Laplacian Deep Hashing for Image Retrieval

Chunzhi Wang, Fangyu Zhou, Lingyu Yan, Zhiwei Ye, Pan Wu and Hanlin Lu

June 2, 2018

Laplacian Deep Hashing for Image Retrieval

Chunzhi Wang, Fangyu Zhou, Lingyu Yan, Zhiwei Ye, Pan Wu, Hanlin Lu
School of Computer Science
Hubei University of Technology
Wuhan, China
chunzhiwang@vip.163.com, fangyuzhoufarry@163.com, yanlingyu@hbut.edu.cn, weizhiye121@163.com, wupan6454@163.com, lupangpang123@163.com

Abstract—**The hash method has been widely applied in the large scale image retrieval to improve the speed and reduce the storage cost in retrieval. The traditional hash methods are based on handcraft features and shallow models. These methods map the massive images to hash codes by visual features. Although the methods have advantage in preserving the similarity relationships of original images in the mapping operation, the performances of retrieval are not excellent. Due to the good performance of CNNs in classification and retrieval, the features extraction become the important element in image retrieval. Consequently, the framework of deep CNNs is introduced to hash methods as component of features extraction to improve the performance in retrieval. However, the over-fitting is problem in the deep CNNs. In this work, we propose the Laplacian Deep Hash method, combining the superiority of deep CNNs and shallow hash models to optimize the hash method in image retrieval. The robust and discriminative semantic features are obtained by deep CNNs to improve hash codes generation. And the shallow hash models as the hash function are applied to decrease over-fitting . Finally, the effectivity and efficiency of this method is presented in the comparative experiment results.**

Index Terms—**deep learning, hash, image retrieval, artificial intelligence.** *(key words)*

I. INTRODUCTION

Because of the rapid advance in Internet, there are more and more new applications of multimedia published on the websites or mobile terminal for customers. Consequently, there are massive data of videos and graphs produced by people every day. Therefore, demand of contend-based image retrieval in big data is more necessary than before.Due to the large scale of data, the fast speed and low storage cost of image retrieval have been widely studied to optimize the methods of image retrieval. The hashing method is the approach of approximate nearest neighbor(ANN) search, which have been used in large scale image retrieval.

The first point of contend-based image retrieval is accuracy. The accuracy of retrieval is based on the similarity of images,which is represented and quantization by the features of images in the state-of-the-art now. The effective features are the promise of good performance of the image retrieval. Since 2012, the convolutional neural networks(CNNs) were widely applied to extract features,because of the good performance of CNNs in image classification[1].And [1] proof that the depth of CNNs is important element for the performance of the model. So the deep CNNs models were applied to extract effective deep semantic features,which improves the accuracy in retrieval..Although the deep CNNs have the high accuracy in the retrieval,the massive storage cost of retrieval is still the problem for the large scale images retrieval. Meanwhile, the efficiency of the CNNs model is limited by the depth.

The second point in contend-based image retrieval is efficiency. To accelerate the speed of retrieval, hash is the common approach. Because the calculation in hash for the retrieval is to calculate the Hamming distance between two images, which is more convenient than the Euclidean distance.There are several hash methods were classic[2,3],which try to preserve the similarity relationship of the original images.They have the comparatively good performance in image retrieval,which are based the shallow models. Although the efficiency and storage cost are apposite in hash approach, the accuracy is the challenge, due to loss of information when the high dimensional image features are mapped to low dimensional hash codes.

For different tasks in image retrieval, there are several methods, which have different ideas. The QaDWH[4] first learn hash codes and corresponding class-wise weight by deep hash network,then generate hash bit weights for classes of images.The retrieval of QaDWH [4]is based on the weighted Hamming distance. The DHCQ[5] is the hash method based on minimizing the classification error and quantization error to solve scalable face image retrieval. There are also several novel hash method used in different domain[6,7,8,9]. [6] applied supervised deep hashing framework with pairwise labels information to intelligent vehicle to improve the performance of environmental recognition. [7] applied deep hash neural networks(DHNNs) to retrieval of large-scale remote sensing images.To solve the two points, several deep hash method for image retrieval [10,11,12] are proposed. The framework of them generally is add the latent layers after the convolutional layers of deep CNNs. These methods applied the large scale data to train the deep CNNs and hash function that is the neural network with massive hyper-parameters. It easily lead to over-fitting. Inspired by above methods , we explore the combination of deep learning and shallow hash function, which extracts the deep semantic features by deep CNNs

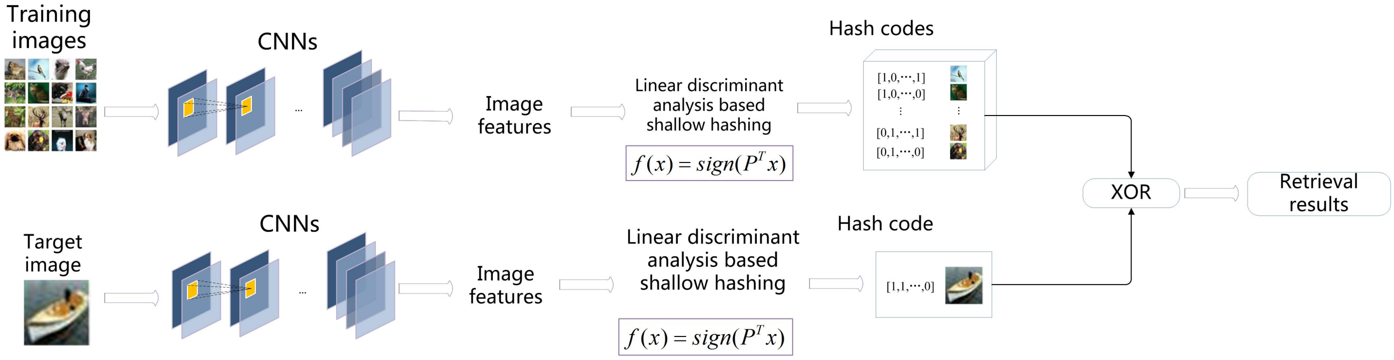

Fig. 1. The framework of LDH

and generated effective hash codes by shallow hash function on the features. We adopt the advantage of deep CNNs that obtaining effective high level semantic information of images. And we adopt the advantages of shallow model in anti-over-fitting and preserving the similarity relationship of image features. We propose the LDH method. The LDH is the hash method, which combine the deep convolutional neural network and shallow hash model. Deep convolutional neural network is applied to extract robust and discriminative features. According to the features, shallow hash model generate the effective and compact hash labels, The rest of this paper is organized as follows. Section 2 is the elaboration of the proposed approach. Experiments results are presented in Section 3. The conclusion is in Section 4.

II. Approach

The proposed Laplacian deep hashing structure is consisted of deep feature extraction stage and Laplacian hashing stage. In the deep features extraction stage, we apply the deep convolutional neural network to extract deep features from the images in database. The framework of deep convolutional neural network is existing model(e.g. VGG-16 or GoogLeNet trained in Image net). By fine-turning the existing model,we get the deep semantic features from the images. The features extracted by this method are better than the handcraft features because they contain deep semantic information of images. In the Laplacian hashing stage, we firstly use the KNN algorithm to construct the graph model of features in the former stage.Moreover, the graph model of features is mapped by using LE algorithm. Then, we apply the binarization to map the results to hash codes, which are the hash labels to the corresponding images. After hash learning period, the LE algorithm was used to generate hash code for images retrieved.. This method combine the advantage of deep learning and shallow hash algorithm, which accelerates query speed,reduces the storage cost and keeps high accuracy. The outline of LDH is shown in Fig. 1. In the following, it is the detailed introduction of LDH.

A. Deep Features Extraction

In this stage, the main process is fine-turning the deep convolutional neural network to get deep semantic features. For the accuracy of features, we introduce a deep CNNs(e.g. VGG-16 or GoogLeNet),which has been trained in large scale dataset(e.g. Image Net). The destination of the fine-turning is to make the deep convolutional neural network adaptive to different datasets.In this work, we select the VGG-16 to extract the features. The VGG-16 published by Visual Geometry Group is comprised of 13 convolutional layers and 3 fully connected layers[4].

B. Laplacian hashing stage

In this stage, we apply LE algorithm and binarization to transform the deep semantic features into hash codes and train the hash function. Consider a training dataset comprised of U images. The $\{x_j\}_{j=1}^{u} \in R^n$ denotes the n dimensional features vectors of U images.The $y_j \in \{0,1\}^h$ represents the hash label of x_j.

This algorithm adopts local similarity to construct the relationship of the data points. The k-Nearest Neighbor(KNN) is applied to construct the relationship graph of the features of images. The relationship graph is

$$N_{ij} \begin{cases} N_{ij} = true & x_i \text{ is the neighbor } x_j \\ N_{ij} = false & \text{otherwise;} (i, j = 1...u) \end{cases} \tag{1}$$

When N_{ij} is true,it means that x_i is the neighbor of x_j. Otherwise, the x_i is not the neighbor of x_j.

The local similarity matrix is

$$W_{ij} = \begin{cases} \dfrac{x_i^T x_j}{\|x_i\| \cdot \|x_j\|} & N_{ij} = true \\ 0 & N_{ij} = false \end{cases} \quad (i, j = 1...u) \tag{2}$$

The diagonal matrix D is

$$D_{ii} = \sum_{j=1}^{u} W_{ij} \ \ (i, j = 1...u) \tag{3}$$

The distance between hash label y_i and y_j is

$$P_{ij} = \left\| y_i - y_j \right\|^2 \quad (i, j = 1...u) \tag{4}$$

.

The objective function is

$$\phi = \min(\sum_{i=1}^{u} \sum_{j=1}^{u} W_{ij} P_{ij}) \tag{5}$$

The goal of the objective function is to minimize the weighted average Hamming distance.

Then the ϕ is transformed into φ ,the φ is

$$\varphi == \min(tr(Y^T LY)) \ \text{ s.t. } Y^T DY = I \tag{6}$$

The $L = D - W$ is the Laplacian matrix and $tr(\bullet)$ is the trace of square matrix.

Finally,the φ is converted into LapEig problem η with slacking constraint $y_i \in \{0,1\}^l$,and acquire the optimal l dimensional real-valued vector v to denote each image. The η is below

$$\eta = \arg\min_V tr(V^T LV) \ \ s.t. \begin{cases} V^T DV = I \\ V^T D1 = 0 \end{cases} \tag{7}$$

The $tr(V^T LV)$ is the real relaxation of the weighted average distance $tr(Y^T LY)$. The solution of the optimization problem is $V = [a_1, ..., a_m]$, where the columns of $V = [a_1, ..., a_m]$ are the m eigenvectors corresponding to the smallest nonzero eigenvalues of generalized eigenvalue problem. Therefore, the η can be transformed to the following equation

$$LV = \lambda DV \tag{8}$$

Next, the m-dimensional real-value vectors $V = [v_1, ..., v_t]$ are transformed to binary codes based on following formula.

$$v_i^b = \begin{cases} 1 & v_i^b \geq \theta^b \\ 0 & otherwise \end{cases} \tag{9}$$

The v_i^b means the b-th element of the i-th real-value vector.The θ^b is the threshold of b-th elements of all vectors . And it is

$$\theta^b = \frac{1}{t} \sum_{i=1}^{t} v_i^b \tag{11}$$

III. EXPERIMENTS

In this section, we design comparative experiments to evaluate the performance of our method LDH in retrieval. In the experiment, we compare our method with LSH and ITQ on the MNIST digit dataset and CIFAR-10 dataset. The measurement selected to evaluate the performance of approaches is the percision-recall curve. Because the categories of two datasets are the 10, we select 10 as the number of nearest neighbor for the KNN.

A. MNIST dataset

The MNIST dataset is the set of handwritten digit with 60,000 samples in the training set and 10,000 samples in the test set, which is the subset of NIST dataset. Each sample is a 28×28 pixels monochromatic image with a corresponding label which means that the image denotes a number in 0 to 9. This dataset is generally regard as the standard dataset to test the effectivity. In the experiment, we select 10,000 images and corresponding labels from the training set, which are respectively randomly selected from 10 categories. These samples are applied to fine-turn the CNN model, then we use the features extracted from the CNN model to generated the corresponding hash codes by LE algorithm, which is also the operation to train the hash function for LE. Afterwards, the 10,000 samples in test set are converted to hash codes to evaluate the performance of methods. Figure.2 shows the results of retrieval of the LDH and comparative methods, where the length of hash labels is 48-bits.

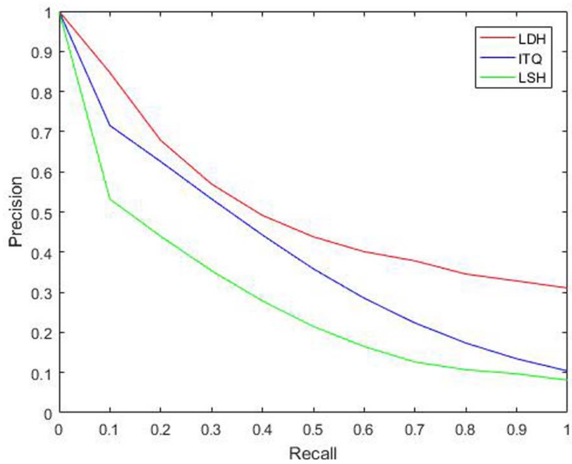

Fig. 2. Precision-Recall curve of comparative methods on MNIST with 48bits

In Figure.2, LDH shows obvious superiority on the Precision-Recall curve comparing with other algorithms. The main reason of the advantage of the LDH is extracting features by deep CNNs. The LSH and ITQ generate hash codes by shallow features, which include less visual information than deep semantic features. Consequently, the hash codes generated by the deep semantic features represent the original images better than the shallow features. However, the semantic information of images in MNIST are comparative less than the images from real world. So, to test the performance of method in complex dataset, we test our method on CIFAR-10.

B.CIFAR-10 dataset

The CIFAR-10 dataset is the subset of 80 million tiny images dataset, of which the images are labeled. The CIFAR-10 dataset consists of 60,000 images in 10 classes. each image is 32×32 pixels color image in RGB model. There are 50,000 training images(5000 images per class) and 10,000(1000 images per class) test images in the CIFAR-10 dataset. We use training images to construct the CNN model and the hash function of LE. Then we apply the test images to test the performance of approaches. In this experiment, we select 48-bit as the length of hash labels. Figure.3 show the results of the comparative experiment.

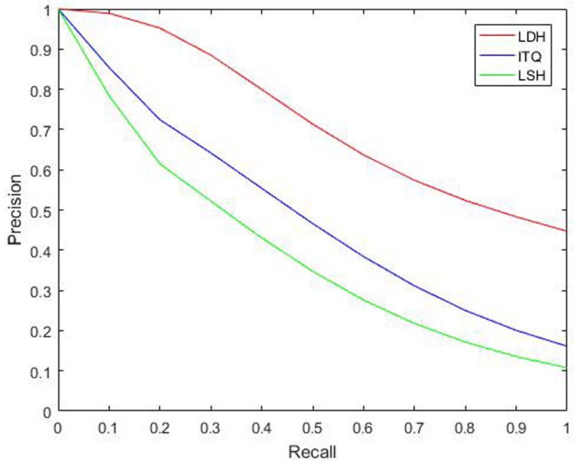

Fig. 3. Precision-Recall curve of comparative methods on CIFAR-10 with 48bits

It can be observed in Fig.3 that the performance is best in the comparative experiment. Because the images in CIFAR-10 are more complex, the performance of retrieval on CIFAR-10 is worse than the performance on MNIST. The result suggests two points: (1) the features extracted by CNNs model have better effect than the hand-craft features. (2) The length of hash codes should be adaptive to the semantic information of the images. For the good performance of retrieval, the features of images are important element. In our approach, we take the advantage of the CNNs model to get the deep semantic features from images to realize the high accuracy of retrieval. Then we take the advantage of the LE algorithm to map the high dimensional features to low dimensional space and generate the hash codes, which improves the speed of data processing and retrieval.

Acknowledgment

Thanks for the funding supported by the National Natural Science Foundation of China (No.61772180, No.61502155), Natural Science Foundation of Hubei University of Technology (No. BSQD15029).

References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems (Vol.60, pp.1097-1105). Curran Associates Inc.

[2] Gong, Y., Lazebnik, S., Gordo, A., & Perronnin, F. (2013). Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval. IEEE Transactions on Pattern Analysis & Machine Intelligence, 35(12), 2916-2929.

[3] Slaney, M., & Casey, M. (2008). Locality-sensitive hashing for finding nearest neighbors [lecture notes]. Signal Processing Magazine IEEE,25(2), 128-131.

[4] Zhang, J., & Peng, Y. (2016). Query-adaptive image retrieval by deep weighted hashing.

[5] Tang, J., Li, Z., & Zhu, X. (2017). Supervised deep hashing for scalable face image retrieval. Pattern Recognition, 75.

[6] Yan, C., Xie, H., Yang, D., Yin, J., Zhang, Y., & Dai, Q., et al. (2017). Supervised hash coding with deep neural network for environment perception of intelligent vehicles. IEEE Transactions on Intelligent Transportation Systems, PP(99), 1-12.

[7] Li, Y., Zhang, Y., Huang, X., Zhu, H., & Ma, J. (2017). Large-scale remote sensing image retrieval by deep hashing neural networks. IEEE Transactions on Geoscience & Remote Sensing, PP(99), 1-16.

[8] Liu, W., Ma, H., Qi, H., Zhao, D., & Chen, Z. (2017). Deep learning hashing for mobile visual search. Eurasip Journal on Image & Video Processing, 2017(1), 17.

[9] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Computer Science.

[10] Yang, H. F., Lin, K., & Chen, C. S. (2015). Supervised learning of semantics-preserving hash via deep convolutional neural networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, PP(99), 1-1.

[11] Lu, X., Song, L., Xie, R., Yang, X., & Zhang, W. (2017). Deep hash learning for efficient image retrieval. IEEE International Conference on Multimedia & Expo Workshops (pp.579-584). IEEE Computer Society.

[12] Qiu, Changyan & Cai, Yiheng & Gao, Xurong & Cui, Yize. (2017). Medical image retrieval based on the deep convolution network and hash coding. 1-6. 10.1109/CISP-BMEI.2017.8302194.

[13] Cao, L., Gao, L., Song, J., Shen, F., & Wang, Y. (2017). Multiple hierarchical deep hashing for large scale image retrieval. Multimedia Tools & Applications(1), 1-14.

[14] Liu, H., Wang, R., Shan, S., & Chen, X. (2016). Deep Supervised Hashing for Fast Image Retrieval. Computer Vision and Pattern Recognition (pp.2064-2072). IEEE.

[15] Do, T. T., Tan, D. K. L., Pham, T. T., & Cheung, N. M. (2017). Simultaneous Feature Aggregating and Hashing for Large-Scale Image Search. Computer Vision and Pattern Recognition (pp.4217-4226). IEEE.

[16] Gkountakos, K., Semertzidis, T., Papadopoulos, G. T., & Daras, P. (2017). Incorporation of semantic segmentation information in deep hashing techniques for image retrieval. International Conference on Engineering, Technology and Innovation (pp.632-638).

[17] Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., & Zhang, Y., et al. (2014). Deep Learning for Content-Based Image Retrieval: A Comprehensive Study. the ACM International Conference (pp.157-166). ACM.

[18] Krizhevsky, A., & Hinton, G. E. (2012). Using very deep autoencoders for content-based image retrieval. Esann 2011, European Symposium on Artificial Neural Networks, Bruges, Belgium, April 27-29, 2011, Proceedings. DBLP.