

A Novel Image Retrieval Method with Improved DCNN and Hash

Yan Zhou, Lili Pan*, Rongyu Chen and Weizhi Shao

Central South University of Forestry and Technology, Changsha, 410000, China

*Corresponding Author: Lili Pan. Email: lily_pan@163.com

Received: 25 May 2020; Accepted: 10 July 2020

Abstract: In large-scale image retrieval, deep features extracted by Convolutional Neural Network (CNN) can effectively express more image information than those extracted by traditional manual methods. However, the deep feature dimensions obtained by Deep Convolutional Neural Network (DCNN) are too high and redundant, which leads to low retrieval efficiency. We propose a novel image retrieval method, which combines deep features selection with improved DCNN and hash transform based on high-dimension features reduction to gain low-dimension deep features and realizes efficient image retrieval. Firstly, the improved network is based on the existing deep model to build a more profound and broader network by adding multiple groups of different branches. Therefore, it is named DFS-Net (Deep Feature Selection Network). The adaptive learning deep features of the Network can effectively alleviate the influence of over-fitting and improve the feature expression of image content. Secondly, the information gain rate method is used to filter the extracted deep features to reduce the feature dimension and ensure the information loss is small. The last step of the method, hash Transform, sparsifies and binarizes this representation to reduce the computation and storage pressure while maintaining the retrieval accuracy. Finally, the scheme is based on the distinguished ResNet50, InceptionV3, and MobileNetV2 models, and studied and evaluated deeply on the CIFAR10 and Caltech256 datasets. The experimental results show that the novel method can train the deep features with stronger recognition ability on limited training samples, and improve the accuracy and efficiency of image retrieval effectively.

Keywords: Deep feature; feature dimensionality reduction; feature selection

1 Introduction

With the rise of artificial intelligence, pattern recognition technology is widely applied in social life. In particular, computer image retrieval technology which is favored by people, because it is intuitive, rapid, and accurate characteristics [1]. In recent years, content-based image retrieval (CBIR) [2–3] has developed quickly and achieved a mass of research works.

The retrieval is usually the process of comparing the identical or similar features of objects to obtains multiple images of the same or similar kind, which involves the extraction method of image features and the selection of features. The computer is susceptible to the effects of the environment when retrieving objects, such as the change of illumination, proportion, angle of view, etc. Meanwhile, there are significant intra-class variations and small inter-class differences in the same category of images, which becomes a problem that must be solved in image retrieval. If the underlying features (color, texture, and shape) are directly selected for description, the difference between the classes is minimal, and the retrieval is inefficient with these features. As thus, for image retrieval of similar objects, invariant local features with excellent anti-interference performance are usually selected, such as SIFT (Scale-Invariant Feature Transform) [4]. But in recent years, these features expressed by the image content will not be able to reflect



human perception high-level semantic concepts, the feature expression model dominated by DCNN (Deep Convolutional Neural Network) as the mainstream can significantly improve the retrieval accuracy when applied to the same category of image retrieval [5–7].

Although the features extracted by DCNN can better reflect the image semantics, its high-dimensional features and redundancy lead to poor retrieval performance. It is necessary to train the network model before feature extraction. To better reflect the high-level content of the extracted image features, this paper proposes a method to improve the structure of DCNN and train the network models to improve the accuracy of the models. Even so, there are still high-dimensional features and high redundancy issues. Therefore, in the experiment, The Information Gain Rate is adopted to filter features according to the correlation between the features, and high-dimensional features reorder from high to low according to the influence degree of the expression image content. Finally, the hash dimension reduction of these features not only improves the efficiency of image retrieval but also reduces the amount of computation and storage space.

The rest of this article is arranged as follows: the second part discusses the relevant work. The third part introduces the DFS module of the deep convolutional neural network and explains how its design principle. The fourth part is the experiment and analysis. Finally, a comprehensive assessment of the proposed novel method is performed by using an improved network and hash in Chapter 5.

2 Related Work

This section describes the current state of research on DCNN and image retrieval.

2.1 Deep Convolutional Neural Network

Since 2003, the features extracted by the SIFT description operator [8] based on the local description have superior performance in the image retrieval direction and translation, so the method has been widely adopted and studied. Krizhevsky et al. proposed an improved DCNN model for image recognition tasks and used ReLU as the activation function of DCNN to solve the problem of gradient dispersion of Sigmoid in the deep network [9]. In the meantime, to avoid the fuzzy effect of average pooling, the whole network uses max-pooling and a smaller pooling core to instead of average pooling, so the output of the average-pooling layer produces overlap and coverage, which enriches the diversity of features. In 2014, Oxford University improved VGG based on AlexNet. The characteristic of the network is to repeatedly use convolution kernels of the same size to extract more complex and expressive features. Besides, the number of network layers is deepened to improve the feature expression effect of image content.

Deepening the structure of the network is beneficial to the extraction of image features, but merely increasing the number of layers of the system will lead to learning stagnation and higher training errors. The residual module in the Residual Network (ResNet) solves such problems well, effectively trains deeper network, and extracts features from different layers more abundantly [10]. Also, the deeper the network features are, the more abstract they are, the more semantic information they have, and the better the retrieval effect is. In GoogLeNet [11] and its improved network, horizontal convolution arrangement design is adopted, grouping and parallel convolution are selected, and multiple convolutions of different sizes are used to extract information of various clusters in the image. In this way, convolution and pooling of different scales are integrated. A module on one layer can obtain information on multiple levels, and then multi-scale features can be fused to form new features. In the next module, new features can also be extracted from different scales for multidimensional feature fusion.

2.2 Image Retrieval

In the early years, it found that image content can be described by transforming colors, shapes, textures [12], or structures as a single global feature representation. For example, Dong proposed an image retrieval method based on the texture features of the target region according to the Otsu algorithm and the Local Binary Pattern (LBP) algorithm [13]. As time goes by, it is found that the traditional local descriptors and their deformation both have a common problem, that is, these features lack learning ability, which limits

the expression ability of their image content and makes it difficult to adapt to a variety of datasets. Sharma [14] proposed a new class of data-independent locality-sensitive hashing (LSH) algorithms based on the fruit fly olfactory circuit. In this method, the original data of images read directly, and a new hash algorithm is used to provide better candidate ranking in a high dimensional space and to quickly find candidate neighbors for ranking in a low dimensional area. However, the image data read by this method is the low-level image pixels recognized by the computer, which fails to extract the deep features of images. Although the calculation speed is fast, it is not suitable for large pixel images. To express the high-level semantic concept of human perception through the extracted features, Sun et al. [15] applied deep learning to image retrieval. In this paper, the author proposed to identify and retrieve the Chinese medicine pictures by using the convolutional neural network and added a ternary loss value to retrieve similar images. Nevertheless, the VGG19 model employed in this scheme in the excessive use of Fully-Connected (FC) layers, which contained a large number of parameters that made the training speed slow, easy to over-fit, and occupied a large number of memory resources. Ren [16] proposed that a deep convolutional neural network was used to extract the deep features of images. PCA algorithm is adopted to reduce the feature dimension and minimize the information loss, which is an excellent solution to the problem of high feature latitude and ample storage space extracted from the network.

In this paper, a modification scheme of network architecture is proposed, that is, adding the DFS module to the network, extracting features of different scales and merge them into new features after the same convolution, to improve the performance of DCNN descriptor in image retrieval task. Through training, the image features extracted by the network and then reordering them according to the degree of relevance with the information gain rate, and the redundancy of image features can reduce through appropriate dimensionality reduction. The objective is to minimize the Euclidean distance between each image representation and its nearest representation. These signals are easy to compress, dimensionality reduction, and even faster retrieval. We aim to generate low-dimensional image representation, improve retrieval accuracy, and reduce memory.

3 Algorithm Implementation

To obtain features with more complementary features and effectively improve the accuracy and efficiency of retrieval, we propose a new network, DFS-Net, to promote the expression ability of the network and reduce the difficulty of extraction, and then hash the extracted features to reduce storage memory and speed up retrieval. For one example, see Fig. 1 below.

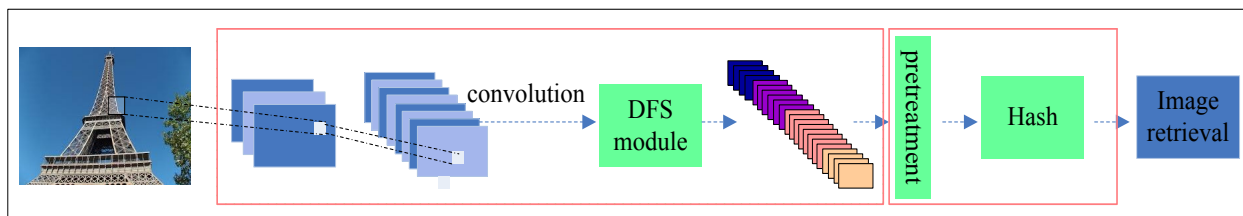


Figure 1: The DFS-Net for image retrieval

The key to DFS-net is to use a multi-scale feature selection module. By learning more and more differentiated features to have the network adaptive learning deep features and effectively improve the retrieval accuracy. To further enhance the retrieval efficiency and reduce the storage requirements, we hash the deep features of the image. However, the direct hashing will reduce retrieval accuracy, so we choose the feature selection algorithm (information gain rate) based on feature relevance to reorder and reduce the dimension of features. On the premise of low information loss, the spatial size is reduced from the perspective of lowering feature redundancy. To ensure the effectiveness of DFS-Net and lessen the cost resources, the network design will follow the following principles:

Principle 1: The network framework must be generic, enabling it to be implemented on classic models (ResNet50, InceptionV3, etc.) and widespread public datasets.

Principle 2: Taking full account of the hardware's carrying capacity, such as memory and video memory overhead.

3.1 DFS-Module

Most of the deep layers of the classical convolutional neural network connect to the full connection layer or the single convolutional pooling layer, which to some extent, will cause the loss of image feature information and is not conducive to the extraction of image features, thus leading to weak image retrieval effect. Therefore, this paper adopts the DFS (deep feature selection module) to convolute simultaneously on multiple branches and scales, which can extract features of different sizes is shown in Fig. 2. Rich features make classification and judgment more accurate.

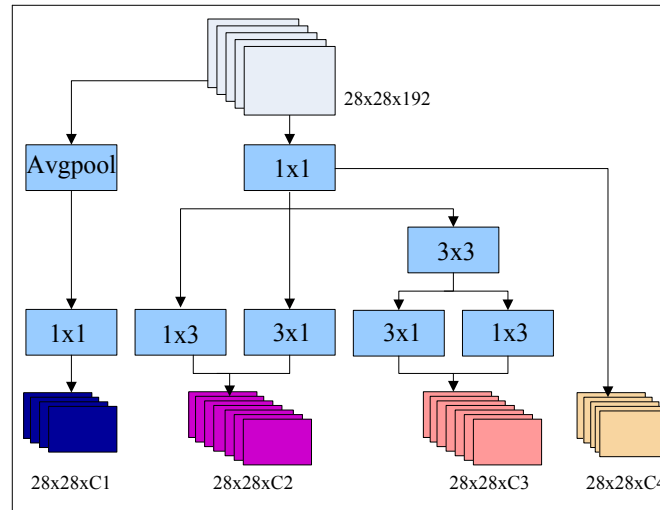


Figure 2: The deep feature selection module

The network structure since VGG16 mainly includes the convolutional computing layer and the max-pooling layer. However, the large-scale convolution kernels require a large number of parameters to be trained in deep convolution. For example, there are more than 1.8 million parameters that need to prepare if $28 \times 28 \times 192$ features convolve with 5×5 convolution with 384 filters, and if 3×3 convolution takes the place of 5×5 convolution is used, require the training of the parameter are halved, even so, the number of parameters to be trained is still quite large. In order to reduce these complex data effectively, a 1×1 convolution operation is first used to compress the output features of the network before entering the DFS module, to reduce the computational complexity of subsequent convolution operation while eliminating redundancy (principle 2). Max-pooling can be used to filter useless information in the shallow network, but it will lose too many high-dimensional details in a deep network. When a convolution neural network, after repeatedly convolution and pooling of the input image, some unique information in the image cannot be collected or extracted, and the information difference is weak. Therefore, average pooling is selected in the module to replace the max-pooling, to prevent high-level information loss caused by deep network pooling and ensure the integrity of information.

By using the principle that the sparse matrix is decomposition into dense matrix calculation, the convergence speed is accelerated, and the decomposition of the feature dimension is realized. When the DFS module extracts features on multiple scales, the output features are no longer uniformly distributed, but highly correlated features cluster together, that is, multiple sub-feature sets of dense distribution. This feature set is a set of features with a strong correlation, which are gathered together to weaken the irrelevant non-key features and reduce the redundant information of the output features. The natural convergence speed is faster when taking the feature set with high classification accuracy as the input of reverse calculation.

3.2 Hash Function

The retrieval time is shortened after the high dimensional features are hashed down directly, but the accuracy is not ideal. Therefore, this work first uses the information gain rate to rearrange and reduce the dimension of features. The information gain rate uses the ratio of the node information gain to the node split information metric; that is, the Gain Rate of each attribute related to the classification evaluation. Therefore, the information gain rate can not only make the features of the partition more different but also make up for the defects of the information gain algorithm (The information gain always tends to choose features with more attribute values).

The relevant formula of Information Gain Rate is as follows:

$$H(D) = - \sum_{n=1}^N \frac{|C_n|}{|D|} \log_2 \frac{|C_n|}{|D|} \quad (1)$$

$$g(D, A) = H(D) - H(D|A) \quad (2)$$

$$g_R(D, A) = \alpha \cdot g(D, A) \quad (3)$$

$$\alpha = \frac{1}{H_A(D)} \quad (4)$$

$$H_A(D) = - \sum_{i=1}^m \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|} \quad (5)$$

In (1), there is the sample set D , assuming that the sample set can be divided into N categories, and the probability of each class is $\frac{|C_n|}{|D|}$, where $|C_n|$ represents the number of samples of type N , and $|D|$ represents the total number of sample sets. In (2), $g(D, A)$, it means using feature A to partition the information gain of the data set. $g_R(D, A)$ represents the information gain rate, which is the product of the penalty parameter α and the information gain. Formulas 4 and 5 are the specific expressions of parameter α .

According to the information gain rate, the feature set is reordered. Then, the ordered features are converted into binary feature information for similarity recognition. The hash feature descriptor can be regarded as a compression expression based on visual content, which has the advantages of small computation and high matching speed. The algorithm flow of hash is shown in Algorithm 1.

Algorithm1: Features of the hash

Inputs: $F = \{F_1, F_2, \dots, F_M\}$

Outputs: The hash form after feature optimization.

1: Hash pretreatment:

2: $F_i = \text{InformationGainRatio}(F)$

3: Feature set reordering

4: Determine the dimension of image features, expressed as $C \in R^{(1 \times b)}$

5: Calculate the mean value of the specified dimension θ .

6: Compare the feature values according to formula (7), $C(i, j) = \begin{cases} 1, & \theta > C(i, j) \\ 0, & \theta \leq C(i, j) \end{cases}$

7: The above result is spliced to form a b -bit binary sequence, expressed as $H \in R^{b \times 1}$.

For example: First, the image features are extracted by the DFS-Net and represented by F ($M = 2048$ or $M = 1280$). Secondly, feature optimization is processed in two steps: First, according to the information gain rate, the deep features are reordered and reduced, then low-dimensionality features are hashed, and the feature values of each dimension are compared with the average value less than or equal to 0 and greater than or equal to 1. Finally, the hamming distance is utilized for image retrieval.

4 Experimental Results and Analysis

4.1 Experimental Setup

4.1.1 Datasets

Dataset 1: CIFAR-10 consists of 10 categories of objects, each of which consists of 6,000 32×32 color images, with 60,000 images in total. It shows in Fig. 3.



Figure 3: CIFAR-10 example

Dataset 2: Caltech256 obtains the data set after removing the cluttered data and divides the images into 256 categories, each of which has more than 80 images. For example, electrical appliances, plants, animals, etc., the number of images of each type is not equal is shown in Fig. 4.



Figure 4: Caltech256 example

4.1.2 Comparative Experiment (Based on the Same Dataset)

Scheme 1: Training the network model (ResNet50, InceptionV3, and MobileNetV2) to extract the deep features of the image and retrieve them directly by using Euclidean distance.

Scheme 2: Training the DFS-Net model, extracting the deep features of the images, and using the Euclidean distance for direct retrieval.

Scheme 3: The DFS-Net model is trained to extract the deep features of the image, processing Information Gain Rate and the hash of the features, and using a hamming distance to retrieve the image.

4.2 Evaluation Indicators

In this paper, Mean Average Precision (MAP) is used as the evaluation index. AP (Average Precision) is an index to measure the accuracy of each kind of image retrieval. MAP, which is used to measure the accuracy of image retrieval of all types, is the average value of AP. The more related images are retrieved, the larger the MAP value may be. If the relevant image can't be found, the MAP defaults to 0.

4.3 Experimental Evaluation and Analysis

In this experiment, the Softmax classifier on the last layer is updated, and all layers of the network are fine-tuned to verify the effectiveness of DFS-Net in a short time. The performance results of the DFS-Nets are shown in Tab. 1 and Tab. 3, indicating the accuracy of each network model (Ori stands for the original network.). Tab. 2 and Tab. 4 show the experimental results based on the MAP evaluation criteria. The methods proposed in this paper are implemented in three different networks, and three comparative methods are proposed. Based on DFS-Net-I and DFS-Net-R, the extraction dimensions are 16, 32, 64, 128, 256, 512, 1024, 2048. The extraction dimensions based on DFS-Net-M is 16, 32, 64, 128, 256, 512, 1024, 1280. Also, the feature dimensions of each method can achieve the best retrieval accuracy. In these three networks, the retrieval accuracy is the highest when the dimension is about half of the original, which proves that our method has the best retrieval accuracy and efficiency.

Table 1: Improved model accuracy of network training with CIFAR-10

Model 1	Accuracy (%)	vs. Ori (%)
DFS-Net-R	87.10	+0.20
DFS-Net-I	88.80	+1.50
DFS-Net-M	83.40	+1.50

Table 2: Experimental retrieval accuracy with CIFAR-10

DCNN	Accuracy (%)	vs. Scheme 1	vs. Scheme 2
DFS-Net-I	87.93	11.45	2.76
DFS-Net-R	84.60	22.32	8.56
DFS-Net-M	77.74	9.55	7.95

Table 3: Experimental retrieval time with CIFAR-10

DCNN	time(s)	vs. Scheme 1	vs. Scheme 2
DFS-Net-I	12.77	-0.08	-0.11
DFS-Net-R	12.52	-0.51	-0.47
DFS-Net-M	12.66	-0.54	-0.47

In the Inception network provided by the official website, the method proposed in this paper has the best retrieval accuracy and efficiency. The retrieval accuracy of DFS-Net-I is 87.93%, which is 16.45% higher than that of scheme 1. In the improved DFS-Net-R and DFS-Net-M, accuracy is also increased by 12.32% and 9.55%, respectively. It can be seen that after the deep feature selection module is added to the network, the training model is optimized, and the retrieval efficiency is significantly improved. Compared with scheme 3, the unimproved network extracts feature and processes them with information gain rate and hash. Its retrieval accuracy is lower than that of DFS-Net, not to mention the retrieval results of the comparative method presented here. It confirmed that the deep features extracted by the improved network could better express the information content of the image.

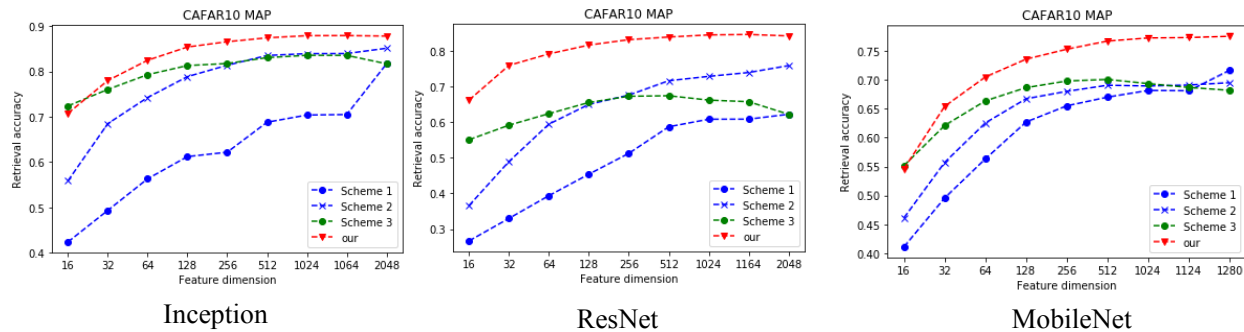


Figure 5: Experimental results based on classical networks

Experiments show that the DFS-Net-I is the optimal model, and the retrieval accuracy reached to 83.91%, which is 6.92% higher than scheme 1 in all networks. The efficiency of DFS-Net-R and DFS-Net-M, increased by 17.1% and 8.03% respectively. It can be seen that after the deep feature selection module is added to the network, the training model is optimized, and the retrieval efficiency is significantly improved. By comparing with experimental Schemes 2 and 3, it can be seen that the improved network has a great contribution to improving the efficiency of feature retrieval.

From the experimental results in Tab. 3 and Tab. 6, it can be found that by comparing the DFS-Net direct features extraction retrieval effect with the unimproved network features extraction retrieval effect, and it is proved that the addition of deep feature selection module in the network is conducive to the extraction of more distinguishing features, thus improving the retrieval accuracy.

Table 4: Improved model accuracy of network training with Caltech256

Model1	Accuracy (%)	vs. Ori (%)
DFS-Net-R	81.29	+0.74
DFS-Net-I	85.54	+0.69
DFS-Net-M	80.07	+0.10

Table 5: Experimental retrieval accuracy with Caltech256

DCNN	Accuracy (%)	vs. Scheme 1	vs. Scheme 2
DFS-Net-I	83.91	+6.92	+1.630
DFS-Net-R	76.74	+17.10	+11.28
DFS-Net-M	74.20	+8.03	+3.29

Table 6: Experimental retrieval time with Caltech256

DCNN	time(s)	vs. Scheme 1	vs. Scheme 2
DFS-Net-I	64.71	-6.41	-3.09
DFS-Net-R	63.33	-4.26	-4.12
DFS-Net-M	64.96	-2.21	-3.81

Our method, which reprocesses the features extracted by the improved network, takes the information gain rate as the standard, tested the strength of each feature recognition ability, and then reordered the features according to the strength order of feature recognition ability, which is helpful to remove redundant features. The features of different network extractions are improved by using Information Gain Rate and hash methods, which enhance the retrieval performance to various degrees. However, generally speaking, after the reduction of the Information Gain Rate and hash dimension, the retrieval efficiency of the feature extracted by the improved network is further improved. Meanwhile, the time needed and feature dimension

required are both low. As shown in Figs. 5 and 6. The experiments show that our method can be applied to various networks.

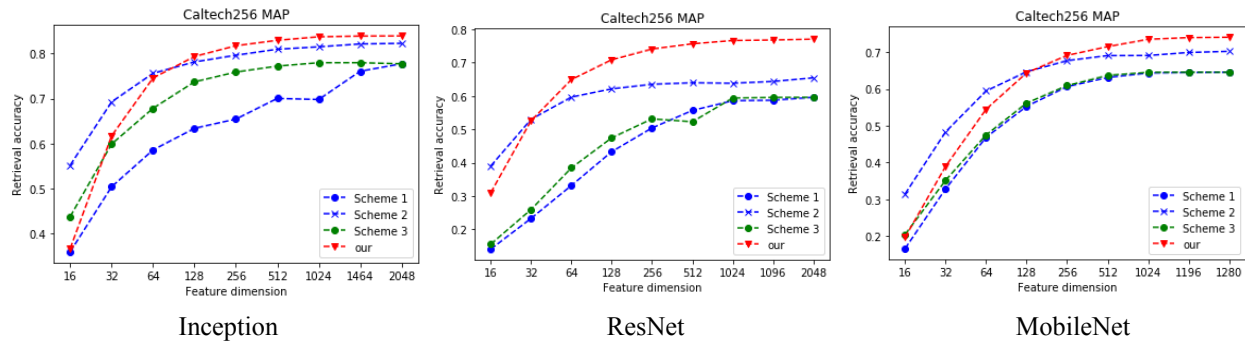


Figure 6: Experimental results based on classical networks

5 Conclusion

This method reconstructs DCNN architecture and uses deep learning to extract features with strong recognition ability. The DFS-Net with multiple convolutional branches, which based on the classical network model, is added to form a deep feature selector so that it can perceive more details of the image. At the same time, the hash transform is introduced, and the corresponding dimensionality reduction strategy is proposed to control the calculation cost. Information Gain Rate, as the criterion of feature selection, can effectively eliminate the redundancy, extraction, and efficient features of the image correlation. The features of improved network extraction are further improved after Information Gain Rate and hash transform. Finally, DFS-Net based on the ResNet50, InceptionV3, and MobileNetV2 is thoroughly investigated and evaluated on the Cifar10 and Caltech256 datasets, the retrieval accuracy is higher than that of unmodified networks. The comprehensive experiments demonstrate that the presented novel method can effectively improve the retrieval accuracy and efficiency of small and medium-size datasets. In future work, we will further study more advanced network improvement strategies and hash methods for image retrieval.

Funding Statement: This work is partially supported by National Natural Foundation of China (Grant No. 61772561), the Key Research & Development Plan of Hunan Province (Grant No. 2018NK2012), Graduate Education and Teaching Reform Project of Central South University of Forestry and Technology (Grant No. 2018JG005), and Teaching Reform Project of Central South University of Forestry and Technology (Grant No. 20180682).

Conflicts of Interest: We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

- [1] Z. H. Zhou, "Learnware, on the future of machine learning," *Frontiers of Computer Science*, vol. 10, no. 4, pp. 589–590, 2016.
- [2] Z. Mehmood, F. Abbas, T. Mahmood, M. A. Javid, A. Rehman *et al.*, "Content-based image retrieval based on visual words fusion versus features fusion of local and global features," *Arabian Journal for Science & Engineering*, vol. 43, no. 12, pp. 7265–7284, 2018.
- [3] C. Li, L. L. Pan, R. Y. Chen, Y. Zhou and W. Z. Shao, "A novel fusion DCNN for image classification," *Computer Engineering & Science*, vol. 41, no. 12, pp. 2179–2186, 2019.
- [4] J. Lin, O. Yang, Y. Z. Liu and H. Z. Shu, "Robust hashing for image authentication using SIFT feature and quaternion Zernike moments," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 2609–2626, 2017.
- [5] S. S. Husain and M. Bober, "Improving large-scale image retrieval through robust aggregation of local descriptors," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 9, pp. 1783–1796, 2017.

- [6] L. L. Pan, J. H. Qin, H. Chen, X. Y. Xiang, C. Li *et al.*, “Image augmentation-based food recognition with convolutional neural networks,” *Computers, Materials & Continua*, vol. 59, no. 1, pp. 297–313, 2019.
- [7] L. L. Pan, C. Li, S. Pouyanfar, R. Y. Chen and Y. Zhou, “Effect of rotation on the propagation of waves in hollow poroelastic circular cylinder with magnetic field,” *Computers, Materials & Continua*, vol. 62, no. 2, pp. 731–746, 2020.
- [8] F. Ye, Y. F. Su, H. Xiao, X. Q. Zhao and W. D. Min, “Remote sensing image registration using convolutional neural network features,” *IEEE Geoscience & Remote Sensing Letters*, vol. 15, no. 2, pp. 232–236, 2018.
- [9] A. Krizhevsky, I. Sutskever and G. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 205–222, 2012.
- [10] K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 9, pp. 346–361, 2014.
- [11] J. J. Lu and K. Y. Tong, “Robust single accelerometer-based activity recognition using modified recurrence plot,” *IEEE Sensors Journal*, vol. 1, no. 1, pp. 99–102, 2019.
- [12] H. Kebapci, B. Yanikoglu and G. Unal, “Plant image retrieval using color, shape, and texture features,” *The Computer Journal*, vol. 54, no. 9, pp. 1475–1490, 2018.
- [13] Y. S. Dong, D. C. Tao, S. Member, X. L. Li, J. W. Ma *et al.*, “Texture classification and retrieval using shearlets and linear regression,” *IEEE Trans Cybern*, vol. 45, no. 3, pp. 358–369, 2017.
- [14] J. Sharma and S. Navlakha, “Improving similarity search with high-dimensional locality-sensitive hashing,” arXiv:1812.01844, 2018.
- [15] S. Xin, H. N. Qian and Z. H. Deng, “Chinese herbal medicine image recognition and retrieval by convolutional neural network,” *PLoS One*, vol. 11, no. 6, pp. 156–168, 2016.
- [16] X. L. Ren, X. G. Chen, J. S. Cao and T. R. Cai, “Image retrieval based on deep learning features,” *Computer Engineering & Design*, vol. 39, no. 2, pp. 503–510, 2018.