Illustration visual communication based on computer vision image retrieval algorithm

H.Z. Zhang [1]
[1] *School of Art and Design, Minnan Science and Technology College, 362000, China, Kangmei Town, Quanzhou, Kangyuan Road, No. 8*

Abstract

In illustration design, good visual communication can make the audience resonate. Computer vision image retrieval algorithm provides important support and assistance for the visual communication of illustration. However, the traditional image retrieval algorithm has problems of subjectivity and inaccuracy in complex image classification. Therefore, this paper optimizes the feature extraction module of convolutional neural network and fuses hash algorithm to improve the efficiency and speed of image retrieval. The experimental results show that the accuracy of the improved convolutional neural network is 82.7 %, which is more than 6 percentage points higher than the traditional algorithm model. The recall rate of the volume neural network model improved by hashing algorithm is 94.1 %. Research is of great significance to the visual communication of illustration design, which helps designers to find relevant materials more accurately, improve the artistic quality and ornamental value of their works, and promote the innovation and development of illustration design.

Keywords: visual communication; image retrieval; convolutional neural network; Hash algorithm
Citation: Zhang HZ. Illustration visual communication based on computer vision image retrieval algorithm. Computer Optics 2025; 49 (1): 132-140. DOI: 10.18287/2412-6179-CO-1449.

Introduction

In modern society, the rapid development of computer vision technology, which uses computers to process, analyze and understand images or videos, has brought new opportunities and challenges to the field of image retrieval [1]. In this context, illustration visual communication based on computer vision image retrieval algorithm has become a field of concern and research [2]. As an art form with strong visual impact and personalized expression, illustration has a wide range of application prospects and important artistic value [3]. Visual communication, as an art form that conveys information, emotion and meaning to the audience through visual elements, is of great significance to the artistic expression of illustration. Through image retrieval algorithms, designers can find materials related to the style or theme of illustration, so as to give specific expression and emotion to visual works and enhance their artistic value and appreciation [4]. However, traditional image retrieval algorithms have some limitations in the face of complex image content. Among them, the main problem is that the label cannot accurately describe the image content, resulting in subjectivity in the classification process, which may lead to inaccurate results. In addition, the search and screening process of image materials takes a lot of time and effort, and designers may not be able to quickly find materials that match the design theme or style, which affects the quality and efficiency of the design work. Therefore, this paper proposes a Convolutional Neural network (CNN) based on attention mechanisms to optimize the extraction of illustrated image features, and further introduces the hash algorithm as the encoding and indexing method of illustrated image features to improve the efficiency and speed of image retrieval. The research aims to contribute to the development of the field of illustration visual communication, promote the integration of illustration design and computer vision technology, and open up new creative possibilities.

This study is divided into five parts. The first part is the introduction, explaining the research background and briefly introducing the research content; The second part summarizes the current research results of image retrieval algorithm model; The third part introduces the image retrieval model which combines deep learning and perceptual hash algorithm. In the fourth part, experiments are carried out on the basis of the methods in the third part and the results are analyzed to verify the performance of the proposed model. The fifth part summarizes the whole article.

1. Related work

With the continuous development and improvement of computer image retrieval technology and algorithm, the accuracy, efficiency and practicability of image retrieval have been improved, and provide strong support for various application scenarios. Image retrieval technology has many branches, including content-based image retrieval technology, image feature extraction technology, large-scale image retrieval and so on. Therefore, many scholars have conducted research on these technologies. Hussain et al. proposed a content-based image retrieval model based on deep convolutional neural network to overcome the difficulty of extracting image retrieval features due to the rapid increase in multimedia content distribution, and directly extracted depth features from images, thereby effectively improving the efficiency of image retrieval technology [5]. Kumar et al. proposed an image feature extraction method based on improved deep convolutional neural networks to solve the problem of low efficiency in large-scale image retrieval. Principal component analysis was introduced to reduce the dimensionality of extracted features, thus improving the performance of image retrieval technology, including accuracy and robustness [6]. Taheri et al., aiming at the challenge faced by image retrieval technology to understand the semantic meaning of image content and accurately find the image related to user query, proposed a method that integrates deep neural network and manual features, and extracts semantic and advanced features from multiple levels in combination with pyramid method, thus enhancing the semantic understanding ability of image retrieval [7].

In addition, aiming at the "semantic gap" between low-level visual features and high-level semantic features in image retrieval, Wu proposed an image retrieval method that combines deep learning semantic feature extraction and regularization Softmax. Experimental results show that this method can effectively extract semantic features with high retrieval accuracy. The classification accuracy of STL-10 image data set reached 60.3 % [8]. Naee et al. proposed a new deep learning technology to address the challenge of obtaining the highest accuracy in image retrieval, which combines the most advanced content-based image retrieval methods by implementing convolutional neural networks with autocorrelation, gradient computation, scaling, filtering and positioning functions. Experimental results on texture data sets of 17 kinds of flowers and tropical fruits show that this technique effectively improves the accuracy of image retrieval [9]. Keisham and Neelima proposed an image retrieval system based on deep search and rescue algorithm to address the challenge of image retrieval caused by the increasing number of multimedia contents on the Internet. By extracting image features and using adaptive sunflower optimization algorithm to cluster fusion features, image retrieval was completed. Thus, the accuracy and recall rate of image retrieval technology are effectively improved [10].

To sum up, image retrieval technology is constantly developing and improving. Researchers have improved the accuracy and efficiency of image retrieval by proposing new methods and technologies, including feature extraction based on deep learning and enhancement of semantic understanding ability. However, the problem of labels not accurately describing visual content in complex illustrated images has not been solved. Therefore, aiming at illustration visual communication, this paper proposes an image retrieval technology that uses attention mechanism to improve GNN and hash algorithm to improve GNN. The innovation of this research is that through the improvement of GNN, image retrieval technology can be combined with illustration design to improve the visual communication effect of illustration.

2. Illustration visual communication based on computer vision image retrieval algorithm

This section is divided into two parts. First, the attention mechanism is used to improve convolutional neural networks, and a feature extraction method for illustration images is proposed. Secondly, hash algorithm is used to improve the attention mechanism, and a feature representation method for database is proposed. Through the combination of the two to complete the illustration image retrieval work.

2.1. Optimization of feature extraction for illustration images based on CNN fusion attention mechanism

The quality of visual communication is of great significance to the development of illustration. With the continuous increase of illustration materials and the diversification of design needs, artists need to use a large number of illustrated images to inspire. Therefore, it is necessary to use computer vision image retrieval algorithm [11]. The primary task of targeted illustration image retrieval is to extract the feature content of illustration. However, the existing image feature extraction methods have the problem of low extraction accuracy [12]. Therefore, this paper proposes an improved convolutional neural network method using attention mechanism to optimize the feature extraction process of illustration images, so as to improve the accuracy and efficiency of feature extraction.

The data of illustration image is taken as input, which includes pixel value and channel information of image, and the output is the feature representation obtained after network processing. The convolution model consists of multiple convolution layers, pooling layers and fully connected layers. In the convolution layer, the model extracts features in the image through a series of convolution operations, which can be edges, colors, textures, etc. [13]. The convolution operation slides over the image through a filter and computes the convolution operation to generate a feature map. The pooling layer is used to reduce the spatial dimension of the feature map, reduce the number of parameters, and enhance the robustness of the model to changes such as translation and rotation. The fully connected layer maps the feature vectors to the space of the output class. The traditional CNN model mainly focuses on local information when extracting features, and does not fully consider the semantic relationship between different regions in the image. This may lead to less accurate features extracted when processing complex scenes or images containing multiple objects, affecting the accuracy and efficiency of image retrieval [14]. Therefore, the study introduces channel attention mechanism and spatial attention mechanism to make up for the deficiency. First, the convolutional neural network is described, and its feature extraction can be described by Formula (1), which is essentially to find the optimal weight value.

$$\theta^* = \arg\max_{\theta} L(f(x,\theta), y) . \tag{1}$$

In Formula (1), L is the loss function of the CNN; and Euclidean distance loss is used in the study. θ contains all parameters in the parameter space of the text CNN model; θ^* is the group with the best weight value among all parameters; $f(x,\theta)$ is the model; x is the input data information; y is the true label of each sample. After calculating the gradient of ownership weight, the weight value will be updated, and the basic update weight in the algorithm is Formula (2).

$$\theta_{iter} = \theta_{iter-1} - \eta g_{iter} . \quad (2)$$

In Formula (2), the number of iterations during the training process is represented by *iter*; θ_{iter} is all parameters iterated to the *iter* th parameter space; deep learning rate is represented by η; g_{iter} is a gradient. This is the most basic update algorithm that can be used for model training, but it has slow convergence speed and long training time. So this study uses the Adam optimization algorithm, which combines the RMSprop method and Momentum method to improve training efficiency, adjust adaptive learning rate, and retain the previous gradient direction information after each iteration update. The process of updating the weight values for each iteration in the Adam optimization algorithm is represented by Formulas (3) and (4).

$$\upsilon_{iter} = \beta_1 \upsilon_{iter-1} + (1-\beta_1) g_{iter} . \quad (3)$$

In Formula (3), the update direction at the *iter* iteration is represented by υ_{iter}, and β_1 is the attenuation coefficient.

$$\omega_{iter} = \beta_2 \omega_{iter-1} + (1-\beta_2) g_{iter} \bullet g_{iter} . \quad (4)$$

In Formula (4), the momentum at the *iter* iteration is represented by ω_{iter}, and β_2 is the attenuation coefficient. The control of learning rate decay is represented by Formulas (5), (6), and (7).

$$\hat{\upsilon}_{iter} = \upsilon_{iter} / (1-\beta_1^{iter}) . \quad (5)$$

In Formula (5), the update direction at the *iter* iteration is represented by υ_{iter}, and β_1^{iter} is the attenuation coefficient.

$$\hat{\omega}_{iter} = \frac{\omega_{iter}}{1-\beta_2^{iter}} . \quad (6)$$

In Formula (6), the momentum at the *iter* iteration is represented by ω_{iter}, and β_2^{iter} is the attenuation coefficient.

$$\theta_{iter} = \theta_{iter-1} - \eta \frac{1}{\sqrt{\hat{\omega}_{iter}} + \varepsilon} \bullet \hat{\upsilon}_{iter} . \quad (7)$$

In Formula (7), ε is a constant. To avoid the denominator value being equal to 0, the value of ε is usually taken as 10^{-8}. In the pre improvement structure, the output data activation operation steps are described by the Relu function using Formula (8).

$$Relu(x) = \max(0, x) . \quad (8)$$

In Formula (8), x is the input value. When $x \geq 0$, the Relu function directly outputs it; On the contrary, the output value is 0. But if the input values into the Relu function are not greater than or equal to 0, that is, if the output values are all 0, it will reduce its learning performance. So this study uses the Celu activation function, described by Formula (9).

$$Celu(x) = \max(0, x) + \min(0, \alpha \times (\exp(x/\alpha) - 1) . \quad (9)$$

In Formula (9), the Celu function outputs $x \geq 0$ normally, and outputs $\alpha \times (\exp(x/\alpha)) - 1$ and $\alpha = 1$ when x is less than 0. This study introduced attention mechanism into the model, and Figure 1 shows the feature capture module structure that combines channel and spatial attention mechanisms. The channel attention mechanism captures effective feature information in an image by learning the relevant information between channels in the network structure. The channel attention mechanism combined with spatial attention mechanism can capture specific target objects in the image, and the most important feature information of the corresponding feature layer can be captured through maximum pooling operation.

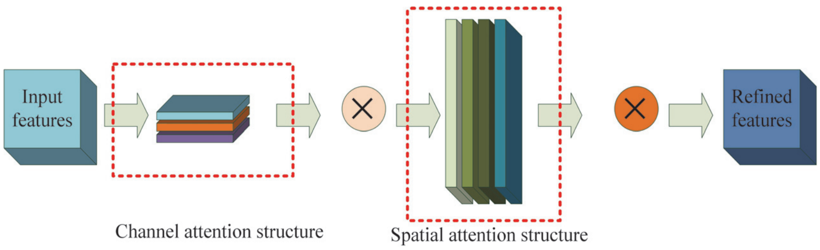


Fig. 1. Channel and spatial attention mechanisms

The combination of channel attention mechanism and spatial attention mechanism has multiple advantages for the convolutional neural network-based illustration feature extraction method [15]. First, the channel attention mechanism can learn the relevant information between channels in the network structure, which helps to capture the effective feature information in the image. At the same time, by combining with the spatial attention mechanism, the channel attention mechanism can further focus on specific target objects, and capture the most important feature information in the corresponding feature layer through the maximum pooling operation, so as to improve the representation ability and discrimination of features. Secondly, the global maximum pooling operation converts the two-dimensional data of the feature layer into a one-dimensional vector, preserving the most important feature data of each channel, which helps to reduce the impact of redundant information. Finally, combining the advantages of channel attention mechanism and spatial attention mechanism, the key information of the image can be captured more accurately in the process of feature extraction of illustration, thus improving the ability of feature expression and the accuracy of image retrieval. The process of the maximum pooling operation is described by Formula (10).

$$\alpha_i^{(0)} = \max_v (\hat{s}_i^l(v)) . \tag{10}$$

In Formula (10), $\alpha_i^{(0)}$ is the vector, and $\hat{s}_i^l(v)$ is the vth feature data of the ith channel in the lth layer. After obtaining the vector, it is input into the fully connected network, and the Celu function activates it, as expressed by Formula (11).

$$\alpha_j^{(1)} = Celu(\sum_i \alpha_i^{(0)} w_{ij}^{(a_1)} + b_j^{a_1}) . \tag{11}$$

In Formula (11), $\alpha_j^{(1)}$ is the hidden layer vector, and $w_{ij}^{(a_1)}$ and $b_j^{a_1}$ are the weight parameters for the update gate operation. After obtaining the hidden layer, it will transmit the results to the output layer. In neural networks, both output and input are vectors with channel quantities equal to length. The activation function in the output layer is sigmoid, and the output result will be a probability value in the (0,1) interval, which is described in Formula (12).

$$\alpha_i^{(2)} = sigmoid(\sum_j \alpha_j^{(1)} w_{ji}^{(a_2)} + b_i^{a_2}) . \tag{12}$$

In Formula (12), $w_{ij}^{(a_2)}$ and $b_j^{a_2}$ are the weight parameters for the update gate operation. Finally, in order to suppress the worthless text data of some channels, it needs to retain important channel text data, which is described by Formula (13).

$$s_i^l = \alpha_i^{(2)} \hat{s}_i^l . \tag{13}$$

In Formula (13), $\hat{s}_i^l$ is the characteristic data of the ith channel in the ith layer.

2.2. Design of image retrieval model based on CNN fusion hash algorithm

Traditional image retrieval methods have some challenges in feature extraction and matching, especially in processing high-dimensional feature vectors, which will consume a lot of storage space and matching time [16]. In order to further improve the efficiency of image retrieval, a convolutional neural network fusion hash algorithm image retrieval model is proposed, which maps high-dimensional feature vectors to low-dimensional binary hash code space to achieve data reduction and efficient index construction, so as to improve the efficiency of feature matching between the illustrated image features in the database and the features of the illustrated image to be retrieved. A hash algorithm is an algorithm that maps input data of arbitrary length to output data of fixed length. Its main purpose is to realize the unique identification and fast retrieval of data. Hash learning includes steps such as feature extraction, feature mapping and feature quantization. Hash learning includes steps such as feature extraction, feature mapping, and feature quantization. Figure 2 shows the hash feature transcoding process. It can be seen that first, feature extraction is performed on the input image to obtain the feature image [17]. Next, the feature image is transformed into an approximate hash feature code through feature mapping. Finally, approximate hash feature codes are transformed into hash codes through feature quantization. This image search method based on hash algorithm can provide efficient search and matching capabilities when processing large-scale image data.

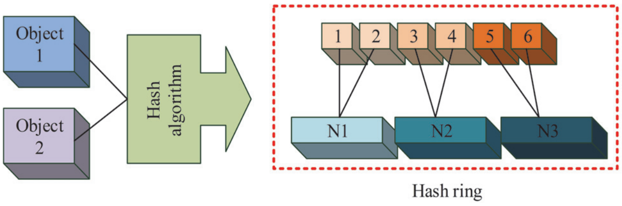


Fig. 2. Hash algorithm transcoding process

Formula (14) describes common hash algorithm models.

$$H = h(m) . \tag{14}$$

In Formula (14), H is the generated hash code, h is the hash function, and m is the input data. After feature extraction, it is necessary to calculate the similarity between images. Similarity measurement is a method of measuring the degree of similarity between features, used to determine whether the visual features of an image are similar. When calculating feature distance, using different similarity measurement algorithms can lead to differences in the presentation and computational efficiency of image similarity. This study conducts similarity judgment by calculating the Hamming distance between feature vectors. The Hamming distance is used to calculate the number of differences in the corresponding positions of two hash codes of the same length. Formula (15) describes the process of calculating the Hamming distance.

$$d(A,B) = \sum_i (A_i \oplus B_i) . \tag{15}$$

In Formula (15), $d(A, B)$ means the Hamming distance between the strings A and B; $\Sigma_i (A_i \oplus B_i)$ represents the sum of all positions; A_i and B_i represent the characters of the string at position i; $\oplus$ represents the bitwise XOR operator. Smaller Hamming distance means more similar two strings; Larger Hamming distance means greater difference between two strings. For example, if two strings are identical, the Hamming distance is 0; If two strings are completely different, the Hamming distance is equal to the length of the string [18].

Figure 3 shows the framework structure of the model. As can be seen from Fig. 3, the model is mainly divided into three parts, namely, the processing of images to be retrieved, the processing of databases, and the retrieval part. Firstly, the illustration images that need to be retrieved are processed, and the feature extraction method based on improved convolutional neural network based on attention mechanism is used to extract the features in the images. Secondly, for the image in the database, the feature vector of the image is extracted, and the feature representation is mapped to K-bit hash code through the hash layer, while the error calculation is carried out in the loss layer to maintain the semantic similarity of the hash code. Finally, in the retrieval stage, the trained model is used to obtain the hash code of the queried images, and the similarity is measured with the hash code of the images in the dataset one by one, and the image results similar to the queried images are finally obtained.

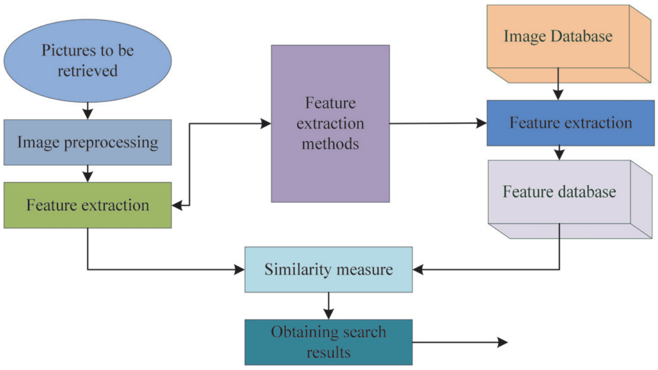


Fig. 3. Structure of improved image retrieval model

Fig. 4 shows the image feature extraction process of the CNN model, and it can be seen that this study added a hash layer between the original fully connected layer and softmax layer of CNN. Hash layer can map the output of the fully connected layer to hash encoding, in order to achieve feature compression and dimensionality reduction. The model contains five convolutional modules, with the first two modules having two convolutional layers and the last three modules having three convolutional layers. Each module contains a maximum pooling layer. The convolutional layers in each module have the same number of convolutional kernels, with 64, 128, 256, 512, and 512 convolutional kernels for each module.

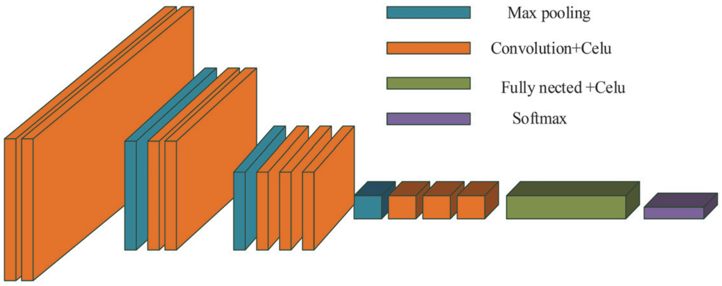


Fig. 4. Image feature extraction process of improved CNN model

3. Experimental results and analysis

In order to study the effectiveness and superiority of the proposed content, experiments and discussions are carried out respectively on the optimization method of illustration image feature extraction based on convolutional neural network fusion attention mechanism and the image retrieval model based on convolutional neural network fusion hash algorithm. The former is mainly to verify the performance of image feature extraction, while the latter is mainly to verify the performance of image retrieval. The operating system version of the experimental platform is Ubuntu 20.04.2, and the GPU is GeForce GTX 1080Ti, which is designed and implemented by Pytorch framework. The initial learning rate was set at 1e-3 for 100 iteration cycles, and the learning rate became one-tenth of the original after every 30 iteration cycles. The optimizer uses the Adam optimizer.

3.1. Experimental results and analysis on performance detection of improved CNNs

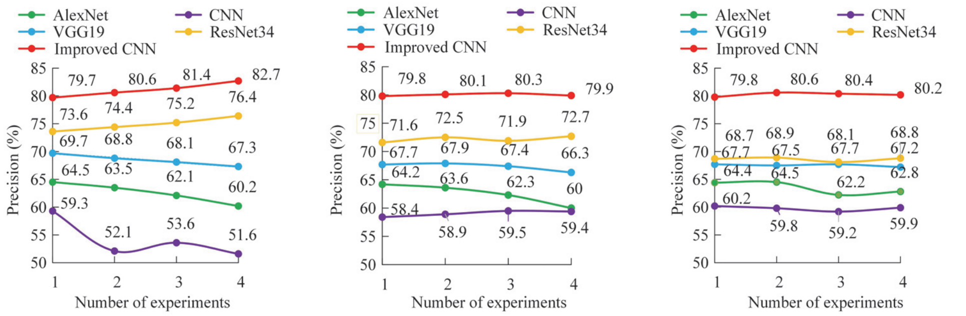
The experiment compared the retrieval performance of different types of CNNs, namely deep CNN AlexNet, Visual Geometry Group (VGG) 19, Residual Network (ResNet) 34, CNN, and improved CNN, on three sets of illustration image sets. The experiment used three sets of 5257 images as samples, with different categories of illustration images as a group, as shown in Tab. 1.

To avoid chance, four image retrieval operations were performed for each of the three groups of samples. Figure 5 shows the change in accuracy of each model for four retrieval tasks on each set of images. In the first set of image tests, the performance of all five models was unstable. The accuracy of CNN and its improved model showed a slight increase of about 3 %, with the highest reaching 76.4 % and 82.7 %, respectively. The accuracy of AlexNet, Res-

Net34, and VGG19 decreased by 2 % to 8 %, especially for VGG19, which plummeted to 52.1 % in the second round of testing. For the second and third sets of images, except for AlexNet, the accuracy of other models has decreased slightly, while the accuracy of other models waas generally stable, with a variation of around 1 %. In the third set of image tests, the optimized CNN accuracy was 80.6 %, 11.7 percentage points higher than the CNN model.

Tab. 1. Various image sets

First set of images	Finance and economics	Physical culture	Entertainment	Society	Historic photos
Number of images	342	452	483	445	229
Second set of images	Iconic image	Documentary photos	Magazine cover	Special images	Book illustrations
Number of images	346	375	423	236	348
Third set of images	Oil painting images	Thangka image	Wash painting	Sketch	Comic
Number of images	472	197	207	341	361



(a) Comparison of precision between different models in the first set of images

(b) Comparison of precision between different models in the second set of images

(c) Comparison of precision between different models in the third set of images

Fig. 5. Accuracy changes of retrieval tasks for each model on three sets of image sets

Tab. 2 provides a comparison of the average accuracy of the five models tested on each image set. It can be seen that regardless of which image set, the improved CNN-based model, ResNet50, and CNN had an average recognition accuracy that was 13 to 27 percentage points higher than the AlexNet and VGG19 models. In the classification model based on residual connection structure, the improved CNN model had the highest average accuracy overall, with the highest accuracy tested in the first image set, which was 81.1 %.

Tab. 2. Comparison of average accuracy of five models tested on each set of images

Algorithm \ Image set	Group 1	Group 2	Group 3
AlexNet	62.56 %	62.52 %	63.48 %
CNN	54.15 %	59.05 %	59.78 %
VGG19	68.48 %	67.33 %	67.52 %
ResNet34	74.90 %	72.18 %	68.63 %
Improved CNN	81.10 %	80.03 %	80.25 %

In this study, the ResNet50 model was improved by replacing the Relu function used by traditional CNN with the Celu function. In order to test whether the improvement can improve the model recognition accuracy, based on the improved CNN model, the retrieval accuracy of three sets of illustration images under the Relu, PRelu, LeakyRelu, Tanh, and Celu activation functions was compared. Similarly, the detection was repeated four times on each set of illustration images. Fig 6 shows the trend of recognition accuracy changes of the improved CNN model for three sets of image sets using different activation functions. It can be seen that regardless of which group of images were recognized and classified, the recognition accuracy under the Celu function and LeakyRelu function environments was 15 to 20 percentage points higher than the other three activation functions. The Celu function had a higher accuracy in detection than the LeakyRelu function, with the former reaching a maximum of 80.2 %.

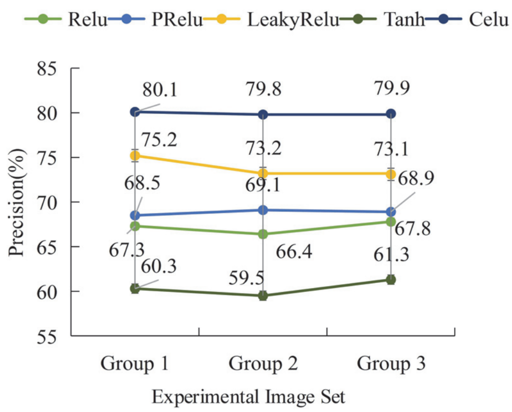


Fig. 6. Comparison of image retrieval accuracy of models under different activation functions

3.2. Experimental results and analysis of performance testing of image retrieval models

In response to the improvement of the improved CNN illustration image retrieval model combined with the hash algorithm, the performance of the image retrieval model was compared and tested on the illustration image set given in Tab. 1. The system evaluation method mainly used the Mean of Average Precision (MAP) of image retrieval to measure the performance. Fig. 7 shows a comparison of the accuracy changes of illustration images for four algorithm models. It can be seen that the Generative Adversarial Network (GAN), Siamese Network (SN), and CNN without the fusion of hash algorithm were used as control algorithms. The accuracy of CNN with the fusion of hash algorithm was higher than 80 % on all three sets of illustration image sets, and fluctuated slightly around 85 %, which was 9 to 25 percentage points higher than the three comparison algorithms.

By adjusting the top_k parameter during the query, the weight of the query results can be affected. When more items need to be queried, the value of top_k will increase, which means the number of results in the query will correspondingly increase, thereby improving the recall rate of the system. The larger the query top_k value, the more relevant information about image features the model will obtain. A comparative experiment was conducted on the accuracy of similar image recognition by repeatedly testing different image retrieval models in a painting image set containing six sets of similar styles. Tab. 3 provides the illustration image set data.

Fig. 8 shows a comparison of the recall changes of four image detection models. As the cumulative amount of retrieved images increased, the recall of all four image retrieval models showed an upward trend. Compared to the three comparison algorithm models, the CNN model improved by combining the hash algorithm had the highest initial recall rate of 85.3 %, which was 15 to 25 percentage points higher than the three comparison algorithms. When the cumulative number of retrieved images reached 1000, the improved CNN algorithm model had the highest recall rate of 94.1 %, which was 16 to 28 percentage points higher than the three heap ratio algorithm models.

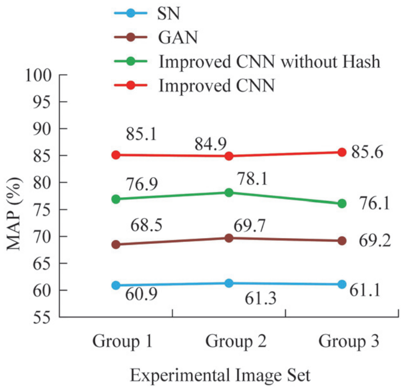


Fig. 7. Comparison of precision changes

Tab. 3. Painting Image Set Data

Image group \ Painting style	Baroque	Rococo	Gothic	Line drawing	Realistic writing	Cartoon
1	108	97	13	22	19	15
2	17	19	156	21	19	15
3	17	18	23	172	14	16
4	11	9	12	99	89	36
5	19	20	88	44	18	62
6	45	55	9	56	65	14

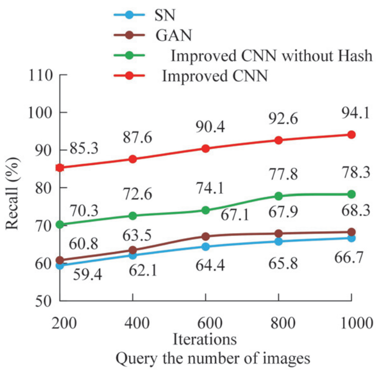


Fig. 8. Comparison of recall changes among four image detection models

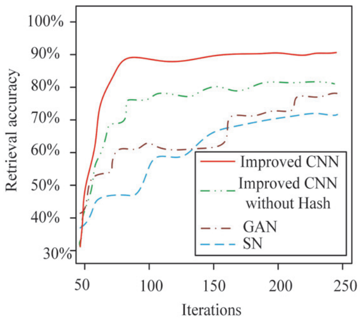
Fig. 9 shows the comparison of convergence speed and response time of four algorithms. It can be seen that in the early stage of iteration, the improved CNN algorithm model quickly found the optimal solution. At this time, the retrieval accuracy was above 90 %, and as the iteration progressed, its image retrieval accuracy gradually stabilized at 93 %. Although the CNN model without the fusion of hash algorithm also found the optimal solution when the number of iterations was less than 100, its optimal image retrieval accuracy was more than 10 percentage points lower than the improved CNN model. The response time of the CNN model that integrates hash algorithm was also the smallest among the four algorithm models, that is, the improved CNN had the fastest execution speed and could be applied to tasks with high real-time requirements or large-scale data processing tasks.

Conclusion

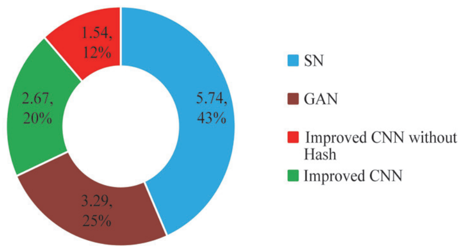
With the advancement of multimedia, mobile networks, and cloud technology, large-scale image search has become a necessary technology in our daily lives. Illustration, as an art with a strong visual effect and unique style, can be used to search for illustrations of similar styles and used in design through image search. However, due to the rapid increase in multimedia content, traditional image search methods have limitations. For example, labels may not fully describe images, which may result in inaccurate search results. In response to this issue, this study optimized the feature extraction of CNNs and combined it with hash algorithms, aiming to improve the effectiveness and speed of image search. The experimental results showed that the improved CNN had the highest accuracy of 82.7 % in image set testing, which was more than 6 percentage points higher than the traditional CNN algorithm model; The CNN model recognition accuracy under the Celu function and LeakyRelu function environments was 15 to 20 percentage points higher than the other three activation functions. The Celu function could achieve higher accuracy in detection than the LeakyRelu function, with the former reaching a maximum of 80.2 %. The CNN model improved by combining the hash algorithm had the highest initial recall rate of 85.3 %, which was 15 to 25 percentage points higher than the three comparison algorithms. When the cumulative number of retrieved images reached 1000, the CNN algorithm model fused with hash algorithm had the highest recall rate of 94.1 %, which was 16 to 28 percentage points higher than the three heap ratio algorithm models. In the early stage of the iteration, the CNN algorithm model that integrates the hash algorithm quickly found the optimal solution. At this time, the retrieval accuracy was above 90 %, and as the iteration progressed, its image retrieval accuracy gradually stabilized at 93 %. The response time of the CNN model that integrates the hash algorithm was also the smallest among the four algorithm models, indicating that the improved CNN had the fastest execution speed. The retrieval accuracy of the illustration image retrieval model designed in this study had significantly improved, but this study only considered differences in image styles and did not consider differences in the quality of each image. This is also a direction for future research.

Acknowledgements

The research was supported by the Fujian Provincial Department of Education’s 2017 Young and Middle-aged Teacher Education Research Project (Social Sciences) “Research on the Combination of Chinese and Western Architectural Decoration in Modern Overseas Chinese Residential Buildings in Quanzhou” (Project No.: JAS170827).



(a) Comparison of convergence states of different algorithms



(b) Comparison of Response Times of Different Algorithms (s)

Fig. 9. Comparison of performance of different algorithms

References

[1] Ünver M, Olgun M, Türkarslan E. Cosine and cotangent similarity measures based on Choquet integral for Spherical fuzzy sets and applications to pattern recognition. J Comput Cogn Eng 2022; 1(1): 21-31. DOI: 10.47852/bonviewJCCE2022010105.

[2] Wang Z, Liu X, Lin J, et al. Multi-attention based cross-domain beauty product image retrieval. Sci China Inf Sci 2020; 63(2): 120112. DOI: 10.1007/s11432-019-2721-0.

[3] Revathi M, Jeya IJS, Deepa SN. Deep learning-based soft computing model for image classification application. Soft Comput 2020; 24(24): 18411-18430. DOI: 10.1007/s00500-020-05048-7.

[4] Xu Y, Wei M, Kamruzzaman MM. Inter/intra-category discriminative features for aerial image classification: A quality-aware selection model. Future Gener Comput Syst 2021; 119(4): 77-83. DOI: 10.1016/j.future.2020.11.015.

[5] Hussain S, Zia MA, Arshad W. Additive deep feature optimization for semantic image retrieval. Expert Syst Appl 2021; 170(1): 114545. DOI: 10.1016/j.eswa.2020.114545.

[6] Kumar S, Singh MK, Mishra M. Efficient deep feature based semantic image retrieval. Neural Process Lett 2023, 55(3): 2225-2248. DOI: 10.1007/s11063-022-11079-y.

[7] Taheri F, Rahbar K, Beheshtifard Z. Content-based image retrieval using handcraft feature fusion in semantic pyramid. Int J Multimed Inf Retr 2023, 12(2): 21. DOI: 10.1007/s13735-023-00292-7.

[8] Wu Q. Image retrieval method based on deep learning semantic feature extraction and regularization softmax. Multimed Tools Appl 2020, 79(13): 9419-9433. DOI: 10.1007/s11042-019-7605-5.

[9] Naee A, Anees T, Ahmed KT, Naqvi RA, Ahmad S, Whangbo T. Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval. Complex Intell Syst 2023, 9(2): 1729-1751. DOI: 10.1007/s40747-022-00866-8.

[10] Keisham N, Neelima A. Efficient content-based image retrieval using deep search and rescue algorithm. Soft Comput 2022, 26(4): 1597-1616. DOI: 10.1007/s00500-021-06660-x.

[11] Wei B, Hao K, Gao L, et al. A biologically inspired visual integrated model for image classification. Neurocomputing 2020; 405(3): 103-113. DOI: 10.1016/j.neucom.2020.04.081.

[12] Huang ZX, Li J, Hua Z. Underwater image enhancement via LBP-based attention residual network. IET Image Process 2022; 16(1): 158-175. DOI: 10.1049/ipr2.12341.
[13] Tang J, Zhang J, Yin J. Temporal consistency two-stream CNN for human motion prediction. Neurocomputing 2022; 468: 245-256. DOI: 10.1016/j.neucom.2021.10.011.
[14] Zhu XM, Chen CX, Zheng B, et al. Automatic recognition of lactating sow postures by refined two-stream RGB-D faster R-CNN. Biosyst Eng 2020; 189: 116-132. DOI: 10.1016/j.biosystemseng.2019.11.013.
[15] Bai T, Guan H, Wang S, et al. Traditional Chinese medicine entity relation extraction based on CNN with segment attention. Neural Comput & Applicat 2022; 34(4): 2739-2748. DOI: 10.1007/s00521-021-05897-9.
[16] Behera A, Wharton Z, Keidel A, et al. Deep CNN, body pose and body-object interaction features for drivers' activity monitoring. IEEE Trans Intell Transp Syst 2020; 23(3): 2874-2881. DOI: 10.1109/TITS.2020.3027240.
[17] Wang H, He J, Zhang X, et al. A short text classification method based on *N*-Gram and CNN. Chin J Electron 2020; 29(2): 248-254. DOI: 10.1049/cje.2020.01.001.
[18] Jiang X, Wang N, Xin J, et al. Image super-resolution via multi-view information fusion networks. Neurocomputing 2020; 402: 29-37. DOI: 10.1016/j.neucom.2020.03.073.

Author's information

Huize Zhang (b. 1984) graduated from Kunming University of Science and Technology in 2012 with School of Art and media, Employed at Minnan Science and Technology College, Current lecturer title. Research interests: graphic design of computer software, art painting. E-mail: *zhzzhz12345@163.com*

Received November 11, 2023. The final version – May 08, 2024.