Journal of Information Hiding and Privacy Protection
DOI:10.32604/jihpp.2020.010472
Article

Image Retrieval Based on Deep Feature Extraction and Reduction with Improved CNN and PCA

Rongyu Chen, Lili Pan*, Yan Zhou and Qianhui Lei

Central South University of Forestry and Technology, Changsha, 410000, China
*Corresponding Author: Lili Pan. Email: lily_pan@163.com
Received: 01 July 2020; Accepted: 28 July 2020

Abstract: With the rapid development of information technology, the speed and efficiency of image retrieval are increasingly required in many fields, and a compelling image retrieval method is critical for the development of information. Feature extraction based on deep learning has become dominant in image retrieval due to their discrimination more complete, information more complementary and higher precision. However, the high-dimension deep features extracted by CNNs (convolutional neural networks) limits the retrieval efficiency and makes it difficult to satisfy the requirements of existing image retrieval. To solving this problem, the high-dimension feature reduction technology is proposed with improved CNN and PCA quadratic dimensionality reduction. Firstly, in the last layer of the classical networks, this study makes a well-designed DR-Module (dimensionality reduction module) to compress the number of channels of the feature map as much as possible, and ensures the amount of information. Secondly, the deep features are compressed again with PCA (Principal Components Analysis), and the compression ratios of the two dimensionality reductions are reduced, respectively. Therefore, the retrieval efficiency is dramatically improved. Finally, it is proved on the Cifar100 and Caltech101 datasets that the novel method not only improves the retrieval accuracy but also enhances the retrieval efficiency. Experimental results strongly demonstrate that the proposed method performs well in small and medium-sized datasets.

Keywords: Image retrieval; deep features; convolutional neural networks; principal components analysis

1 Introduction

With the improvement of living quality and the arrival of the era of big data, the human demand for image retrieval and computer performance are increasingly high. To pursue better retrieval performance, the research on image retrieval has shifted from Text-based Image Retrieval to CBIR (Content-based Image Retrieval). The original CBIR tends to extract SIFT (Scale-invariant feature transform), HSV (Hue, Saturation, Value), YCbCr (Color Space), and other local features [1–3]. Juan et al. [4] compared the features of SIFT, PCA-SIFT (Principal Components Analysis-Scale-invariant feature transform), and SURF (Speeded Up Robust Features), which laid the foundation for subsequent studies. The rapid development of deep learning makes features extracted no longer limited to the shallow features, and CNN are also widely used to obtain global features of images.

Krizhevsky et al. [5] designed AlexNet which won the Champion in the ImageNet data contest with an error rate of 15.4%, and the error rate was well below the second (26.2%). The success of AlexNet shows that the way of computer automatic learning mode features and features learning into the establishment of the model can reduce the imperfection caused by artificial design features, and the globality of the extracted features can better reflect the correlation between pictures. Radenović et al. [6] used fine-tuning CNN in a

fully automated manner, employed state-of-the-art retrieval and SfM (Structure-from-Motion) methods to obtain 3D models. 3D models are used to guide the selection of the training data for CNN fine-tuning and conduct the training of the fine-tuning of the VGG of the classical convolutional neural network. The features extracted by the fine-tuning VGG were used for image retrieval. Although this method can inherit the globality of VGG network features, its high dimensional features will consume a large amount of computing, storage resources, and search time for image retrieval.

With the coming of big data era, the higher dimensional deep features for classification task cannot satisfy image retrieval research, while image retrieval is facing the entire image databases, not just the test sets. It is essential to calculate the similarity distance between test sets and whole image databases. If too many features represent each image, the storage time and retrieval time will be significantly increased. Recently, with the development of CNN architecture, models have emerged which can extract low dimensional features for classification tasks. However, the use of 1024 d feature expression in each image will still generate a lot of consumption and reduce the retrieval efficiency for large-scale image retrieval. Therefore, this paper firstly designs a DR-Module in classic networks, which fully plays the performance of the network to learn the compression features autonomously, retains high correlation features in the process of autonomous learning. Secondly, PCA is used to dimensionality reduction of extracted features and represented by binary codes to realize the most effective expression force with the least data. Finally, the features after DR-Module was calculated by Hamming distance. The first top_k images were taken (the average accuracy of the first top_k images) was calculated.

2 Related Work

This section will focus on deep learning and image retrieval.

2.1 Deep Learning

Deep learning is a technology to realize machine learning, which is derived from the research of artificial neural networks. Its purpose is to build a neural network that can simulate the human brain to analyze and understand data. Currently, the performance of existing algorithms has been surpassed in recognition or classification in deep learning application scenarios that meet specific conditions. CNN is also applied in various fields of life, such as food ingredients [7–8], image recognition, target detection, and image retrieval.

The rise of VGG laid a foundation for the development of classic CNNs in 2014. The large convolution kernel of VGG is replaced by the small convolution kernel. Its multi-layer nonlinear layer ensures the learning of more complex patterns by increasing the deep of the network at a low cost, but its high dimensional features can not conducive to image retrieval. ResNet, a residual network proposed by [9], not only solved the vanishing gradient problem of the network but also reduced features to half of the VGG features, reducing storage consumption. The DenseNet network proposed by [10] could reduce the amount of computation per layer and enhance the utilization of features. The extracted feature dimension was one half of ResNet, achieving high classification accuracy with low dimensional features. The latest MobileNet proposed by [11] used the lightweight network to make a flat error rate, which solved the high training time and became a new standard design of the network in contemporary.

CNN automatically learns the features (color, brightness, edge, texture, etc.) of images at all levels through convolution and pooling operation, which consistent with the common sense of human understanding of images. At this stage, CNN for the classification training model has been widely used in the domain of image retrieval.

2.2 Image Retrieval

The study of image retrieval methods began with text-based image retrieval in the 1970s. With the advent of the age of big data, text-based image retrieval methods can no longer meet the requirements of the times, research has shifted from text-based to content-based on image retrieval.

The core of CBIR lies in features extraction. Color, shape, texture [12], and structure are usually transformed into a single feature to describe image content in early CBIR algorithms and systems. The manual feature descriptors of CBIR, such as Gabor filter, SIFT, HOG (Histogram of Oriented Gridients) [13] have been deeply studied in Search Engine, Medicine, E-commerce, and other fields.

To obtain the implicit relationship of a large amount of image data, generate more distinguishing and representative features, some researches begin to focus on the close connection between deep learning and retrieval tasks, such as trademark image retrieval [14] and medical image retrieval [15]. For the phenomenon that the features extracted by classic CNN may not achieve the best effect, Zhou et al. [16] effectively encoded the complementary clues of SIFT features and CNN features into the matching kernel by integrating the matching functions of SIFT and CNN features. Li et al. [17–18] integrated features extracted from different networks through the fusion network, which improved the classification accuracy. Although it improves the ability of image recognition, its high dimensional features will consume more memory. As people pursuit of the efficiency of retrieval, the Hash Code and PCA method are widely used in image retrieval. Zhang et al. [19] proposed a weighted cross-entropy loss and loss to the minimum mean square error structure loss function method, which can realize feature learning and Hash code simultaneously. Hervé et al. [20] improved the PCA dimensionality reduction method and reduced the dimensionality through the co-occurrence and relevance of visual words, so that its low-dimensional features could be applied to large-scale image retrieval. Although Zhang and Herve embody the high efficiency of image retrieval, it does not have the self-learning image performance based on CNN. Guo et al. [21] proposed a CNN based hash method, which binarized the activation of the fully connected layer with a threshold of 0 and outputted the binary results as hash codes. Then Gao et al. [22] proposed a quantitative analysis method based on the Infomax principle in neural network for the optimization of sum pooling, which matched well with the principal feature vector of PCA with remarkable effect. It is proved that the improved network matching modified PCA algorithm is feasible in the field of image retrieval.

The image information expressed by the deep features obtained by the classification task is too redundant for image retrieval and will consume a lot of resources. Therefore, it is not advisable to directly conduct image retrieval with the deep features extracted by the existing CNN. Based on the retrieval accuracy and retrieval efficiency, this paper proposes to use DRNet (dimensionality reduction CNN) +PCA to obtain compression features containing more abundant image information for image retrieval, so as to achieve higher average accuracy and retrieval efficiency.

3 The Framework of Deep Feature Secondary Dimensionality Reduction

Since 2012, CNNs have achieved remarkable results in the field of image classification through constant enhancement in depth and structure. Based on classic CNN (ResNet50, DenseNet121, MobileNetV2), a new network DRNet is constructed by adding DR-Module, and the features extracted are reduced by PCA again in this study.

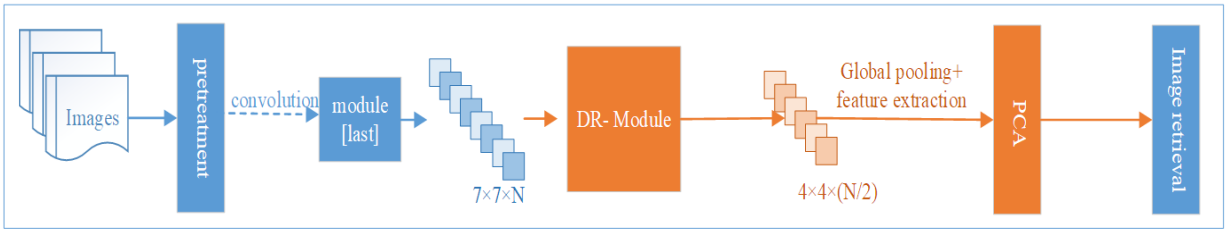

Figure 1: with the framework of secondary dimension reduction

The architecture is shown in Fig. 1. DR-Module adopts channel shuffle, group convolution, mean-pooling long-jump connection, and other operations to express the information of the image through the fusion and compression of deep features. Exceedingly reducing the feature dimension in CNN is not only bad for network training, but also can reduce the retrieval accuracy. Therefore, DRNet reduces the feature to 1/2 of the original through DR-Module, which not only reduces the feature dimension but also

improves the retrieval accuracy after the dimension is reduced in this paper. Although the image feature is reduced to 1/2 of the original value in the network, the storage memory and retrieval time of the 1024 d feature are too high for image retrieval. Therefore, in this paper, the features extracted by DRNet are dimensionalized again by PCA, and the effective information expression of a picture features as low as 128 d. In this paper, DRNet and PCA are used together for two dimensionality reduction to achieve higher retrieval accuracy, lower memory consumption, and faster image retrieval efficiency.

3.1 DR-Module (Dimensionality Reduction Module)

The development of classic CNN is not only to improve the retrieval accuracy by deepening the network length and widening the network width, but also to take into account the computation amount of CNN in the convolution process and the memory of extracting features. A large number of parameters will generate a huge amount of computation, and high-dimensional features will consume a huge amount of resources.

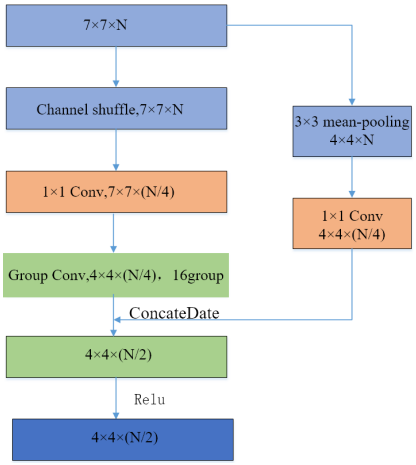

Figure 2: DR-Module design

Assuming the size of the input feature map is 7 × 7 × N, the DR-Module combines with the advantages of classic CNNs is shown in Fig. 2, which is aiming to achieve higher retrieval accuracy and retrieval efficiency with fewer layers and parameters. The DR- Module focuses on the benefits of channel shuffle, group convolution, mean-pooling long-jump connection, and other operations.

In this paper, the DR-Module of CNN adopts the branching mode. The trunk uses channel shuffle to divide the input features into four groups, where the four groups of features are compressed to be 1/4 of the original input features. Another branch firstly mean-pooling is carried out to deep feature with convolution kernel of 3 step size of 2, and then compress the number of channels to 1/4 of the original number with a convolution kernel of 1 × 1; Finally, the feature map generated by splicing the trunk and branch is 4 × 4 × (N/2).

There are several reasons for the DR-Module structure:

(1) The first advantage of grouping convolution is obtaining compression features by effective training. Group convolution divides the datasets into several batches and then trains each batch separately in DRNet, which is crucial for training speed when training DRNet. The second advantage is that the model is more efficient for the reduction of parameters, and the amount of computation between features and features is decreased.

(2) Channel shuffle strengthens the connection between features and features, enriches the information of each channel, and makes up for the information loss caused by grouping convolution. For example: for N d features, {1,2,3,4,5,6,7, ... N}. Since 1, 2, 3 and 4 belong to adjacent features, the information contained in them may be the same, if convolution is carried out directly, only the adjacent

features are compressed, and the connection between features is not enhanced. If it is divided into four groups and channel shuffle, the features are {1,5,...2,6,...3,7,...4,8, ... N}, the features are not the adjacent feature when convolution is done, which strengthens the relationship between features and features.

(3) The branch reduces the deep feature dimension with the mean-pool long- jump connection and 1 × 1 convolution kernel, maximum inherits deep features for the classification task and solves the Vanishing Gradient problem of the network. Fourthly, the ReLU activation function is used to find the optimal loss function by adding nonlinear factors, which improves the expression ability of the model.

In summary, the DR-Module enables networks to learn how to compress features with the minimum amount of computation, obtain highly relevant features, adequately express image information with low dimensional features, and improve retrieval efficiency and accuracy.

3.2 PCA-01

Principal Components Analysis, also known as karhunen-loeve Transform, is a technique used to explore high dimensional data structures. PCA is used for exploration and visualization of high dimensional datasets commonly. It can also be used for data compression, data preprocessing, and so on. Although the dimension of decreased after PCA, it takes up eight bytes, which is too much storage memory for the massive features extracted from numerous pictures. The features of PCA will be standardized after dimensionality reduction, and the average value is 0. For this phenomenon, each feature of the image is represented by either a 0 or a 1 after PCA, each feature occupies only one byte. With the DR-Module proposed in this paper, not only can the storage space be saved, but also the retrieval efficiency can be improved. In this paper, the features of CNNs in 1024 d will be reduced to 512,256,128,64,32,16 to find the best performance of retrieval results.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$c = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{2}$$

$$F_{PCA} = \bar{x} \cdot W \tag{3}$$

$$F_{PCA-01} = \begin{cases} 0 F_{PCA} \le 0 \\ 1 F_{PCA} > 0 \end{cases} \tag{4}$$

$$\frac{\sum_{i=1}^{d'} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \ge t \tag{5}$$

$x = \{x_1, x_2, ..., x_n\}$ represents the features of the datasets，n represents n features of each image. Firstly, calculate the mean of the features of the datasets, as shown in Formula (1); Secondly, the covariance is calculated, such as Formula (2). The eigenvalue and eigenvectors of the covariance matrix are calculated based on the covariance and sorted in order from large to small; Thirdly, Calculate the value of, such as Formula (3); Finally, as is normalized, its mean value is 0, according to which can be calculated, as shown in Formula (4). In general, if the first d' components are retained, a reconstruction threshold t should be set so that the selection can be established as the minimum d' value，as shown in Formula (5). In this paper, d’ = 1024,512,256,128,64,32,16.

3.3 Image Retrieval Algorithm with Quadratic Dimension Reduction

The image retrieval algorithms generally include extracting image features, reducing PCA dimensionality, calculating similarity distance, and outputting image retrieval results. Based on the

quadratic dimensionality reduction method in this paper, the algorithm flow of image retrieval is shown in algorithm 1, including the precision calculation method.

Algorithm 1: Image retrieval algorithm

Inputs: datasets
Outputs: Acc.
1: Carry out datasets
2: F_M = model(datasets)
3: F_{PCA} = PCA (F_M)
4: F_{PCA_01} = F_{PCA} (0, 1)
5: d = Hamming Distances (F_{PCA_01})
6: For K = 0:k
7: rank_list2 = argsort (d)
8: Class_R = class_data(rank_list)
9: End For
10: Result = where (Class_i == Class_R, 1, 0)
11: Acc. = T_r/T_t
12: Return Acc.

In Algorithm 1, input datasets (image 1, image 2, image 3 and so on), call the trained model of DRNet to extract the image features for F_M. Then, obtain the features as F_{PCA} by Formula (3), and the F_{PCA} is represented by 0 or 1 based on the average value 0 get F_{PCA_01} which is used to image retrieve, such as Formula (4). The Hamming distance (d) is calculated according to xor for features (F_{PCA_01}), and Line 7 to Line 10 show that the image sequence number(rank_list2) and image category (Class_R) of the first k closest images are returned. Finally, the retrieval accuracy (Acc) is calculated according to the number of similar images in the recovered images, Tr is denoted as No.k of relevant images, Tt is denoted as a total No. of images.

4 Experimental Analysis

In the experiment, firstly, classic CNNs (MobileNetV2, ResNet50, and DenseNet121) were fine-tuned to obtain the best model (mini_batch was set to 16 during the training, 30 rounds of training in total); then the best model of classical CNNs and DRNet were combined with PCA for image retrieval; finally, and retrieval time and retrieval accuracy were compared. This paper bases on two datasets of Cifar100 and Caltech101, which highlights that the proposed DR-Module and PCA achieve higher retrieval precision and retrieval efficiency.

All the networks are implemented in Keras, a popular deep learning framework, and trained/tested on a GPU server with 16G server memory, NVIDIA GeForce RTX 2060 graphics card, 6G video memory, 240 Tensor cores, and 1920 CUDA cores.

4.1 Cifar100 Dataset

Cifar100 dataset [23] contains a total of 60,000 images in 100 categories, including seals, whales, roses, sunflowers, bottles, bowls, camels, cattle, etc. Sixty images were randomly selected for each category, 50 for training and 10 for testing. The average accuracy of the top10 was calculated.

This dataset bases on classic networks (ResNet50, DenseNet121, and MobileNetV2). In this study, four extraction feature methods are classic CNN fine-tuning optimal model extraction feature (Ori), proposed DRNet optimal model extraction feature (DR), classical CNN+PCA extraction feature (Ori-PCA), DRNet+PCA extraction feature (DR-PCA).

Table 1: The accuracy of Cifar100

	Ori	DR	Ori-PCA	DR-PCA
ResNet50	0.6356	0.6290	0.6344	0.6748
DenseNet121	0.6407	0.6554	0.5469	0.6594
MobileNetV2	0.5792	0.5930	0.5454	0.6013

In the Cifar100 dataset, our experiment confirmed that the deep features extracted by classical CNN+PCA do not improve the retrieval accuracy. On the contrary, the DRNet+PCA method proposed in this paper has enhanced the accuracy of each network. As shown in Tab. 1, the DRNet+PCA has the best effect in ResNet50, 4% higher than the classic CNN, and 5% higher than the classic CNN+PCA. In terms of MobileNetV2, the DRNet accuracy increases by 1.4% compared with the typical CNN, and it also increases by 1% when combined with PCA. The DenseNet121 DRNet+PCA retrieval accuracy is not significantly improved compared with the DRNet and the classic CNN, but it is 11% higher than the classic CNN+PCA, which fully demonstrates the superior effect of the combination of the two. Thus, DRNet+PCA can achieve higher retrieval accuracy.

Table 2: The storage memory of Cifar100 (MB)

	Ori	DR	Ori-PCA	DR-PCA
ResNet50	201	107	23.4	11.7
DenseNet121	106	51.6	11.7	5.90
MobileNetV2	110	64.2	14.6	7.36

It can be seen that features extracted by DRNet occupy 1/2 of the memory of features obtained by classic CNN in Tab. 2, which can save half of the storage memory and improve the retrieval efficiency. The storage space of the DRNet+PCA method only accounts for 1/18 of the classical CNN method, which significantly reduces the memory storage space, shows the superiority of the technique in saving storage space in this paper.

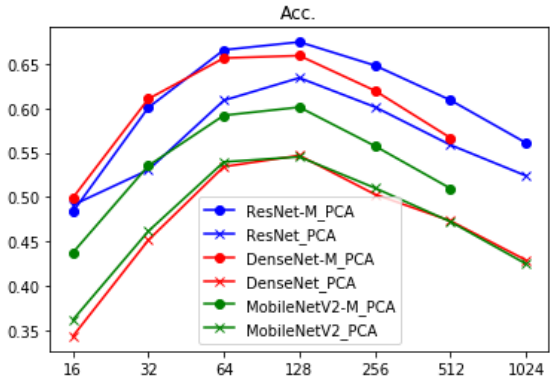

Figure 3: Comparison between classic CNN and DRNet based on PCA in cifar100

Fig. 3 shows the image retrieval accuracy of DRNet+PCA and classic CNN+PCA achieve the highest when it is 128 d. Whether it is ResNet50, DenseNet121, or MobileNetV2, classic CNN+PCA is far lower than that of DRNet+PCA. For example, ResNet50, the dimension of classic CNN is 2048 dimension, and the feature extracted by DRNet is 1024 d, so features storage memory is reduced by half. The PCA and DRNet achieve the highest retrieval accuracy when the dimension is 128 d, which is 1/32 of features extraction of classic CNN, significantly reducing the image features storage memory. It can be seen that the novel method realizes high retrieval accuracy by low dimensional features in this paper.

4.2 Caltech101 Dataset

The Caltech101 dataset was collected by [24]. It contains 9146 images, 101 classes of pictures, including camera, chair, football, panda, big tree, watch and so on, each category contains approximately 40 to 800 images. According to the experimental method of [25], 30 images were randomly selected in each category, 20 of which were used for model training and 10 for testing, and the average accuracy of top10 each image of the search was calculated.

In the experiment, four feature extraction methods are classical CNN optimal model extraction feature (Ori); DRNet best model extraction feature (DR); classical CNN+PCA extraction feature (Ori-PCA); DRNet+PCA extraction feature (DR-PCA).

Table 3: The accuracy of Caltech101

	Ori	DR	Ori-PCA	DR-PCA
ResNet50	0.8776	0.8861	0.8849	0.9067
DenseNet121	0.9130	0.9259	0.8697	0.9285
MobileNetV2	0.8870	0.9172	0.8815	0.9143

Table 4: Accuracy comparison with existing methods

Method	Acc. (%)
Das	85
Walaa	91.5
M_PCA	92.85

The experimental results of the Caltech101 dataset on the classical CNN and the DRNet are shown in Tab. 3, and comparison with previous studies are shown in Tab. 4. Tab. 3 shows that on Caltech101 dataset, the accuracy of DRNet is higher than that of classic CNN, and the accuracy of DRNet+PCA is higher than that of classic CNN+PCA.

According to Tab. 4, the average accuracy of the image retrieval method proposed in this paper reached 92.85% on the Caltech101 dataset, 8% higher than that of [26], and 1.4% higher than that of [25]. The experimental results demonstrate the proposed method has higher retrieval accuracy than the previous image retrieval research on the Caltech dataset.

The research by Walaa et al. did not consider the retrieval efficiency and storage memory of image retrieval. It can be seen the storage memory studied in this paper is much lower than the conventional features storage memory in Tab. 5. Fig. 4 shows, the DRNet+PCA in dimension for 128 d of the image retrieval accuracy is the highest, retrieval accuracy is much higher than 1024 d, achieving higher precision by lower dimensions. Compared with the classic CNN, the DRNet not only reduces the size but also improves the retrieval accuracy.

Table 5: The storage memory of Caltech101 (MB)

	Ori	DR	Ori-PCA	DR-PCA
ResNet50	98.4	54.4	11.7	5.94
DenseNet121	55.4	25.6	5.94	2.98
MobileNetV2	51	32.2	7.43	3.72

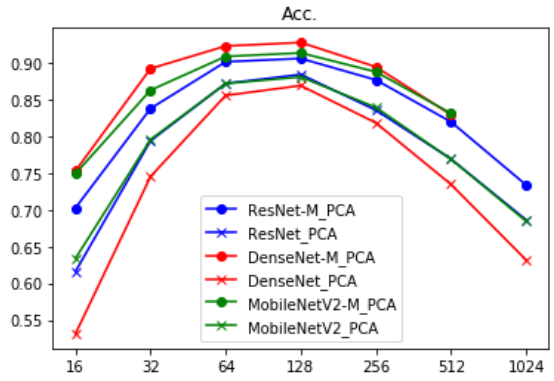

Figure 4: Comparison between classic CNN and DRNet based on PCA in Caltech101

5 Conclusion

In this paper, a new DR-Module is set based on classic CNNs, which achieves high retrieval accuracy by low-dimension features. The PCA-01 method enhances the efficiency of image retrieval by reducing memory cost. In this scheme, firstly the DR-Module is used for features fusion of image features, and the compression ratios are 50% as before. Then PCA is utilized to reduce the dimension of the DRNet features again, and the dimension reduction features are set to 0 or 1 at the same time. The storage memory of the 0 or 1 is much lower than that of conventional features. The average accuracy of the novel framework is 92.85% in this paper when the feature dimension is 128 for the small dataset Caltech101, which is higher than the previous works. Experimental results strongly demonstrate our method achieves higher retrieval accuracy and costs lower storage space. In future work, the research group will continue to improve features extraction and optimize the retrieval features for image retrieval, to achieve better retrieval performance for massive datasets.

Funding Statement: This work is partially supported by National Natural Foundation of China (Grant No. 61772561), and the Key Research & Development Plan of Hunan Province (Grant No. 2018NK2012).

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] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 4, no. 2, pp. 506–513, 2004.

[2] S. Sural, Q. Gang and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Int. Conf. on IEEE*, vol. 2, pp. 589–592, 2002.

[3] M. H. Saad, H. I. Saleh, H. Konbor and M. Ashour, "Image retrieval based on integration between YCbCr color histogram and shape feature," in *Computer Engineering Conf.*, pp. 97–102, 2011.

[4] L. Juan and O. Gwun, "A comparison of sift, PCA-sift and surf," *International Journal of Signal Processing*, vol. 3, no. 4, pp. 143–152, 2009.

[5] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[6] F. Radenović, G. Tolias and O. Chum, "CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples," in *European Conf. on Computer Vision*, pp. 3–20, 2016.

[7] L. Pan, J. Qin, H. Chen, X. 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.

[8] L. Pan, S. Pouyanfar, H. Chen, J. Qin and S. C. Chen, "Deepfood: Automatic multi-class classification of food

ingredients using deep learning," in *2017 IEEE 3rd Int. Conf. on Collaboration and Internet Computing (CIC)*, pp. 181–189, 2017.

[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[10] G. Huang, Z. Liu, D. M. Van, Laurens, K. Q. Weinberger *et al.,* "Densely connected convolutional networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

[11] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang *et al.,* "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[12] S. Kaur and V. K. Banga, "Content based image retrieval: Survey and comparison between RGB and HSV model," *International Journal of Engineering Trends & Technology*, vol. 4, no. 4, pp. 575–579, 2013.

[13] R. Hu and J. Collomosse, "A performance evaluation of gradient field HOG descriptor for sketch based image retrieval," *Computer Vision & Image Understanding*, vol. 117, no. 7, pp. 790–806, 2013.

[14] T. Lan, X. Feng, Z. Xia, S. Pan and J. Peng, "Similar trademark image retrieval integrating LBP and convolutional neural network," in *Int. Conf. on Image and Graphics*, pp. 231–242, 2017.

[15] A. Qayyum, S. M. Anwar, M. Awais and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, 2017.

[16] D. Zhou, X. Li and Y. J. Zhang, "A novel CNN-based match kernel for image retrieval," in *IEEE Int. Conf. on Image Processing*, pp. 2445–2449, 2016.

[17] 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.

[18] L. L. Pan, C. Li, S. Pouyanfar, R. Y. Chen and Y. Zhou, "A novel combinational convolutional neural network for automatic food-ingredient classification," *Computers, Materials & Continua*, vol. 62, no. 2, pp. 731–746, 2020.

[19] Z. Zhang, Q. Zou, Q. Wang, Y. Lin and Q. Li, "Instance similarity deep hashing for multi-label image retrieval," arXiv preprint arXiv:1803.02987, 2018.

[20] J. Hervé and C. Ondřej, "Negative evidences and co-occurences in image retrieval: The benefit of PCA and whitening," in *European Conf. on Computer Vision*, pp. 774–787, 2012.

[21] J. Guo and J. Li, "CNN based hashing for image retrieval," arXiv preprint arXiv:1509.01354, 2015.

[22] Z. Gao, L. Wang, L. Zhou and M. Yang, "Infomax principle based pooling of deep convolutional activations for image retrieval," in *IEEE Int. Conf. on Multimedia and Expo (ICME)*, pp. 457–462, 2017.

[23] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Technical Report*, University of Toronto, 2009.

[24] L. Fei-Fei, R. Fergus and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *2004 Conf. on Computer Vision and Pattern Recognition Workshop*, 2004.

[25] E. Walaa, K. Abdelahab and Y. Shady, "CBIR based on weighted multi-feature voting technique," *International Journal of Imaging and Robotics,* vol. 18, no. 2, pp. 39–52, 2018.

[26] R. Das, S. Thepade and S. Ghosh, "Novel feature extraction technique for content-based image recognition with query classification," *International Journal of Computational Vision and Robotics*, vol. 7, no. 1–2, pp. 123–147, 2017.