

An Attribute-Assisted Reranking Model for Web Image Search

Junjie Cai, Zheng-Jun Zha, *Member, IEEE*, Meng Wang, Shiliang Zhang, and Qi Tian, *Senior Member, IEEE*

Abstract—Image search reranking is an effective approach to refine the text-based image search result. Most existing reranking approaches are based on low-level visual features. In this paper, we propose to exploit semantic attributes for image search reranking. Based on the classifiers for all the predefined attributes, each image is represented by an attribute feature consisting of the responses from these classifiers. A hypergraph is then used to model the relationship between images by integrating low-level visual features and attribute features. Hypergraph ranking is then performed to order the images. Its basic principle is that visually similar images should have similar ranking scores. In this paper, we propose a visual-attribute joint hypergraph learning approach to simultaneously explore two information sources. A hypergraph is constructed to model the relationship of all images. We conduct experiments on more than 1,000 queries in MSRA-MMV2.0 data set. The experimental results demonstrate the effectiveness of our approach.

Index Terms—Search, hypergraph, attribute-assisted.

I. INTRODUCTION

WITH the dramatic increase of online images, image retrieval has attracted significant attention in both academia and industry [31]–[38]. Many image search engines such as Google and Bing have relied on matching textual information of the images against queries given by users. However, text-based image retrieval suffers from essential difficulties that are caused mainly by the incapability of the associated text to appropriately describe the image content.

Recently, visual reranking has been proposed to refine text-based search results by exploiting the visual information contained in the images [1]–[3]. The existing visual reranking methods can be typically categorized into three categories as the clustering based, classification based and graph based methods. The clustering based reranking methods stem from the key observation that a wealth of visual

characteristics can be shared by relevant images. With intelligent clustering algorithms (e.g., mean-shift, K -means, and K -medoids), initial search results from text-based retrieval can be grouped by visual closeness. However, for queries that return highly diverse results or without clear visual patterns, the performance of the clustering-based methods is not guaranteed. In the classification based methods, visual reranking is formulated as binary classification problem aiming to identify whether each search result is relevant or not. Pseudo Relevance Feedback (PRF) is applied to select training images to learn a classifier or a ranking model. However, in many real scenarios, representative examples obtained via PRF for the training dataset are very noisy and might not be adequate for constructing effective classifiers. Graph based methods have been proposed recently and received increasing attention as demonstrated to be effective. The multimedia entities in top ranks and their visual relationship can be represented as a collection of nodes and edges. The local patterns or salient features discovered using graph analysis are very powerful to improve the effectiveness of rank lists. Nevertheless, the reranking algorithms mentioned above are purely based on low-level visual features while generally do not consider any semantic relationship among initial ranked list. The high level semantic concepts which are crucial to capture property of images could deliver more clear semantic messages between various nodes in the graph. Thus, in this paper, we propose to exploit stronger semantic relationship in the graph for image search reranking.

On the other hand, semantic attributes have received tremendous attention recently, where their effectiveness was demonstrated in broad applications, including face verification [6], object recognition [5]–[11], fine-grained visual categorization [16], classification with humans-in-the-loop [22] and image search [4]. Semantic attributes could be shape, color, texture, material, or part of objects, such as “round,” “red,” “mental,” “wheel” and “leg” etc. As a kind of intermediate-level descriptor, an attribute has semantic meaning as opposed to low-level visual features, but it is easy to model compared to a full object, e.g., “car”. Thus, attributes are expected to narrow down the semantic gap between low-level visual features and high-level semantic meanings. Furthermore, attribute-based image representation has also shown great promises for discriminative and descriptive ability due to intuitive interpretation and cross-category generalization property. They describe image regions that are common within an object category but rare outside of it. Hence, attribute-based visual descriptor has achieved good

Manuscript received February 3, 2014; revised May 24, 2014, July 29, 2014, and October 6, 2014; accepted November 14, 2014. Date of publication November 20, 2014; date of current version December 16, 2014. This work was supported by the National Science Foundation of China under Grant 61429201. The work of Q. Tian was supported in part by the U.S. Army Research Laboratory under Grant W911NF-12-1-0057 and in part by the Faculty Research Awards through NEC Laboratories of America. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dimitrios Tzovaras.

J. Cai, S. Zhang, and Q. Tian are with the University of Texas at San Antonio, San Antonio, TX 78249 USA (e-mail: caijunjieustc@gmail.com; slzhang.jdl@gmail.com; qitian@cs.utsa.edu).

Z.-J. Zha is with the Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 100080, China (e-mail: junzzustc@gmail.com).

M. Wang is with the Hefei University of Technology, Hefei 230009, China (e-mail: eric.mengwang@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2014.2372616

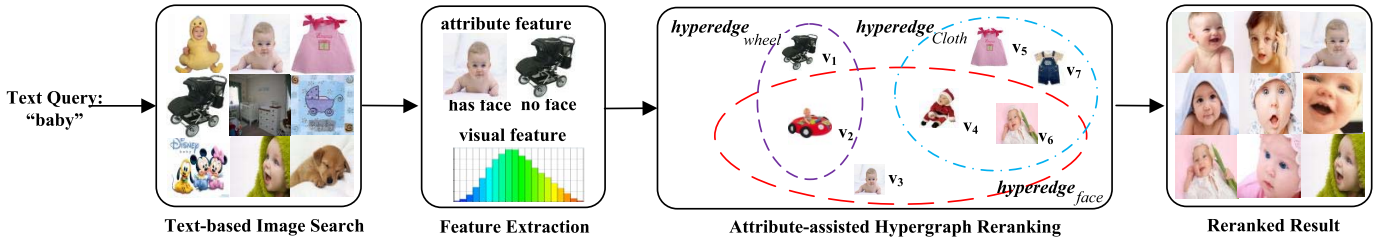


Fig. 1. Flowchart of the proposed attribute-assisted hypergraph reranking method. The search engine returns the images related to the textual query “baby” and then our proposed approach is applied to reorder the result with attribute feature. We show the top-9 ranked images in the text-based search results and the reranked results in the first and last block, respectively.

performance in assisting the task of image classification. Besides that, an attribute is potentially any visual property that humans can precisely communicate or understand, even if it does not correspond to a traditionally-defined object part. For instance, “red-dot in center of wings” is a valid attribute, even though there is not a single butterfly part that corresponds to it. Furthermore, the type of the most effective features should vary across queries. For example, for queries that are related to color distribution, such as sunset, sunrise and beach, color features will be useful. For some queries like building and street, edge and texture features will be more effective. It can be understood that semantic attribute could also be viewed a description or modality of image data. Using multimodal features can guarantee that the useful features for different queries are contained. Therefore, all these superiorities drive us to exploit semantic attributes for image representation in the task of web image search reranking.

Motivated by the above observations, we move one step ahead of visual reranking and propose an attribute-assisted reranking approach. Fig. 1 illustrates the flowchart of our proposed method. After a query “baby” is submitted, an initial result is obtained via a text-based search engine. It is observed that text-based search often returns “inconsistent” results. Some visually similar images are scattered in the result while other irrelevant results are filled between them, such as “dog” and “disney baby”. Based on the returned images, both visual features and attribute features are extracted. In particular, the attribute feature of each image consists of the responses from the binary classifiers for all the attributes. These classifiers are learned offline. Visual representation and semantic description are simultaneously exploited in a unified model called *hypergraph*. The preliminary version of this work, which integrates attribute feature and visual feature to improve the reranking performance is published in [28]. In this paper, we extend this work to analyze semantic attributes characteristic and discover that only limited attributes distributed in each image. Hence we propose that the selection of attribute features could be conducted simultaneously through the process of hypergraph learning such that the effects of semantic attributes could be further tapped and incorporated in the reranking framework. Compared with the previous method, a hypergraph is reconstructed to model the relationship of all the images, in which each vertex denotes an image and a hyperedge represents an attribute and a hyperedge connects to multiple vertices. We define the weight of each edge based on the

visual and attribute similarities of images which belongs to the edge. The relevance scores of images are learned based on the hypergraph. The advantage of hypergraph can be summarized that not only does it take into account pairwise relationship between two vertices, but also higher order relationship among three or more vertices containing grouping information. Essentially, modeling relationship among more close samples will be able to preserve the stronger semantic similarity and thus facilitate ranking performance.

However, even these merits, we have to face the challenges it brings to this unified formulation at the same time: 1) simply learning classifiers by fitting them to all visual features often fails to generalize the semantics of the attributes correctly because low-level features are extracted by region or interest point detector instead of aiming to depict specific attribute. So we need to select representative features which are in favor to describe current semantic attributes. We propose to take advantage of ℓ_1 regularized logistic regression trained for each attribute within each class. 2) as attribute features are formed by prediction of several classifiers, semantic description of each image might be inaccurate and noisy. Hence we propose a regularizer on the hyperedge weights which performs a weighting or selection on the hyperedges. In this way, for attributes or hyperedges that are informative, higher weights will be assigned. In contrast, noisy hyperedges will be implicit removed when the weights converges to zeros after hypergraph learning. Finally, we can obtain the reranked list of the images with respect to relevance scores in descending order. We conduct experiments on MSRA-MM V2.0 dataset [7]. The experimental results demonstrate superiority of the proposed attribute-assisted reranking approach over other state-of-the-art reranking methods and their attribute-assisted variants.

The main contribution of this paper can be summarized as follows:

- We propose a new attribute-assisted reranking method based on hypergraph learning. We first train several classifiers for all the pre-defined attributes and each image is represented by attribute feature consisting of the responses from these classifiers. Different from the existing methods, a hypergraph is then used to model the relationship between images by integrating low-level features and attribute features.
- We improve the hypergraph learning method approach presented in [28] by adding a regularizer on the hyperedge weights which performs an implicit selection on the semantic attributes. This makes our approach much more

robust and discriminative for image representation as noisy attributes will be removed and informative ones will be selected.

- We conduct comprehensive experiments to empirically analyze our method on more than 1,000 queries and 1 million images. The experimental results validate the effectiveness of our method.

The rest of the paper is organized as follows. Firstly, we provide a review of the related work on Web image search reranking, semantic attribute and hypergraph learning in Section II. The proposed attribute-assisted reranking approach is elaborated in Section III. Specifically, we elaborate the image features and attribute learning methods in Section III-A and Section III-B, respectively. The proposed hypergraph construction strategy is detailed in Section III-C. In Section IV, we report our experimental results on a Web image dataset: MSRA-MM V2.0 dataset, followed by the conclusions in Section V.

II. RELATED WORK

In this section, we provide a brief description of the existing visual search reranking approaches, review the semantic attributes exploited in recent literature, and describe the hypergraph learning theory.

A. Web Image Search Reranking

Web image search reranking is emerging as one of the promising techniques for automative boosting of retrieval precision [2]. The basic functionality is to reorder the retrieved multimedia entities to achieve the optimal rank list by exploiting visual content in a second step. In particular, given a textual query, an initial list of multimedia entities is returned using the text-based retrieval scheme. Subsequently, the most relevant results are moved to the top of the result list while the less relevant ones are reordered to the lower ranks. As such, the overall search precision at the top ranks can be enhanced dramatically. According to the statistical analysis model used, the existing reranking approaches can roughly be categorized into three categories including the clustering based, classification based and graph based methods.

1) *Clustering-Based Methods*: Clustering analysis is very useful to estimate the inter-entity similarity. One good example of clustering based reranking algorithms is the Information Bottle based scheme developed by Hsu *et al.* [9]. In this method, the images in the initial results are primarily grouped automatically into several clusters. Then the re-ranked result list is created first by ordering the clusters according to the cluster conditional probability and next by ordering the samples within a cluster based on their cluster membership value. In [18], a fast and accurate scheme is proposed for grouping Web image search results into semantic clusters. It is obvious that the clustering based reranking methods can work well when the initial search results contain many near-duplicate media documents. However, for queries that return highly diverse results or without clear visual patterns, the performance is not guaranteed.

2) *Classification-Based Methods*: In the classification based methods, visual reranking is formulated as binary classification problem aiming to identify whether each search result is relevant or not. For instance, a classifier or a ranking model is learned with the pseudo relevance feedback (PRF) [12]. However, in many real scenarios, training examples obtained via PRF are very noisy and might not be adequate for training effective classifier. To address this issue, Schroff *et al.* [3] learned a query independent text based re-ranker. The top ranked results from the text based reranking are then selected as positive training examples. Negative training examples are picked randomly from the other queries. A binary SVM classifier is then used to re-rank the results on the basis of visual features.

3) *Graph-Based Methods*: Graph based methods have been proposed recently and received increased attention as demonstrated to be effective. Jing and Baluja proposed a VisualRank framework to efficiently model similarity of Google image search results with graph [19]. The framework casts the reranking problem as random walk on an affinity graph and reorders images according to the visual similarities. The final result list is generated via sorting the images based on graph nodes' weights. In [2], Tian *et al.* presented a Bayesian reranking framework formulating the reranking process as an energy minimization problem. The objective is to optimize the consistency of ranking scores over visually similar samples and minimize the inconsistency between the optimal list and the initial list. Thus, the performance is significantly dependent on the statistical properties of top ranked search results. Motivated by this observation, Wang *et al.* proposed a semi-supervised framework to refine the text based image retrieval results via leveraging the data distribution and the partial supervision information obtained from the top ranked images [20].

B. Semantic Attributes

Semantic attributes can be regarded as a set of mid-level semantic preserving concepts. Different from low-level visual features, each attribute has an explicit semantic meaning, *e.g.*, "animals". Attribute concepts also differ from specific semantics since they are relatively more general and easier to model, *e.g.*, attributes "animal" and "car" are easier to model and distinguish than the concrete semantic concepts "Husky" and "Gray Wolves". Due to the advantages of being semantic-aware and easier to model, attributes have been studied recently and are revealing their power in various applications such as object recognition [5], [6], [11], and image/video search [4]. Thus, attributes are expected to narrow down the semantic gap between low-level visual features and high-level semantic meanings. By using attribute classifiers, Su *et al.* [27] propose to alleviate the semantic gap between visual words and high level concept, focusing on polysemy phenomenon of particular visual words. By randomly splitting the training data, Farhadi *et al.* [5] exhaustively train thousands of classifiers and then chose some of the discriminative ones as attributes (*e.g.*, attributes that "cat" and "dog" have but "sheep" and "horse" do not). Kumar *et al.* [6] define a set of binary attributes called

TABLE I
NOTATIONS AND DEFINITIONS

Notation	Definition
$\mathcal{X}=(x_1, x_2, \dots, x_n)$	\mathcal{X} indicates the image set, and x_i indicates the i -th image.
$G = (V, E, w)$	G indicates a hypergraph, and V , E and w indicate the set of vertices, the set of edges, and the weights of hyperedges, respectively.
V	The vertex set of the hypergraph
E	The edge set of the hypergraph
W	The diagonal weight matrix and its (i, i) -th element is the weight of the i -th hyperedge.
D	The mean value of elements in the distance matrix.
$\delta(e)$	The degree of edge e .
\mathbf{D}_v	The diagonal matrix of the vertex degrees.
\mathbf{D}_e	The diagonal matrix of the hyperedge degrees.
\mathbf{H}	The incident matrix of the hypergraph.
y	The label vector for hypergraph learning.
f	The relevance score to be learned.

smiles for face verifications. Each attribute detector in *smiles* are exclusively trained for one specific category, *e.g.*, “the Angelina Jolie’s mouth”. However, such category-specific attribute detectors are contrary to the spirit of attributes, *e.g.*, concept that is generalizable and transferrable. Recently, Parikh and Grauman [17] propose a new strategy to compare the relative strength of attributes, *e.g.*, “while A and B are both *shiny*, A is *shinier* than B”. Instead of being trained by binary classifiers, relative attributes are learned by ranking functions (*e.g.*, the ranking SVM). The output of a ranking function indicates the relative presence of the corresponding attribute. In addition, there are works aiming to discover attributes either automatically through the web or semi-automatically with human in the loop [16], and to explore new types of attributes [17]. To date, they explore active learning strategies for training relative attribute ranking functions, such as Learning to rank model to achieve relative attributes [29]. Zhang et al. [30] proposed an attribute-augmented semantic hierarchy for content-based image retrieval. They employ attributes to describe the multiple facts of the concept and hybrid feedbacks of attributes and images are also collected. Such superiority motivates us to exploit attributes for visual reranking.

C. Hypergraph Learning

Before presenting our approach, we first briefly introduce the hypergraph learning theory.

In a simple graph, samples are represented by vertices and an edge links the two related vertices. Learning tasks can be performed on a simple graph. Assuming that samples are represented by feature vectors in a feature space, an undirected graph can be constructed by using their pairwise distances, and graph-based semi-supervised learning approaches can be performed on this graph to categorize objects. It is noted that this simple graph cannot reflect higher-order information. Compared with the edge of a simple graph, a hyperedge in a hypergraph is able to link more than two vertices. For clarity, we first illustrate several important notations and their definitions throughout the paper in Table 1.

A hypergraph $G = (V, E, w)$ is composed by a vertex set V , an edge set E , and the weights of the edges w . Each edge e is given a weight $w(e)$. The hypergraph G can be

denoted by a $|V| \times |E|$ incidence matrix \mathbf{H} with entries defined as:

$$h(v_i, e_j) = \begin{cases} 1 & \text{if } v_i \in e_j, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

For a vertex $v \in V$, its vertex degree can be estimated by:

$$d(v) = \sum_{e \in E} w(e)h(v, e) \quad (2)$$

For a hyperedge $e \in E$, its hyperedge degree can be estimated by:

$$\delta(e) = \sum_{v \in V} h(v, e) \quad (3)$$

We use D_v and D_e to denote the diagonal matrices containing the vertex and hyperedge degrees respectively, and let W denote the diagonal matrix containing the weights of hyperedges.

$$W(i, j) = \begin{cases} w(i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In a hypergraph, many machine learning tasks can be performed, *i.e.* clustering [15] and classification [13]. The binary classification is taken as an example here. For hypergraph learning, the Normalized Laplacian method proposed in [15] is employed, and it is formulated as a regularization framework:

$$\arg \min_f \{\lambda R_{emp}(f) + \Omega(f)\} \quad (5)$$

where f is the relevance score to be learned, $\Omega(f)$ is the normalized cost function, $R_{emp}(f)$ is empirical loss, and λ is a weighting parameter. By minimizing this cost function, vertices sharing many incidental hyperedges are guaranteed to obtain similar relevance scores. The regularizer on the hypergraph is defined as:

$$\Omega(f) = \frac{1}{2} \sum_{e \in E} \sum_{u, v \in e} \frac{w(e)h(u, e)h(v, e)}{\delta(e)} \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2 \quad (6)$$

Let $\Theta = D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}}$, and $\Delta = \mathbf{I} - \Theta$. Here the normalized cost function can be rewritten as:

$$\Omega(f) = f^T \Delta f \quad (7)$$

Here Δ is a positive semi-definite matrix called hypergraph Laplacian.

III. ATTRIBUTE-ASSISTED IMAGE SEARCH RERANKING

In this section, we elaborate the proposed attribute-assisted image search reranking framework. We elaborate image features in Section III-A, and then introduce the proposed attribute learning method in Section III-B. Finally, we describe our hypergraph construction algorithm in Section III-C.

A. Image Features

We used four types of features, including color and texture, which are good for material attributes; edge, which is useful for shape attributes; and scale-invariant feature transform (SIFT) descriptor, which is useful for part attributes. We used a bag-of-words style feature for each of these four feature types.

Color descriptors were densely extracted for each pixel as the 3-channel LAB values. We performed K-means clustering with 128 clusters. The color descriptors of each image were then quantized into a 128-bin histogram. Texture descriptors were computed for each pixel as the 48-dimensional responses of texon filter banks. The texture descriptors of each image were then quantized into a 256-bin histogram. Edges were found using a standard canny edge detector and their orientations were quantized into 8 unsigned bins. This gives rise to a 8-bin edge histogram for each image. SIFT descriptors were densely extracted from the 8×8 neighboring block of each pixel with 4 pixel step size. The descriptors were quantized into a 1,000-dimensional bag-of-words feature.

Since semantic attributes usually appear in one or more certain regions in an image, we further split each image into 2×3 grids and extracted the above four kinds of features from each grid respectively. Finally, we obtained a 9,744-dimensional feature for each image, consisting of a $1,392 \times 6$ -dimensional feature from the grids and a 1,392-dimensional feature from the image. This feature was then used for learning attribute classifiers.

B. Attribute Learning

We learn a Support Vector Machine (SVM)¹ classifier for each attribute. However, simply learning classifiers by fitting them to all visual features often fails to generalize the semantics of the attributes correctly. For each attribute, we need to select the features that are most effective in modeling this attribute. It is necessary to conduct this selection based on the following two observations: 1) such a wealth of low level features are extracted by region or interest point detector, which means these extraction may not aim to depict the specific attribute and include redundant information. Hence we need select representative and discriminative features which are in favor to describe current semantic attributes. 2) the process of selecting a subset of relevant features has been playing an important role in speeding up the learning process and alleviating the effect of the *curse of dimensionality*. We here apply the feature selection method as described in [5]. In particular, if we want to learn a “wheel” classifier, we select features that perform well at distinguishing examples of cars with

“wheels” and cars without “wheels”. By doing so, we help the classifier avoid being confused about “metallic”, as both types of example for this “wheel” classifier have “metallic” surfaces. We select the features using an ℓ_1 -regularized logistic regression trained for each attribute within each class, then pool examples over all classes and train using the selected features. Such regression model is utilized as the preliminary classifiers to learn sparse parameters. The features are then selected by pooling the union of indices of the sparse non-zeros entries in those parameters. The regularization parameters of ℓ_1 -norm regression was set to 0.01 empirically and the parameters of SVM classifiers were determined by five-fold cross validation. For example, we first select features that are good at distinguishing cars with and without “wheel” by fitting an ℓ_1 -regularized logistic regression to those examples. We then use the same procedure to select features that are good at separating motorbikes with and without wheels, buses with and without wheels, and trains with and without wheels. We then pool all those selected features and learn the “wheel” classifier over all classes using those selected features. In this way, we select effective features for each attribute and the selected features are then used for learning the SVM classifier.

C. Attribute-Assisted Hypergraph Construction

We propose an attribute-assisted hypergraph learning method to reorder the ranked images which returned from search engine based on textual query. Different from the typical hypergraph [10], [25], [26], it presents not only whether a vertex v belongs to a hyperedge e , but also the prediction score that v is affiliated to a specific e . The weight is incorporated into graph construction as tradeoff parameters among various features. Our modified hypergraph is thus able to improve reranking performance by mining visual feature as well as attribute information.

The hypergraph model has been widely used to exploit the correlation information among images. In this paper, we regard each image in the data set as a vertex on hypergraph $G = (V, E, w)$. Assume there are n images in the data set, and thus, the generated hypergraph contains n vertices. Let $V = \{v_1, v_2, \dots, v_n\}$ denote n vertices and $E = \{e_1, e_2, \dots, e_m\}$ represent m hyperedges where the images sharing the same attribute are connected by one hyperedge. For various hyperedges, we set the weight vector to be $w = [w_1, w_2, \dots, w_m]$ in the hypergraph, where $\sum_{i=1}^m w_i = 1$. In each hyperedge, we select K images which offer more preference to corresponding attribute based on the descending order of classifier scores. So, the size of a hyperedge in our framework is K . The incidence matrix H of a hypergraph is defined as follows:

$$h(v_i, e_j) = \begin{cases} s_{v_i} & \text{if } v_i \in e_j, \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where s_{v_i} represents the j -attribute classifier score of image v_i . Then for a vertex $v \in V$, we define vertex degree to be $d(v) = \sum_{e \in E} w(e)h(v, e)$. For a hyperedge $e \in E$, its degree is defined as $\delta(e) = \sum_{v \in V} h(v, e)$. We use D_v and D_e to denote the diagonal matrices containing the vertex and hyperedge degrees respectively, and let W denote the diagonal

¹<http://www.csie.ntu.edu.tw/~cjlin/libsvm>

matrix containing the weights of hyperedges. For the weight initialization, we define the similarity matrix A between two images as follows:

$$\begin{aligned} A(i, j) &= \gamma_1 A_{\text{attribute}} + \gamma_2 A_{\text{local}} + \gamma_3 A_{\text{global}} \\ &= \gamma_1 \exp\left(-\frac{Dis_{\text{attribute}}(i, j)}{\bar{D}_{\text{attribute}}}\right) + \gamma_2 \exp\left(-\frac{Dis_{\text{local}}(i, j)}{\bar{D}_{\text{local}}}\right) \\ &\quad + \gamma_3 \exp\left(-\left(\sum_{k=1}^3 \frac{Dis_{gk}(i, j)}{\bar{D}_{gk}}\right)\right) \end{aligned} \quad (9)$$

where $Dis_g(i, j)$, $Dis_{\text{local}}(i, j)$ and $Dis_{\text{attribute}}(i, j)$ are pairwise Euclidean distance between v_i and v_j of global feature, local feature and attribute feature, respectively. For instance, $Dis_{\text{attribute}}(i, j)$ can be denoted as:

$$Dis_{\text{attribute}}(i, j) = \|\mathbf{i}_a - \mathbf{j}_a\|_{\ell_2} = \sqrt{(\mathbf{i}_a - \mathbf{j}_a)^T (\mathbf{i}_a - \mathbf{j}_a)} \quad (10)$$

\mathbf{i}_a and \mathbf{j}_a indicate semantic attribute features of two images. After all the distance matrices for three kinds of features are obtained, the similarity matrix A between two images can be obtained and $\gamma_1, \gamma_2, \gamma_3$ are normalizing constants to keep the proportion of various features to be 1. And \bar{D} is the mean value of elements in the distance matrix. The initial weight $w(e_i)$ is set to $w(e_i) = \sum_{v_j \in e_i} A(i, j)$.

For hypergraph learning, it is formulated as a regularization framework:

$$\arg \min_{f, w} \{\lambda R_{\text{emp}}(f) + \Omega(f) + \mu \Psi(w)\} \quad (11)$$

where f is the relevance score to be learned, $\Omega(f)$ is the normalized cost function, $R_{\text{emp}}(f)$ is empirical loss and $\Psi(w)$ is a regularizer on the weights. Instead of fixing hyperedge weights, we assume that they have Gaussian distribution, such that the weights w can be learned together with the relevance score f .

By minimizing this cost function, vertices sharing many incidental hyperedges are guaranteed to obtain similar relevance scores. Defining $\Theta = D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}}$, we can derive the cost function as

$$\begin{aligned} \Omega(f) &= \frac{1}{2} \sum_{e \in E} \sum_{u, v \in e} \frac{w(e) h(u, e) h(v, e)}{\delta(e)} \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2 \\ &= \frac{1}{2} \sum_{e \in E} \sum_{u, v \in e} \frac{w(e) h(u, e) h(v, e)}{\delta(e)} \left(\frac{f^2(u)}{\sqrt{d(u)}} - \frac{f(u) f(v)}{\sqrt{d(u) d(v)}} \right) \\ &= \sum_{u \in V} f^2(u) \sum_{e \in E} \frac{w(e) h(u, e)}{d(u)} \sum_{v \in V} \frac{h(v, e)}{\delta(e)} \\ &\quad - \sum_{e \in E} \sum_{u, v \in e} \frac{f(u) h(u, e) w(e) h(v, e) f(v)}{\sqrt{d(u) d(v)} \delta(e)} \\ &= f^T (I - \Theta) f \\ &= f^T \Delta f \end{aligned} \quad (12)$$

where I is the identity matrix, $\Delta = I - \Theta$ is a positive semi-definite matrix called hypergraph Laplacian. To force the

assigned relevance score to approach the initial relevance score of y , a regularization term is defined as follows:

$$R_{\text{emp}}(f) = \|f - y\|^2 = \sum_{u \in V} (f(u) - y(u))^2 \quad (13)$$

In the constructed hypergraph, the hyperedges are with different effects as there exists a lot of uninformative attributes for a given query. Therefore, performing a weighting or selection on the hyperedges will be helpful. Here we integrate the learning of the hyperedge weights into the formulation.

$$\Psi(w) = \|w\|_q \quad (14)$$

Different from the conventional hypergraph learning described in Section II-C, we add a ℓ_2 norm regularizer on the weights $\|w\|_{\ell_2}^2$. This strategy is popular to avoid overfitting [13]. The regularization framework then becomes

$$\arg \min_{f, w} \Phi(f) = \arg \min_{f, w} \{f^T \Delta f + \lambda \|f - y\|^2 + \mu \sum_{i=1}^m w_i^2\} \quad (15)$$

where λ, μ are the regularization parameters. Although the function is not jointly convex with respect to f and w , the function is convex with respect to f with fixed w or vice versa. Thus, we can use the alternating optimization to efficiently solve the problem.

We first fix w by differentiating $\Phi(f)$ with respect to f and we have:

$$\frac{\partial \Phi(f)}{\partial f} = \Delta f + \lambda(f - y) = 0 \quad (16)$$

Following some simple algebraic steps, we have

$$f = \frac{\lambda}{1 + \lambda} (\mathbf{I} - \frac{1}{1 + \lambda} \Theta)^{-1} y \quad (17)$$

We define $\alpha = 1/(1 + \lambda)$. Noticing that $\lambda/(1 + \lambda)$ is a constant and does not change the ranking results, we can rewrite f as follows:

$$f = (1 - \alpha)(I - \alpha \Theta)^{-1} y \quad (18)$$

Note that, the matrix $\mathbf{I} - \alpha \Theta$ is highly sparse. Therefore, the computation can be very efficient. We then fix f and minimize with respect to w ,

$$\begin{aligned} \arg \min_w \Phi(f) &= \arg \min_w \{f^T \Delta f + \mu \sum_{i=1}^m w_i^2\} \\ \text{s.t. } \sum_{i=1}^m w_i &= 1, \quad \mu > 0 \end{aligned} \quad (19)$$

We adopt alternating optimization to solve the above problem. Each time we only update one set of variables (we have $m + 2$ sets of variables in all). By using a Lagrange multiplier η to take the constraint $\sum_{i=1}^m w_i = 1$ into consideration, we get the objective function:

$$\arg \min_{w, \eta} \Phi(f) = \arg \min_{w, \eta} \{f^T \Delta f + \mu \sum_{i=1}^m w_i^2 + \eta (\sum_{i=1}^m w_i - 1)\} \quad (20)$$

Algorithm 1 Attribute-Assisted Hypergraph Learning**Step 1:** Initialization.1.1 Set W as a diagonal matrix with initial values.1.2 Construct the hypergraph Laplacian Δ and compute the matrices D_v, D_e and H accordingly.**Step 2:** Label Update.Compute the optimal f based on the equation 17, which is:

$$f = (1 - \alpha)(I - \alpha\Theta)^{-1}y$$

Step 3: Weight Update.Update the weights w_i with the iterative gradient descent method introduced.**Step 4:**After obtaining W , update the matrix Θ accordingly.**Step 5:**Let $t = t + 1$. If $t > T$, quit iteration and output the results, otherwise go to step 2.

By setting the derivative of $\Phi(f)$ with respect to w_i , we can obtain,

$$\begin{aligned} \frac{\partial \Phi(f)}{\partial w_i} &= \frac{\partial}{\partial w_i} f^T (I - D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}}) f + \mu \sum_{i=1}^m w_i^2 \\ &\quad + \eta \left(\sum_{i=1}^m w_i - 1 \right) \\ &= -f^T D_v^{-\frac{1}{2}} H D_e^{-1} H^T D_v^{-\frac{1}{2}} f + 2m\mu + m\eta = 0 \end{aligned} \quad (21)$$

Let $\Gamma = D_v^{-\frac{1}{2}} H$, we have

$$\eta = \frac{f^T \Gamma D_e^{-1} \Gamma^T f - 2\mu}{m} \quad (22)$$

and

$$\begin{aligned} w_i &= -\frac{-f^T \Gamma D_e^{-1} \Gamma^T f + \eta}{2\mu} \\ &= -\frac{-m f^T \Gamma D_e^{-1} \Gamma^T f + f^T \Gamma D_e^{-1} \Gamma^T f - 2\mu}{2\mu m} \end{aligned} \quad (23)$$

In the aforementioned alternating optimization process, since each step decreases the objective function, which has a lower bound 0, the convergence of the alternating optimization is guaranteed. Algorithm 1 summarizes the implementation of the alternating optimization.

D. Utilization of Text-Based Search Prior

Since in reranking, the text-based search provides original ranking lists instead of quantized scores, a necessary step is to turn the ranking positions into scores. Traditional methods usually associate y_i with the position i using heuristic strategies, such as normalized rank $y_i = 1 - i/N$ or rank $y_i = N - i$. In this work, we investigate the relationship between y_i and the position i with a large number of queries. Actually, we can define

$$y_i = E_{q \in Q}[q, i] \quad (24)$$

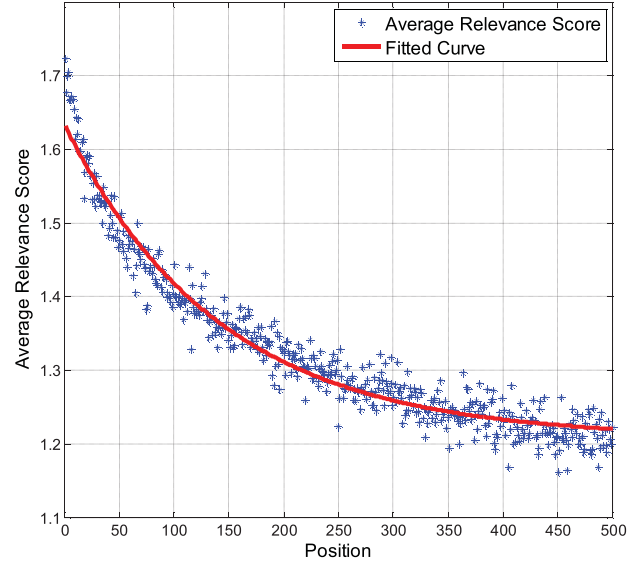


Fig. 2. The average relevance scores at different ranking positions and the fitted curve [26].

where Q means the set of all possible queries, $E_{q \in Q}$ means the expectation over the query set Q , and $y(q, i)$ indicates the relevance ground truth of the i -th search result for query q . Therefore, the most intuitive approach is to estimate y_i by averaging $y(q, i)$ over a large query set. Fig. 2 illustrates the results obtained by using more than 1,000 queries. Details about the queries and the dataset will be introduced in Section IV. However, as shown in Fig. 2, the average relevance score curve with respect to the ranking position is not smooth enough even after using more than 1,000 queries. A prior knowledge can be that the expected relevance score should be decreasing with respect to ranking position. Therefore, we further smooth the curve with a parametric approach. We assume $y = a + be^{-i/c}$ and then fit this function with the non-smooth curve. In this way, with mean squared loss criterion we estimate the parameters a , b and c to be 1, 0.4, 141, respectively. Fig. 2 shows the fitted curve, and we can see that it reasonably preserves the original information.

IV. EXPERIMENTS

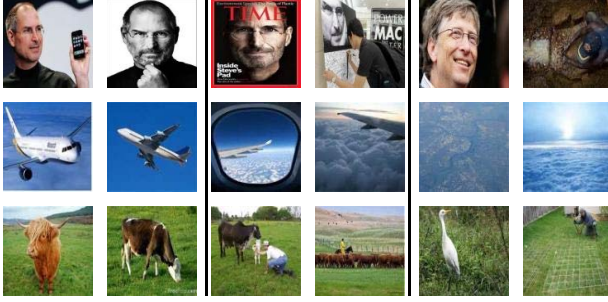
In this section we first describe the experimental dataset that we used to evaluate our proposed approach. Then we present the results of evaluation at different levels and verify the effectiveness of our method.

A. Dataset

We use the MSRA-MM V2.0 dataset as our experimental data. This dataset consists of about 1 million images from 1,097 diverse yet representative queries collected from the query log of Bing [7]. We choose this dataset to evaluate our approach for the following reasons: (1) it is a real-world web image dataset; (2) it contains the original ranking information of a popular search engine, and thus we can easily evaluate whether our approach can improve the performance of the search engine; (3) it is publicly available. There are roughly

TABLE II
THE STATISTICAL INFORMATION ABOUT IMAGE DATASET

Category	Number of Queries	Examples
Animal	100	Alligator, Bat, Cattle
Cartoon	92	Air gear, Final fantasy
Event	78	Olympic, Wedding, WWE
Object	295	Airplane, Bed, Toy
People	68	Girls, Snowman, Baby
Person	40	Tom Hanks, Will Smith
Scene	48	Desert, Rainbow
Time08	88	Barack Obama, Steve Jobs
Misc	288	Japan, Titanic, Adidas



Very Relevant Relevant Irrelevant

Fig. 3. Several example images with different relevance levels with respect to “Steve Jobs”, “Airplane”, “Cattle”.

900 images for each query. For each image, its relevance to the corresponding query is labeled with three levels: very relevant, relevant and irrelevant. These three levels are indicated by scores 2, 1 and 0, respectively. Table II shows the number of queries at various categories and Fig. 3 illustrates several example images of “Steve Jobs”, “Airplane”, “Cattle” with different relevance levels. The queries are used for reranking performance evaluation. We first randomly select 40 queries from two categories “animal” and “object”. All the very relevant images within these queries are kept for attribute classifiers training. We defined 108 attributes by referring to the attributes in [14] as listed in Table III, such as “Beak”, “Engine”. The groundtruth of attributes are manually annotated. As aforementioned in Section III-A, there are four types of features extracted including color, texture, edge and SIFT descriptor. We adopt a bag-of-words style for each of these four features and obtained 9,744-dimensional feature for each image.

B. Evaluation Measure

We adopt Normalized Discounted Cumulative Gain (NDCG) [8], which is a standard evaluation in information retrieval when there are more than two relevance levels, to measure the performance. Given a ranking list, the NDCG score at position n is defined as,

$$NDCG@n = Z_n \sum_{j=1}^n \frac{2^{r(j)} - 1}{\log(1 + j)} \quad (25)$$

where $r(j)$ is the the relevance score of j th image in the ranking list, Z_n is the normalization factor which is chosen to

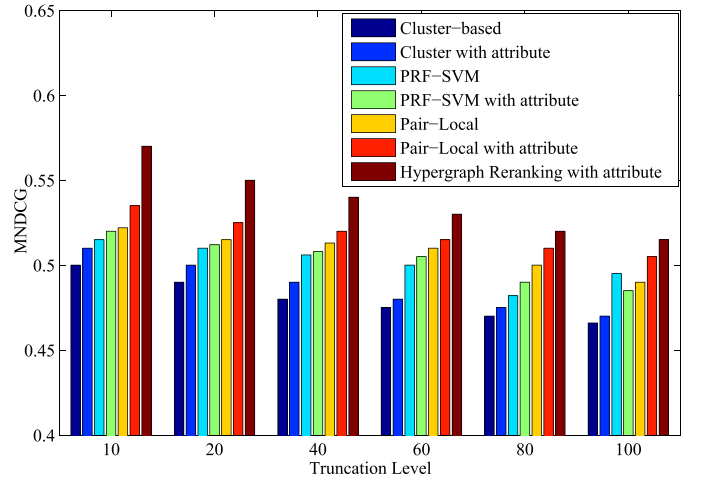


Fig. 4. Performance comparison between our method and conventional methods on MSRA-MM dataset.

guarantee the perfect ranking list’s NDCG@ n is 1, n is truncation level. To evaluate the overall performance, we average the NDCGs over all queries to obtain the Mean NDCG.

C. Experimental Results and Analysis

We first evaluate the effectiveness of our attribute classifiers on the 1,097 testing queries. Regarding the outputs of each attribute classifier on all the images, we use both binary classification decision and continuous confidence scores. The binary outputs are used to evaluate the classifier performance and the confidence score are used to calculate probabilistic hyperedge. Due to the high cost of manual labeling, we only label the 108 attributes on the top 10 images from initial text search for each query. We adopt the widely used metric AUC (area under ROC curve) value for evaluating the accuracies of attribute classifiers. The true positive (tp) and true negative (tn) are divided by the total number of samples (n) and classification accuracy is denoted as:

$$Accuracy = \frac{tp + tn}{n} \quad (26)$$

Fig. 5 shows classification accuracy on each attribute. The experimental results demonstrate that semantic description of each image might be inaccurate and noisy, as the attribute features are formed by prediction of several classifiers.

1) *Performance Comparison Among Cluster-Based, Classification-Based, Graph-Based:* To verify the effectiveness of the proposed attribute-assisted reranking method, we compare the following approaches for performance evaluation:

- Information Bottle [9]. The reranking approach applies information bottleneck clustering over visual features with the help of a smoothed initial ranking. The method is denoted as “cluster-based”.
- Pseudo Relevance Feedback (PRF) [12]. Given a query, we use top 100 search results in the original ranking list as positive samples, and randomly collect 100 images from the whole dataset and use them as negative samples. We adopt RBF kernel based on these samples and learn

TABLE III
108 SEMANTIC ATTRIBUTES USED IN THE MSRA-MM DATASET

Ear	<i>Hand</i>	Knit	Glass	Cylinder	Stripped	StemorTrunk	Minicomputer	Single Doer Action
Eye	Foot	Round	Paper	Exhaust	Propeller	Spider Hole	Tactical Line	Arriving At a Place
<i>Arm</i>	<i>Wing</i>	Snout	Shiny	FurnLeg	Jetengine	Cloudy Area	Measuring Cup	Terrain High Ground
<i>Leg</i>	Door	Mouth	Square	FurnArm	<i>Headlight</i>	Power Plant	Ballet Dancer	Vertical Protrusion
Bar	Sail	Torso	Hollow	Plastic	Taillight	Septic Tank	Rescue Vehicle	Homogeneous Texture
<i>Tail</i>	Mast	<i>Wheel</i>	Window	<i>Feather</i>	ToughSkin	Rice Cooker	Celestial Event	Geographical Region
<i>Beak</i>	Text	Pedal	Engine	Leathre	HumanSkin	Sanatoriums	Destroyer Ship	Domesticated Animal
Head	<i>Rock</i>	Label	<i>Smooth</i>	Triangle	Transpare	<i>Vehicle Part</i>	Computer Joystick	Vertical Protrusion
<i>Nose</i>	<i>Leaf</i>	Metal	<i>Flower</i>	Row Wind	Sidemirror	Communicator	Campaigning Event	Flow Control Device
Hair	Plot	<i>Cloth</i>	Screen	FurnBack	Handlebars	Fire Station	Single Family Home	Semiconductor Device
<i>Face</i>	<i>Wood</i>	<i>Furry</i>	<i>Dotted</i>	FurnSeat	Vegetation	Retail Store	Overhead Projector	Reconnaissance Plane
Snow	Bird	River	Dancer	Chimney	Sound Card	Bush Pilot	Partially Tangible	Master Sergeant Rank

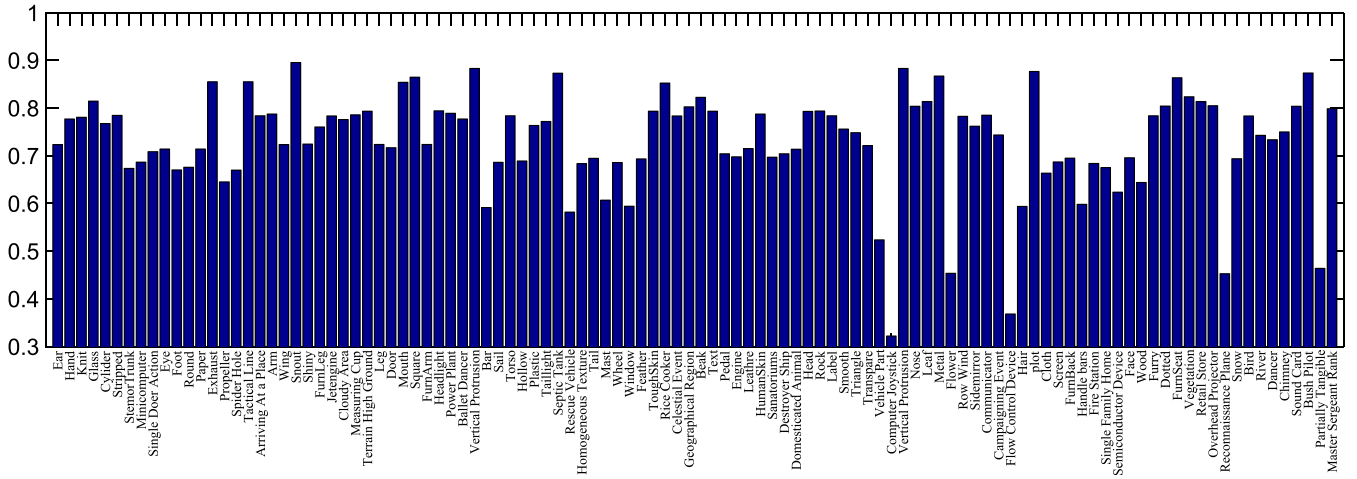


Fig. 5. Attribute classification comparison on MSRA-MM dataset.

SVM classifier to rerank the search results. The approach is denoted as “PRF-SVM”.

- Bayesian Reranking [2]. Since the Local-Pair variant of Bayesian reranking [2] performs the best among the six Bayesian variants reranking approaches, we will use it as the representative of Bayesian reranking methods. The method is denoted as “Pair-Local”.

We fixed the weights in Equation 9 as follows: $\gamma_1, \gamma_2, \gamma_3$ all equal to $1/3$ when attribute feature has been explored; $\gamma_2 = \gamma_3 = 1/2$ and $\gamma_1 = 0$ when independent visual feature has been evaluated in the experiment.

The overall performance is shown in Fig. 4. From this result, we can see that the proposed attribute-assisted reranking method consistently outperforms Bayesian reranking at all truncation levels. Specifically, our method obtains 7.3% and 3.9% relative improvements on MNDG@100 compared with the Bayesian reranking. In addition, for attributed-assisted Pair-Local method, it has relatively improved 1.8% on MNDG@40 in comparison with original reranking approach. The main reason is that our method involves beneficial attribute features throughout the reranking framework. Additionally, we also explore and create visual neighborhood relationship for each hyperedge instead of isolated visual similarity mainly used in the Bayesian reranking.

TABLE IV
PERFORMANCE COMPARISON FOR HYPERGRAPH RERANKING

MNDG	20	40	60	80
Text baseline	0.45	0.43	0.42	0.40
Hypergraph [35]	0.53	0.52	0.51	0.50
Hypergraph with ℓ_1 regularizer	0.53	0.51	0.49	0.48
Hypergraph with ℓ_2 regularizer	0.55	0.54	0.53	0.52

2) Performance Comparison for Hypergraph Reranking:

In the above, we have verified that our proposed hypergraph reranking approach performs the best in comparison with the conventional reranking strategies. In this section, we will further confirm the superiority of joint hypergraph learning approach by adding a robust regularizer on hyperedge weights. We denote the method published in preliminary work [28] as “Hypergraph Reranking”, where we concatenate all features into a long vector and construct hypergraph based on their visual similarity. The performance in terms of MNDG of the two reranking methods as well as the text baseline are illustrated in Table IV. We can see that the presented hypergraph learning approach assisted with ℓ_2 regularizer performs better than Hypergraph Reranking. It achieves around 3.8% improvements at MNDG@20. Moreover, to demonstrate the robustness of the proposed regularizer, we also conduct exper-

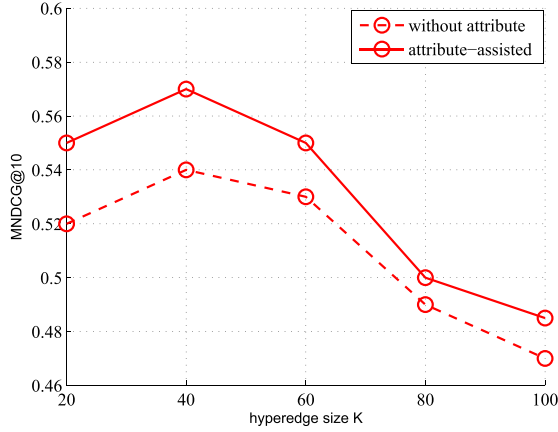


Fig. 6. Performance comparison with various hyperedge size K.

iments with a ℓ_1 regularizer that encourages the sparsity in the $\|w\|_{\ell_1}$ representation. From the experimental comparison, we could see that our approach is more favorable in the task of Web image search reranking, which improves the baseline steadily and outperforms the other reranking strategies.

The good performance of the novel approach for Web image search reranking could be attributed to the following appealing properties: (1) since attribute features are formed by prediction of several classifiers, semantic attributes description might be noisy and only limited semantic attributes could be distributed in a single image. Such implicit attribute selection could make hypergraph based approach much more robust to improve the reranking performance, as inaccurate attributes could be removed and informative ones have been selected for image representation. (2) our proposed iterative regularization framework could further explore the semantic similarity between images by aggregating their local, global similarities instead of simple fusion with concatenation.

As the hyperedges in the graph are formed by K images sharing the common semantic attribute, we perform evaluation on reranking performance with various sizes of hyperedge K . For each K value plotted, we perform the reranking comparison with attribute feature and without its help. Fig. 6 shows the comparison of mean NDCG@10. It achieves the best performance when we set K as 40 among all selected values. It seems that larger or smaller size of hyperedge may result in lower reranking performance in terms of involving more harmful or less useful attributes. We could also see that our proposed reranking approach with local and global feature is boosted by mining attribute information in the experiment.

We also evaluate two weighting parameters λ and μ in our formulation. They modulate the effects of the loss term $\|f - y\|^2$ and the regularizer term $\|w\|_{\ell_2}^2$ respectively. The parameter λ is widely used in graph or hypergraph learning algorithms, and its value determines the closeness of f and y . For parameter μ , if its value tends to be zero, then the proposed algorithm will degenerate to Equation 5, i.e., our preliminary proposed algorithm in [28]. If its value tends to be infinite, then the optimal result is that only one weight is 1 and all others are 0. This extreme case means that there will be only one hyperedge weight used. In Fig. 7, we demonstrate the

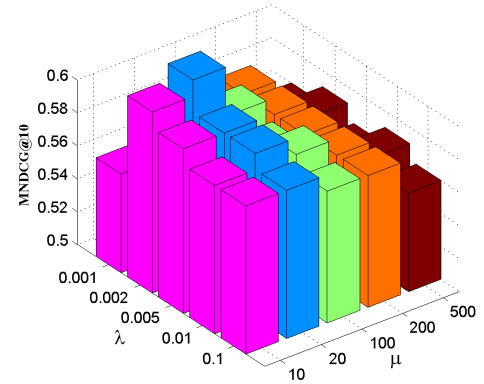


Fig. 7. The average NDCG@10 performance with respect to the variation of λ and μ .



Fig. 8. Some example images with semantic attributes of top 5 highest weights.

mean NDCG@10 performance bar chart with respect to the two parameters μ and λ . It can be seen that the mNDCG is generally higher when we fix λ to be 20. The experimental results illustrate that our approach is able to outperform the other methods when the parameters vary in a wide range.

3) Performance Evaluation for Selected Semantic Attributes: To further illustrate the importance role of semantic attributes on the reranking framework, we conduct experimental evaluation on the hyperedges that have the highest weights. In Table III, it has been pointed out that the top 20 attributes with higher weights in the proposed approach and such semantic attributes are highlighted in italic blue: “Face”, “Arm”, “Leg”, “Hand”, “Tail”, “Beak”, “Wheel”, “Cloth”, “Furry”, “Nose”, “Wing”, “Wood”, “Rock”, “Leaf”, “Headlight”, “Feather”, “Flower”, “Vehicle Part”, “Dotted”, “Smooth”. We also illustrate the semantic attributes that have top 5 highest weights on some images in Fig. 8. We could see that beneficial attributes are distributed in exemplar image which preserve the stronger semantic similarity and thus facilitate ranking performance. Moreover, it is also observed that the semantic attributes with a lower classification score might give a low or high weight in the hypergraph, in regardless of noisiness of predicted attributes. Hence, the experimental results demonstrate the robustness of our proposed approach.

Fig. 9 shows some visualized results of our Attribute-assisted Hypergraph Reranking in comparison to Bayesian reranking and Cluster-based approach on MSRA-MM dataset. It is obvious that that our proposed approach significantly outperforms the baselines methods.



Fig. 9. Case study of the visualized search examples between the proposed method (first row in each example) and baselines of Bayesian reranking [2] (second row in each example) and Cluster-based approach [9] (third row in each example) on MSRA-MM dataset.

V. CONCLUSIONS

Image search reranking has been studied for several years and various approaches have been developed recently to boost the performance of text-based image search engine for general queries. This paper serves as a first attempt to include the attributes in reranking framework. We observe that semantic attributes are expected to narrow down the semantic gap between low-level visual features and high-level semantic meanings. Motivated by that, we propose a novel attribute-assisted retrieval model for reranking images. Based on the classifiers for all the predefined attributes, each image is represented by an attribute feature consisting of the responses from these classifiers. A hypergraph is then used to model the relationship between images by integrating low-level visual features and semantic attribute features. We perform hypergraph ranking to re-order the images, which is also constructed to model the relationship of all images. Its basic principle is that visually similar images should have similar ranking scores and a visual-attribute joint hypergraph learning approach has been proposed to simultaneously explore two information sources. We conduct extensive experiments on 1,000 queries in MSRA-MM V2.0 dataset. The experimental results demonstrate the effectiveness of our proposed attribute-assisted Web image search reranking method.

REFERENCES

- [1] L. Yang and A. Hanjalic, "Supervised reranking for web image search," in *Proc. Int. ACM Conf. Multimedia*, 2010, pp. 183–192.
- [2] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua, "Bayesian visual reranking," *Trans. Multimedia*, vol. 13, no. 4, pp. 639–652, 2012.
- [3] F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting image databases from the web," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
- [4] B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based on multi-attribute queries," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 801–808.
- [5] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1778–1785.
- [6] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 365–372.
- [7] M. Wang, L. Yang, and X.-S. Hua, "MSRA-MM: Bridging research and industrial societies for multimedia," Tech. Rep. MSR-TR-2009-30, 2009.
- [8] K. Järvelin and J. Kekäläinen, "IR evaluation methods for retrieving highly relevant documents," in *Proc. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2000, pp. 41–48.
- [9] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, "Video search reranking via information bottleneck principle," in *Proc. ACM Conf. Multimedia*, 2006, pp. 35–44.
- [10] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas, "Image retrieval via probabilistic hypergraph ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3376–3383.
- [11] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 951–958.

- [12] R. Yan, A. Hauptmann, and R. Jin, "Multimedia search with pseudo-relevance feedback," in *Proc. ACM Int. Conf. Image Video Retr.*, 2003, pp. 238–247.
- [13] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3262–3272, Jul. 2012.
- [14] F. X. Yu, R. Ji, M.-H. Tsai, G. Ye, and S.-F. Chang, "Weak attributes for large-scale image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2949–2956.
- [15] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1601–1608.
- [16] D. Parikh and K. Grauman, "Interactively building a discriminative vocabulary of nameable attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1681–1688.
- [17] D. Parikh and K. Grauman, "Relative attributes," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 503–510.
- [18] F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W.-Y. Ma, "Igroup: Web image search results clustering," in *Proc. 14th Annu. ACM Int. Conf. Multimedia*, 2006, pp. 377–384.
- [19] F. Jing and S. Baluja, "VisualRank: Applying pagerank to large-scale image search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1877–1890, Nov. 2008.
- [20] J. Wang, Y.-G. Jiang, and S.-F. Chang, "Label diagnosis through self tuning for web image search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1390–1397.
- [21] M. Douze, A. Ramisa, and C. Schmid, "Combining attributes and Fisher vectors for efficient image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 745–752.
- [22] A. Kovashka, D. Parikh, and K. Grauman, "WhittleSearch: Image search with relative attribute feedback," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2973–2980.
- [23] Y.-H. Kuo, H.-T. Lin, W.-T. Cheng, Y.-H. Yang, and W. H. Hsu, "Unsupervised auxiliary visual words discovery for large-scale image object retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 905–912.
- [24] J. Wang, K. Lu, D. Pan, N. He, and B.-K. Bao, "Robust object removal with an exemplar-based image inpainting approach," *Neurocomputing*, vol. 123, no. 10, pp. 150–155, Jan. 2014.
- [25] P. Muthukrishnan, D. Radev, and Q. Mei, "Edge weight regularization over multiple graphs for similarity learning," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 374–383.
- [26] M. Wang, H. Li, D. Tao, K. Lu, and X. Wu, "Multimodal graph-based reranking for web image search," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4649–4661, Nov. 2012.
- [27] Y. Su and F. Jurie, "Improving image classification using semantic attributes," *Int. J. Comput. Vis.*, vol. 100, no. 1, pp. 59–77, 2012.
- [28] J. Cai, Z.-J. Zha, W. Zhou, and Q. Tian, "Attribute-assisted reranking for web image retrieval," in *Proc. ACM Int. Conf. Multimedia*, 2012, pp. 873–876.
- [29] L. Liang and K. Grauman, "Beyond comparing image pairs: Setwise active learning for relative attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 208–215.
- [30] H. Zhang, Z.-J. Zha, Y. Yang, S. Yan, Y. Gao, and T.-S. Chua, "Attribute-augmented semantic hierarchy: Towards bridging semantic gap and intention gap in image retrieval," in *Proc. ACM Conf. Multimedia*, 2013, pp. 33–42.
- [31] Q. Tian, N. Sebe, M. S. Lew, E. Loupas, and T. S. Huang, "Image retrieval using wavelet-based salient points," *J. Elect. Imag.*, vol. 10, no. 4, pp. 835–849, 2001.
- [32] L. Zheng and S. Wang, "Visual phraselet: Refining spatial constraints for large scale image search," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 391–394, Apr. 2013.
- [33] J. Cai, Z.-J. Zha, Q. Tian, and Z. Wang, "Semi-automatic Flickr group suggestion," in *Proc. Adv. Multimedia Modeling*, 2011, pp. 77–87.
- [34] J. Cai, Z.-J. Zha, Y. Zhao, and Z. Wang, "Evaluation of histogram based interest point detector in web image classification and search," in *Proc. IEEE Conf. Multimedia Expo*, Jul. 2010, pp. 613–618.
- [35] J. Cai, Z.-J. Zha, H. Luan, S. Zhang, and Q. Tian, "Learning attribute-aware dictionary for image classification and search," in *Proc. 3rd ACM Conf. Int. Conf. Multimedia Retr.*, 2013, pp. 33–40.
- [36] T. Mei, Y. Rui, S. Li, and Q. Tian, "Multimedia search reranking: A literature survey," in *Proc. ACM Comput. Surveys*, 2014.
- [37] K. Lu, N. He, and J. Xue, "Contentbased similarity for 3D model retrieval and classification," in *Proc. Prog. Natural Sci.*, 2009.
- [38] S. Zhang, Q. Huang, S. Jiang, W. Gao, and Q. Tian, "Affective visualization and retrieval for music video," *IEEE Trans. Multimedia*, 2010.

Junjie Cai is currently pursuing the Ph.D. degree with the Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA. His research interests include large-scale media search, social media sharing and management, computer vision, and machine learning.

Zheng-Jun Zha (M'08) received the B.E. and Ph.D. degrees from the Department of Automation, University of Science and Technology of China, Hefei, China, in 2004 and 2009, respectively. He is currently a Professor with the Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei. His current research interests include multimedia content analysis, computer vision, and multimedia applications, such as search, recommendation, and social networking.

Meng Wang is currently a Professor with the Hefei University of Technology, Hefei, China. He received the Ph.D. degree from the University of Science and Technology of China, Hefei. His current research interests include multimedia content analysis, search, mining, recommendation, and large-scale computing.

Shiliang Zhang is currently a Research Fellow with the Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA. He was with Microsoft Research Asia, Beijing, China, as a Research Intern, from 2008 to 2009. He returned to the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2009. His research interests include large-scale image and video retrieval, image/video processing, and multimedia content affective analysis.

Qi Tian (M'96–SM'03) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 1992, the M.S. degree in electrical and computer engineering from Drexel University, Philadelphia, PA, USA, in 1996, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2002. He is currently a Professor with the Department of Computer Science, University of Texas at San Antonio (UTSA), San Antonio, TX, USA. He took a one-year faculty leave with Microsoft Research Asia, Beijing, from 2008 to 2009. His research interests include multimedia information retrieval and computer vision. He has authored over 260 refereed journal and conference papers. His research projects were funded by the NSF, ARO, DHS, SALS, CIAS, and UTSA, and he also received faculty research awards from Google, NEC Laboratories of America, FXPAL, Akiira Media Systems, and HP Labs. He received the Best Paper Awards in PCM 2013, MMM 2013, and ICIMCS 2012, the Top 10% Paper Award in MMSP 2011, the Best Student Paper in ICASSP 2006, and the Best Paper Candidate in PCM 2007. He received the 2010 ACM Service Award. He is the Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, *Journal of Computer Vision and Image Understanding*, *Pattern Recognition Letters*, *EURASIP Journal on Advances in Signal Processing*, *Journal of Visual Communication and Image Representation*, and is on the Editorial Board of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUIT AND SYSTEMS FOR VIDEO TECHNOLOGY, *Multimedia Systems Journal*, *Journal of Multimedia*, and *Journal of Machine Vision and Applications*.