identify and match complex design elements and cultural connotations in CCPs, resulting in low accuracy of retrieval results. CCP images often contain rich color and texture information, and existing technologies may not be able to fully explore and utilize the useful information in the images when processing such high-dimensional data. Moreover, existing research has mostly focused on specific types of images, lacking the generalization ability for image retrieval in the special field of CCPs, which limits the universality of the technology. A single CNN model may have biases in extracting global and local features, and cannot fully express the complexity and diversity of CCPs. Moreover, a single CNN model usually requires a lot of computing resources. For large-scale image retrieval tasks, a single model may be difficult to meet the real-time requirements [14]. Deep forest (DF) is an ensemble learning method based on decision trees, which inherits the advantages of decision trees in processing high-dimensional data and improves the model's generalization ability and robustness by integrating multiple decision trees. The problems that the research attempts to solve and the gaps that will be achieved with other studies include: (1) developing an image retrieval technology that can accurately identify the design elements and cultural connotations of CCP images, to improve the relevance of retrieval results and make the research method more adaptable to CCP images than other methods. (2) Design an efficient image retrieval algorithm to meet the retrieval needs of large-scale image data and achieve rapid response. (3) Research on an image retrieval model with good generalization ability, which can adapt to CCP images from different sources and styles, and enable the research method to intelligently adapt to the constantly changing content of CCP images. In this context, the study attempts to innovatively combine DF and CNN, and constructs a cascaded forest structure. By combining feature vectors and similarity calculation, a new image retrieval technology for CCPs is designed. It is expected to provide certain technical references for the CCP industry.

The research is mainly conducted in four sections. Section II part is the design of image attribute feature extraction technology based on DF algorithm and CCP image retrieval method combined with CNN. Section III is to analyze the effectiveness of the research method through performance testing and application analysis. Section IV discusses and summarizes the entire text.

II. Methods and Materials

A. *Design of Image Attribute Feature Extraction Technology Based on DF Algorithm*

When designing, searching, and purchasing CCPs, a large amount of retrieval is required, among which image retrieval is a commonly used retrieval method [15]. However, CCPs often have rich designs and diverse styles, which may make it difficult to accurately match and recognize images during image retrieval. The DF algorithm can handle high-dimensional data and complex problems during computation, and can extract complex and diverse attribute features of images [16]. Research uses DF algorithm to design image attribute feature extraction technology for CCP image retrieval technology. In response to the complexity of information contained in CCPs, a cascaded forest is proposed to implement layer by layer processing of DF networks, and its structure is constructed as shown in Fig. 1.

In Fig. 1, each layer in the cascaded forest receives the feature information from the previous layer and the original input feature information, and each layer is an integrated structure of a decision tree forest. To obtain better diversity, a study is conducted to construct a composite forest structure using random forests and completely random forests. The construction of each decision tree begins with a randomly selected set of features. At each node of the tree, the system randomly selects a feature to segment the dataset. The ultimate goal is to grow each leaf node to be pure, meaning that the samples in the leaf nodes belong to the same category. The number of decision trees in the forest is a hyperparameter that can be changed according to different tasks of image retrieval for CCPs. In the entire forest system, class vectors are studied as the main transfer content, and the generation method is shown in Fig. 2.

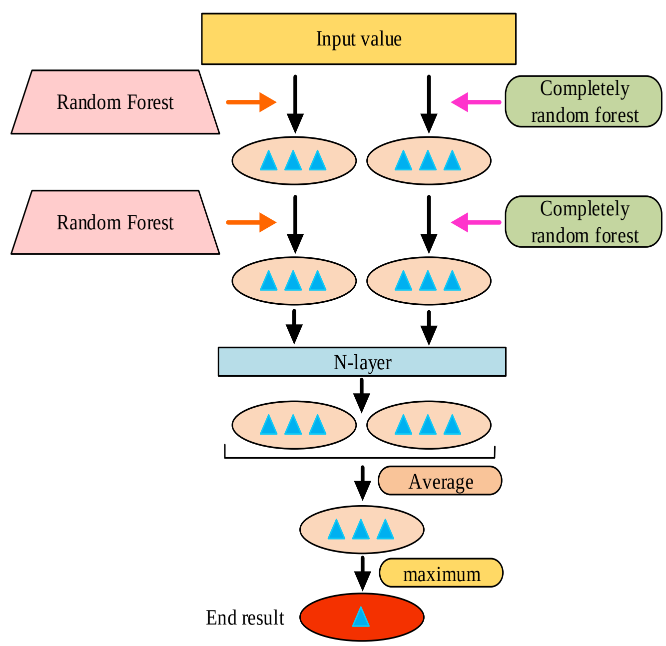


Fig. 1. Cascade forest structure.

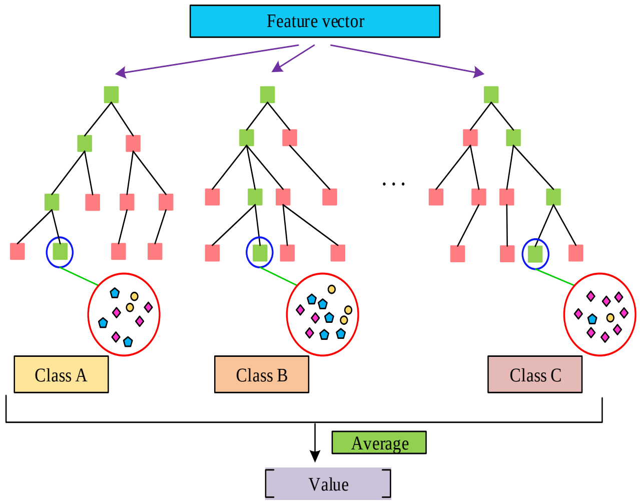


Fig. 2. Class vector generation method.

In Fig. 2, when generating class vectors, at each leaf node of each tree, the class vector is calculated based on the class distribution of the training instance, representing the relative proportion of each type in that node. For each tree in the forest, the separately calculated class vector will be averaged to form the final output of the tree. In addition to the initial layer, the forest at subsequent levels will fuse the class vectors generated by the previous layer with the original input features to form new inputs [17, 18]. Starting from the second layer of forest, regardless of how the number of trees in the forest changes, the dimension of the generated class vector will remain unchanged. The calculation method is to multiply the number of categories by the number of trees in the forest, and then add the dimension of the original features. If there is no significant improvement in performance on the validation set, it will stop the iteration and automatically select the optimal number of layers. When extracting image attribute features, involves the extraction of different attribute features. To improve the universality of the method, extreme gradient enhancement is introduced to optimize the DF algorithm. After a given training set, a tree ensemble model is generated, and then the expected value is obtained, as shown in Eq. (1).

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \tag{1}$$

In Eq. (1), $\hat{y}_i$ represents the expected value. F stands for regression tree. k represents the number of trees. x_i represents the content of the training set. f_k represents leaf nodes. The objective function is calculated from the expected value, as shown in Eq. (2).

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{2}$$

In Eq. (2), $L(\phi)$ represents the objective function. l represents a differentiable loss function. y_i represents compliance true value. Ω is a regularization function. The loss function is minimized after generating a new tree, as shown in Eq. (3).

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) + \Omega(f_t) \tag{3}$$

In Eq. (3), $L^{(t)}$ represents the loss function after generating a new tree. f_t stands for new tree. Taylor expansion is performed and different steps of the original function are calculated to simplify the loss function, as shown in Eq. (4).

$$\tilde{L}^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{4}$$

In Eq. (4), $\tilde{L}^{(t)}$ represents the simplified loss function. g_i represents the first-order degree of the original loss function. h_i represents the second-order degree of the original loss function. The optimal weight of the leaf node is calculated, as shown in Eq. (5).

$$w_j^* = -\frac{\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda} \tag{5}$$

In Eq. (5), w_j^* represents the optimal weight of the leaf node. λ represents a constant term that prevents the denominator from being 0. Then the quality of the evaluation tree is calculated, as shown in Eq. (6).

$$\tilde{L}^{i}(q) = -\frac{1}{2}\sum_{j=1}^{T} \frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T \tag{6}$$

In Eq. (6), $\tilde{L}^{i}(q)$ represents the quality value of the tree. T represents the number of leaf nodes in the tree. γ represents the leaf node coefficient. The process of extracting image attribute features by introducing extreme gradient enhancement is shown in Fig. 3.

In Fig. 3, when extracting image attribute features for CCPs, the original image data is first read and preprocessed before starting feature extraction. The preprocessed image data is input into the DF model for preliminary feature extraction. The initially extracted features are combined with the original image data to form an enhanced feature class vector, which is then fed into the next layer of DF model. After each feature extraction, the effectiveness of the features is evaluated through classification testing. After averaging and maximizing operations, the final feature extraction result of CCP image attributes is obtained.

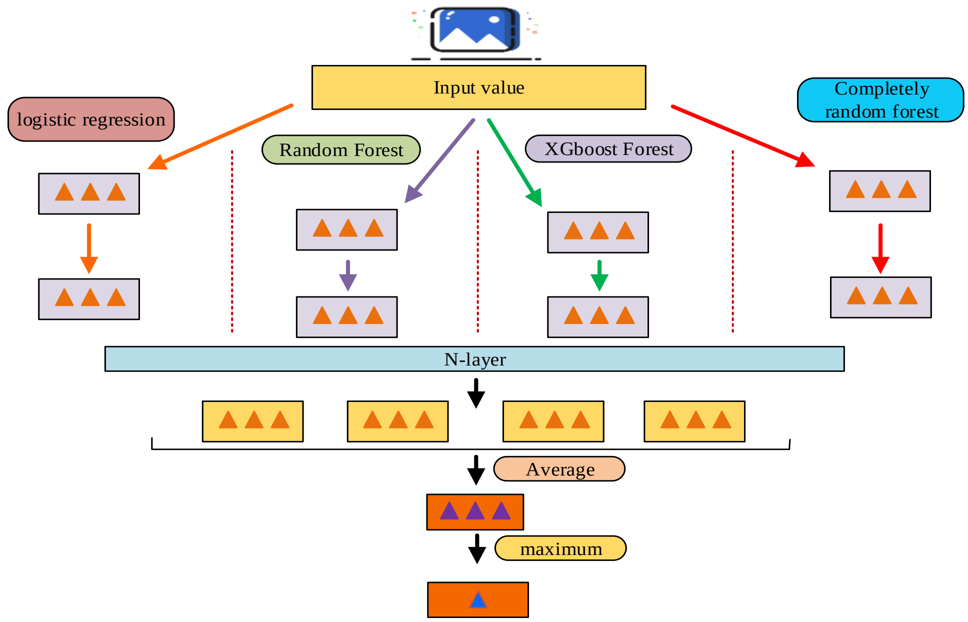


Fig. 3. Image attribute feature extraction process.

B. Image Retrieval Method for Cultural and Creative Products Combined with CNN

After extracting image attribute features, image attribute features can be used as the core data for image retrieval. CNN deep features are a data-driven high-level image feature with good image feature expression ability. They have certain advantages over traditional methods in extracting global, local features, and contextual information from images [19, 20]. Based on the DF algorithm, research designs the image retrieval methods for CCPs combined with CNN. The research uses the aggregation method of multi-scale and multi-region features in CNN and combines them with the spatial and channel weights of depth features to achieve multi-dimensional feature aggregation. When performing multi-region feature aggregation on CCP images, it needs to define several regions and calculate the aggregated feature vector for each region, as shown in Eq. (7).

$$f_{\Re} = \left[f_{\Re,1}, \dots, f_{\Re,\ell}, \dots, f_{\Re,\Im} \right]^{tra} \tag{7}$$

In Eq. (7), $f_{\Re}$ represents the region aggregation feature vector. tra stands for transpose matrix. $f_{\Re,\ell}$ represents the maximum value in the $\Re$ region on the ℓ th feature map. After obtaining the feature vectors of all regions, the normalized features are summed to form a multidimensional aggregated feature vector. When aggregating multi-dimensional features through spatial and channel weights of deep features, feature vectors are generated as shown in Eq. (8).

$$F_{crow} = \left[f_1, \dots, f_\ell, \dots, f_\Im \right], f_\ell = \sum_{y=1}^{H} \sum_{x=1}^{W} \alpha_{xy} \beta_\ell S_\ell (x, y) \tag{8}$$

In Eq. (8), F_{crow} represents the deep feature aggregation feature vector. α_{xy} represents spatial weight. β_ℓ represents channel weight. The feature aggregation process established in the study is shown in Fig. 4.

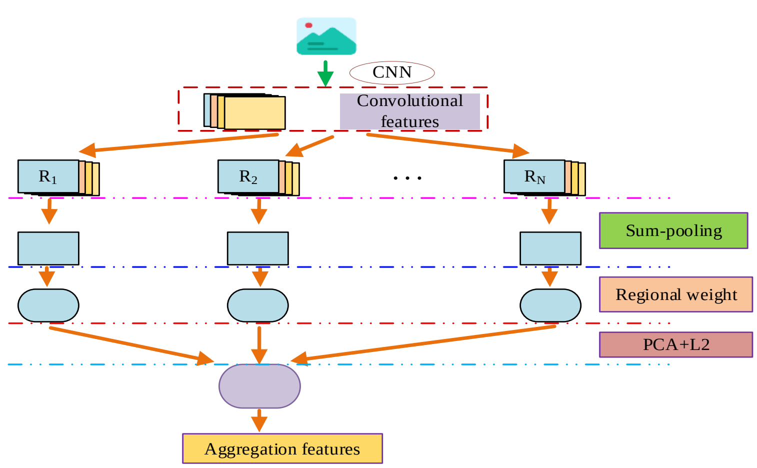


Fig. 4. Characteristic polymerization process.

In Fig. 4, when aggregating image features, the feature maps are first divided into sub regions, and the channel sensitivity weights are calculated based on the sparsity and intensity of the non-zero response values of the channels. After obtaining the significance weights of different regions, it calculates the feature vectors of different regions. The feature vectors of each region are normalized, one principal component analysis and whitening process are completed, and then a second normalization is performed. Sum-pooling is performed on the region features that have been normalized twice, and the aggregation results are normalized once to obtain the aggregated feature vectors. To more effectively extract effective information from feature maps, research is being conducted to incorporate channel sensitivity weights for optimization. The channel sensitivity is shown in Eq. (9).

$$\mathbb{N} = 1 - \gamma_\ell \tag{9}$$

In the above equations, $\mathbb{N}$ represents channel sensitivity. γ_ℓ represents the sum of the intensity amplitudes and positive response values of all non-zero response values on each channel. In the image of CCPs, the importance of different regions varies, and the corresponding weights also vary, as shown in Fig. 5.

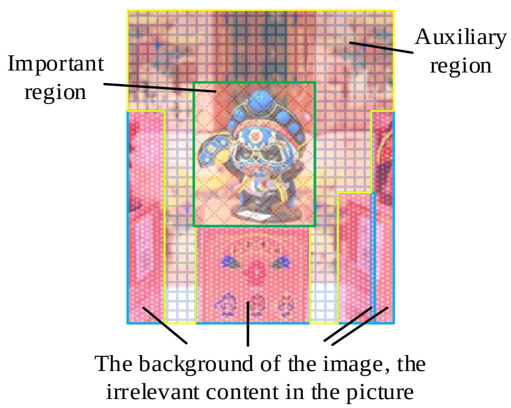


Fig. 5. Image area weight difference example.

As shown in Fig. 5, there are important areas in the image of CCPs that express the main body of the product, occupying the highest importance. In addition to important areas, there are auxiliary areas that play a secondary role in creating a visual atmosphere and enriching image details. The second most important factors are the background of the image and irrelevant content in the image. The significance weight of a region is defined as shown in Eq. (10).

$$\beta_r = \frac{\sqrt{\sum_{i\in R_r}\sum_{j\in R_r} S'(i,j)^2}}{\sum\sum_{ij} S' * A_r} \tag{10}$$

In Eq. (10), β_r represents the significance weight of the region. A_r represents the proportion of the size of the target area to the entire feature map. S' represents the sum of the values of all feature maps at each position. (i, j) represents the location of the feature map. The proportion of the target area to the entire feature map is calculated as shown in Eq. (11).

$$A_r = \frac{w_r \times h_r}{W \times H} \tag{11}$$

In Eq. (11), $W \times H$ represents the size of the feature map. $w_r \times h_r$ represents the size of the target area. Region feature vectors are generated as shown in Eq. (12).

$$\hat{f}_r^{\,\ell} = \left(\sum_{i\in R_r}\sum_{j\in R_r} \lambda_\ell S_\ell(i,j)\right)^{\alpha} \tag{12}$$

In Eq. (12), $\hat{f}_r^{\,\ell}$ represents the numerical value of the region feature vector on the ℓth feature map. λ_ℓ represents channel sensitivity weight. S_ℓ represents the sum of the values of the feature map at its position. The calculation of channel sensitivity weights is shown in Eq. (13).

$$\lambda_\ell = log\left(\frac{L\mathbb{C}+\sum_{\ell=1}^{L}\gamma_\ell}{\mathbb{C}+\gamma_\ell}\right) \tag{13}$$

In Eq. (13), $\mathbb{C}$ represents a small constant that ensures numerical stability. After extracting the image features of CCPs, the image retrieval module is constructed through feature semantic similarity measurement. In image retrieval, precise matching is often difficult to achieve due to the rich design elements and cultural connotations of CCPs. Fuzzy semantics provides a more flexible matching method, allowing the system to retrieve based on the similarity between image features and query conditions. The semantic similarity between two samples are calculated using fuzzy semantics as shown in Eq. (14).

$$s_{i,j} = \sum_{k=1}^{K}\left(\mu_{\xi_{P_i}}\left(P_j^k\right) + \mu_{\xi_{P_j}}\left(P_i^k\right)\right) \tag{14}$$

In Eq. (14), $s_{i,j}$ represents semantic similarity. $\mu_{\xi_{P_i}}$ represents the fuzzy semantics of templates. P_j^k represents the k nearest neighbor sample among the sample neighbors. The retrieval results are output based on semantic similarity. The operation process of the deep forest convolutional neural network (DF-CNN) CCP image retrieval module designed for research is shown in Fig. 6.

In Fig. 6, when conducting image retrieval for CCPs, after preprocessing the image and query information, the feature representation of the image is extracted. Image features are input into the feature space and an index with the image library is established. The similarity between query information and image library images is measured and matched to obtain corresponding retrieval images and complete sorting. At the same time, the results are fed back to the image library to enrich the search information of the image library. Finally, a list of content for image retrieval of CCPs is obtained, and the retrieval is completed.

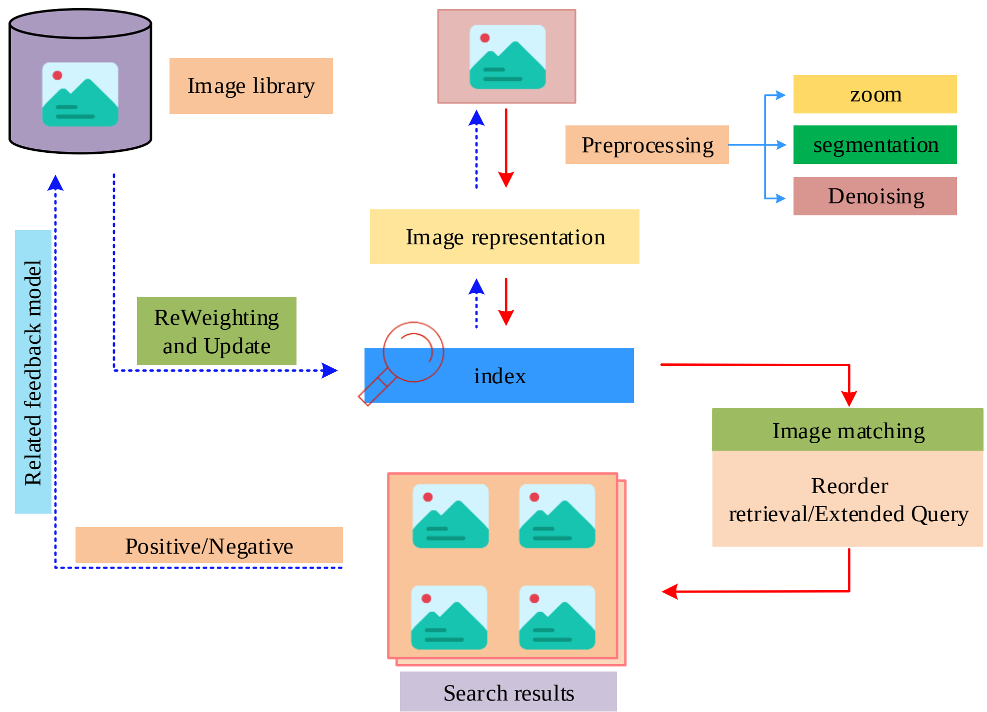


Fig. 6. Cultural and creative products image retrieval module operation process.

III. Results

A. Performance Testing of Image Retrieval Technology for Cultural and Creative Products

To analyze the performance of the image retrieval technology for CCPs designed for research at runtime, ImageNet and Open Images Dataset datasets were selected for testing. When conducting performance testing, the software and hardware environments are shown in Table I.

During testing, the core objects in the dataset images were equated with important areas in the CCP images, and images containing only a single core object and multiple core objects were divided. DF-CNN was compared with the currently advanced Inception Networks and Speeded Up Robust Features methods during testing. The convergence performance of different methods was compared, as shown in Fig. 7.

TABLE I. Experimental Software and Hardware Environmental Parameters

Software and hardware environment name	Parameter specification
Device type	Deep learning server
Processor	13th Gen Intel(R) Core(TM) i5-13490F 2.50 GHz
Graphics card	Nvidia GTX 4080
Internal memory	64GB
Hard disk space	1TB
Operating system	Ubuntu 14.04
Experimental platform	Caffe, Python 2.7.12, Matlab R2015b

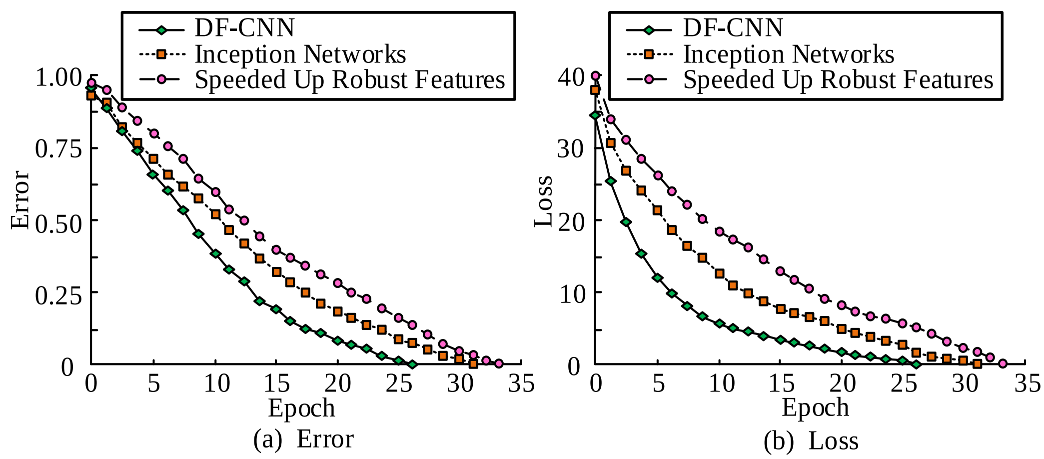


Fig. 7. Convergence performance test.

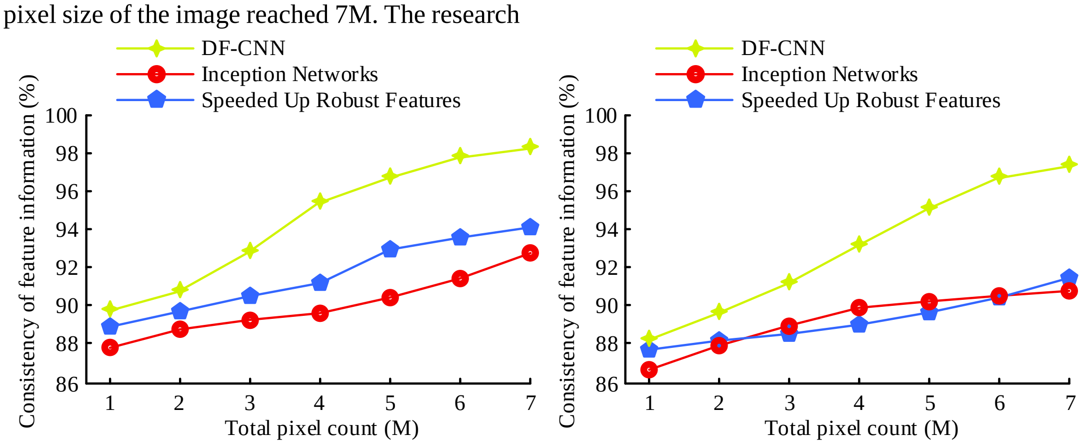
In Fig. 7, the overall convergence trend of performance for different methods during training was consistent. Fig. 7(a) shows that during the error convergence process, the error value of DF-CNN at the beginning of the iteration was lower than that of Speeded Up Robust Features and higher than that of Inception Networks. Within 5 iterations, the error value of DF-CNN rapidly decreased and was already lower than Inception Networks by the 5th iteration. In Fig. 7 (b), during the loss convergence process, the error value of DF-CNN at the beginning of the iteration was 34, which was lower than the Speeded Up Robust Features and Inception Networks. The loss value of DF-CNN decreased rapidly in the early stage and gradually slowed down in the later stage until the training was completed. The DF-CNN proposed in the study only required 26 iterations of training to achieve the optimal error and loss values, which were 7 and 10 fewer than Speeded Up Robust Features and Inception Networks, respectively. This indicates that the research method has better convergence performance and high training efficiency. The consistency of extracting image feature information was tested using different methods in images containing different numbers of pixels, as shown in Fig. 8.

From Fig. 8, the consistency of the results obtained by different methods in extracting image feature information increased with the total number of pixels in the image. Fig. 8 (a) shows that when the image was set to only contain a single cultural and creative object, the image feature information consistency of Speeded Up Robust Features was 94.1% when the total pixel amount of the image reached 7M. The similarity of image feature information in DF-CNN increased rapidly as the total pixel size of the image increased from 1M to 4M, and then gradually slowed down the rate of increase. During the entire process, the similarity of image feature information was 89.8% when the total pixel size of the image was 1M. When the total pixel count of the image reached 7M, it increased to 98.4%. From Fig. 8(b), when the image contained three composite cultural and creative objects, the image feature information consistency of Speeded Up Robust Features was 91.6% when the total pixel amount of the image reached 7M. The image feature information consistency of Inception Networks was 90.7% when the total pixel size of the image reached 7M. The image feature information consistency of DF-CNN was 97.6% when the total pixel size of the image reached 7M. The research method experienced a slight decrease in performance when extracting image features from multiple objects, but the magnitude of the decrease was smaller than other methods. This indicates that the research method has good performance in extracting image feature information.

B. Analysis of the Practical Application Effect of Image Retrieval Technology for Cultural and Creative Products

When analyzing the practical application of image retrieval technology for CCPs designed in research, the study extracted CCP images from a CCP design website as a retrieval image library. The running time of different methods was analyzed, and to improve the accuracy of statistical results, each scheme was repeated 10 times, presented in the form of mean and standard deviation. Seven different search fields were set and abbreviated as A, B, C, D, E, F, and G. The runtime was divided into two different stages: feature extraction and retrieval result generation, as shown in Fig. 9.

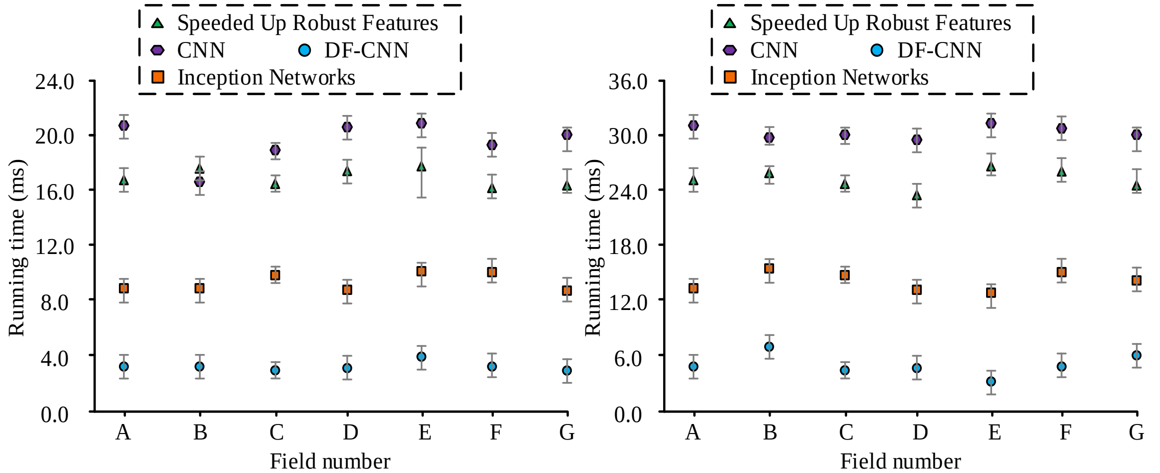
In Fig. 9, the runtime time of different methods was related to the fields and generally remained within a certain range. Fig. 9 (a) shows that during feature extraction, CNN had significantly higher runtime in six out of seven fields compared to other methods. The running time of Inception Networks was significantly lower than that of Speeded Up Robust Features, reaching its lowest point in the D field, between 7.0ms and 8.0ms. The running time difference of DF-CNN in the seven fields was relatively small, with a minimum of 2.3ms and a maximum of only 4.7ms. The fluctuation in time during multiple runs was also maintained within 2.0ms. As shown in Fig. 9 (b), different methods exhibited significant efficiency stratification when generating search results. Among the seven fields, the results were ranked according to the longest running time of CNN, the second longest running time of Speeded Up Robust Features, the third longest running time of Inception Networks, and the shortest running time of DF-CNN. The search result generation time of DF-CNN in seven fields was the highest at only 7.4ms, and the lowest at 1.6ms. The research method had higher operational efficiency at runtime and could complete image retrieval of CCPs at a faster speed. The proportion of false retrieval content in the retrieval results of different methods in practical applications was analyzed, as shown in Fig. 10.



(a) Single cultural and creative object

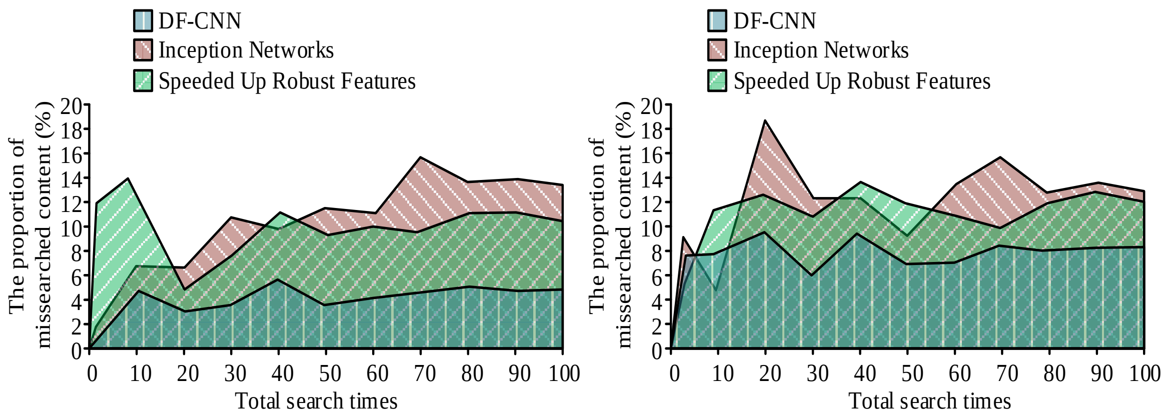
(b) 3 composite cultural and creative objects

Fig. 8. Extracting the coincidence degree of image feature information.



(a) Feature extraction (b) Search result generation

Fig. 9. Runtime analysis.



(a) Single cultural and creative object (b) 3 composite cultural and creative objects

Fig. 10. Analysis of the proportion of false retrieved content.

As shown in Fig. 10, different methods tended to stabilize the proportion of false retrieval content as the number of searches increased when generating retrieval content. Fig. 10 (a) shows that when searching for a single CCP object, Inception Networks had the highest proportion of false searches in 100 searches, reaching 15.8%. When the number of searches reached 80, it tended to stabilize and fluctuated slightly around 13.7%. The false search rate of Speeded Up Robust Features reached a maximum of 14.1% in 100 searches, and stabilized when the search frequency reached 50, mainly fluctuating in the range of 10.0% to 11.0%. The proportion of false searches in DF-CNN remained within 6.0% during 100 search cycles, and tended to stabilize at around 5.0% when the search cycle reached 50. From Fig. 10(b), when searching for three comprehensive CCPs, the proportion of false searches by Inception Networks in 100 search processes ultimately fluctuated around 13.4%. The proportion of false searches in Speeded Up Robust Features ultimately fluctuated around 12.3%. The proportion of false searches in DF-CNN tended to stabilize when the number of searches reached 50, fluctuating around 8.2%. The research method can better exclude irrelevant content when conducting image retrieval of CCPs. The overall content matching of the generated search result list was analyzed, and 20 searches on each key field were performed, as shown in Fig. 11.

In Fig. 11, in the analysis of content matching in search results, the matching degree of the results obtained from multiple searches could be integrated into a range of intervals. When the length of the search result list was 20, the content matching degree of Speeded Up Robust Features mainly fluctuated around 78%. The content matching degree of the retrieval results of Inception Networks mainly fluctuated around 85%. The content matching degree of DF-CNN search results mainly fluctuated around 92%. As the length of the search result list increased, there was a certain degree of decrease in the content matching of the search results for Speeded Up Robust Features and Inception Networks. However, the degree of decrease in the content matching degree of DF-CNN search results was minimal. When the length of the search result list reached 100, the content matching degree of the search results still fluctuated around 91%. The research method could maintain high operational performance and retrieval accuracy for a long time when conducting a large number of CCP image retrieval tasks.

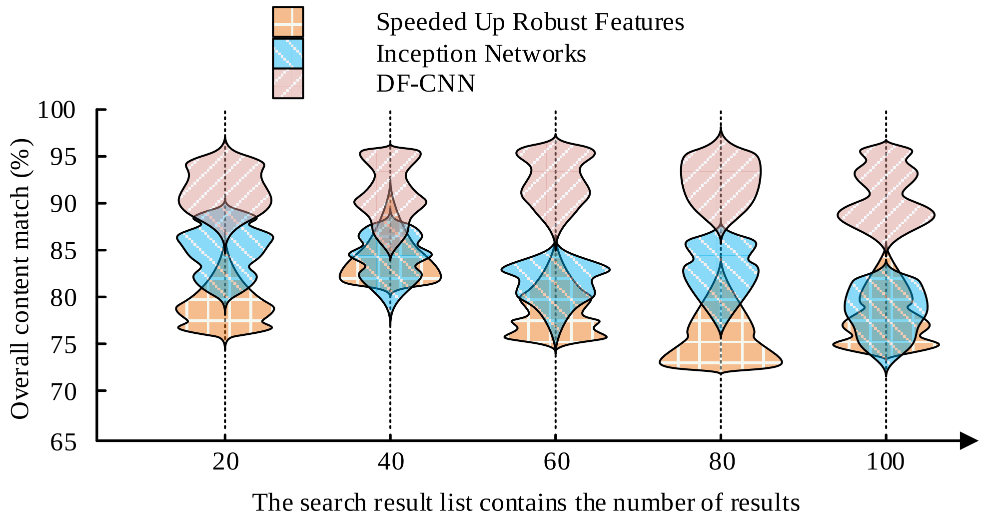


Fig. 11. Search results content matching degree analysis.

IV. Discussion and Conclusion

A CCP search assistance technology based on DF-CNN image retrieval module was studied and designed to enhance the exposure ability of CCPs. In the process, the number of decision trees in the forest was used as a hyperparameter to evaluate the effectiveness of features through classification testing. The aggregated feature vectors were calculated for each region, and region features which completed twice normalization were conducted sum-pooling aggregation. The significance weight of the region was defined. After extracting the image features of CCPs, the image retrieval module was constructed using feature semantic similarity measurement. The similarity measurement was performed on the query information and image library images to match the corresponding retrieval images. Finally, the effectiveness of the method was analyzed. The experimental results showed that in convergence performance testing, the research method only required 26 iterations of training to achieve the optimal error and loss values. When setting the image to only contain a single cultural and creative object for image feature information matching analysis, it increased to 98.4% when the total pixel size of the image reached 7M, indicating that DF-CNN could effectively capture the detailed features of the image when processing high-resolution images, providing richer information for image retrieval. When searching for three comprehensive CCPs, the proportion of false searches in the research method ultimately fluctuated around 8.2%, indicating that DF-CNN could effectively handle the complexity and diversity of features in multi-object images. The content matching degree of research method retrieval results has been fluctuating around 91% for a long time, which means that when using DF-CNN for image retrieval, users can obtain highly relevant results to the query, thereby improving the user experience of retrieval. The research method could generate higher quality retrieval results when conducting image retrieval of CCPs. With the development of technology, the application of 3D images and dynamic images was becoming increasingly widespread. Future work can consider extending DF-CNN to these new image types to meet a wider range of application requirements. However, in actual deployment and application, the scalability and robustness of the model also need to be considered. Future work can explore the combination of image retrieval technology and natural language processing to achieve text-based image retrieval, to improve the accuracy of retrieval and the naturalness of user interaction. At the same time, research can also be conducted on retrieval algorithms that are suitable for 3D models and video content, to meet the market's demand for dynamic and stereoscopic visual content retrieval. User privacy protection in the process of image retrieval should be strengthened and how to provide personalized retrieval services without leaking user data needs to be studied.

Declaration of Competing Interest

We declare that we have no conflict of interest.

References

[1] Zhao W X, Liu J, Ren R, Wen J R. Dense text retrieval based on pretrained language models: A survey. ACM Transactions on Information Systems, 2024, 42(4): 1-60.

[2] Rana M, Bhushan M. Machine learning and deep learning approach for medical image analysis: diagnosis to detection. Multimedia Tools and Applications, 2023, 82(17): 26731-26769.

[3] Zhang F, Xu M, Xu C. Tell, imagine, and search: End-to-end learning for composing text and image to image retrieval. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2022, 18(2): 1-23.

[4] Humenberger M, Cabon Y, Pion N, Weinzaepfel P, Lee D, Guérin N, et al. Investigating the role of image retrieval for visual localization: An exhaustive benchmark. International Journal of Computer Vision, 2022, 130(7): 1811-1836.

[5] Salih S F, Abdulla A A. An effective bi-layer content-based image retrieval technique. The Journal of Supercomputing, 2023, 79(2): 2308-2331.

[6] Wang H, Qu H, Xu J, Wang J, Wei Y, Zhang Z. Texture image retrieval based on fusion of local and global features. Multimedia Tools and Applications, 2022, 81(10): 14081-14104.

[7] Li Q, Tian X, Ng W W Y, Kwong S. Recent development of hashing-based image retrieval in non-stationary environments. International Journal of Machine Learning and Cybernetics, 2022, 13(12): 3867-3886.

[8] Khan S U, Hussain T, Ullah A, Baik S W. Deep-ReID: Deep features and autoencoder assisted image patching strategy for person re-identification in smart cities surveillance. Multimedia Tools and Applications, 2024, 83(5): 15079-15100.

[9] Xin B, Zheng Y, Wang T, Chen L, & Wang Y. A diagnosis method based on depthwise separable convolutional neural network for the attachment on the blade of marine current turbine. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2021, 235(10):1916-1926.

[10] Kumar S, Mankame D P. Optimization driven deep convolution neural network for brain tumor classification. Biocybernetics and Biomedical Engineering, 2020, 40(3): 1190-1204.

[11] Thirusangu N, Subramanian T, Almekkawy M. Segmentation of induced substantia nigra from transcranial ultrasound images using deep convolutional neural network. The Journal of the Acoustical Society of America, 2020, 148(4):2636-2637.

[12] Ma Y, Song Q, Hu M, Zhu X. Correction: A Lightweight Neural Learning Algorithm for Real-Time Facial Feature Tracking System via Split-Attention and Heterogeneous Convolution. Neural Processing Letters, 2023, 55(2): 1581-1581.

[13] Ashtiani F, Geers A J, Aflatouni F. An on-chip photonic deep neural network for image classification. Nature, 2022, 606(7914): 501-506.

[14] Naeem A, Anees T, Ahmed K T, Naqvi R A, Ahmad S, Whangbo T. Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval. Complex & Intelligent Systems, 2023, 9(2): 1729-1751.

[15] Rajwar K, Deep K, Das S. An exhaustive review of the metaheuristic algorithms for search and optimization: Taxonomy, applications, and open challenges. Artificial Intelligence Review, 2023, 56(11): 13187-13257.

[16] Gheisari M, Hamidpour H, Liu Y, Saedi P, Raza A, Jalili A, Rokhsati H, Amin R. Data Mining Techniques for Web Mining: A Survey. Artificial Intelligence and Applications, 2023, 1(1): 3-10.

[17] Bhosale Y H, Patnaik K S. Application of deep learning techniques in diagnosis of covid-19 (coronavirus): a systematic review. Neural processing letters, 2023, 55(3): 3551-3603.

[18] Godwin E C, Izuchukwu C, Mewomo O T. Image restorations using a modified relaxed inertial technique for generalized split feasibility problems. Mathematical Methods in the Applied Sciences, 2023, 46(5): 5521-5544.

[19] Mohammed A, Kora R. A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences, 2023, 35(2): 757-774.

[20] Wu C, Khishe M, Mohammadi M, Taher Karim S H, Rashid T A. RETRACTED ARTICLE: Evolving deep convolutional neutral network by hybrid sine–cosine and extreme learning machine for real-time COVID19 diagnosis from X-ray images. Soft Computing, 2023, 27(6): 3307-3326.