

Supervised Deep Semantics-Preserving Hashing for Real-Time Pulmonary Nodule Image Retrieval

Yongjun Qi^{1,2,3} · Junhua Gu^{1,2,4,6} · Yajuan Zhang^{4,6} · Gengshen Wu⁵ · Feng Wang⁴

Received: date / Accepted: date

Abstract Hashing-based medical image retrieval has drawn extensive attention recently, which aims at providing effective aided diagnosis for medical personnel. In the paper, a novel deep hashing framework is proposed in the medical image retrieval, where the processes of deep feature extraction, binary code learning, and deep hash function learning are jointly carried out in supervised fashion. Particularly, the discrete constrained objective function in the hash code learning is optimized iteratively, where the binary code can be directly solved with no need for relaxation. In the meantime, the semantic similarity is maintained by fully exploring supervision information during the discrete optimization, where the neighborhood structure of training data is preserved by applying a graph regulariza-

tion term. Additionally, to gain the fine-grained ranking of the returned medical images sharing the same Hamming distance, a novel image re-ranking scheme is proposed to refine the similarity measurement by jointly considering Euclidean distance between the real-valued feature descriptors and their category information between those images. Extensive experiments on the pulmonary nodule image dataset demonstrate that the proposed method can achieve better retrieval performance over the state-of-the-arts.

Keywords Deep Learning · Semantics-Preserving Hashing · Pulmonary Nodule · Real-Time Image Retrieval

1 Introduction

Lung cancer is one of the most fatal diseases in modern times and yields extremely high cancer morbidity and mortality [1]. Recent medical reports show that early detection is the most effective way in lung cancer treatment, which improves the patient survival rate significantly from 14% to 49% [2]. Benefiting from the rapid development of medical imaging technology, the accurate early detection/diagnosis is becoming a reality because of the massive amount of digital medical data like pulmonary CT images [3,4]. The professional medical staff can make quick and precise diagnosis regarding the lung cancer by comparing the existing cases to the similar previous ones directly. With the explosive growth on the quantity of medical images, efficient medical image retrieval becomes an urgent demand, which aims at finding similar disease cases from the gallery to the query and providing effective aided diagnosis information for medical practitioners. Earlier retrieval methods are generally performed based

Yongjun Qi
qyjun@189.cn
✉Junhua Gu
jhgu_hebut@163.com
Yajuan Zhang
zhangyajuan@scse.hebut.edu.cn
Gengshen Wu
gengshen.wu@lancaster.ac.uk
Feng Wang
wangfengidea@163.com

¹State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin, China

²Laboratory of Electromagnetic Field and Electrical Apparatus Reliability, Hebei University of Technology, Tianjin, China

³Information Technology Center, North China Institute of Aerospace Engineering, Langfang, China

⁴School of Artificial Intelligence, Hebei University of Technology, Tianjin, China

⁵School of Computing and Communications, Lancaster University, Lancaster, UK

⁶Hebei Province Key Laboratory of Big Data Calculation, Hebei University of Technology, Tianjin, China

on the floating-point level calculations on medical images, namely extracting their real-valued handcrafted features (e.g., SIFT [5]) first and computing the distances (e.g., Euclidean distance) exhaustively between them. That usually leads to low efficiency and high storage requirement, which is extremely challenging when tackling *large-scale* retrieval tasks. Consequently, hashing that represents data with compact binary code (i.e., $\{0, 1\}$ or $\{-1, 1\}$) has attracted extensive attention in conducting efficient retrieval in the large-scale datasets via bit-wise calculations on the Hamming distance [6, 7].

Specifically, the hashing algorithm can be divided into two different domains: data-independent (e.g., Locality Sensitive Hashing (LSH) [6]) and data-dependent (e.g., Spectral Hashing (SH) [8]). Particularly, the data-independent hashing builds hash function without the need for training data, whereas the data-dependent hashing (i.e., learning to hash) involves a dedicated training process that links the training data to the to-be-learned hash function closely [6]. The latter hashing methods will explore and conserve the data structure along with the hash function learning, thus achieving better performance compared to the data-dependent ones [9, 10]. Subsequently, traditional learning to hash can be further categorized as: supervised hashing (e.g., Binary Reconstructive Embedding (BRE) [11], Minimal Loss Hashing (MLH) [12], Supervised Hashing with Kernels (KSH) [13]) and unsupervised hashing (e.g., Iterative Quantization (ITQ) [14], Spectral Hashing (SH) [8] and Anchor Graph Hashing (AGH) [15]), based on using prior knowledge (e.g., labels) or not during the offline training process. This paper will focus on applying supervised hashing, which usually obtains better performance against those unsupervised ones, in the medical image retrieval. In fact, most previous medical images have been carefully labeled (diagnosed) owing to great efforts from medical personnel, which makes the supervised fashion practical in such medical-related retrieval tasks.

Without loss of generality, previous supervised methods can be adopted directly in the medical image retrieval. However, several drawbacks in the exiting frameworks make the retrieval performance less favorable. Firstly, the traditional hashing methods adopt two-step learning paradigm in the hash code learning, namely perform optimization on the real-valued features from the training data and obtain its binary code directly by thresholding (e.g., $sign(x) = 1$ if $x \geq 0$, otherwise -1). However, arbitrary relaxation is required for the floating-level optimization and the quantization error would be accumulated during such optimization, which yields suboptimal results [16, 17].

Moreover, existing methods fail to preserve the neighborhood structure in the hash code learning, which reduces the code quality in the retrieval [7]. Although some deep-based hashing methods use the class label in the deep network training to provide better feature representation before binarization, the binary codes for similar images could be totally different if directly quantizing the output vectors of deep networks with a certain threshold [18–22].

In addition, the traditional hashing methods measure the similarity between images by calculating the Hamming distance between their hash codes. Compared to the floating-point descriptors that adopt Euclidean distance as measurement, there is a high probability that many candidates sharing the same Hamming distance would be returned and they are ranked equally to a query instance. This could be acceptable when dealing with general image retrieval, However, when regarding the medical cases, the returned results are expected to be as accurate as possible to offer better aided diagnosis. A recent study shows that image re-ranking algorithm could be a solution to refine the retrieval results in this case [18].

To address the above limitations, we propose a novel supervised deep hashing method, termed Supervised Deep Semantics-Preserving Hashing. Particularly, it first trains a deep network with the class labels to provide robust deep representation for the upcoming code learning. In the binary code learning, the reconstruction loss between the original feature and its binary code is minimized iteratively. On top of that, a graph regularization term is constructed based on the supervision information and applied in the code learning so as to preserve the neighborhood relationships with the fully-utilized prior knowledge. Moreover, with an alternating optimization scheme, the binary code can be solved directly without relaxation. In the online retrieval phase, the returned images with the same Hamming distance are re-ranked based on its real-valued descriptors and class labels, thus obtaining a more accurate ranking. The proposed medical image retrieval process is shown in Fig. 1. Compared to the existing works, our contributions can be summarized as follows:

- A novel end-to-end supervised deep hashing framework is proposed to generate effective binary code in the large-scale medical image retrieval. The supervised deep architecture provides robust deep features and enables the close engagement of the deep feature learning and the upcoming code learning. Particularly, with an alternating discrete optimization scheme, the binary code can be solved directly without the need for relaxation. To make full use of prior knowledge, a graph regularization term con-

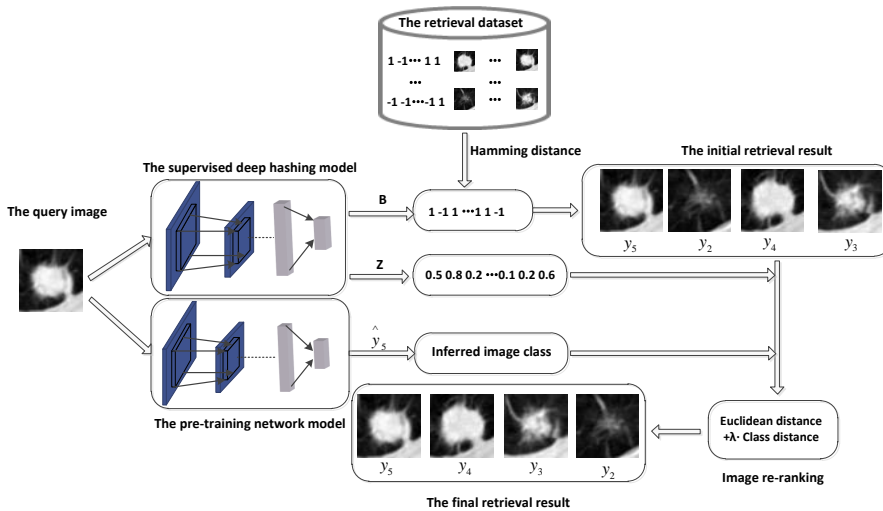


Fig. 1 The proposed medical image retrieval. The binary codes for both query and gallery data are generated by the proposed deep hashing framework. By comparing the binary codes from query and gallery, the initial retrieval result is obtained based on the calculated Hamming distances in ascending order. Then the images with the same Hamming distance are re-ranked by the Euclidean distance between image features from the latent layer with the class distance as a regularization term.

structured on the supervision information is further applied to preserve the neighborhood structure during the code learning, thus improving the retrieval performance with better code quality.

- A novel image re-ranking scheme is proposed to refine the retrieval results sharing the same Hamming distance. Specifically, the Euclidean distance between the real-valued feature descriptors and the category information are jointly taken into account when refining the similarity measurement between those returned medical images, thus providing better ranking results to assist the diagnosis.
- Extensive experiments on public medical dataset show that the proposed method outperforms the state-of-the-arts significantly in terms of retrieval accuracy.

The rest of this paper is organized as follows. In Section 2, we briefly present and discuss some related works on supervised hashing and image re-ranking. Then detailed descriptions of our method are provided in section 3. In Section 4, we report experimental results from the proposed method and some baselines to consolidate the contributions of our work. Finally, we draw the conclusion of this paper in Section 5.

2 Related Works

2.1 Supervised Hashing

Supervised hashing methods can fully take advantage of the supervision information (e.g., labels), thus obtaining higher retrieval accuracy than the unsupervised methods usually. Some representative supervised methods are briefly discussed. For instance, Binary Reconstructive Embedding (BRE) is proposed in [11], where hash functions are learned by minimizing the reconstruction loss between the original features and the binary codes. However, the pairwise similarity between data samples is required in BRE, which is storage-expensive on large-scale datasets. Minimal Loss Hashing (MLH) [12] is proposed based on latent structural SVM [23], which builds hash functions via optimizing the upper bound of a hinge-like loss. It turns out to be more efficient than BRE on the large-scale dataset. Subsequently, Latent Factor Hashing (LFH) is presented in [24] to learn similarity-preserved binary codes, where the stochastic learning method is used to further improve the training efficiency.

To deal with linearly inseparable data, nonlinear hash functions with the powerful generalization capability have attracted much attention recently. For example, Kernel-based Supervised Hashing (KSH) [13] introduces nonlinearity in learning hash functions, which are built in their kernel spaces. Two-Step Hashing (TSH) method is proposed in [25] that unifies hash function

and hash code learning in an effective and flexible hashing framework, which yields better retrieval performance.

In recent years, the success of convolutional neural network(CNN) in many visual recognition tasks has inspired the supervised deep hashing algorithms. Particularly, CNNH and CNNH+ [26] decompose the pairwise similarity matrices into approximate binary codes based on class labels, and then CNN is trained with the supervisory information to build hash functions. However, the learning processes of binary code and deep feature, in this case, is a typically one-way interaction, which inevitably produces sub-optimal solutions because of large quantization errors. To tackle this limitation, several end-to-end supervised hashing frameworks are proposed in [18,21,20], where the image representation learning and the binary code optimization are jointly combined, thus improving the retrieval performance significantly.

Supervised deep hashing methods are also widely used in medical image retrieval due to its fast query speed and low memory cost. A supervised hashing framework is developed in [27] with kernel learning for histopathological image analysis. Consequently, a supervised deep framework is constructed in [28] to extract the medical signs of lung nodules, and the hash code is learned after reducing the dimensionality with PCA. In [29], the siamese network is utilized to learn the hash code with pairwise inputs, while a multiple instance deep hashing is proposed in [30] for large-scale retrieval with weak bag-level supervision as guidance.

However, these methods adopt relaxation strategies by discarding the discrete constraints when learning the hash functions, which causes large quantization errors. Moreover, the class labels cannot retain the neighborhood relationships between images, which will affect the retrieval performance. To alleviate this issue, the pairwise or triplet images are fed into the learning model simultaneously in the training stage, such as [20–22,31]. However, the process of input construction is computationally expensive and storage extensive.

2.2 Image Re-Ranking

Compared to the image retrieval that adopts real-valued level distance calculation, it is more likely to return many candidates with the same Hamming distance to a query when using binary codes [18]. When regarding large-scale retrieval, there is a higher probability that hundreds of images are ranked equally and returned in the result lists. It is very important to have a finer-grained ranking of those images sharing the same Hamming distance to offer a more accurate aided diagnosis. Earlier work on image re-ranking as [18] uses

Euclidean distance as similarity measure, then the images are re-ranked according to the refined distance. In [32], a weighted-distance based re-ranking termed Bit Importance based Re-ranking (BIR) is proposed, where the bit importance is considered and used to refine the distance by re-ranking. High weights are assigned to important bits and vice versa. A re-ranking method for remote sensing image retrieval is presented in [33], where the images are re-ranked via calculating two different image-to-class distances: image-to-training-class distance and image-to-query-class distances. In fact, the classes of this method does not participate in distance calculation, but samples several images from different classes in the training set and query set to calculate the Euclidean distance from the query image. In general, the distance function of features cannot accurately measure the similarity between two images, especially for medical images with rich and complex visual content.

In our proposed re-ranking strategy, the Euclidean distance and supervision information is combined in refining the similarity measurement between the returned medical images sharing the same Hamming distance, thus leading to better ranking performance.

3 Methodology

In this section, we will present the proposed supervised deep hashing framework in details.

3.1 Framework Overview

The work flowchart of the supervised hashing framework is shown in Fig. 2. Firstly, a backbone of the pre-trained VGG-16 [34], which includes a total of thirteen convolutional layers, five maximum pooling layers, and three fully-connected (*fc*) layers, is finetuned on the medical image dataset to provide robust features for the ongoing retrieval task. Particularly, we modify the neuron number of the last two *fc* layers, FC7¹ and the output layer, to 1024 and the length of the binary codes *m* in this step respectively, $m \in \{16, 32, 48, 64\}$. Then the deep features of training data are extracted from FC7 layer of the pre-trained deep network and sent to the binary code learning model.

In the code learning phase, the binary code is learned by minimizing the reconstruction loss against the original image feature, while the neighborhood structure of

¹ It is renamed as 'latent layer' in this paper for more intuitive understanding.

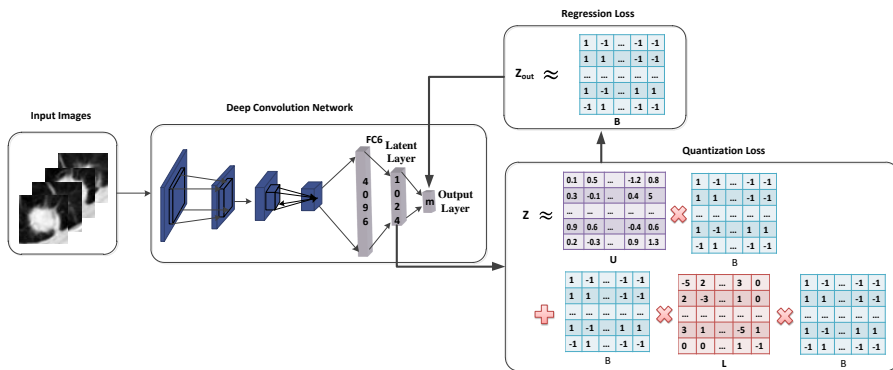


Fig. 2 Supervised deep hashing framework. There are three components in the training process: deep feature learning, binary code learning, and deep hash function learning. The latent layer generates the deep features for the upcoming code learning process. In the hash code learning, the binary code is learned by approximating the original feature with a projection matrix, while the neighborhood structure of the original manifold is preserved by applying graph regularization. The deep hash function is built with the learned binary code as the guidance.

original data is also preserved by applying graph regularization to the binary codes. With an alternating optimization, the binary code can be solved directly with no need for relaxation, thus improving the code quality.

In the hash function learning stage, the learned binary code is used as a pseudo label to re-train the deep network, thus generating the binary representations for the out-of-sample (i.e., non-training) data.

Finally, the retrieval result can be obtained by sorting the Hamming distances between the query and gallery set in ascending order. While the returned images sharing the same Hamming distance are re-ranked by jointly considering the Euclidean distance between image features and the class distance to obtain better ranking results.

In the next subsections, each of the above modules is detailed to provide further insights into the proposed hashing framework.

3.2 Supervised Deep Feature Learning

To provide more discriminative features for the subsequent image retrieval process, supervised pre-training on the training data is performed via minimizing the classification loss, as shown in Fig. 3. In this case, the output number of the last fc layer is set to 6^2 and then a cross-entropy loss is applied in the classification process. By doing so, the finetuned deep model can be more compatible in representing the medical images with robust deep features.

Without loss of generality, for an input image x , we denote $p(i)$ as the true probability distribution (i.e., label) and $q(i)$ as the predictive probability distribution for the i -th ($i \in [1, 6]$) neuron from the output layer of the deep model, the cross-entropy loss is defined as below:

$$\sum_{i=1}^6 p(i) \cdot \log\left(\frac{1}{q(i)}\right) = -\sum_{i=1}^6 p(i) \cdot \log(q(i)). \quad (1)$$

3.3 Binary Code learning and Discrete Optimization

After the supervised pre-training process, the deep feature Z are extracted from the latent layer and fed into the binary code learning, as shown in Fig. 2. Particularly, the objective function of this process consists of two loss terms: reconstruction loss and graph loss. Following previous works, Z can be approximated simply by the product of the projection matrix U and its binary code B , namely $Z \approx UB$, which is formulated as the first loss term below:

$$\min_{U, B} \|Z - UB\|_F^2, \quad (2)$$

where $Z \in \mathbb{R}^{1024 \times n}$, n represents the number of images. $U \in \mathbb{R}^{1024 \times m}$, m denotes the binary code length and $m \in \{16, 32, 48, 64\}$, $B \in \{-1, 1\}^{m \times n}$.

Moreover, the unsupervised graph learning is introduced and used as a regularization term in the objective function so as to restrict the neighboring relationship from the original manifold in the binary code generation [35]. Specifically, the graph regularity can be formulated the following [35, 8]:

² In the training set, pulmonary nodules are classified into six categories according to their malignant degree, See Section 4.

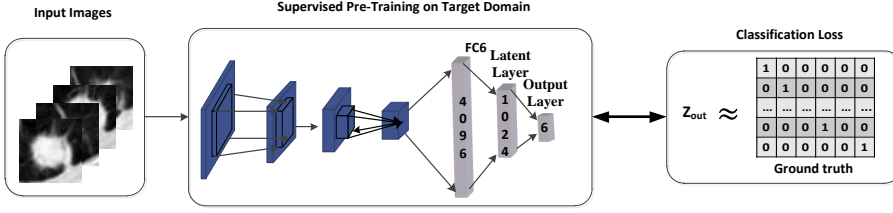


Fig. 3 Pre-trained model. The output layer consists of six hidden nodes (six classes in the dataset) and produces the probability distribution for each input. Cross-entropy loss is applied when finetuning the deep networks on the target dataset.

$$\begin{aligned}
\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|b_i - b_j\|_F^2 S_{ij} &= \sum_{i=1}^n b_i^T b_j D_{ii} - \sum_{i,j} b_i^T b_j S_{ij} \\
&= \text{Tr}(B^T D B) - \text{Tr}(B^T S B) \\
&= \text{Tr}(B L B^T), \tag{3}
\end{aligned}$$

where $\text{Tr}(\cdot)$ is the trace of a matrix. $S \in \mathbb{R}^{n \times n}$ denotes the similarity matrix that could be easily constructed by manual scoring [21, 20, 36]. In this work, if the labels of the i -th and j -th images are the same, then $s_{ij} = 1$ and otherwise $s_{ij} = 0$. D is a diagonal matrix, where the entries are the sum of columns from S . L denotes the Laplacian matrix, where $L = D - S$.

Combining Eq. (2) and (3), the overall objective function of the binary code learning is formulated as below:

$$\min_{U, B} \|Z - UB\|_F^2 + \beta \text{Tr}(B L B^T), \tag{4}$$

where β is used to balance the two loss terms.

However, the above objective function is an NP-hard problem and cannot be solved directly because of the applied binary constraints. Subsequently, an alternating optimization method is proposed to tackle this problem, where each variable is optimized iteratively by fixing other parameters. The step-by-step optimization is illustrated as below.

1. **Fixing other parameters and update U .** By fixing other parameters, Eq. (4) with respect to U can be rewritten as follow:

$$\min_U \|Z - UB\|_F^2, \text{ s.t., } B \in \{-1, 1\}^{m \times n}. \tag{5}$$

The above equation can be further expanded as:

$$\begin{aligned}
&\min_U \text{Tr}((Z - UB)^T (Z - UB)) \\
&= \min_U \text{Tr}(Z^T Z - Z^T U B - B^T U^T Z + B^T U^T U B) \\
&= \min_U \text{Tr}(Z^T Z) - 2\text{Tr}(Z^T U B) + \text{Tr}(B^T U^T U B) \\
&= \min_U \text{Tr}(Z^T Z) - 2\text{Tr}(Z^T U B) + \text{Tr}(B B^T U^T U). \tag{6}
\end{aligned}$$

According to the matrix properties in [37], which are:

$$\frac{\partial}{\partial X} \text{Tr}(A X B) = A^T B^T, \tag{7}$$

$$\frac{\partial}{\partial X} \text{Tr}(B X^T X) = X B^T + X B, \tag{8}$$

we can calculate the derivation of Eq. (6) with respect to U . By setting the derivative as 0, the optimal solution of U can be obtained:

$$U = Z B^T (B B^T)^{-1}. \tag{9}$$

2. **Fixing other parameters and update B .** The Eq. (4) can be expanded out and rewritten with only B as the target variable:

$$\begin{aligned}
&\min_B (\text{Tr}(B^T U^T U B) - 2B^T U^T Z) + \beta \text{Tr}(B L B^T) \\
&= \min_B \text{Tr}(B^T U^T U B) - 2\text{Tr}(B^T U^T Z) + \beta \text{Tr}(B L B^T) \\
&= \min_B \|G^T B\|_F^2 - 2\text{Tr}(B^T Q) + \text{Tr}(B K B^T), \tag{10}
\end{aligned}$$

where $G = U^T$, $Q = U^T Z$ and $K = \beta L$. Then $B \in \{-1, 1\}^{m \times n}$ can be solved bit by bit by following [38]. Particularly, let b^T , g^T , q^T be the j -th row of B , G , Q respectively, and B' , G' , Q' be the matrix of B , G , Q excluding b , g , q respectively. Eq. (10) can be written as follows:

$$\begin{aligned}
&\min_b (\text{Tr}(B^T G G^T B) - 2\text{Tr}(B^T Q) + \text{Tr}(B^T B K)) \\
&= \min_b \text{Tr} \left(\begin{pmatrix} B' \\ b^T \end{pmatrix}^T \begin{pmatrix} G' \\ g^T \end{pmatrix} \begin{pmatrix} G' \\ g^T \end{pmatrix}^T \begin{pmatrix} B' \\ b^T \end{pmatrix} \right) \\
&\quad - 2\text{Tr} \left(\begin{pmatrix} B' \\ b^T \end{pmatrix}^T \begin{pmatrix} Q' \\ q^T \end{pmatrix} \right) + \text{Tr} \left(\begin{pmatrix} B' \\ b^T \end{pmatrix}^T \begin{pmatrix} B' \\ b^T \end{pmatrix} K \right) \\
&= \min_b \text{Tr} ((B'^T G' + b g^T) (G'^T B' + g b^T)) \\
&\quad - 2\text{Tr}(B'^T Q' + b q^T) + \text{Tr}((B'^T B' + b b^T) K) \\
&= \min_b g^T G'^T B' b - q^T b + b^T K b \\
&= \min_b p^T b + b^T K b, \tag{11}
\end{aligned}$$

where $p = B'^T G' g - q$. As can be seen, Eq. (11) is a Binary Quadratic Programming (BQP) problem and it can be solved by flipping each entry of b

sequentially by following traditional coordinate descent based optimization methods [38, 7]. By optimizing Eq.(9), Eq.(10) and Eq. (11) after T iterations, the binary code for the training data can be obtained.

3.4 Deep Hash Function Learning

With the learned binary code B , it can be used as the guidance in training the deep hash function. Particularly, the Euclidean loss between the output of the deep network and the corresponding binary code is minimized, which can be formulated as follow:

$$\min_{\Theta} \|H(X; \Theta) - B\|_F^2, s.t., B \in \{-1, 1\}^{m \times n}, \quad (12)$$

where X denotes the original input image. $H(X; \Theta)$ denotes the deep hash function and Θ denotes the deep network parameters including weights and biases. The deep hash function can be built via deep network training, where Θ is optimized iteratively in a batch-wise fashion until convergence [7]. Finally, the whole learning process is summarized in Algorithm 1.

Algorithm 1 Supervised Deep Semantics-Preserving Hashing

Input: Input image feature X , binary code length m , affinity matrix S , randomly initialized parameters B , U , Θ , Hyper-parameters β , maximum iteration number T

Output: parameters Θ

#Binary code learning

Extract the features Z from latent layer of the deep network;

for $i=1$ to T **do**

 Fix all other variables and update U by Eq.(9);

 Fix all other variables and update B from Eq.(10) to Eq.(11);

end for

#Hash function learning

Update Θ by Eq. (12);

return Θ

3.5 Image Re-Ranking

As discussed in previous sections, many candidates sharing the same Hamming distance could be returned to the query instance in the performance evaluation of binary code. Regarding this issue, the popular solution is to use Euclidean distance in the similarity measurement as [18], where the images are re-ranked based on the refined distances. In this paper, the dataset is classified according to the malignant degree of pulmonary

nodules, so the distance between different classes is different, and the distance between adjacent classes is the closest. For example, the first class (i.e. non-nodules) and the second class (i.e., highly unlikely for cancer) should be much closer than the sixth class (i.e., highly suspicious for cancer) in the feature space. That is also applied in our re-ranking method as the backbone. Moreover, to make the full use of the prior knowledge, the category information (i.e., label) is also adopted as an extra regularization term in refining the similarity. In another word, our method re-ranks the images by refining the similarity between images from two aspects: visual similarity (i.e., the Euclidean distance between features) and semantic similarity (i.e., the category information), jointly. The refined similarity S_r is defined as below:

$$S_r = \frac{1}{1 + \|x^q - x^g\|_2} + \lambda \frac{1}{1 + |\hat{y} - y|}, \quad (13)$$

where x^q and x^g represent the feature vector of query instance and a candidate from the gallery set, respectively. λ is a balance parameter. y denotes the true class of the query instance in the pulmonary nodule dataset, where $y \in \{1, 2, 3, 4, 5, 6\}$. \hat{y} denotes the predictive class of the query instance inferred from the pre-trained deep network and $\hat{y} \in \{1, 2, 3, 4, 5, 6\}$. $\|x^q - x^g\|_2$ calculates the visual distance between image features and $|\hat{y} - y|$ measures the distance between image classes. Both of the terms in Eq. (13) are normalized within the range of $(0, 1]$ such that the values can be processed under the same order of magnitude. At the same time, λ is used to adjust the contribution of semantic similarity in computing S_r . In the re-ranking phase, S_r between a query instance and its returned candidates sharing the same Hamming distance are calculated sequentially and then they are re-ranked in descending order of S_r . The larger S_r indicates higher similarity. By doing so, better ranking results (see Section 4) can be obtained for medical personnel to make a more accurate diagnosis.

4 Experiments

In this section, we provide detailed descriptions on the experiments, including the dataset, implementation details and comparison results with several baselines.

4.1 Data Set

Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [39] is adopted in our experiments to make the performance evaluation. It is a publicly available lung images dataset that contains

1,018 medical research cases, where each of them includes the CT images and their associated XML files recording annotation information. The dataset includes a total of 244,527 CT slices, each of which is 512×512 pixels in size. Some original CT images are provided in Figure 4.

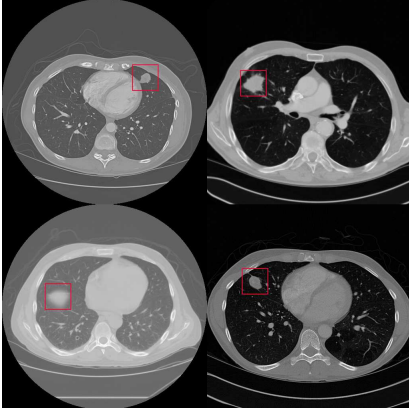


Fig. 4 Examples of CT image on lung cancer in the dataset. There are four unprocessed different CT images. The red rectangles identify the locations of pulmonary nodules, which are cut and placed in the upper corners of each image. Image block size of pulmonary nodules are 64×64 .

In addition, a total of 36,378 nodules were labeled including 11,509 nodules with diameter $< 3mm$ and 19,004 non-nodule. The statistics on the diameter distribution of pulmonary nodules are shown in Figure 5

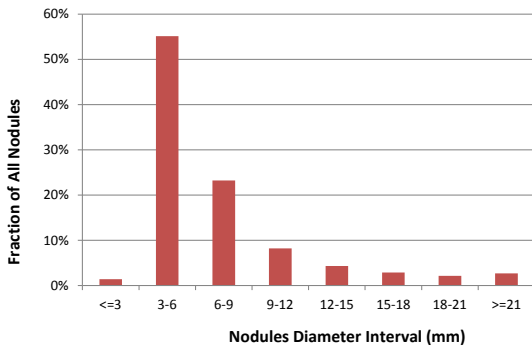


Fig. 5 Statistics on the diameter distribution of pulmonary nodules.

In LIDC-IDRI, pulmonary nodules are labeled independently by four authoritative radiologists in two stages, which are further divided into six categories: *non-nodules*, *highly unlikely for cancer*, *moderately unlikely for cancer*, *indeterminate likelihood*, *moderately suspicious for cancer*, and *highly suspicious for cancer*. However, the original CT images cannot be applied directly in the experiments. There are a lot of noises in

the dataset that should be removed to facilitate the upcoming tasks. More importantly, most of the regions (i.e., non-nodule area) in the CT images are not useful, which may disturb the retrieval performance on pulmonary nodules to some extent. To make the best use of the data, some preprocessing actions are carried out to preserve the areas of pulmonary nodule within CT images to reduce the potential risks in the following retrieval tasks. The diameter of most pulmonary nodules is smaller than 60 mm in LIDC-IDRI [39]. By fixing the size of each pulmonary nodule image as 64×64 , it enables to cover most areas of all pulmonary nodule and greatly improve the proportion of useful information in CT images. Without loss of generality, pulmonary nodules can be obtained by advanced object detection algorithms, such as Faster-RCNN [40], FPN [41], and DSOD [42], etc., where the regions of interest are detected and ready to be used for the upcoming retrieval tasks.

Then data augmentation is used for the categories with fewer images to keep the balance of image number in each category. Finally, a total number of 43,038 images are included in the pulmonary nodules dataset. We selected 30,000 as training data, while the remaining 13,038 is left as testing data. 50 images are randomly selected from the testing data as query instances and the rest of testing data forms the gallery in the online retrieval.

4.2 Implementation Details

In the experiment, several classical hashing methods, such as SSDH [43], DSH [31], KSH [13] and CCA-ITQ [14], are adopted as baselines in the evaluations of retrieval performance. We adopt mean Average Precision at the top K ($mAP@K$) and Precision-Recall (PR) curve as the main performance metrics as in [44, 18]. Particularly, the mAP is used to evaluate the overall retrieval performance of the hashing algorithm, which is defined as below:

$$mAP = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \left(\frac{1}{k} \sum_{j=1}^k Rel(i, j) \right), \quad (14)$$

where Q represents the query instance set, k represents the number of retrieved images related to the query instance. $Rel(i, j)$ represents the precision of the top j -th retrieved image of i -th query instance and $Rel(i, j) \in \{0, 1\}$, where $Rel(\cdot)$ is 1 if a query instance and the retrieved image share the same label and otherwise 0. Moreover, P-R curve shows the relationship between the retrieval precision and the recall rate.

To investigate the effectiveness of the proposed image re-ranking strategy, two baseline methods: Lin et al. [18] and BIR [32], are adopted in the performance evaluation against our method. Specifically, the image re-ranking methods are evaluated by using Precision@ K , which is defined as follow:

$$Precision@K = \frac{\sum_{i=1}^K Rel(i)}{K}. \quad (15)$$

Four different bit sizes: 16, 32, 48 and 64, are implemented for all those hashing methods. The hyper-parameters T is set to 10 in updating the binary codes. β and λ are both set to 1, respectively. The experiments are carried out on the pulmonary nodules dataset. The algorithm is executed under Tensorflow³, which is an open-source python-based framework for machine learning. We adopt Adam optimizer in the network training with the basic learning rate of 10^{-3} , The batch size is set as 50 and the maximum iteration is set to 150. All weights are initialized from a zero-centered normal distribution with a standard deviation of 0.02. For the fair comparison, identical train and testing sets are used in the performance evaluation for all state-of-the-arts. The server configurations are: Intel Xeon E5-2683v4 CPU, 128GB RAM, and two NVIDIA Quadro P5000 GPUs.

4.3 Comparisons with The State-of-The-Arts

On the pulmonary nodule dataset, our method is compared with those baselines extensively under a given retrieval task. Three different forms of retrieval results are provided in the following subsections.

1. **mAP@ K .** Fig. 6 shows the mAP@ K results under different binary code lengths. As can be seen, our method has achieved better performance in all settings against that from baselines. Particularly, the mAP@ K results from the most competitive SSDH are about 2% in average lower than the proposed method. Moreover, slight improvements have been achieved on mAP@ K with the increasing code length, where the best performance is given at the code length of 64.
2. **P-R curves.** For the comprehensive evaluation, Fig. 7 shows the Precision-Recall curves of all methods at 64 bit on the dataset. As observed from the figure, our method has achieved the best performance again over all the baselines, which are consistent with the mAP@ K results.
3. **Top-5 retrieved images.** Furthermore, Fig. 8 shows the top-5 returned images obtained by all the hashing methods at the code length of 64. The red rectangles indicate incorrect retrieval results. Given one query image, five similar images for each of them can be correctly returned by the proposed method, while other baselines make several mistakes in those results. Compared to those baselines, the main reason of the superior performance by our method is that the learning paradigm learns the binary code directly without relaxation, thus improving the code quality dramatically by avoiding quantization errors. In the meanwhile, the similarity between the binary code data is preserved to the maximum extent in the Hamming space via using supervision information properly.

4.4 Investigation on Image Re-Ranking

More experiments are conducted in this section to evaluate the performance of the proposed re-ranking strategy. Two baselines, Lin et al. [18] and BIR [32] are included in the evaluation. Fig. 9 shows the top 10 retrieved images with the same Hamming distance by using different re-ranking methods. In the experiment, there are more than 10 candidates with the same Hamming distance are returned. Only the top 10 images are provided in Fig. 9. Images in the red rectangle belong to different classes against the query. As can be seen, our method achieves the lowest error rate in the top-10 retrieved results compared to the baselines after re-ranking, which indicates its effectiveness.

Fig. 10 shows the Precision@ K results from all methods on the dataset, where our ranking method achieves the best performance. Particularly, the precision@1000 from our method is at least 10% higher than Lin et al., which demonstrates the superiority of the proposed re-ranking strategy. Compared to those two baselines, the performance gain on our method lies on two aspects: 1) the proposed binary code learning enables to generate more effective hash code in the retrieval process, which provides highly competitive initial results; 2) the proposed image re-ranking scheme is carried out on two aspects simultaneously: the real-valued distance of image feature and the category distance, which is more advanced than Lin et al. and BIR. In [18], only the Euclidean distance is involved in refining the distance between the query and gallery set, while the category distance is applied in our method to provide more reference information. BIR assigns different weights to each bit in the binary code based on the bit importance. However, the weights rely on the quality of the learned binary code, which could be less discriminative.

³ <https://www.tensorflow.org/>

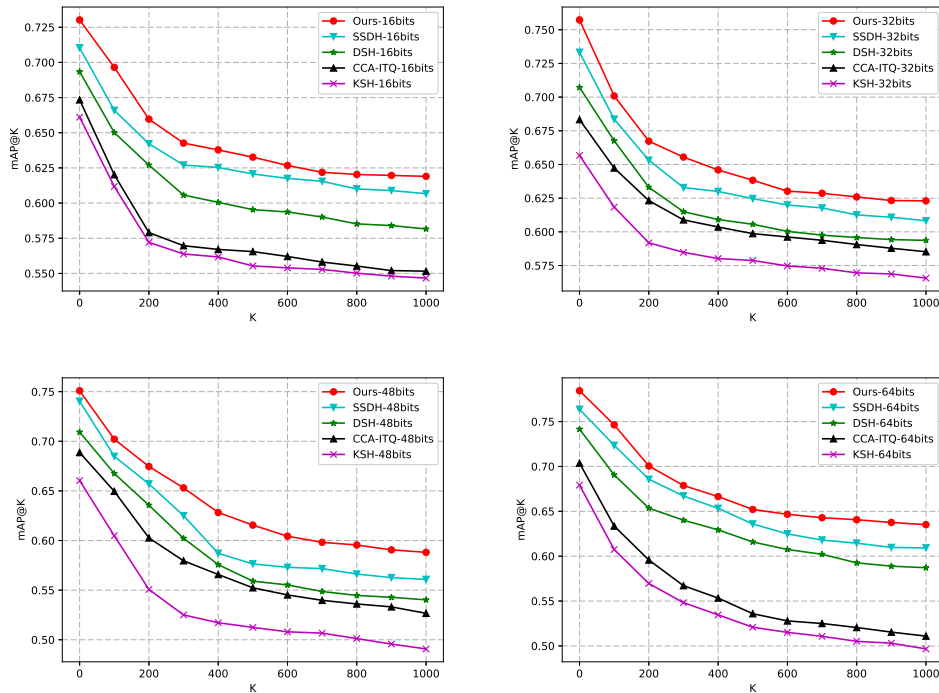


Fig. 6 The mAP@K curves using various hashing methods at different code lengths.

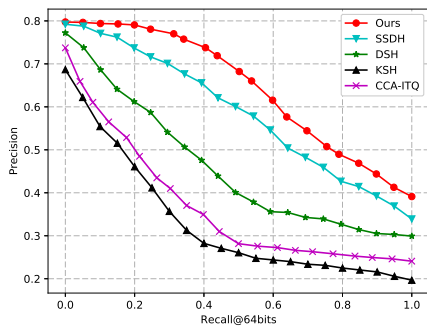


Fig. 7 Precision-Recall curves using various hashing methods at the code lengths of 64 bits on the pulmonary nodules dataset.

5 Conclusions

In the paper, a novel end-to-end supervised deep hashing method has been proposed in the medical image retrieval. Firstly, the discrete constrained objective function in the hash code learning is optimized such that the binary code can be directly solved with no need for relaxation. Moreover, the semantic similarity is maintained by fully utilizing the supervision information, where the neighborhood structure of training data is preserved by applying a graph regularization term during the discrete optimization. Additionally, to gain the fine-grained ranking of the returned medical images shar-

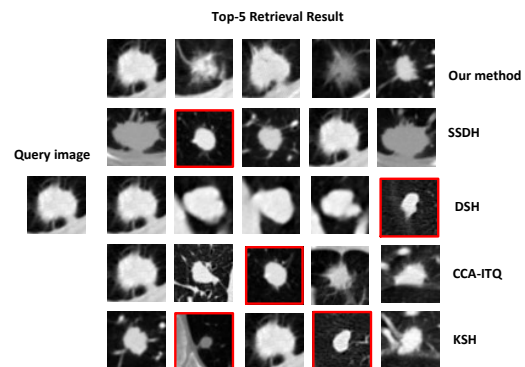


Fig. 8 The top-5 retrieval results of 64 bits using various hashing methods.

ing the same Hamming distance, a novel image re-ranking scheme is proposed to refine the similarity measurement by jointly considering Euclidean distance between the real-valued feature descriptors and their category information between those images. Extensive experiments on the pulmonary nodule image dataset show the superiority of the proposed method over the state-of-the-arts, and the $mAP@100$ value can achieve 63.51%, which is 2.59% higher than 60.92% achieved by SSDH when the length of hash code is 64 bits. Moreover, the retrieved similar lung nodule images provide an effective aided diagnosis for medical personnel.

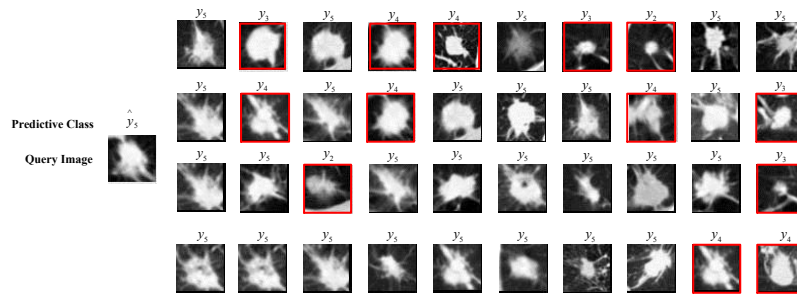


Fig. 9 Image re-ranked with the same Hamming distance. The top row shows the top 10 retrieved images with the same Hamming distance without using any re-ranking schemes. The second to fourth rows are the re-ranked results by Lin et al. [18], BIR [32] and our method.

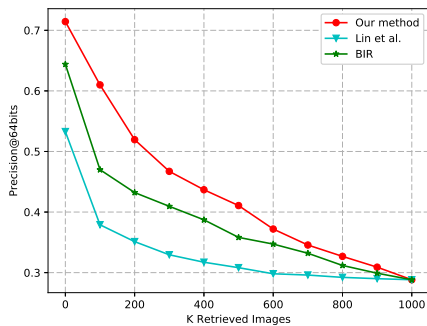


Fig. 10 The precision curve of three image re-ranking methods for top 1,000 returned candidates at 64 bits.

Declarations

Not applicable

Acknowledgments

This work is supported in part by the Natural Science Foundation of China under Grant 61702157 and 41804118, in part by Hebei Province Department of Education Fund under Grant QN2018085, in part by Innovation Capacity Improvement Project of Hebei Province 199676146H, and in part by the Key Program from NSF of North China Institute of Aerospace Engineering under Grant ZD-2013-05.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Torre, L.A., Siegel, R.L., Jemal, A.: Lung cancer statistics. *Advances in Experimental Medicine & Biology* **893**, 1–19 (2016)
2. Bach, P.B., Mirkin, J.N., Oliver, T.K., Azzoli, C.G., Berry, D., Brawley, O.W., Byers, T., Colditz, G.A., Gould, M.K., Jett, J.R.: Benefits and harms of ct screening for lung cancer: A systematic review. *Jama the Journal of the American Medical Association* **307**(22), 2418 (2012)
3. Li, Z., Zhang, X., Müller, H., Zhang, S.: Large-scale retrieval for medical image analytics: A comprehensive review. *Medical Image Analysis* **43**, 66–84 (2018)
4. Sundararajan, K., Woodard, D.L.: Deep learning for biometrics: A survey. *ACM Computing Surveys* **51**(3), 65–16534 (2018)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **60**(2), 91–110 (2004)
6. Gionis, A., Indyk, P., Motwani, R.: Similarity search in high dimensions via hashing. In: *International Conference on Very Large Data Bases(VLDB)*, Edinburgh, Scotland, UK, September 7-10, pp. 518–529 (1999)
7. Wu, G., Han, J., Lin, Z., Ding, G., Zhang, B., Ni, Q.: Joint image-text hashing for fast large-scale cross-media retrieval using self-supervised deep learning. *IEEE Transactions on Industrial Electronics* **66**(12), 9868–9877 (2019)
8. Weiss, Y., Torralba, A., Fergus, R.: Spectral hashing. In: *Twenty-Second Annual Conference on Neural Information Processing Systems(NIPS)*, Vancouver, British Columbia, Canada, December 8-11, pp. 1753–1760 (2008)
9. Wang, J., Zhang, T., Song, J., Sebe, N., Shen, H.T.: A survey on learning to hash. *IEEE transactions on pattern analysis and machine intelligence* **40**(4), 769–790 (2018)
10. Chi, L., Zhu, X.: Hashing techniques: A survey and taxonomy. *ACM Computing Surveys* **50**(1), 11–11136 (2017)
11. Kulis, B., Darrell, T.: Learning to hash with binary reconstructive embeddings. In: *Annual Conference on Neural Information Processing Systems(NIPS)*, Vancouver, British Columbia, Canada, December 7-10, pp. 1042–1050 (2009)
12. Norouzi, M., Fleet, D.J.: Minimal loss hashing for compact binary codes. In: *International Conference on Machine Learning(ICML)*, Bellevue, Washington, USA, June 28 - July 2, pp. 353–360 (2011)

13. Liu, W., Wang, J., Ji, R., Jiang, Y., Chang, S.: Supervised hashing with kernels. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Providence, RI, USA, June 16-21, pp. 2074–2081 (2012)
14. Gong, Y., Lazebnik, S., Gordo, A., Perronnin, F.: Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(12), 2916–2929 (2013)
15. Liu, W., Wang, J., Kumar, S., Chang, S.: Hashing with graphs. In: International Conference on Machine Learning(ICML), Bellevue, Washington, USA, June 28 - July 2, pp. 1–8 (2011)
16. Shen, F., Shen, C., Liu, W., Tao Shen, H.: Supervised discrete hashing. In: IEEE Conference on Computer Vision and Pattern recognition(CVPR), Boston, MA, USA, June 7-12, pp. 37–45 (2015)
17. Jiang, Q.-Y., Li, W.-J.: Discrete latent factor model for cross-modal hashing. *IEEE Transactions on Image Processing* **28**(7), 3490–3501 (2019)
18. Lin, K., Yang, H., Hsiao, J., Chen, C.: Deep learning of binary hash codes for fast image retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Boston, MA, USA, June 7-12, pp. 27–35 (2015)
19. Yang, H., Lin, K., Chen, C.: Supervised learning of semantics-preserving hash via deep convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence* **40**(2), 437–451 (2018)
20. Lai, H., Pan, Y., Liu, Y., Yan, S.: Simultaneous feature learning and hash coding with deep neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Boston, MA, USA, June 7-12, pp. 3270–3278 (2015)
21. Cao, Z., Long, M., Wang, J., Yu, P.S.: Hashnet: Deep learning to hash by continuation. In: IEEE International Conference on Computer Vision(ICCV), Venice, Italy, October 22-29, pp. 5609–5618 (2017)
22. Cai, Y., Li, Y., Qiu, C., Ma, J., Gao, X.: Medical image retrieval based on convolutional neural network and supervised hashing. *IEEE Access* **7**, 51877–51885 (2019)
23. Yu, C.J., Joachims, T.: Learning structural svms with latent variables. In: International Conference on Machine Learning(ICML), Montreal, Quebec, Canada, June 14-18, pp. 1169–1176 (2009)
24. Zhang, P., Zhang, W., Li, W., Guo, M.: Supervised hashing with latent factor models. In: International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR), Gold Coast, QLD, Australia, July 06-11, pp. 173–182 (2014)
25. Lin, G., Shen, C., Suter, D., van den Hengel, A.: A general two-step approach to learning-based hashing. In: IEEE International Conference on Computer Vision(ICCV), Sydney, Australia, December 1-8, pp. 2552–2559 (2013)
26. Xia, R., Pan, Y., Lai, H., Liu, C., Yan, S.: Supervised hashing for image retrieval via image representation learning. In: Twenty-Eighth AAAI Conference on Artificial Intelligence(AAAI), Québec City, Québec, Canada, July 27 -31, pp. 2156–2162 (2014)
27. Zhang, X., Liu, W., Dundar, M., Badve, S., Zhang, S.: Towards large-scale histopathological image analysis: Hashing-based image retrieval. *IEEE Transactions on Medical Imaging* **34**(2), 496–506 (2015)
28. Zhao, J., Pan, L., Zhao, P., Tang, X.: Medical sign recognition of lung nodules based on image retrieval with semantic features and supervised hashing. *Journal of Computer Science and Technology* **32**(3), 457–469 (2017)
29. Cai, Y., Li, Y., Qiu, C., Ma, J., Gao, X.: Medical image retrieval based on convolutional neural network and supervised hashing. *IEEE Access* **7**, 51877–51885 (2019)
30. Conjeti, S., Paschali, M., Katouzian, A., Navab, N.: Deep multiple instance hashing for scalable medical image retrieval. In: Medical Image Computing and Computer Assisted Intervention(MICCAI), Quebec City, QC, Canada, September 11-13, pp. 550–558 (2017)
31. Liu, H., Wang, R., Shan, S., Chen, X.: Deep supervised hashing for fast image retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Las Vegas, NV, USA, June 27-30, pp. 2064–2072 (2016)
32. Fu, H., Kong, X., Wang, Z.: Binary code reranking method with weighted hamming distance. *Multimedia tools and applications* **75**(3), 1391–1408 (2016)
33. Ye, F., Dong, M., Luo, W., Chen, X., Min, W.: A new re-ranking method based on convolutional neural network and two image-to-class distances for remote sensing image retrieval. *IEEE Access* **7**, 141498–141507 (2019)
34. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations(ICLR), San Diego, CA, USA, May 7-9 (2015)
35. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans Pattern Anal Mach Intell* **33**(8), 1548–1560 (2011)
36. Lin, Z., Ding, G., Hu, M., Wang, J.: Semantics-preserving hashing for cross-view retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Boston, MA, USA, June 7-12, pp. 3864–3872 (2015)
37. Petersen, K.B., Pedersen, M.S.: *The Matrix Cookbook*, p. 6. Technical University of Denmark, Cambridge MA (2012)
38. Shen, F., Shen, C., Liu, W., Shen, H.T.: Supervised discrete hashing. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 7-12, pp. 37–45 (2015)
39. Armato, S.G.I., Mclellan, G., Bidaut, L., Mcnitt-Gray, M.F., Clarke, L.P.: The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics* **38**(2), 915–931 (2011)
40. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Annual Conference on Neural Information Processing Systems(NIPS), Montreal, Quebec, Canada, December 7-12, pp. 91–99 (2015)
41. Lin, T., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J.: Feature pyramid networks for object detection. In: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Honolulu, HI, USA, July 21-26, pp. 936–944 (2017)
42. Shen, Z., Liu, Z., Li, J., Jiang, Y., Chen, Y., Xue, X.: DSOD: learning deeply supervised object detectors from scratch. In: IEEE International Conference on Computer Vision(ICCV), Venice, Italy, October 22-29, pp. 1937–1945 (2017)
43. Yang, H., Lin, K., Chen, C.: Supervised learning of semantics-preserving hash via deep convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence* **40**(2), 437–451 (2018)
44. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: International Conference on Machine Learning(ICML), Pittsburgh, Pennsylvania, USA, June 25-29, pp. 233–240 (2006)