

Image Retrieval Using Deep Convolutional Neural Networks and Regularized Locality Preserving Indexing Strategy

Xiaoxiao Ma, Jiajun Wang*

School of Electric and Information Engineering, Soochow University, Suzhou, China

Email: *jjwang@suda.edu.cn

How to cite this paper: Ma, X.X. and Wang, J.J. (2017) Image Retrieval Using Deep Convolutional Neural Networks and Regularized Locality Preserving Indexing Strategy. *Journal of Computer and Communications*, 5, 33-39.

<https://doi.org/10.4236/jcc.2017.53004>

Received: February 17, 2017

Accepted: March 10, 2017

Published: March 13, 2017

Abstract

Convolutional Neural Networks (CNN) has been a very popular area in large scale data processing and many works have demonstrate that CNN is a very promising tool in many field, e.g., image classification and image retrieval. Theoretically, CNN features can become better and better with the increase of CNN layers. But on the other side more layers can dramatically increase the computational cost on the same condition of other devices. In addition to CNN features, how to dig out the potential information contained in the features is also an important aspect. In this paper, we propose a novel approach utilize deep CNN to extract image features and then introduce a Regularized Locality Preserving Indexing (RLPI) method which can make features more differentiated through learning a new space of the data space. First, we apply deep networks (VGG-net) to extract image features and then introduce Regularized Locality Preserving Indexing (RLPI) method to train a model. Finally, the new feature space can be generated through this model and then can be used to image retrieval.

Keywords

Image Retrieval, CNN, RLPI, Hash

1. Introduction

In traditional CBIR systems, low-level features such as the color, shape and texture features are usually extracted to construct a feature vector for describing images and then, based on a proper similarity measure, images are retrieved by comparing the feature vector corresponding to the query image and those corresponding to images in the data set. Generally, there are three key issues in CBIR systems, (1) selecting appropriate feature extraction method, (2) extracting

appropriate image features and (3) matching features with effective method. Many researchers devote most of their attention to the first issue. However, they usually fail to extract the internal structure contained in the features which is crucial for distinguishing data points. In our paper, we aim to find this internal structure from the original data space. Moreover, the convolutional neural network has been developing rapidly since 2012 when Krizhevsky et al. won the championship on the classification of the Image Net based on CNN [1]. Thus, it is a good tool for us to extract more abstract image features. Recently, some works prefer to combine CNN with hash code and a few hashing methods are proposed to measure the relevance between the query and the database. Hash methods can be divided into two categories: supervised hash and unsupervised hash. Since a pair-wise similarity matrix is needed for hash methods, the storage and computational cost can dramatically increase especially for the case when the database is very large. In order to resolve such a problem, we propose a method which utilizes the abstract features of CNN and the efficiency of RLPI.

Indeed, how to dig out the potential information contained in these features is another critical issue. We believe that there is a certain internal structural link between similar features. Thus, our main purpose is to find out this link and RLPI is a good choice in helping us with this research [2]. RLPI is fundamentally based on LPI which is proposed to find out the discriminative inner structure of the document space. However, in our paper, we apply it to the image feature space and get a good performance.

2. Method

2.1. VGG-Net

As mentioned before, we utilize the deep CNN for extracting abstract features from images. In our work, we utilize a VGG-net model with 5 nets for our image retrieval purpose. In here, these five nets are represented with five alphabets from A to E, respectively. The width of convolution layers starts from 64 in the first layer and then increases by a factor of 2 after each max-polling layer, until it reaches 512 and then maintains. In addition to convolution layers, there are five max-polling layers. Although VGG-net contains five nets, the convolution layers and the pooling layers in these five nets have the same parametric settings. This strategy can make sure that the shape comes out of each convolution layer group is consistent, no matter how many convolution layers are added in the convolution group.

Many studies demonstrate that deeper networks can achieve better performance. However, training deeper networks not only dramatically increases the computational requirements but also needs stringent hardware support. In our work, we utilize VGG-net model to extract image features. In order to implement this network with moderate computing requirements, each image is re-scaled to the same size of 224×224 , which is then represented with a vector of 4096 in dimension in terms of the network after removing the FC-layer.

2.2. A Brief Review of LPI and RLPI

2.2.1. Locality Preserving Indexing

LPI is proposed to find out the discriminative inner structure of the document space and extract the most discriminative features hidden in the data space. Given a set of data points and a similarity matrix. Then LPI can be obtained through solving the following minimization problem:

$$\begin{aligned} a^* &= \arg \min_{a^T XDX^T a=1} \sum_{i=1}^m \sum_{j=1}^m (a^T x_i - a^T x_j)^2 W_{ij} \\ &= \arg \min_{a^T XDX^T a=1} a^T X(D-W)X^T a \end{aligned} \quad (1)$$

where $D = \sum_j W_{ij}$ is a diagonal matrix and its entries are column sums of W . The W can be constructed as:

$$W_{ij} = \begin{cases} \frac{x_i^T x_j}{\|x_i\| \|x_j\|}, & \text{if } x_i \in N_p(x_j) \text{ or } x_j \in N_p(x_i) \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

where $N_p(x_i)$ is a set of p nearest neighbors of x_i .

As the objective function will generate a heavy penalty if neighboring data points x_i and x_j are mapped far apart. Thus, to get the objective function minimization is an attempt to ensure that if x_i and x_j are close then $y_i = a^T x_i$ and $y_j = a^T x_j$ are close as well [3]. Then, the basis function of LPI are the eigenvectors associated with the smallest eigenvalues of the following generalized eigen-problem:

$$X(D-W)X^T a = \lambda XDX^T a \quad (3)$$

Thus, the minimization problem in Equation (1) can be changed to the following problem:

$$a^* = \arg \max_{a^T XDX^T a=1} a^T XWX^T a \quad (4)$$

and the optimal a 's are also the maximum eigenvectors of eigen-problem:

$$XWX^T a = \lambda XDX^T a \quad (5)$$

However, according to [2], the high computational cost limits the application of LPI on large datasets. RLPI aims to solve this drawback through solving the eigen-problem in Equation (5) efficiently.

2.2.2. Regularized Locality Preserving Indexing

The following theorem can be used to solve the eigen-problem in Equation (5) efficiently:

Let y be the eigenvector of eigen-problem:

$$Wy = \lambda Dy \quad (6)$$

with eigenvalue λ . If $X^T a = y$, and a is the eigenvector of eigen-problem in Equation (5) with the same eigenvalue λ [2].

Based on this theorem, the direct computation of the eigen-problem in Equation (5) can be avoided and the LPI basis function can be acquired through the

following two steps:

- 1) Solve the eigen-problem in Equation (6) to get y .
- 2) Find a which can satisfy the equation $X^T a = y$ and a possible method to find a which can best fit the following equation in the least squares sense:

$$a = \arg \min_a \sum_{i=1}^m (a^T x_i - y_i)^2 \quad (7)$$

where y_i is the i -th element of y . A penalty can be imposed on the norm of a to avoid infinite many solutions for the linear equations system $X^T a = y$ through the following method:

$$a = \arg \min_a \left(\sum_{i=1}^m (a^T x_i - y_i)^2 + \alpha \|a\|^2 \right) \quad (8)$$

where $\alpha \geq 0$ is a parameter to control the amounts of shrinkage. This is the regularization which can be well studied in [4]. Thus, we called this method as Regularized Locality Preserving Indexing (RLPI). Let $A = [a_1, a_2, \dots, a_d]$, and the RLPI embedding is as follows:

$$x \rightarrow z = A^T x$$

where z is a d -dimensional new representation of the image x and this is our main purpose. From this point of view, RLPI is also a dimensionality reduction method. A is the transformation matrix. When we perform image retrieval, we can use this new feature representation to retrieval the similar images in the database through cosine distance.

3. Experiments Results

In order to evaluate the performance of our proposed method, experiments are conducted on Caltech-256 dataset. For the purpose of comparisons, results from other methods are also presented.

3.1. Image Datasets

Caltech-256 dataset contains 29780 images in 256 categories. We select images from the first 70 classes of the caltech-256 dataset to construct a smaller dataset (referred to as Caltech-70 here and after) consisting of 7674 images for our experiment. We select 500 images randomly as the queries and the remaining as the targets for search.

3.2. Quantitative Evaluation

Three measures are used to evaluate the performance of different algorithms. The first one is the precision defined as:

$$P@K = \frac{N_{correct}}{K}, \quad (9)$$

where $N_{correct}$ is the number of correct relevant returns among the K retrieved images.

The precision tells us the rate of relevant images in total retrieved images in a

particular search. However, sometimes, we want to get more relevant images from the database rather than just a very high precision. Thus, the recall is also an important measure of the performance of different algorithms. We define the second measure recall as:

$$R @ K = \frac{N_{correct}}{N_{relevant}}, \quad (10)$$

where $N_{relevant}$ is the total number of relevant images in the dataset.

3.3. Results

A. Comparisons with Hash Type Methods

In this section, we compare our proposed method (VGG-RLPI) with many hash type methods [5] [6] [7]. In [6], Iterative Quantization (ITQ) and PCA random rotation (PCA-RR) methods are proposed which represents state-of-art in 2013. And many classic hash type methods such as locality sensitive hashing (LSH), PCA hash (PCAH) and SKLSH [7] are also in comparison with our proposed method and from the results, we can find that our proposed method is superior to these methods.

Figure 1 shows the retrieval precision and recall of different methods averaged among 500 queries in the Caltech-70 dataset. For the purpose of comparisons, results from other hash type methods are also presented. In these methods, features are extracted with the VGG-net. From the curves in this figure we can see that our proposed method (VGG-RLPI) performs the best among others.

B. Comparisons with Dimension Reduction Methods

As has mentioned above, the RLPI is also a dimension reduction method. Thus, comparisons are also performed with two other dimension reduction methods: the Principle Component Analysis (PCA) method and the Linear Graph Embedding (LGE) method. For fair comparisons, retrieval experiments are first performed with reduced feature vectors obtained from these three methods to determine the optimal dimension. **Figure 2** gives the mean precision of the three methods at different number of returns in the Caltech-70 dataset. From the figure, we can see that the optimal dimension is 100 both for LGE and RLPI. However, for PCA method, the curves almost coincide when feature dimension exceed 50. Thus, for the following comparison, we select the precision corresponding to 100 dimension for the three dimension reduction methods. From the results in **Figure 3** we can find that our proposed method superior to the other two ones.

4. Conclusion

In this paper, a novel method utilizing the deep CNN and the RLPI is proposed for image retrieval. Since the CNN features have both abstract and global properties in FC-layers, it can well represent an image and also has a good discrimination ability in both classification and information retrieval tasks. However, using the features extracted from CNN to perform pattern matching directly is inefficient. On the other hand, the RLPI can learn a new feature space which is

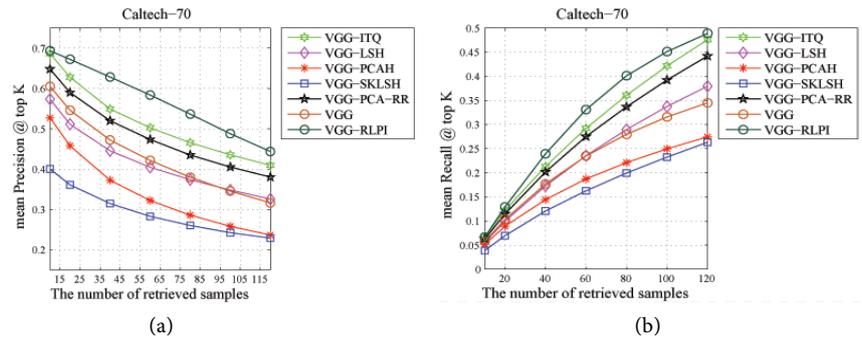


Figure 1. Image retrieval results: (a) Precision of Image Retrieval on Caltech-70 dataset; (b) recall of Image Retrieval on Caltech-70 dataset.

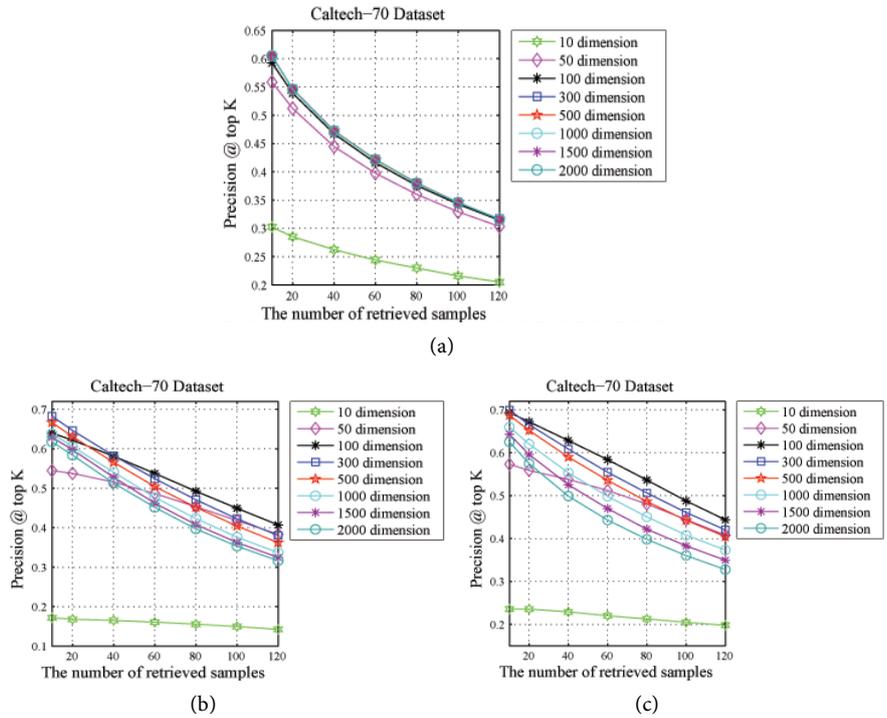


Figure 2. Mean retrieval precision with reduced feature vectors from two dimension reduction methods: (a) PCA method and (b) LGE method and (c) RLPI method.

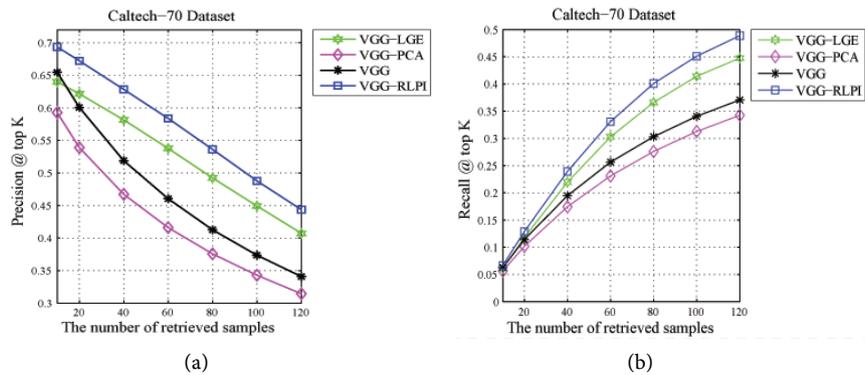


Figure 3. Comparison of VGG-RLPI and dimension reduction methods with optimal dimension in the Caltech-70 dataset: (a) Mean retrieval precision; (b) mean retrieval recall.

more discriminative compare to the original features. Experiments results in the Caltech-70 datasets show that our proposed method outperforms existing hash based methods and two other popular dimension reduction methods.

References

- [1] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. In: *Advances in Neural Information Processing Systems*.
- [2] Cai, D. and He, X.F. and Zhang, W.V. and Han, J.W. (2007) Regularized Locality Preserving Indexing via Spectral Regression. *Proceedings of the 16th ACM Conference on Conference on Information and Knowledge Management*, 741-750.
- [3] He, X., Cai, D., Liu, H. and Ma, W.-Y. (2004) Locality Preserving Indexing for Document Representation. In: *Proc. 2004 Int. Conf. on Research and Development in Information Retrieval (SIGIR'04)*, Sheffield, July 2004, 96-103.
<https://doi.org/10.1145/1008992.1009012>
- [4] Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York.
<https://doi.org/10.1007/978-0-387-21606-5>
- [5] Kafai, M., Eshghi, K. and Bhanu, B. (2014) Discrete Cosine Transform Locality-Sensitive Hashes for Face Retrieval. *IEEE Transactions on Multimedia*, **16**, 1090-1103. <https://doi.org/10.1109/TMM.2014.2305633>
- [6] Gong, Y., Lazebnik, S., Gordo, A. and Perronnin, F. (2013) Iterative Quantization: A Procrustean Approach to Learning Binary Codes for Large-Scale Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 2916-2929.
<https://doi.org/10.1109/TPAMI.2012.193>
- [7] Raginsky, M. and Lazebnik, S. (2009) Locality-Sensitive Binary Codes from Shift-Invariant Kernels. In: Bengio, Y., Schuurmans, D., Lafferty, J.D., Williams, C.K.I. and Culotta, A., Eds. *Advances in Neural Information Processing Systems 22*, Curran Associates, Inc., 1509-1517.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jcc@scirp.org