ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20210925093644
Original scientific paper

Image Retrieval Method Combining Bayes and SVM Classifier Based on Relevance Feedback with Application to Small-scale Datasets

Siyu LAI, Qinghua YANG, Wenjin HE, Yuanzhong ZHU, Juan WANG*

Abstract: A vast amount of images has been generated due to the diversity and digitalization of devices for image acquisition. However, the gap between low-level visual features and high-level semantic representations has been a major concern that hinders retrieval accuracy. A retrieval method based on the transfer learning model and the relevance feedback technique was formulated in this study to optimize the dynamic trade-off between the structural complexity and retrieval performance of the small- and medium-scale content-based image retrieval (CBIR) system. First, the pretrained deep learning model was fine-tuned to extract features from target datasets. Then, the target dataset was clustered into the relative and irrelative image library by exploring the Bayes classifier. Next, the support vector machine (SVM) classifier was used to retrieve similar images in the relative library. Finally, the relevance feedback technique was employed to update the parameters of both classifiers iteratively until the request for the retrieval was met. Results demonstrate that the proposed method achieves 95.87% in classification index *F*1 - Score, which surpasses that of the suboptimal approach DCNN-BSVM by 6.76%. The performance of the proposed method is superior to that of other approaches considering retrieval criteria as average precision, average recall, and mean average precision. The study indicates that the Bayes + SVM combined classifier accomplishes the optimal quantities more efficiently than only either Bayes or SVM classifier under the transfer learning framework. Transfer learning skillfully excels training from scratch considering the feature extraction modes. This study provides a certain reference for other insights on applications of small- and medium-scale CBIR systems with inadequate samples.

Keywords: Bayes classifier; image retrieval; relevance feedback; support vector machine classifier; transfer learning

1 INTRODUCTION

With the development of multimedia and the pervasiveness of mobile devices, a diverse array of images has been accreted in various professional fields, such as medicine, military affairs, criminal investigation, e-commerce, and smart logistics [1]. The organization, management, and retrieval of these massive images with high efficiency have become popular topics for academics worldwide. Content-based image retrieval (CBIR) systems provide indispensable opportunities to address the problem of extracting and sharing information. Therefore, these systems have been actively studied and applied [2].

A general CBIR system comprises two components: offline and online phases. In the first phase, feature vectors are extracted from the image library to construct a local feature database. This stage is a time-consuming process and depends extremely highly on the computing power, the number of images, and the algorithm used. In the second phase, the same features are extracted from the query image, and the similarity measure is calculated between the features of the query image and that of the local feature database. Images with high similarity are returned to users. The traditional CBIR systems work with low-level visual features (e.g., color, texture, and shape), which often fail to provide sufficient semantic representations in certain fields, thus yielding only a limited number of examples [3-5]. The contradiction of this kind inevitably leads to the semantic gap between low-level features and high-level representations because of the shortage in semantic features, which cannot be easily eliminated. In the field of image processing and semantic recognition, integrating multiple studies, such as feature extraction [6, 7], classification [8, 9], and annotation [10, 11], is useful in mapping from low-level visual features to high-level semantic concepts. Currently, one practical approach to achieve this objective is reducing the number of images involved by using the well-trained classifier(s). However, this approach is problematic because the accuracy markedly depends on the level to which the features are extracted and the gap is bridged, which is quite challenging [12, 13].

The above analysis indicates that extensive and in-depth studies on feature extraction and semantic gap bridging have been conducted [14, 15]. However, unified standards on the kinds of appropriate features in a specific field, especially on reaching a balance between the complexity of feature extraction and retrieval performance, are unavailable. Therefore, optimizing the feature extraction mode and improving the dynamic balance between structural complexity and retrieval efficiency are becoming urgent issues that must be tackled.

Thus, a model for feature extraction is established through model transferring. A joint classifier combining theories of Bayes and support vector machine (SVM) is also constructed to update the classification parameters using the strategy of relevance feedback to retrieve target images accurately and further provide a reference for the design and optimization of small-scale CBIR systems under deep learning environments.

2 STATE OF THE ART

Numerous studies on CBIR systems related to industries, such as remote sensing, medicine, commercial promotion, and construction, are available. Chang et al. [16] proposed a novel architecture, which combines segmentation and grid module, as well as K-means and K-nearest neighbor clustering algorithms, to build an effective CBIR system. However, this architecture is unsuitable for high-dimensional feature spaces. Nowaková et al. [17] presented a novel method for fuzzy medical image retrieval using vector quantization with fuzzy signatures in conjunction with fuzzy S-trees to attain not only the list of similar images but also the nature of the medical findings. However, the disadvantage of a system with this technology is heavy calculation. Zhu et al. [18] formulated a unified hashing framework to preserve visual similarities and perform semantic transfer simultaneously to address the embedded learned hash codes with limited

discriminative semantics. However, this framework fails to involve online learning. Wang et al. [19] proposed a new SVM-based relevance feedback approach using probabilistic feature and weighted kernel function to overcome the limitations due to the excessive number of questions regarding feedback iteration. In this approach, the SVM kernel function is modified dynamically according to the weight values of feedback samples. However, this approach is sensitive to parameter adjustment and kernel selection. A novel active learning approach was developed by Bhosle et al. [20] based on a random forest classifier and feature reweighting technique to tackle the problem of imbalanced training datasets. However, the operation of manual labelling would increase the time cost. Traditional K-means is sensitive to the selection of initial clustering centers. Therefore, Yin et al. [21] combined improved decision-directed algorithm with information entropy to enhance the performance of K-means, but obtaining sufficient computing power is difficult. Maeda et al. [22] suggested a method for insect image retrieval based on supervised local regression and global alignment with relevance feedback. Their approach estimates ranking scores by preserving the neighborhood structure of feature space. However, the contribution of some terms to the performance improvement remains unclear. Mustaffa et al. [23] introduced a new method for CBIR by integrating the color models with linear discriminant analysis, which not only provides effective representation for low-level features but also allows optimal linear transformation. Nevertheless, the performance of the system would be poor when dealing with nonlinear data. Pandey et al. [24] designed a semantic CBIR system that works close to human perception by utilizing the branch selection and pruning algorithms. However, this system may easily get stuck in local optima. Ye et al. [25] developed a retrieval method based on weighted distance and basic features of the convolutional neural network (CNN), which simplifies the architecture of the system and improves the retrieval performance simultaneously. However, high computing power is required for such a method. Yang et al. [26] regarded an image as a "bag" of salient regions by analyzing the characteristics of aurora and quantized each CNN feature to its nearest center for indexing and retrieval. However, this method is quite sensitive to parameter settings. Wang et al. [27] optimized the retrieval mode in the variational Bayes framework by preserving the original input information while imposing extra constraints to correct the corrupted bits and deal with the possible retrieval performance deterioration of unknown noise pattern. However, this mode depends heavily on feature extraction. Korytkowski et al. [28] provided a powerful version of the differential evolution algorithm with effective embedded mechanisms for strong exploration and preservation of the population diversity. However, such a condition is heavy maintenance required by CBIR systems. Seetharaman et al. [29] proposed a system based on a full range Gaussian Markov random field model, which enables users to fix the significance level for the test statistic at the desired level and retrieves only the required images. Nonetheless, the ambiguous presence of meta-parameter and network topology persisted. The food-domain representative databases were adopted by Ciocca et al. [30] to have robust features for food-related tasks to evaluate the performance of the proposed food retrieval system. The results demonstrate that high representativeness of the database for the food domain leads to accurate recognition and retrieval of features.

The aforementioned works, namely techniques of image retrieval with machine or deep learning, mainly focus on feature extraction, dimension reduction, similarity measure, and structural optimization in the pursuit of the balance of system, where public datasets are generally available. However, applications to small-scale datasets for a specific domain are rarely considered. Many studies have found that the models trained by deep learning have powerful transferability, and the models trained on large datasets can be reused on small- or medium-scale datasets without training from scratch. An image retrieval method with model transfer based on parameter sharing and fine-tuning is formulated in this study. In this method, the Bayes classifier is used to compress the initial image library, and then the strategy of relevance feedback is employed to update the SVM classifier parameters iteratively to bridge the semantic gap in the field of image retrieval. Moreover, a discussion on the classification and retrieval of radiology and floating debris images is conducted, which would provide experimental evidence with application to optimization of the CBIR system based on the small-scale datasets.

The main contributions of this paper comprise the following three parts. (1) A multimodality radiology image dataset (classified by body parts) and a floating debris dataset (classified by categories) are collected. (2) A deep learning framework is constructed and trained using the transfer learning technique based on the pretrained GoogLeNet network. (3) An image retrieval system with relevance feedback is implemented by utilizing the features learned from the proposed framework.

The remainder of this study is organized as follows. Section 3 establishes a model for feature transfer learning and depicts the integration mechanism of the proposed joint classifier with relevance feedback technique. Section 4 conducts a detailed assessment of the performance of the proposed method based on the two abovementioned small-scale datasets. Finally, Section 5 summarizes the study and provides relevant conclusions.

3 METHODOLOGY

The framework proposed in this study is illustrated in Fig. 1. The first part is the offline phase. In this phase, the ILSVRC image dataset containing 1000 object categories is inputted as shown in Fig. 1a. Then, a series of functions, such as convolution, pooling, and activation, is implemented to extract features from the dataset, as shown in Fig. 1b. Finally, the obtained features are transformed into predictable probabilities with the coaction of the fully connected layer and Softmax function as demonstrated in Fig. 1c. In the online phase, the last learnable layer (i.e., loss3-classifier) and the output layer of the pretrained GoogLeNet (Fig. 1d) are replaced by new layers suitable for the target dataset, as shown in Fig. 1e. Fig. 1f reveals the retrieval strategy of the proposed framework, that is, the target dataset is first adopted to train the pretrained network considering specified training options and

parameters. Afterward, the Bayes classifier is exploited to filter the target dataset to obtain the compressed but relevant library by calculating the feature similarities between the query image and the feature database. Finally, the SVM classifier with the relevance feedback strategy is utilized to retrieve the query in the relevant library, and the parameters of both classifiers are updated iteratively until the achieved accuracy holds for the query requirement.

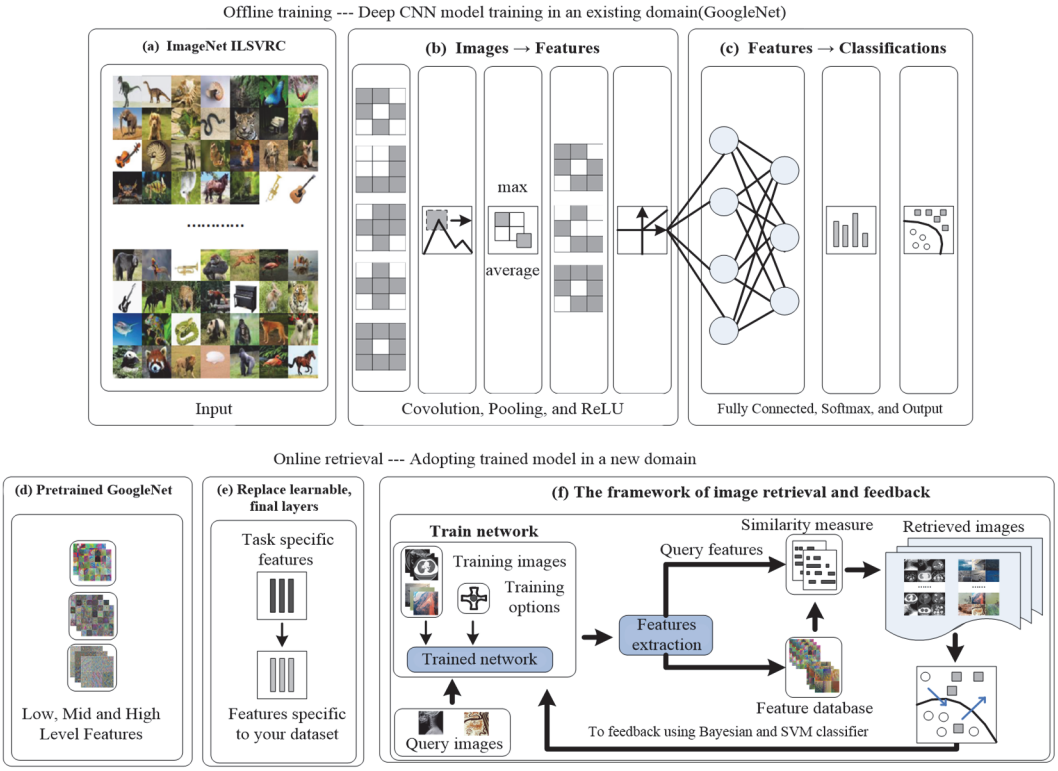


Figure 1 Overview of the proposed framework

3.1 Architecture of CNN for Feature Extraction

GoogLeNet, which is characterized by deeper layers, simpler parameter settings, lower hardware requirements, and higher performance in contrast to CNNs (such as LeNet, VGG16, and AlexNet), is considered in this study as a baseline to extract features from the target dataset.

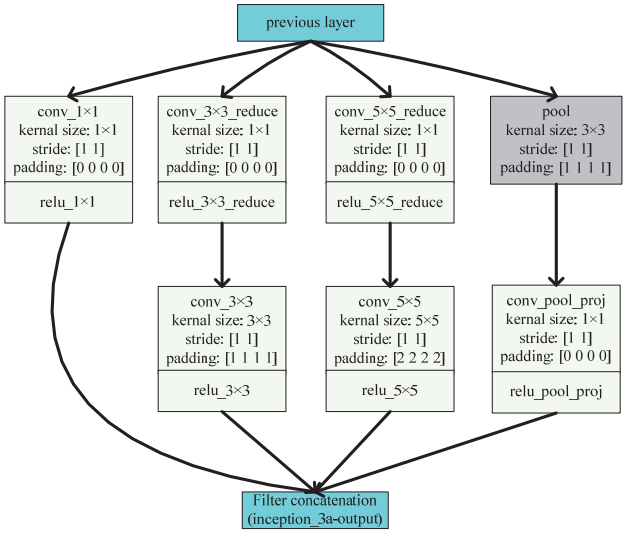


Figure 2 Illustration of Inception_3a

The GoogLeNet architecture consists of 22 layers, and part of these layers are inception modules, which are neural networks that leverage feature detection at different scales and reduce the computational budget through dimensional reduction, as shown in Fig. 2.

3.2 Bayes Classifier

Gaussian distribution is a universal probability model, which is closely associated with many events in the real world. Given a vector x in the n - dimensional space R^n follows Gaussian distribution, the probability density function of x is then expressed as:

$$P(x)=\frac{1}{(2\pi)^{d/2}\,|\sum|^{1/2}}\mathrm{e}^{-1/2(x-\varepsilon)^{\mathrm{T}}\Sigma^{-1}(x-\varepsilon)} \tag{1}$$

where $x=[x_1,\ldots,x_n]$ provides the d - dimensional eigenvector, $\varepsilon=[\varepsilon(x_1),\ldots,\varepsilon(x_n)]$ provides the d - dimensional mean vector, $\Sigma=E\{(x-\varepsilon)(x-\varepsilon)^{\mathrm{T}}\}$ is the covariance matrix with the size of $d\times d$, Σ^{-1} and $|\Sigma|$ represent the inverse and determinant of Σ, respectively.

Bayes classifier is usually in conformity with Gaussian distribution in practice. The posterior probability for each category can be calculated by training sufficient samples. The category with maximum posterior probability would be reported as the predicted class for a given example, and the discriminant function of class w_i can be defined as:

$$g_i(x)=P(x\,|\,w_i)P(w_i),\ i=1,2,\ldots,c \tag{2}$$

where probability $P(w_i)$ is a constant independent from eigenvector. Suppose the class-conditional probability density $P(x|w_i)$ satisfies d - dimensional normal distribution. Eq. (2) can then be rewritten as follows:

$$g_i(x) = \frac{P(w_i)}{(2\pi)^{d/2} |\sum_i|^{1/2}} e^{-1/2(x-\varepsilon_i)^T \sum_i^{-1}(x-\varepsilon_i)} \quad (3)$$

The natural logarithm with base-e is applied for Eq. (3). Thus, the entries that are irrelevant to the categorization task can be removed, and a revised (reduced) version of class-conditional probability density considering x belonging to w_i can then be obtained as:

$$g_i(x) = -\frac{1}{2}(x-\varepsilon_i)^T \sum_i^{-1}(x-\varepsilon_i) + \ln P(w_i) + c_i \quad (4)$$

3.3 SVM Classifier

SVM, which was created on the basis of Vapnik Chervonenkis (VC)-dimension and structural risk minimization and is a machine learning algorithm based on statistics, is expected to provide an effective generalization of new data by balancing model complexity against its learning capability with only limited samples.

Assuming that a training dataset is given in the form of $\{(x_1, y_1), (x_2, y_2), ..., (x_l, y_l)\}$, where $x \in R^n$, $y \in \{-1, 1\}$ denotes the class label to which the feature vector x belongs. $g(x) = \omega \cdot x + b$ represents the linear discriminant function under n-dimensional space. Thus, the hyperplane that divides the group of feature vectors x can be written as $\omega \cdot x + b$, and parameter $2/||\omega||$defines the margin considering the hyperplane. In most cases, the solution to the optimal hyperplane can be converted into the following quadratic programming problem:

$$\text{Min}\phi(\omega) = \frac{\|\omega\|^2}{2} \quad (5)$$

which is subject to $y_i[(\omega \cdot x_i) + b] \geq 1$, $i = 1, 2, \ldots, l$, where l is the number of samples. An optimal solution α_i^* to this problem is achieved by introducing Lagrange multiplier $a = (a_1, a_2, ..., a_l)$, where $\omega = \sum_{i-1}^{l} a_i^* y_i x_i$, and x_i is the sample that lies on the margin, which is also known as support vector. The new discriminant function is then understood to be:

$$f(x) = \sum_{i-1}^{l} a_i^* y_i (x_i, x) + b^* \quad (6)$$

By opting for some nonlinear kernel tricks, that is, $K(x, y) = (\phi(x), \phi(y))$, the original input space is transformed into a new high-dimensional feature space, in which the optimal hyperplane is also obtained. The classification function is produced as shown in Eq. (7).

$$f(x) = \sum_{i-1}^{l} a_i^* y_i K(x_i, x) + b^* \quad (7)$$

3.4 Relevance Feedback Strategy Integrating Bayes and SVM Classifier

Bayes algorithm frequently suffers from problems of small sample size and sampling asymmetry. Thus, the relevance feedback strategy integrating Bayes and SVM classifier is suggested in response to this drawback. Bayes classifier is first used to cluster the target dataset to achieve a compressed but relevant library. The SVM classifier is then employed to retrieve the query followed by rounds of parameter updating using the relevance feedback technique. Consequently, the retrieval performance may be improved due to the continuous use of parameter correction and reduction of the amount of images to be compared.

Eq. (4) shows that the classifier for class w_i has three parameters: ε_i, Σ_i, and $P(w_i)$. In some circumstances, especially for CBIR with small or medium size datasets, the number of images of a certain category is incomparable to that of feature dimensions, thus complicating the accurate estimation of Σ_{k_i}. Therefore, Σ_{k_i} can be simplified into a diagonal matrix $diag\{\sigma_{k_i}^2\}$, where $\sigma_{k_i}(m) = \Sigma_{k_i}(m, m)$. Infinite approximation of the optimal classification model is possible when the involved parameters are iteratively updated considering interactions from users. I_k denotes the query image, and $C_p = \{I_{p_1}, ..., I_{p_q}\}$ denotes the newly returned image collection of positive feedback, where q represents the number of positive images that are fed back. The parameters could be updated in accordance with Eq. (8) by merging the newly returned collection with that corresponding to the category it belongs, where n_k indicates the number of the existing positive images.

$$\begin{aligned} &(n_k + q)\sigma_{k_i}^2 = \\ &= n_k\sigma_{k_i}^2 + \frac{n_k q\varepsilon_{k_i}^2 - 2n_k\varepsilon_{k_i}\sum I_{p_i}}{n_k + q} + \sum I_{p_i}^2 - \frac{\left(\sum I_{p_i}\right)^2}{n_k + q}, \\ &\varepsilon_{k_i} = \frac{n_k \times \varepsilon_{k_i} + sum(C_p)}{n_k + q}, n_k = n_k + q \end{aligned} \quad (8)$$

The classification function of SVM (Eq. (7)) computes the distance between samples and the hyperplane, that is, the uncertainty degree of the classification result. The value of the function is close to zero (considerable uncertainty and vice versa) when the sample is close to the hyperplane. Thus, the classifier would be discriminable after features considering the reiterative learning of a specific category as long as samples with large uncertainty are returned before others and taken as part of training sets for the subsequent training process. The obtained classifier would be predictable as the number of the returned samples with large uncertainty decreases. Returning samples with large uncertainty adaptively enable retrieval systems to have high initiative. Therefore, the classification function corresponding to the sample that may be fed back should

be weighted to quantify the uncertainty degree. Consequently, samples with large uncertainty are returned preferentially during the feedback process to avoid disturbances due to other irrelevant samples.

Historical experiences typically reflect the knowledge considering the characteristics of the returned samples and the classification results learned from previous training. Comprehensively drawing upon historical experiences by considering the obtained data in the last feedback during determination of samples that must be returned is helpful. Thus, each image is assigned an initial weight to control the extent to which the historical experience could be merged, as shown in Eq. (9).

$$c(i) = (1-\beta)c(i) + f(x_i) \tag{9}$$

where $c(i)$ is the weighting function for similarity measurement (greater dissimilarity means less uncertainty), the decay coefficient β is used to adjust the degree to which historical information is preserved, and $f(x_i)$ is the current classification function.

Thus, Eq. (9) can be rewritten as shown below to maximize the use of the current retrieval result.

$$c(i) = (1-\beta)c(i) + \alpha f(x_i) \tag{10}$$

where α is a decay coefficient used to manage the proportion of the current classification result that should be adopted. Consequently, the weighting function avoids falling into the local solution by not only preserving the existing historical experience but also by using the newly generated retrieval result.

Table 1 Structure of RF-SVM algorithm

Algorithm: Improved RF-SVM

Require: Initial positive and negative training sets T_p and T_n, respectively; validation set T_v; returned positive and negative sample sets F_p and F_n, respectively; weighted base positive and negative sample sets $Fw_p(k)$ and $Fw_n(k)$, respectively; weighted up-to-date positive and negative sample sets $F'w_p(k)$ and $F'w_n(k)$, respectively, where k denotes for the order of the feedbacks. Let $c(i) = 0$ for each sample in T_v. In addition to T_p, T_n, and T_v, all other sets are equal $\{\varnothing\}$.

while stopping condition is not met **do**

Update T_p and T_n according to Eq. (11), with which the old classifie is trained to be discriminative and considered to be the current classifier.

$$T_p = T_p \cup F_p,\ T_n = T_n \cup F_n \tag{11}$$

Classify T_v with the new classifier. The indexes of the misclassified samples are recorded and appended into the returned $F'w_p(k)$ and $F'w_n(k)$. Additionally, the base sample sets $Fw_p(k)$ and $Fw_n(k)$ are renewed on the basis of Eq. (12).

$$Fw_p(k) = \left(F'w_p(k) \cup F_p\right) - F_n,\ Fw_n(k) = \left(F'w_n(k) \cup F_n\right) - F_p \tag{12}$$

Compute the uncertainty of each misclassified sample in $Fw_p(k)$ and $Fw_n(k)$ according to Eqs. (7) and Eq. (10). The first T ranked samples are recommended to be merged into F_p or F_n.

end while

Integrating the idea of active feedback [31], an improved method for relevance feedback based on SVM with "one-vs-rest" multi-class strategy (RF-SVM) is proposed. Tab. 1 shows the brief process.

Combining the aforementioned Bayes classifier and the RF-SVM algorithm, the workflow of the retrieval process of this study can be drawn as follows. 1) The first T images of the returned collection are annotated into positive and negative sample sets (no image returned at the beginning). Relevant and irrelevant classes considering the query image are updated using the sub-formula of Eq. (12), and the parameters of the Bayes classifier are refreshed in accordance with Eq. (8). Afterward, the renewed training sets according to Eq. (11) are used to retrain the updated SVM classifier. 2) The updated Bayes classifier is employed to reclassify the dataset to obtain the relevant library, to which the new SVM classifier is applied subsequently, and the first T samples with large uncertainty are then returned. 3) The above steps are repeated until the users terminate the procedure.

4 RESULT ANALYSIS AND DISCUSSION

The experiments are conducted on an HP 15 G2 Laptop with Intel Core i7 4810 MQ 2.8 GHz, 24 GB RAM under Win10 and Matlab2019a. The last learnable layer (loss3-classifier) and the final classification layer (output) of GoogLeNet are replaced with two new layers adapted to our datasets to retrain the pretrained network. The training options mini-batch-size, max-epochs and learning-rate are set as 10, 6 and $3e^{-4}$, respectively. The radial basis function is selected as the kernel of SVM, and the mean of Euclidean, Markov, Cosine, and Minkowski distance is employed as a distance measure. α and β are set as 0.3, σ_{k_i} is initialized to the identity matrix $\boldsymbol{I}$, and $\varepsilon_{k_i} = \vec{x}_{k_i}$ and $n_k = 1$.

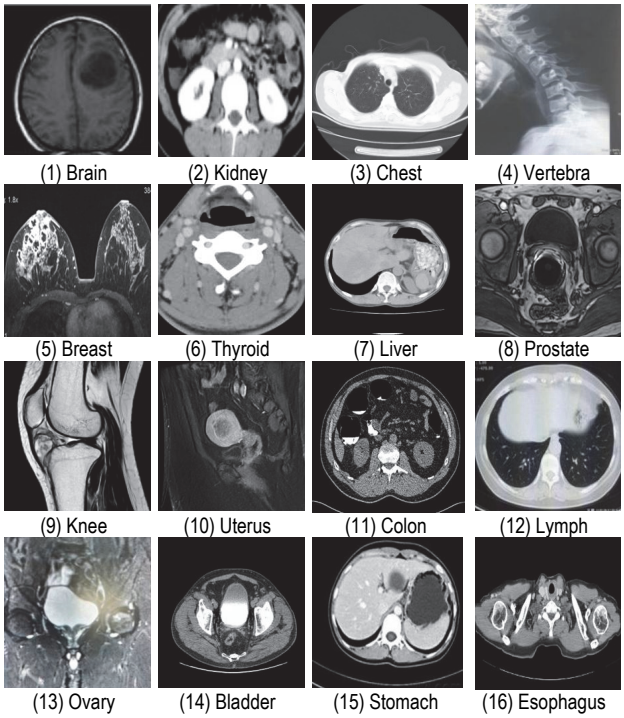


Figure 3 Examples of the radiology images

4.1 Datasets

The datasets adopted in this study are twofold. One part is the radiology image set, which is collected from the public obtainable medical database (MR, PT, CT, and PET) [32, 33]. This set is categorized into 16 classes, such as brain, chest, vertebra, and thyroid, according to the human body organ. The other part is the floating debris image set gathered during the development of this study. This set comprises eight categories, such as eutrophication, oil, wastewater, and scum. All the images are resized to 224 × 224 and converted into JPG. Part of the examples from the two datasets is illustrated in Fig. 3 and Fig. 4.

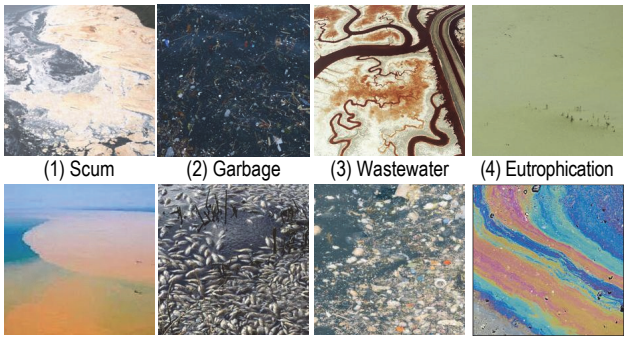


Figure 4 Examples of the floating debris images

4.2 Classification Performance

The classification performance is crucial for model evaluation. In this case, Macro Precision (*MP*), Macro Recall (*MR*), Macro Accuracy (*MA*), and *F*1-Score are selected as evaluation metrics, and the definitions of these indicators are shown in Eq. (13) to Eq. (16).

$$MP = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \tag{13}$$

$$MR = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \tag{14}$$

$$MA = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{15}$$

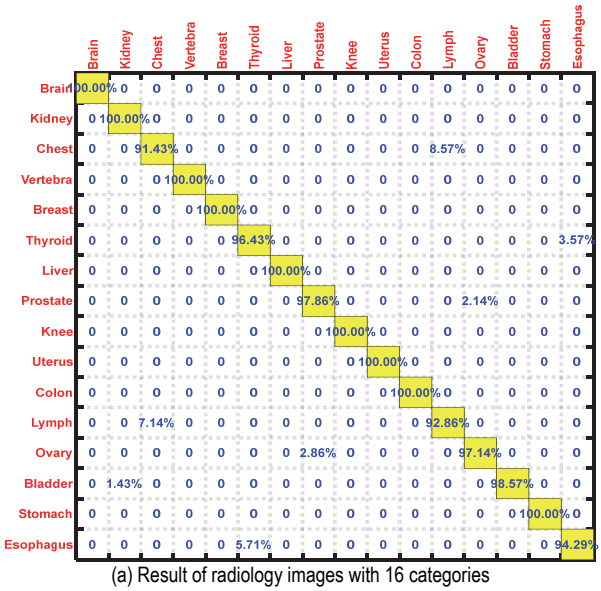
$$F1\text{-}Score = 2 \times \frac{MP \times MR}{MP + MR} \tag{16}$$

where *MP* is the average of the fraction of true positive instances among all the predicted positive ones. *MR* is the average of the fraction of the total amount of true positive instances that were predicted. *MA* indicates the average percentage of correct predictions from the model over all kinds of predictions.*F*1-Score is the harmonic mean of *MP* and *MR*. *TP* is true positive and shows the number of instances from class *C*, which are correctly predicted. *FP* is false positive and shows the number of instances of non-class *C*, which are incorrectly predicted. *TN* is true negative and denotes the number of instances that are correctly predicted not of class *C*. *FN* is false negative and denotes the number of instances from class *C*, which is incorrectly predicted, and *N* provides the total number of classes.

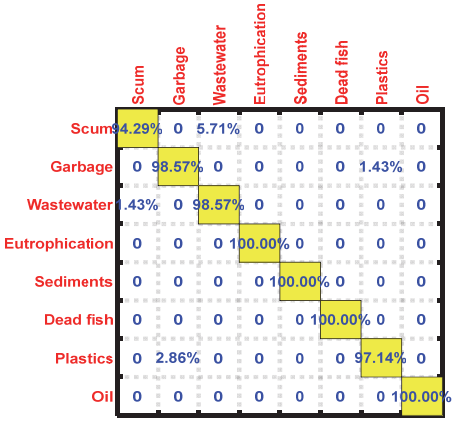
For high confidence, 10-fold cross-validation is conducted for 20 times, yielding *MP*, *MR*, and *F*1-Score of 96.53%, 95.22%, and 95.87% in experiment 1 and 98.76%, 99.54%, and 99.15% in experiment 2, respectively. Fig. 5a shows that a majority of *MAs* in experiment 1 achieve 100% except for that of class chest, thyroid, prostate, lymph, ovary, bladder, and esophagus, which are 91.43%, 96.43%, 97.86%, 92.86%, 97.14%, 98.57%, and 94.29%, respectively. In experiment 2, in addition to class foam, garbage, wastewater, and plastics, which respectively achieve 94.29%, 98.57%, 98.57%, and 97.14% in *MAs*, all other categories reach 100%, as shown in Fig. 5b. Furthermore, the comparison result between the proposed method and basic CNN (BCNN) algorithm in [34] is conducted, as shown in Tab. 2.

Table 2 Comparison of classification performance between the proposed method and BCNN in article [34]

Method	Trainingset	Testingset	Modality	View	Class	*F*1
Proposed	2240	960	MR,PT, CT,PET	Sagittal, Axial, Coronal	16	95.87%
BCNN	2340	4561	CT	Axial	12	89.8%



(a) Result of radiology images with 16 categories



(b) Result of floating debris images with 8 categories

Figure 5 Confusion matrices of categorization

Tab. 2 shows that the number of the training set used in experiment 1 is comparable to that of its counterpart, but with only approximately one-fifth of the testing set in data size. In addition to CT, images with other multimodal formats, such as MR, PT, and PET, are also conducted using the proposed method. Furthermore, the number of views and the number of image categories used by the proposed method are more than that of the BCNN. The classification accuracy of the proposed method exceeds that of its rival by 6.76% for index *F*1-Score.

4.3 Retrieval Settings

In experiment 1, 225 positive examples (from the category where the query image belongs to) and 225 negative ones (15 from each of the rest of the 15 categories) are randomly selected. Similarly, the positive and negative examples exploited in experiment 2 are 100 and 105 (15 from each of the rest of the 7 categories), respectively. The comparison with the four algorithms (transferred GoogLeNet + Bayes classifier-based method (TGCNN-Bayes, M1), transferred GoogLeNet + SVM classifier-based method (TGCNN-SVM, M2), deep CNN + joint Bayes + SVM classifier-based method (DCNN-BSVM, M3), and the semiautomatic joint Bayes + SVM classifier-based method with hand-crafted feature extraction (SA-BSVM, M4)) would help understand the efficiency of the proposed transferred GoogLeNet + joint Bayes + SVM classifier-based method (TGCNN-BSVM) presented in this study. For convenience, feedback is implemented only once for all methods, and DCNN-BSVM adopts a deep CNN architecture trained from scratch. Meanwhile, SA-BSVM uses features, such as histogram, kurtosis, skewness, and color moment, which are manually contoured and extracted. Additionally, a set of evaluation metrics, namely average recall (*AR*), average precision (*AP*), precision-recall (*PVR*), and mean average precision (*mAP*), is selected to analyze the retrieval performance of the proposed method, where the precision and recall are calculated as follows:

$$Precision = \frac{No.\ relevant\ images\ retrieved}{Total\ No.\ images\ retrieved} \times 100\% \tag{17}$$

$$Recall = \frac{No.\ relevant\ images\ retrieved}{Total\ No.\ relevant\ images} \times 100\% \tag{18}$$

4.4 Retrieval Performance

The joint classifiers used in the experiments are constructed according to the presented description in Sections. 3.2 and 3.3, while the feedback strategy is implemented as indicated in Section. 3.4. Fig. 6a, Fig. 6b, and Fig. 6c illustrate the recalls of using the proposed method to retrieve the query "Brain" from the radiology dataset at a feedback time of zero, one, and two, respectively. Fig. 6a shows that the recall without feedback reaches nearly 100% when returning 300 images. Meanwhile, the recalls with one and two feedbacks have only approximately 250 and 180 returned images to obtain the same percentages, as respectively shown in Fig. 6b and Fig. 7c. Similarly, the query "Oil"from the floating debris dataset is retrieved in experiment 2, the number of returned images is around 170, 140, and 110 when the recalls reach 100%. Samples with large uncertainty are gradually labelled for both experiments, which leads to the emergence of positive samples at places where negative ones initially exist. Consequently, the arguments of classifiers can be updated through feedback, and classifiers rapidly converge. Therefore, the number of returned images decreases when achieving the identical recall accompanied by successive feedbacks.

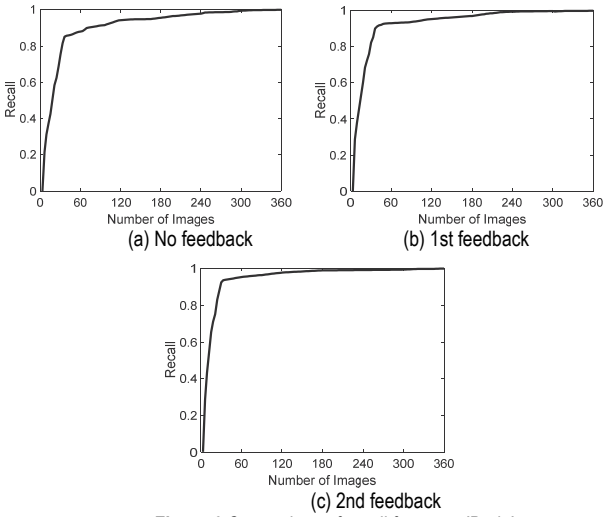


Figure 6 Comparison of recall for query 'Brain'

Off-the-shelf standard radiology and floating debris datasets are currently unavailable to benchmark the retrieval system. In many cases, samples gathered from non-standard datasets are highly resembled, which affects the reliability of retrieval performance. Thus, multiple categories are selected to diminish the possibility of uncertainty and evaluate the performance of the system in both experiments. For simplicity, the evaluation is conducted under the condition that the feedback is performed only once. Fig. 7 and Fig. 8 are the illustrations of retrieving queries "Brain" and "Oil" in experiments 1 and 2, respectively. The figures indicate that the images highlighted by red boxes denote misclassification, and the numbers in the first line represent the ranked order in which the returned images are arranged. Fig. 7 reveals that all methods can return the query image itself in the first place (query image existing in the dataset), and the 10 presented images retrieved by the proposed method are of the same category as the query. By contrast, images of different categories are returned in an irregular order by the four other methods. Among the four methods, SA-BSVM returns the irrelevant image first, DCNN-BSVM is the last one, and TGCNN-Bayes and TGCNN-SVM rank second and third, respectively. The results of experiment 2 (query image not involved) presented in Fig. 8 reveal that all five methods can return similar images. The order of precedence in which the irrelevant images are returned is similar to that of experiment 1. Specifically, irrelevant images appear at the same position, that is, No. 30, for TGCNN-SVM and TGCNN-Bayes, which partly mirrors their comparable performance. Considering an intuitive viewpoint, the retrieval performance of the proposed method exceeds that of the four other methods in both experiments.

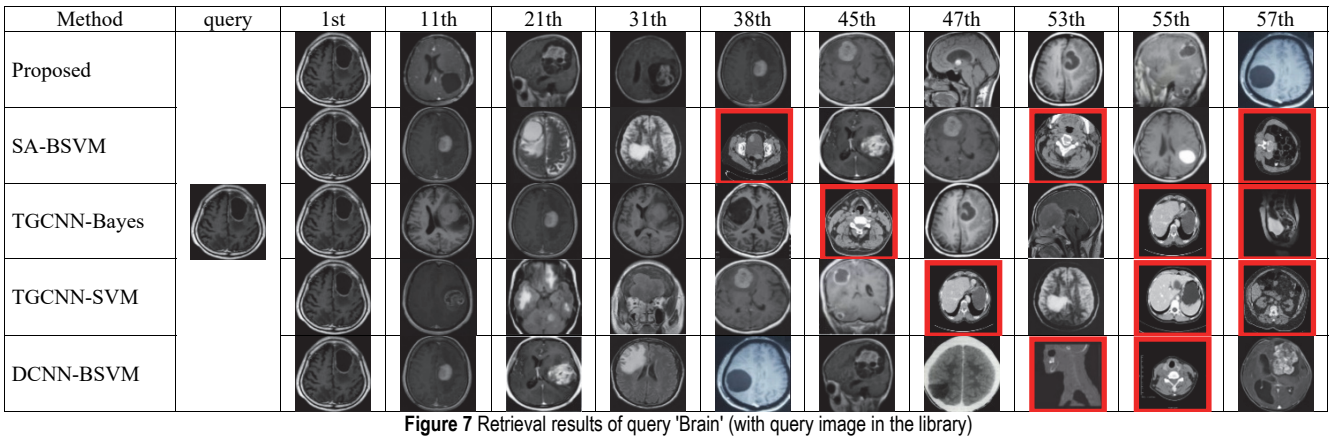


Figure 7 Retrieval results of query 'Brain' (with query image in the library)

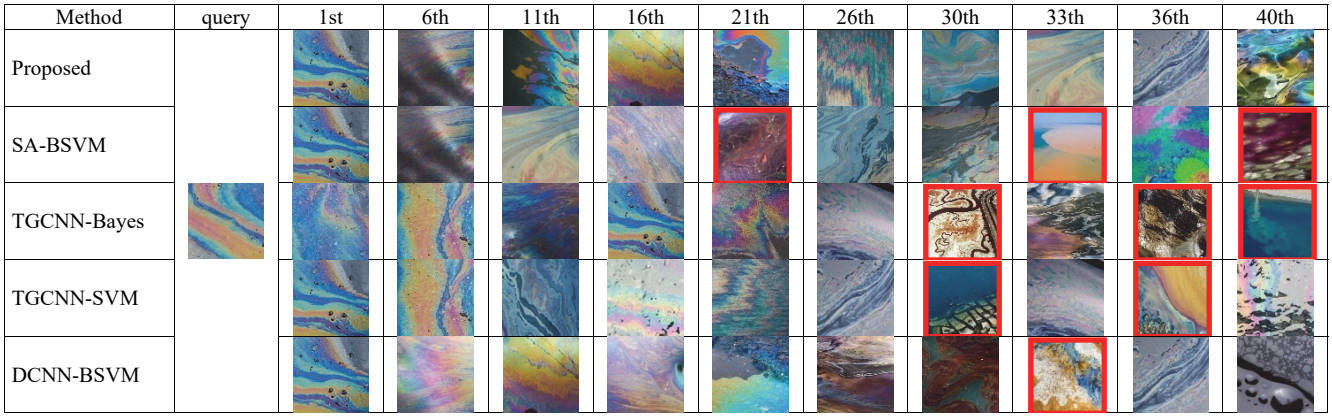
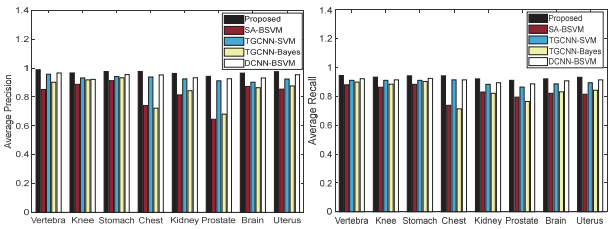


Figure 8 Retrieval results of query 'Oil' (without query image in the library)

Eight categories, such as vertebrate, knee, stomach, chest, kidney, prostate, brain, and uterus, are selected to conduct 10 retrievals to verify the effectiveness of the proposed method further. Fig. 9 charts the comparison of the *AP* and *AR* of the five aforementioned methods retrieving 50 images from a 500 randomly sampled image library. Fig. 9a shows the following results for the trends of *APs*. The *APs* of the proposed method, DCNN-BSVM, and TGCNN-SVM is relatively stable. DCNN-BSVM lies between the proposed and TGCNN-SVM. The *APs* of TGCNN-Bayes and SA-BSVM fluctuates, especially for the three categories of chest, kidney, and prostate. This finding is partially due to the adjacent anatomical positions of the three category pairs: chest–lymph, kidney-bladder, prostate-ovary. Moreover, the number of features used in SA-BSVM is substantially less than that of the four other deep learning-based methods, and the Bayes classifier alone frequently suffers from the problem of insufficient samples. Therefore, achieving such a result is reasonable for the two methods. Fig. 9b shows that the trends of *ARs* resemble that of *APs*. However, variations could be found among the eight categories due to random sampling and inconsistency of examples. Moreover, considering floating debris dataset, the closeness of visual similarities between scum and wastewater, garbage and plastics introduces a relatively low *ARs* for the four categories. Overall, the proposed method is superior to the other methods in *AP* and *AR*.

Take the radiology dataset for example. The results of 10 retrievals for the 8 categories are summarized in Tab. 3. These results indicate that the *ARs*, *APs*, and *mAPs* of the proposed method give optimal quantities when returning 25, 50, and 100 images, respectively. The overall performance is only slightly affected despite the occurrence of unstable retrievals in the categories of chest, kidney, and prostate. The retrieval accuracies in *ARs* and *mAPs* increase with the rise of the number of returned images. However, the retrieval performance for *APs* decreases as the number of returned images increases. Comprehensively, the proposed method takes the first place, while SA-BSVM is the last one. Meanwhile, DCNN-BSVM, TGCNN-SVM, and TGCNN-Bayes are ranked second, third, and fourth, respectively.



(a) Average precision (b) Average recall

Figure 9 Comparison of accuracy of 10 retrievals on radiology dataset

Table 3 Comparison of the average accuracy for different methods (the best value is in boldface)

Index	Images returned	Proposed	M1	M2	M3	M4
APs	25	**0.985**	0.872	0.946	0.973	0.854
	50	**0.968**	0.842	0.928	0.942	0.822
	100	**0.786**	0.694	0.721	0.745	0.701
ARs	25	**0.458**	0.387	0.402	0.421	0.376
	50	**0.931**	0.833	0.878	0.895	0.828
	100	**0.978**	0.890	0.925	0.965	0. 868
mAPs	25	**0.663**	0.532	0.598	0.619	0.496
	50	**0.863**	0.688	0.794	0.829	0.632
	100	**0.904**	0.767	0.841	0.875	0.715

Furthermore, Fig. 10 illustrates the detailed performance comparison between the aforementioned retrieval methods. Fig. 10a shows the trends of the *ARs* of five methods for all radiology categories. All methods can retrieve images that are semantically similar to the query as much as possible with the increase in the number of returned images. Therefore, all *AR* curves rise, but the *AR* of the proposed method remains optimal at all scales of axis X. For example, the *AR* of the proposed method reaches 93.18% when the number of the returned image is 60, while the *ARs* of other methods do not even reach 90%, which indicates that the proposed method enjoys high accuracy. Images that are semantically dissimilar to the query are retrieved continuously for all methods for *AP* with the expansion of the number of returned images. Therefore, all *AP* curves tend to decline, but that of the proposed method maintains the optimal curve, as shown in Fig. 10b. Specifically, the *AP* of the proposed method, which is 88.92%, is higher than any of the other methods when the number of the returned image is 30. Despite the slight differences between the proposed and the sub-optimal method (DCNN-BSVM), the curve of the latter tends to decline more rapidly than the former as the number of the returned images increases from 30 to 120. In addition, Fig. 10c shows the comparison of the five methods in PVR, wherein the *APs* gradually decrease as the *ARs* expands. The area under curve (AUC) enclosed by axes *X* and *Y* and the PVR curve plotted by the proposed method surpasses that of other methods; a large AUC generally implies an effective retrieval performance. The tendency of all curves stabilizes for mAP when the number of returned images reaches 60, and the quantity of the proposed method is higher than that of other approaches at sharp rising and stable parts of the curve, as shown in Fig. 10d. Similarly, the experimental results on the floating debris dataset are similar to that of experiment 1; that is, all four indexes of the proposed method achieve global optimum.

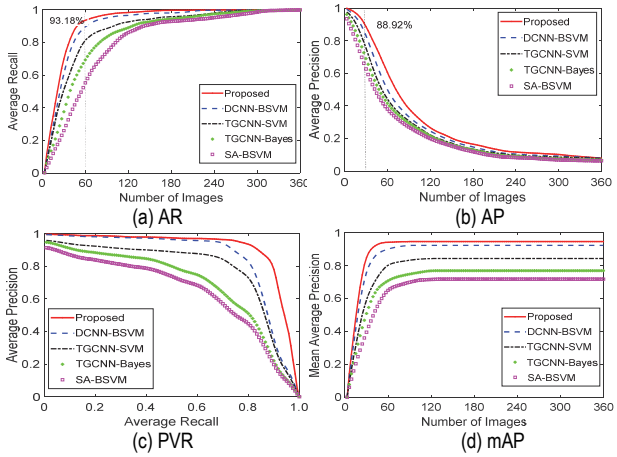


Figure 10 Comparison of average accuracy of radiology dataset with 16 categories

5 CONCLUSION

A novel retrieval model based on a joint classifier combining theories of Bayes and SVM and relevance feedback technique was formulated to optimize the dynamic trade-off between the structural complexity and retrieval performance of small- and medium-scale CBIR systems by maximizing the features derived from large-scale datasets. A case study conducted on two small-scale datasets was analyzed to compare the classification and retrieval performance results of the proposed CBIR system and traditional methods. The following conclusions could be drawn.

(1) Models developed from transfer learning would benefit from the common features between the original and target retrieval contexts. Rather than training models from scratch, the pretrained model trained on a large-scale base dataset could be repurposed to a target model if fine-tuned to a small-scale dataset. Using this approach on small-scale datasets would accomplish better performance than adopting traditional methods.

(2) Considering factors that affect retrieval accuracy, the transferred model has more scalability than that trained from scratch, deep features outperform low-level features in semantic representation, and the classification performance of composite classifiers is superior to that of independent classifiers for small-scale CBIR systems.

(3) In the absence of sufficient training samples, initialization of features and stable parameters inherited from the pre-trained model is crucial in developing the target model, especially for the first low convolutional layers. Therefore, achieving high retrieval accuracy without additional computing load would be possible. However, training from scratch remains to be the first choice for domain-related retrieval tasks with the expectation for eminent performance.

Motivated by the successful implementations of CBIR in social life, a retrieval framework of transfer learning based on a joint classifier combining theories of Bayes and SVM and relevance feedback technique was proposed and applied to small-scale datasets. Owing to the insufficiency in abundant training samples, the proposed framework attains superior performance in classification and retrieval by fine-tuning the pretrained model compared with traditional techniques. Thus, this framework may be used as a reference to the continuous study and development of CBIR systems. The conclusions from this study are not pervasive due to the insufficient training samples. The next step of this work will focus on the optimization of the proposed method with application to large-scale datasets and high computing power.

Acknowledgements

This study was supported in part by the National Natural Science Foundation of China under Grant 62176217, the Foundation of Sichuan Educational Committee under Grant 18ZA0201, the Foundation of Applied Basic Research Program of Sichuan Province under Grant 2019YJ0342, the Innovation Team Funds of China West Normal University under Grant KCXTD2022-3, and in part by Educational Reform Project of North Sichuan Medical College under Grant 21-31-095.

6 REFERENCES

[1] Meharban, M. S. & Priya, S. (2016). A review on image retrieval techniques. *Bonfring International Journal of Advances in Image Processing*, *6*(2), 7-10. https://doi.org/10.9756/BIJAIP.8136

[2] Jagtap, J. & Bhosle, N. (2021). A comprehensive survey on the reduction of the semantic gap in content-based image retrieval. *International Journal of Applied Pattern Recognition*, *6*(3), 254-271. https://doi.org/10.1504/ijapr.2021.10040334

[3] Sun, J., Ding, E., Sun, B., Chen, L., & Kerns, M. K. (2020). Image salient object detection algorithm based on adaptive multi-feature template. *DYNA-Ingeniería e Industria*, *95*(6), 646-653. https://doi.org/10.6036/9844

[4] Van, T. & Le, T. (2016). Content-based image retrieval using a signature graph and a self-organizing map. *International Journal of Applied Mathematics and Computer Science*, *26*(2), 423-438. https://doi.org/10.1515/amcs-2016-0030

[5] Vonghirandecha, P., Karnjanadecha, M., & Intajag, S. (2019). Contrast and color balance enhancement for non-uniform illumination retinal images. *Tehnički glasnik*, *13*(4), 291-296. https://doi.org/10.31803/tg-20191104185229

[6] Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., & Khalil, T. (2019). Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical Problems in Engineering*, 2019. https://doi.org/10.1155/2019/9658350

[7] Elazquez B. J. S., Cavas Martinez, F., Campuzano Brando, V. A., Alio Del Barrio, J., Fernandez Cañavate, F. J., & Alio, J. L. (2020). Automatic image processing applied to corneal endothelium cell count and shape characterization. *DYNA*, *95*(1), 170-174. https://doi.org/10.6036/9275

[8] Otto, C., Springstein, M., Anand, A., & Ewerth, R. (2020). Characterization and classification of semantic image-text relations. *International Journal of Multimedia Information Retrieval*, *9*(1), 31-45. https://doi.org/10.1007/s13735-019-00187-6

[9] Salçin, K. (2019). Detection and classification of brain tumours from MRI images using faster R-CNN. *Tehnički glasnik*, *13*(4), 337-342. https://doi.org/10.31803/tg-20190712095507

[10] Im, D. H. & Park, G. D. (2015). Linked tag: image annotation using semantic relationships between image tags. *Multimedia Tools and Applications*, *74*(7), 2273-2287. https://doi.org/10.1007/s11042-014-1855-z

[11] Sun, J., Ding, E., Li, D., Akram, A., & Kerns, M. K. (2020). Long-term Object Tracking Based on Improved Continuously Adaptive Mean Shift Algorithm. *Journal of Engineering Science & Technology Review*, *13*(5), 33-41. https://doi.org/10.25103/jestr.135.05

[12] Rawat, W. & Wang, Z. (2017). Deep convolutional neural networks for image classification: *A comprehensive review. Neural computation*, *29*(9), 2352-2449. https://doi.org/10.1162/neco_a_00990

[13] Virmani, D., Jain, N., Parikh, K., & Upadhyaya, S. (2018). Boundary Outlier Centroid Based Reduced Overlapping Image Segmentation. *Journal of Engineering Science & Technology Review*, *11*(5), 1-9. https://doi.org/10.25103/jestr.115.01

[14] Tyagi, V. (2017). Content-based image retrieval techniques: a review. *Content-Based Image Retrieval*, 29-48. https://doi.org/10.1007/978-981-10-6759-4_2

[15] Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, *32*, 20-54. https://doi.org/10.1016/j.jvcir.2015.07.012

[16] Chang, R. I., Lin, S. Y., Ho, J. M., Fann, C. W., & Wang, Y. C. (2012). A novel content based image retrieval system using k-means/knn with feature extraction. *Computer Science and Information Systems*, *9*(4), 1645-1661. https://doi.org/10.2298/CSIS120122047C

[17] Nowaková, J., Prílepok, M., & Snášel, V. (2017). Medical image retrieval using vector quantization and fuzzy S-tree. *Journal of medical systems*, *41*(2), 1-16. https://doi.org/10.1007/s10916-016-0659-2

[18] Zhu, L., Huang, Z., Li, Z., Xie, L., & Shen, H. T. (2018). Exploring auxiliary context: discrete semantic transfer hashing for scalable image retrieval. *IEEE transactions on neural networks and learning systems*, *29*(11), 5264-5276. https://doi.org/10.1109/tnnls.2018.2797248

[19] Wang, X. Y., Liang, L. L., Li, W. Y., Li, D. M., & Yang, H. Y. (2016). A new SVM-based relevance feedback image retrieval using probabilistic feature and weighted kernel function. *Journal of Visual Communication and Image Representation*, *100*(38), 256-275. https://doi.org/10.1016/j.jvcir.2016.03.008

[20] Bhosle, N. & Kokare, M. (2020). Random forest-based active learning for content-based image retrieval. *International Journal of Intelligent Information and Database Systems*, *13*(1), 72-88. https://doi.org/10.1504/IJIIDS.2020.108223

[21] Yin, C. & Zhang, S. (2017). Parallel implementing improved k-means applied for image retrieval and anomaly detection. *Multimedia Tools and Applications*, *76*(16), 16911-16927. https://doi.org/10.1007/s11042-016-3638-1

[22] Maeda, K., Genma, S., Ogawa, T., & Haseyama, M. (2020). Image Retrieval Based on Supervised Local Regression and Global Alignment with Relevance Feedback for Insect Identification. *ITE Transactions on Media Technology and Applications*, *8*(3), 140-150. https://doi.org/10.3169/mta.8.140

[23] Mustaffa, M. R., Azman, A., & Kunesegeran, G. (2017). Content-Based Image Retrieval Using Color Models and Linear Discriminant Analysis. *Advanced Science Letters*, *23*(6), 5387-5390. https://doi.org/10.1166/asl.2017.7382

[24] Pandey, S., Khanna, P., & Yokota, H. (2016). A semantics and image retrieval system for hierarchical image databases. *Information Processing & Management*, *52*(4), 571-591. https://doi.org/10.1016/j.ipm.2015.12.005

[25] Ye, F., Xiao, H., Zhao, X., Dong, M., Luo, W., & Min, W. (2018). Remote sensing image retrieval using convolutional neural network features and weighted distance. *IEEE geoscience and remote sensing letters*, *15*(10), 1535-1539. https://doi.org/10.1109/lgrs.2018.2847303

[26] Yang, X., Wang, N., Song, B., & Gao, X. (2019). BoSR: A CNN-based aurora image retrieval method. *Neural Networks*, *116*, 188-197. https://doi.org/10.1016/j.neunet.2019.04.012

[27] Wang, D., Song, G., & Tan, X. (2019). Bayesian denoising hashing for robust image retrieval. *Pattern Recognition*, *86*, 134-142. https://doi.org/10.1016/j.patcog.2018.09.006

[28] Korytkowski, M., Šenkeřík, R., Scherer, M. M., Angryk, R. A., Kordos, M., & Siwocha, A. (2020). Efficient image retrieval by fuzzy rules from boosting and metaheuristic. *Journal of Artificial Intelligence and Soft Computing Research*, *10*(1), 57-69. https://doi.org/10.2478/jaiscr-2020-0005

[29] Seetharaman, K. (2015). Image retrieval based on micro-level spatial structure features and content analysis using Full Range Gaussian Markov Random Field model. *Engineering Applications of Artificial Intelligence*, *40*, 103-116. https://doi.org/10.1016/j.engappai.2015.01.008

[30] Ciocca, G., Napoletano, P., & Schettini, R. (2018). CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding*, *176*, 70-77. https://doi.org/10.1016/j.cviu.2018.09.001

[31] Wang, X. Y., Yang, H. Y., Li, Y. W., Li, W. Y., & Chen, J. W. (2015). A new SVM-based active feedback scheme for image retrieval. *Engineering Applications of Artificial Intelligence*, *37*, 43-53. https://doi.org/10.1016/j.engappai.2014.08.012

[32] Prior, F., Smith, K., Sharma, A., Kirby, J., Tarbox, L., Clark, K., Bennett, W.,Nolan, T., & Freymann, J. (2017). The public cancer radiology imaging collections of The Cancer Imaging Archive. *Scientific data*, *4*(1), 1-7. https://doi.org/10.1038/sdata.2017.124

[33] Buendía, F., Gayoso-Cabada, J., & Sierra, J. L. (2018). From Digital Medical Collections to Radiology Training E-Learning Courses. In *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*, Salamanca, Spain. 488-494. https://doi.org/10.1145/3284179.3284262

[34] Yan, Z., Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Metaxas, D. N., & Zhou, X. S. (2015). Bodypart recognition using multi-stage deep learning. In *International conference on information processing in medical imaging*, Sabhal Mor Ostaig, Isle of Skye, UK, 449-461. https://doi.org/10.1007/978-3-319-19992-4_35

Contact information:

Siyu LAI, Master, Associate Professor
Department of Medical Imaging, North Sichuan Medical College, China
E-mail: lsy_791211@126.com

Qinghua YANG, Master, Associate Professor
Department of Medical Imaging, North Sichuan Medical College, China
E-mail: forbyyoung@126.com

Wenjin HE, Master, Associate Professor
Department of Medical Imaging, North Sichuan Medical College, China
E-mail: yaya42134@163.com

Yuanzhong ZHU, Master, Professor
Department of Medical Imaging, North Sichuan Medical College, China
E-mail: yz_zhu@126.com

Juan WANG,PhD, Professor
(Corresponding author)
1) College of Computer Science, China West Normal University, China
2) Department of Computer and Information Sciences, Temple University, USA
College of Computer Science, China West Normal University, Shida Road, Nanchong, 637002, Sichuan Province, China
E-mail: wjuan0712@cwnu.edu.cn