2015 IEEE International Conference on Systems, Man, and Cybernetics

Query-adaptive image search re-ranking using deep convolutional neural network feature

Bin Lin
Department of Electronic Engineering and Information Science, University of Science and Technology of China
Hefei Anhui, China
Email: lb1991@mail.ustc.edu.cn

Xinmei Tian
Department of Electronic Engineering and Information Science, University of Science and Technology of China
Hefei Anhui, China
Email: xinmei@ustc.edu.cn

***Abstract*—Image search re-ranking, as an effective tool to improve the text-based image search result, has been adopted by many commercial search engines nowadays. Given a query keyword, images are first retrieved based on the textual information. Then visual features are extracted from images to re-order them by mining their visual patterns. However, the popular visual features applied in re-ranking are not informative enough. Besides, the parameters for the re-ranking models are set equally for all queries, which fails to cope with the variability of different queries. In this paper, we propose a novel re-ranking method which adopts informative visual features for image representation and adaptively re-rank the images. Specifically, we adopt a proven successful DCNN feature (deep convolutional neural network), which shows the excellent performance in many computer vision fields, to calculate the visual similarities between images. For each query, the parameters for the image search re-ranking model is adaptively determined using the QDE (query difficulty estimation) method. Experiments are conducted on the INRIA web353 dataset. The experimental results demonstrate that our method achieves significant improvement over state-of-the-art methods.**

***Index Terms*—Image search re-ranking, Deep convolutional neural network, Query adaptive re-ranking**

I. Introduction

Web image search engine relies on the surrounding textual information to search images. However, the text-based ranking results are often filled with noise. To improve the text-based image search results, image search re-ranking is often conducted by using both the visual information and the textual information. It is defined as re-ordering the visual documents based on the initial text-based search results and their visual patterns. It can be treated as the post process of the search activity [1].

Image search re-ranking is mainly based on two assumptions. (1) The top ranked images are expected to possess the same semantic meaning with the query. (2) Images relevant to the query are expected to share the similar visual patterns [2]. In order to find those query-relevant images, two stages are applied to conduct the image search re-ranking, which are the extraction of the visual features and the construction of the re-ranking model.

Researchers proposed different methods to build the image search re-ranking model[3], [7], [8], [9], [10], [11], [12], [13]. Krupac et al. extracted SIFT features [4] on the dense grid and applied the BOW (bag of visual words) model [5] to describe the image visual information [3]. Click infromation can describe the relevance between images and queries accurately, thus the multimodel sparse coding method was proposed by Yu to predict the click data [6]. Lee et al. clustered the retrieved images based on the assumption that the relevant visual documents tend to be more similar to each other than to irrelevant ones and those visual documents clusters were re-ranked after the clustering [8]. Yan et al. adopted pseudo-relevance feedback to assume the top-ranked images to be the few relevant ones, and those pseudo-relevant samples were further used in SVM to classify the remaining images into relevant or irrelevant classes [9]. Motivated by the well-known PageRank technique [21], Jing and Baluja proposed the VisualRank algorithm to analyze the visual link structures among images [7]. Tian et al. treated the re-ranking as a global optimization problem and proposed a Bayesian framework to derive the re-ranking model [10]. Yang and Hanjalic were inspired by the learning-to-rank paradigm and derived a re-ranking function in a supervised way from the human-labeled training data [11]. Tian et al. proposed re-ranking selection with a preference learning model to automatically select the best search result list from a number of candidates [12]. Yang and Hanjalic proposed two stage learning method combining the unsupervised offline search engine and the supervised online human supervision to build the re-ranking model [13]. Luo and Tao proposed a manifold regularized multi-task method to learn a discriminative subspace to deal with multiple labels, thus images with different labels are divided [23]. Besides, multi-view learning has proven its usefulness on image retrieval field. Motivated by the success of the vector-valued function, Luo et.al employed multi-view vector-valued manifold regularization to integrate multiple features for image retrieval [24]. Multiview features and click features can be both used in the re-ranking model, thus Yu proposed a re-ranking method using click constraints and multi-view features to improve the retrieval performance [25].

However, there are some drawbacks in these re-ranking methods. First, the features applied in image search re-ranking

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.382

aren't informative enough. Therefore, a more discriminative feature needs to be discovered. Second, due to the huge variance among different queries, the parameters for the re-ranking models shouldn't remain the same for different queries. Therefore, it is not optimal to use the query-independent model for different queries.

Based on these observations, contributions are made to solve these drawbacks. Firstly, we apply the deep convolutional neutral network, which can reserve the main information of the image, to compress the image into a low-dimensional feature vector. This informative feature can describe the image well and thus fit the re-ranking model. Secondly, query difficulty estimation is applied to find images which can best describe the query so that suitable query-adaptive parameters are determined to learn the image search re-ranking model. This adaptive learning method can cope with the huge variance between queries.

The remainder of this paper is organized as follows: Section 2 introduces our query-adaptive re-ranking algorithm in details. Section 3 describes the experiment on the multimedia dataset and proves the effectiveness of our methods. Section 4 concludes the paper and raises a suggestion for future work.

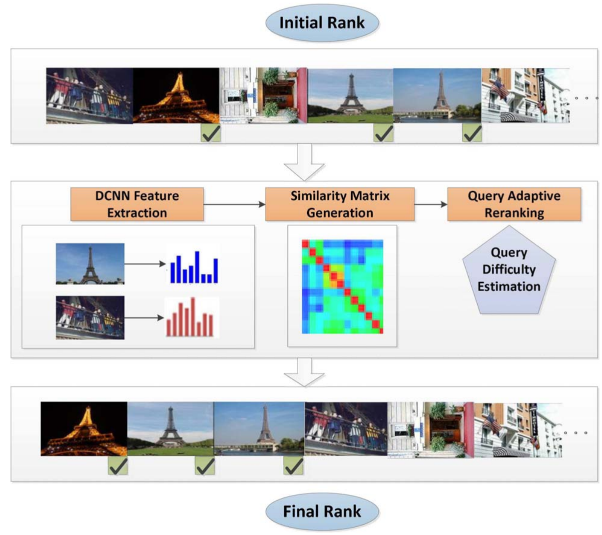

Fig. 1. The framework of image search re-ranking model in this paper.

II. Informative Feature Extraction and Adaptive Re-ranking

As aforementioned, the traditional re-ranking system is constituted of two key parts. In this paper, we propose a method to contribute on both parts. Our architecture contains three stages as in Fig. 1. (1) Feature extraction: DCNN features are extracted from the images. (2) Visual similarity description: In this part, visual similarities are computed based on the DCNN feature, and a particular similarity matrix is generated for each query. (3) Re-ranking model learning: In this part, QDE is applied to learn the model adaptively for different queries. Given an initial rank of one query, we apply the algorithm to get the final rank. Images with the tick mark are query-relevant. After the re-ranking, the relevant images are ranked higher.

A. Informative Features Extraction

Discriminative and effective visual features play a crucial role in image search re-ranking. Deep convolutional neural network has proven its ability to describe the image information in image classification field [15]. Due to its well descriptive power on images, this network is able to be applied to other computer vision fields including image re-ranking. So in this paper, we extract feature based on DCNN for each image.

Convolutional neural network was first proposed by LeCun et al. [14] to deal with handwritten digit recognition problem. The convolutional neural network can well intimate the biological vision system and thus maintain the strong ability to describe the image. But using the entire image as an input is both time-consuming and complex. Huge amount of parameters are needed to be trained. It is almost impossible to train the network due to the limitation of the computing power and lack of methods to avoid over-fitting. Luckily, in recent years, deep learning methods have gained a significant improvement in several ways, including the unsupervised and layer-wised pre-training, better activation functions and new training methods. The usage of multi-core computers and the implement of GPU also significantly reduce the training time. Based on these, Krizhevsky et al. [15] proposed a deep convolutional neural network with millions of parameters and applied it on the ImageNet dataset. The overall architecture includes eight layers, where the first five are convolutional and the remaining three are fully connected. The output of the final layer is fed to a 1000-way soft-max which produces the distribution over the labels. The high classification result shows its significant power on describing all kinds of images.

First, we implement the deep convolutional neural network as in [15]. This neural network is carefully trained on the ILSVRC-2012 (ImageNet Large-Scale Visual Recognition Challenge) dataset [22], which contains 1.2 million images that cover 1000 categories. With the huge amount number of training images and the vast diversities of the different categories, this deep convolutional neural network model can well describe different images. Second, in order to probe the visual knowledge of the image, we consider the feature activations induced at the last, 4096-dimensional hidden layer. If two images produce feature activation vectors with a small distance separation, the higher levels of the neural network consider them to be similar. So the external 1000-way soft-max is removed in the original network. We normalize those activations as our feature for a single image. The normalization function is defined in Eq. (1) where f is the input DCNN feature, x is the normalized feature vector and N denotes the dimensionality of the feature vector x_i and x_j.

$$x = [x_1, x_2, ..., x_N], x_i = \frac{\sqrt{f_i}}{\sum_{j=1}^{N} f_j} \tag{1}$$

After feature extraction, the measurement for the visual similarity between two images needs to be defined. The chi-square distance d_{ij} is adopted to calculate the visual distance between image i and image j.

$$d_{ij} = \frac{\sum_{k=1}^{N} \frac{(x_{ik}-x_{jk})^2}{x_{ik}+x_{jk}}}{2} \tag{2}$$

Since the similarity and the distance are in inverse relation, an inverse proportional function is applied to transfer the chi-square distance into the similarity value. The formulation is given in Eq. (3)

$$s_{ij} = \frac{1}{d_{ij} + \lambda} \tag{3}$$

where s_{ij} denotes the similarity between image i and image j, and d_{ij} denotes the distance between the feature vector of the image i and j. Empirically, $\lambda = 0.5$ is chosen to avoid the situation in which $d_{ij} = 0$. For each query, visual similarities are computed between every two returned images and thus form the similarity matrix S.

B. Query-Adaptive Re-ranking

In order to implement our query-adaptive re-ranking algorithm, VisualRank [7] is applied as the re-ranking model with parameters adaptively determined for each query. VisualRank employs the random walk intuition based on the visual hyperlinks among the images. The intuition of using visual hyperlinks is that if an image is viewed by a user, the related images may also be of interest. There are two assumptions about VisualRank. (1) If image i has a visual hyperlink to image j, there is certain probability that the user will jump from i to j. And the images that are visited often are important. (2) If a large correlation exists between an important image i and another image j, we can deduce that j is likely to be important as well [7].

The VisualRank algorithm is defined as an iterative formulation

$$VR = dS^* \times VR + (1-d)p \tag{4}$$

where S^* is the column normalized adjacency similarity matrix S. VR rank is the ranking vector which indicates the ranking score for each image. d is called the damping value which represents the probability that a random walk process goes to other images in the graph. The weighting vector p reflects the different importance of images in the query.

As aforementioned, query-independent parameters are used in existing re-ranking methods. Due to the huge variance among different queries, query-adaptive parameters should be selected instead. In VisualRank, there are two major parameters, which are the weighting vector and the damping value.

1) Weighting Vector Adaptive Selection: The value in the weighting vector reflects the reliability of images in the initial text-based search lists. The random walk will walk to the random images with the probability of $(1-d)$. It is better for those random images to be relevant to the query. As aforementioned, the images ranked higher are more likely to be relevant to the query, so a non-uniform vector p is applied to bias the computation. The weighting vector p is defined as in Eq. (5).

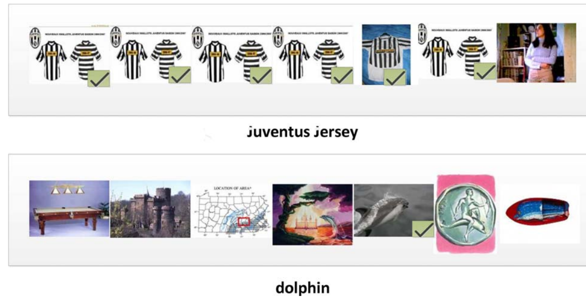

Fig. 2. Top-7-ranked images returned by a text-based image search engine for two queries: "juventus jersey" and "dolphin", ordered left to right. Query-relevant images are marked by the ticking sign. It illustrates that this image search engine suffers from a radical variance in retrieval performance over different queries.

$$p_i = \begin{cases} \frac{1}{T_{rel}} & i \leq T_{rel} \\ 0 & \text{else} \end{cases} \tag{5}$$

A relevance threshold T_{rel} is defined to separate the important and the unimportant images. For images ranked higher than T_{rel}, we set the weight $\frac{1}{T_{rel}}$ to them. For those unimportant images, the weight is set to 0 to filter out their influence on the weighting vector.

The relevance threshold is determined by query difficulty estimation (QDE), which predicts the quality of the search results without knowing the ground truth information. For those "difficult" queries, the search engine performs poorly, which means there are few relevant images on the front of the initial rank. On the contrary, for those "easy" queries, great performance is achieved and the returned images are mostly relevant ones. In Fig. 2, the examples of "difficult" queries and "easy queries" are given.

Coherence score is applied as the measurement of QDE, which is calculated based on assumption that relevant images share common visual patterns with each other [19]. So the "easy" query can get a high coherence score. It was firstly applied in predicting textual query difficulty [16]. And its application in predicting image search query difficulty was a success as well [17]. The coherent image pairs are those whose visual similarities are larger than a certain threshold. For images within top-T-ranked, the CoS is given in Eq. (6)

$$CoS = \frac{1}{|T(T-1)|} \sum_{i,j=1,\ldots,T; i \neq j} \delta(x_i, x_j) \tag{6}$$

where $\delta(x_i, x_j)$ is a binary function measuring the relevance between feature vector x_i and feature vector x_j. And this function is defined in Eq. (7)

$$\delta(x_i, x_j) = \begin{cases} 1, & s_{ij} > Tr_{sim} \\ 0, & \text{else} \end{cases} \tag{7}$$

and Tr_{sim} is defined so that 80% of image pairs in the dataset should have smaller similarities than this value.

If the $CoS@T$ value is large, the difficulty of this query is relatively low, which means the images relevant to the query mainly lie on the front of the initial rank list. We calculate the $CoS@T$ value under different Ts ranging from 1 to 100. When the $CoS@T$ reaches its maximum value, this query is considered to be the easiest under the corresponding T. So this top-T-ranked images can well represent the query. Thus, the relevance threshold is set to this T. The formulation is given in Eq. (8).

$$T_{rel} = \underset{T, 1 \leq T \leq 100}{\operatorname{argmax}} \; CoS@T \tag{8}$$

2) Damping Value Adaptive Selection: After determining the weighting vector for different queries, damping value should be selected as well. Damping value d measures the possibilities that the random walk will walk through the hyperlink, while the random image will be visited with a probability of $(1 - d)$. For queries with a small relevance threshold, the random image is restricted to those "important images", thus the damping value should be lower to create more chances for those images to be visited. Based on this observation, the damping value should have a direct proportion to the relevance threshold. We conduct different experiments to determine the best relation between these two parameters, and the result is given in Eq. (9).

$$d = \begin{cases} 0.15, & T_{rel} \leq 10 \\ 0.4, & 10 < T_{rel} \leq 50 \\ 0.8, & T_{rel} > 50 \end{cases} \tag{9}$$

3) Summary: In conclusion, we follow these steps to choose query-adaptive parameters for re-ranking models. Firstly, CoS is applied to conduct the query difficulty estimation. Secondly, the relevance threshold is selected according to the CoS, therefore the weighting vector is automatically determined for each query. Thirdly, the proper damping value d is selected based on the relevance threshold. Finally, these query-adaptive parameters are chosen to adaptively build the image search re-ranking model using VisualRank algorithm.

III. Experiment

In this section, we aim to test the effectiveness of our query-adaptive re-ranking method on a Web image search dataset.

A. Dataset

We conduct our experiment on the INRIA web353 dataset collected by Krapac et al. [3]. This dataset includes 353 queries, where the original textual query is also included. Fig. 3 shows some example images in this dataset. For 80% of queries, there are more than 200 images. Each image is resized to 150*150 pixel square. The ground-truth relevance label for every image is divided to two levels, which are "relevant" and "irrelevant". The 353 queries are diverse in topics, covering people ("Will Smith"), flag ("Italy flag"), landmarks ("tower"), animal ("tiger") and etc. Some queries are "easy" queries, such as "Juventus Torino jersey", "Justin Timberlake". Their mean average precisions are all higher than 0.85. However, some queries are "difficult" queries including "logo fc Barcelona", "dolphin" and "PSG jersey". In all, there are about 43.86% images in this dataset labeled as relevant samples.

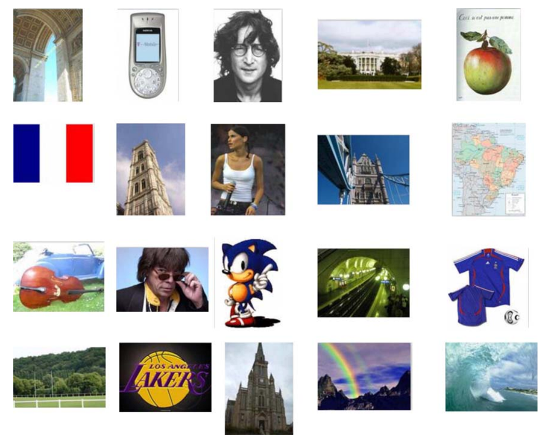

Fig. 3. Example pictures in INRIA web353 dataset.

B. Performance Metrics

Average precision (AP), which can reflect the occurrence of the relevant images, is adopted as our criterion to measure the effectiveness of the algorithm [18]. We compute the average precisions at several truncation levels of T, i.e. $AP@T, T = 20, 40, 60, 80$, which reflect the precision for the top-T-ranked images, and they are defined in Eq. (10).

$$AP@T = \frac{1}{Z_T} \sum_{i=1}^{T} [precision(i) \times rel(i)] \tag{10}$$

where $rel(i)$ is a binary function which reflects the relevance of the ith-ranked image. The precision value is the precision of top-i-ranked images.

$$precision(i) = \frac{1}{i} \sum_{j=1}^{i} rel(j) \tag{11}$$

The parameter Z_T makes sure that the $AP@T = 1$ for the best ranking result.

In order to get a full comprehension of the re-ranking system performance, several Ts are chosen to calculate the $MAP@T$ on the dataset. The mean average precision is conducted over 353 queries to get the final results.

C. Experiments for Feature Comparison

To evaluate the effectiveness of the deep convolutional neural network feature in re-ranking, we compare it with the most popular local features, including (1) SIFT feature with interest point detection and (2) Dense SIFT. The implementation of local feature extraction is given below.

For the extraction of SIFT feature with interest point detection, we follow the setting in [7]. The visual similarity between image i and image j is defined in Eq. (12). If the

TABLE I
MAP VALUE USING DIFFERENT FEATURES.

	Search Engine	SIFT	Dense SIFT	DCNN
MAP@5	0.611	0.577	0.691	**0.799**
MAP@10	0.553	0.560	0.633	**0.743**
MAP@20	0.503	0.535	0.564	**0.656**
MAP@40	0.452	0.498	0.492	**0.552**
MAP@60	0.431	0.479	0.479	**0.557**
MAP@80	0.426	0.473	0.477	**0.567**
MAP@ALL	0.569	0.601	0.607	**0.680**

two regions are matched, the ratio of distances between the closest neighbor and the second-closest neighbor should be greater than 0.8, which eliminates 90% of the false matches while discarding less than 5% of the correct matches at the same time [4].

$$s_{ij} = \frac{match\_features}{0.5 \times (num\_feature(i) + num\_feature(j))} \quad (12)$$

For the dense SIFT features, they are extracted on a 6x6 dense grid as in [3]. We adopt the BOVW model and train the 1000-dimensional vocabulary. After quantizing local descriptors into visual words, each image can be treated as a visual document filled with visual words. TF-IDF model is applied to measure the importance of the visual words [20]. The histogram for the image is defined in Eq. (13)

$$\begin{aligned} x_i &= [x_{i1}, x_{i2}, x_{i3}, ..., x_{i1000}]^T; \\ x_{ij} &= tf_{ij} \times idf_{ij}; \\ i &= 1, 2, 3, ..., N; j = 1, 2, 3, .., 1000 \end{aligned} \quad (13)$$

where tf_{ij} is the frequency of visual words w_j in image i and idf_{ij} is the inverse document frequency which quantifies the importance of visual words w_j over the whole image dataset. Cosine similarity is utilized to define the relations between two tf-idf histograms x_i and x_j as in Eq. (14).

$$s_{ij} = \frac{x_i \cdot x_j}{\|x_i\| \cdot \|x_j\|} \quad (14)$$

The 4096-dimensional deep convolutional neural network features are computed for every images in the dataset. Then the visual similarity between two images is computed via Eq. (2) and Eq. (3).

We use those three different visual features to calculate the similarity matrices in VisualRank. The weighting vector and dumping value is empirically selected for each query. The MAP results are given in Table I.

The "Search Engine" result is conducted based on the textual information on the meta-data file. It is listed as the baseline search engine result for re-ranking. The MAP of the SIFT feature is given in the third column. It turns out that the image search result is relatively poor using local SIFT feature. For $MAP@5$, the result is even worse than the initial ranking result. There are two main reasons why this feature gives poor result. (1) The quality of images in this database is quite low. The image file is only 4kb approximately each, which means the image has been compressed to a high level. So we lost a lot of gradient information, thus the descriptiveness of SIFT is relatively weak. (2) The number of SIFT features for each image is quite few, which makes it even harder to measure the similarity between the image pair.

TABLE II
MAP VALUE USING DIFFERENT PARAMETER SELECTION STRATEGIES.

	Search Engine	Strategy1	Strategy2	Strategy3
MAP@5	0.611	**0.799**	0.779	0.793
MAP@10	0.553	0.743	0.729	**0.748**
MAP@20	0.503	0.656	0.679	**0.704**
MAP@40	0.452	0.552	0.628	**0.658**
MAP@60	0.431	0.557	0.611	**0.637**
MAP@80	0.426	0.567	0.606	**0.631**
MAP@ALL	0.569	0.680	0.705	**0.724**

The MAP of dense SIFT is given in the fourth column. It achieves better results than the initial rank for every T, which proves the effectiveness of this method. But when T is larger, the results show little improvement than using Local SIFT, which demonstrates the weakness of dense SIFT.

We obtain the best performance so far by using deep convolutional neural network feature. The results are the best for every T when calculating the MAP value. $MAP@ALL$ value gets as high as 0.680, which is 11% better than the initial rank result.

D. Experiments for Adaptive Re-ranking Model

In order to verify the effectiveness of the proposed query adaptive re-ranking model, we conduct VisualRank with three different strategies to set parameters..

(1) The Empirical Parameter Setting

We experiment on different damping values and relevance thresholds. $d = 0.85$ and $T_{rel} = 30$ are selected respectively as the parameters for the re-ranking model, because we achieve the best re-ranking performance based on these settings.

(2) Adaptive Weighting Vector Selection

We conduct the query difficulty estimation for every query to compute the relevance threshold. Then, the weighting vector is automatically selected for each query. For all queries, the damping value is set to 0.85 as in strategy (1) does.

(3) Adaptive Weighting Vector and Damping Value Selection

QDE is applied to every query and thus the damping value and the weighting vector are both chosen adaptively as aforementioned. The MAP results for those three strategies are given in Table II. The second column demonstrates the result using empirical parameters. The third column shows the result adding partial adaptive re-ranking which gains improvement on MAP over strategy 1 while $T = 20, 40, 60, 80, ALL$. After adaptively selecting both parameters at the same time, the performance reaches a higher value for every T except for $T = 5$. And $MAP@ALL$ achieves as high as 0.724. From Table II, we can come to a conclusion. Adaptively selecting

TABLE III
RESULTS COMPARING TO STATE-OF-THE-ART METHODS.

Methods	MAP
Search Engine	0.569
PRF[9]	0.658
Bayesian[10]	0.665
Query Relative[3]	0.666
Two-stage Learning[13]	0.705
Our Method	**0.724**

the two parameters in the re-ranking model at the same time can significantly improve the image search performance, and it proves the effectiveness of our query adaptive re-ranking method.

E. Comparison with State-of-the-art Re-ranking Methods

In this part, we compare our results with state-of-the-art re-ranking methods, including pseudo relevance feedback [9], Bayesian re-ranking [10], supervised re-ranking [11], query relative re-ranking [3], and two stage learning [13] . Since most of them only reported the results on MAP@ALL, we compare the results under this metric. The results are obtained from [13] and listed in Table III. Our method demonstrates better performance than the state-of-the-art re-ranking methods.

IV. CONCLUSION

In this paper, we propose an effective query adaptive re-ranking method using deep convolutional neural network feature. In the feature extraction and description stage, the DCNN feature is utilized to help describing the image in a more informative way. In the re-ranking model stage, query difficulty estimation technique is applied to help us select the query-adaptive parameters for re-ranking model. Experiments are conducted on a Web image dataset, which demonstrate the effectiveness of our method.

There are many avenues for future explorations. Firstly, we would like to study whether we can design a different scheme to set the query-adaptive parameters. Secondly, we can use the state-of-the-art object recognition technique to help us detect the regions which are more likely to be query-relevant. In the end, other re-ranking models can be utilized to test the effectiveness of the adaptive parameters selecting scheme.

ACKNOWLEDGMENT

This work is supported by the 973 project under the contract No.2015CB351803, the NSFC under the contract No.61390514 and No.61201413, the Fundamental Research Funds for the Central Universities No. WK2100060011 and No.WK2100100021, the Specialized Research Fund for the Doctoral Program of Higher Education No. WJ2100060003.

REFERENCES

[1] T. Mei, Y. Rui, S. Li, and Q. Tian, "Multimedia Search Reranking: A Literature Survey," ACM Computing Surveys. Vol. 46, no. 3, pp. 1-38, 2014.

[2] X. Wang, S. Qiu, K. Liu, and X. Tang, "Web Image Re-Ranking Using Query-Specific Semantic Signatures," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 4, pp. 810-823, 2014.

[3] J. Krapac, M. Allan, J. Verbeek, and F. Jurie, "Improving Web Image Search Results Using Query-relative Classifiers," In Computer Vision and Pattern Recognition, pp. 1094-1101, 2010.

[4] D. G. Lowe. "Distinctive Image Features From Scale-Invariant Keypoints, "International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[5] Y. Yang, Y.G. Jiang, A.G. Hauptmann, and C.-W. Ngo, "Evaluating Bag-of-Visual-Words Representations in Scene Classification," In Proceedings of the international workshop on multimedia information retrieval, pp. 197-206, 2007.

[6] J. Yu, Y. R, and D. Tao, "Click Prediction for Web Image Reranking using Multimodal Sparse Coding," IEEE Transactions on Image Processing vol.23, no.5, pp. 2019-2032, 2014.

[7] Y. Jing and S. Baluja. "Visualrank: Applying Pagerank to Large-Scale Image Search," IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 30, no. 1, pp. 1877-1890, 2008.

[8] K.-S. Lee, Y.-C. Park, and K.-S. Choi, "Re-ranking Model Based on Document Clusters," Information Processing & Management, vol. 37, no. 1, pp. 1-14, 2001.

[9] R. Yan, A. Hauptmann, and R. Jin, "Multimedia Search with Pseudo-Relevance Feedback," In Image and Video Retrieval, pp. 238-247, 2003.

[10] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua, "Bayesian Video Search Re-ranking," In Proceedings of the 16th ACM international conference on Multimedia, pp. 131-140, 2008.

[11] L. Yang and A. Hanjalic, "Supervised Re-ranking for Web Image Search," In Proceedings of the international conference on Multimedia, pp. 183-192, 2010.

[12] X. Tian, Y. Lu, L. Yang, and Q. Tian, "Learning to Judge Image Search Results," In Proceedings of the 19th ACM international conference on Multimedia, pp. 363-372, 2011.

[13] L. Yang and A. Hanjalic, "Learning to Rerank Web Images," IEEE MultiMedia, pp. 13-21, 2013.

[14] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel, "Handwritten Digit Recognition With a Back-Propagation Network," In Advances in Neural Information Processing Systems, pp. 396-404, 1990.

[15] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet Classification With Deep Convolutional Neural Networks," In Advances in Neural Information Processing Systems, pp. 1097-1105, 2012.

[16] S. Rudinac, M. Larson, and A. Hanjalic, "Exploiting Result Consistency to Select Query Expansions for Spoken Content Retrieval," In Advances in Information Retrieval, pp. 645-648, 2010.

[17] X. Tian, Y. Lu, and L. Yang, "Query difficulty prediction for web image search," Multimedia, IEEE Transactions, vol 14, no. 4, pp. 951-962, 2012.

[18] A.G. Hauptmann and M.G. Christel, "Successful Approaches in the TREC Video Retrieval Evaluations," In Proceedings of the 12th annual ACM international conference on Multimedia, pp. 668-675, 2004.

[19] J. He, M. Larson, and M.D. Rijke, "Using Coherence-Based Measures to Predict Query Difficulty," In Advances in Information Retrieval, pp.689-694, 2008.

[20] G. Salton and C. Buckle, "Term-weighting Approaches in Automatic Text Retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513-523, 1988.

[21] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to The Web," Stanford Univeristy, Stanford, CA, Tech. Report, pp. 1-17 ,1999.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. L, "Imagenet: A Large-Scale Hierarchical Image Database," In Computer Vision and Pattern Recognition, pp. 248-255, 2009.

[23] Y. Luo, D. Tao, B. Geng, C. Xu, and S.J Maybank, "Manifold regularized multitask learning for semi-supervised multilabel image classification," Image Processing, IEEE Transactions, pp. 523-536, 2013.

[24] Y. Luo, D. Tao, B. Geng, C. Xu, and S.J Maybank, "Multiview vector-valued manifold regularization for multilabel image classification," Neural Networks and Learning Systems, IEEE Transactions, pp. 709-722, 2013.

[25] J. Yu, Y. Rui, and B. Chen, "Exploiting Click Constraints and Multi-viewFeatures for Image Reranking," IEEE Transactions on Multimedia, vol. 16, no. 1, pp.159-168, 2014.