



# Visual Re-Ranking via Adaptive Collaborative Hypergraph Learning for Image Retrieval

Noura Bouhlel<sup>(✉)</sup>, Ghada Feki, and Chokri Ben Amar

REGIM: Research Groups in Intelligent Machines, University of Sfax,  
National Engineering School of Sfax (ENIS), BP 1173, 3038 Sfax, Tunisia  
[noura.bouhlel.tn@ieee.org](mailto:noura.bouhlel.tn@ieee.org)

**Abstract.** Visual re-ranking has received considerable attention in recent years. It aims to enhance the performance of text-based image retrieval by boosting the rank of relevant images using visual information. Hypergraph has been widely used for relevance estimation, where textual results are taken as vertices and the re-ranking problem is formulated as a transductive learning on the hypergraph. The potential of the hypergraph learning is essentially determined by the hypergraph construction scheme. To this end, in this paper, we introduce a novel data representation technique named adaptive collaborative representation for hypergraph learning. Compared to the conventional collaborative representation, we consider the data locality to adaptively select relevant and close samples for a test sample and discard irrelevant and faraway ones. Moreover, at the feature level, we impose a weight matrix on the representation errors to adaptively highlight the important features and reduce the effect of redundant/noisy ones. Finally, we also add a nonnegativity constraint on the representation coefficients to enhance the hypergraph interpretability. These attractive properties allow constructing a more informative and quality hypergraph, thereby achieving better retrieval performance than other hypergraph models. Extensive experiments on the public MediaEval benchmarks demonstrate that our re-ranking method achieves consistently superior results, compared to state-of-the-art methods.

**Keywords:** Image retrieval · Visual re-ranking · Hypergraph learning · Collaborative representation · Ridge regression

## 1 Introduction

Empowered by the ubiquitous access to computer devices and the Internet, an ever-growing amount of digital images has been emerged [25]. In light of this, image retrieval is considered as an active research topic that aims at retrieving relevant images to a user query from a large database of digital images [11, 14, 21, 26]. Until recently, most of the popular search engines (e.g., Flickr) are built upon the textual information associated with images [4, 7, 24]. Nevertheless, they cannot comprehensively describe the rich content of images since

they totally ignore the visual information [10]. Besides, they suffer from the fact that the textual information is often noisy, ambiguous and language-dependent [8, 12]. As a consequence, the retrieved results may be noisy and irrelevant which may affect the retrieval performance [17, 24]. To tackle those issues, visual re-ranking has been introduced to refine the text-based retrieval results using the visual information [4, 19, 32, 35]. Namely, it attempts to boost the rank of relevant images with respect to the textual query [24]. Recently, the hypergraph learning has been widely used in many applications for its capability in capturing complex relationships among samples [4, 15, 23]. In case of visual re-ranking, the textual results are taken as vertices and the re-ranking problem is formulated as a transductive learning on the hypergraph [2, 9]. The potential of the hypergraph learning is essentially determined by the hypergraph construction scheme [22]. Most of previous hypergraph learning methods adopt a neighborhood-based strategy to build the hypergraph, in which textual results are taken as vertices and each vertex is linked to its  $k$  nearest neighbors by an hyperedge. While obvious, this method suffers from the following drawbacks: (1) it is sensitive to noise (2) lacks the ability to discover the real neighborhood structure (3) the parameter  $k$  is fixed as global parameter for all samples regardless their local data distribution. To tackle those issues, recent works have proposed to leverage the regularized regression models, namely the sparse representation and the ridge regression for hypergraph construction [22]. Compared to the neighborhood-based hypergraph, the sparse hypergraph achieves superior performance in revealing the local data structure and handling the noisy data. However, it cannot discover related samples to one hyperedge centroid as thoroughly as possible. Moreover, the sparse constraint makes the hypergraph construction very expensive [41]. Recently, the ridge regression has gained considerable attention not only for its effectiveness in data representation but also for its computational efficiency [41]. In contrast to sparse representation which aims at encouraging the competition between samples to represent a datum, the ridge regression attempt to include all samples in the representation process. That's why this framework is often called the collaborative representation. Owing to these desirable properties, in this paper, we put a particular emphasis on the collaborative representation and we propose an adaptive collaborative hypergraph learning for visual re-ranking. The proposed data representation technique adaptively preserve the locality structure and discard irrelevant/outlier samples with respect to a test sample by integrating a distance-regularizer on the representation coefficients. At the feature level, we impose a weight matrix on the representation errors to adaptively highlight the important features and reduce the effect of redundant/noisy ones. Moreover, to enhance the representation interpretability, a nonnegativity constraint is added in such a way that the representation coefficients can directly reveal the similarity among samples. This way, we obtain a more informative and quality hypergraph which not only captures the grouping information but also reveal the local neighborhood structure and exhibit more discriminative power and robustness to noisy data. Extensive experiments on the public MediaEval benchmarks

demonstrate that our re-ranking method achieves consistently superior results, compared to state-of-the-art methods.

## 2 Related Works

In recent years, many visual re-ranking methods have been proposed in the literature. According to the statistical analysis model used, they could be classified as supervised and unsupervised methods. The former cast the re-ranking to a classification problem that aims at separating relevant from irrelevant images using data from the initial results as training samples. For instance, authors in [30] built a supervised classification model using expert annotations to assign a relevance score to each image. The latter assumes that relevant samples are probably to be close to each other than to irrelevant ones. It aims at discovering and mining patterns using pair-wise similarities. Clearly, there are two paramount ways. The first is to leverage clustering to group images with respect to their visual closeness. For instance, a Hierarchical Clustering is applied in [1] and [29] to cluster samples by relevance. Authors in [28] apply a graph-based clustering method where a similarity graph is initially built to represent relationships among images. The second way is to adopt the graph-based learning for its effectiveness in modeling the intrinsic structure within data. VisualRank proposed by Jing and Baluja [20] is the most popular graph-based re-ranking method. It applies a random walk on an affinity graph where images are taken as nodes and their visual similarities as probabilistic hyper-links. In [39], a manifold ranking process is applied over the data manifold, with the aim of naturally finding the most relevant images. Although promising results are achieved, how to represent complex and high-order relationships hidden in data still the performance bottleneck for graph-based re-ranking. As a generalization of the graph learning, the hypergraph learning is receiving increasing attention in recent years owing to its ability in modeling complex data structure in a more flexible and elegant way [3, 23]. Considering the visual re-ranking, the hypergraph learning is widely used for relevance estimation. For instance, in [2], authors construct a  $k$ -nearest neighbor graphs based on the visual similarity between images. Then, a hypergraph ranking is performed to learn the images' relevance scores. Although efficient, this method suffers from some drawbacks. First, the neighborhood strategy cannot capture the local data distribution of each datum since it uses a fixed number of neighbors  $k$  for all samples [35]. Second, the neighborhood strategy is very sensitive to noisy data due to the use of the Euclidean distance as similarity measure [22, 37]. To address those limitations, some researchers have proposed to exploit the regression models for data representation. The most widely used model is the sparse representation (SR) in which each sample is represented as a linear combination of the remaining samples [15, 36]. Compared to the neighborhood-based hypergraph, the sparse hypergraph achieves superior performance in revealing the local data structure and handling the noisy data. However, it cannot discover related samples to one hyperedge centroid as thoroughly as possible. Moreover, the sparse constraint makes the hypergraph construction very expensive.

Recently, the collaborative representation has gained considerable attention not only for its effectiveness in data representation but also for its computational efficiency [41]. Therefore, in this paper, we put a particular emphasis on the collaborative representation and we propose an adaptive collaborative hypergraph learning for visual re-ranking.

### 3 The Proposed Hypergraph Model for Visual Re-Ranking

#### 3.1 Adaptive Collaborative Representation Representation

For clarity, we first introduce some important notations used throughout this paper. The matrix  $X = [x_1, \dots, x_N] \in \mathbb{R}^{d \times N}$  is a collection of  $N$  data samples where  $x_i \in \mathbb{R}^d$  denotes the  $i$ -th data sample.  $\|Z\|_F$  is the Frobenius norm of matrix  $Z$ .  $\mathbf{1}$  and  $\mathbf{1}$  are a matrix and a vector whose elements are equal to 1,  $\odot$  denotes the element-wise multiplication. For a scalar  $v$ , we define  $(v)_+$  as  $(v)_+ = \max(v, 0)$  [27].

**Problem Formulation.** Conventionally, the collaborative representation aims to solve the following least square problem:

$$\underset{Z}{\operatorname{argmin}} \|X - XZ\|_2^2 + \lambda \|Z\|_2^2 \quad (1)$$

In this paper, we propose an adaptive collaborative representation formulated as follows:

$$\underset{Z, W}{\operatorname{argmin}} \left\| W^{1/2} \odot (X - XZ) \right\|_F^2 + \frac{\beta}{2} \|W\|_F^2 + \lambda \|Z\|_F^2 + \gamma \operatorname{tr}(D^T Z)$$

$$\text{s.t } W \geq 0, W^T \mathbf{1} = \mathbf{1}, Z \geq 0, \operatorname{diag}(Z) = 0, Z\mathbf{1} = \mathbf{1} \quad (2)$$

Specifically, the objective function contains the following terms:

1. The self-representation term: It represents the reconstruction error between the estimated and the real data. Many references have pointed out that redundant/noisy features are likely to have large reconstruction errors [23, 40]. Based on this assumption, we regularize the reconstruction errors by a non-negative weight matrix  $W$ . Hence, we adaptively highlight the important features while reducing the effect of redundant/noisy ones.
2. The  $\ell_2$ -regularizer on the weight matrix: This term as well as the constraint  $W^T \mathbf{1} = \mathbf{1}$  are imposed to avoid the trivial solution of  $W$  as in [42].
3. The regularization term on the representation matrix: It shrinks the representation coefficients towards zero by imposing an  $\ell_2$ -regularizer on their sizes. Indeed, all samples will collaborate during the representation process of a test sample since their coefficients will never become exactly zero.

4. The locality-preserving term: The collaborative representation does not consider the data locality which has been observed to be critical for many learning tasks [34]. For this purpose, we incorporate a locality-preserving term in our model so that (1) the local structure is preserved (i.e, close samples will have close representation) and (2) irrelevant/outliers samples are discarded. Mathematically, each element of the distance matrix  $D$  is defined as:  $d_{ij} = \|x_i - x_j\|_2^2$ .
5. Finally, we add the following constraints on the representation matrix  $Z$ :
  - $Z \geq 0$ : A non-negative representation coefficient  $z_{ij}$  can directly reveal the similarity between the samples  $x_i$  and  $x_j$  [45].
  - $\text{diag}(Z) = 0$ : this constraint is used to avoid that a sample is represented as a linear combination of itself.
  - $Z\mathbf{1} = \mathbf{1}$ : the sum of each row of  $Z$  is set to be equal to 1 which ensure that all samples are selected in the joint representation.

**The ADMM-Based Optimization.** There are two unknown variables in the problem (2), e.g.,  $Z$  and  $W$ . To make the problem (2) separable, some auxiliary variables are added as follows:

$$\begin{aligned} & \underset{Z, W}{\operatorname{argmin}} \left\| W^{1/2} \odot E \right\|_F^2 + \frac{\beta}{2} \|W\|_F^2 + \lambda \|J\|_F^2 + \gamma \operatorname{tr}(D^T Z) \\ \text{s.t } & W \geq 0, W^T \mathbf{1} = \mathbf{1}, Z \geq 0, \text{diag}(Z) = 0, Z\mathbf{1} = \mathbf{1}, E = X - XZ, J = Z \end{aligned} \quad (3)$$

Considering the problem (3) as a two-block optimization problem, we adopt the alternating direction method (ADM) to solve it [38]. Thus, we define the augmented Lagrangian function as:

$$\begin{aligned} \mathfrak{L}(Z, W, E, J, C_1, C_2) = & \left\| W^{1/2} \odot E \right\|_F^2 + \frac{\beta}{2} \|W\|_F^2 + \lambda \|J\|_F^2 + \gamma \operatorname{tr}(D^T Z) \\ & + \frac{\mu}{2} \left( \left\| X - XZ - E + \frac{C_1}{\mu} \right\|_F^2 + \left\| Z - J + \frac{C_2}{\mu} \right\|_F^2 \right) \end{aligned} \quad (4)$$

where  $C_1, C_2$  are the Lagrangian multipliers and  $\mu$  is a penalty parameter.

Then, we solve each unknown variable while fixing the other variables in an alternate way.

**Step 1:** The variable  $W$  is obtained by minimizing the following problem while fixing the other variables:

$$\underset{W}{\min} \left\| W^{1/2} \odot E \right\|_F^2 + \frac{\beta}{2} \|W\|_F^2 \quad \text{s.t } W \geq 0, W^T \mathbf{1} = \mathbf{1} \quad (5)$$

Solving the problem (5) is equivalent to solve:

$$\underset{w_{ij} \geq 0, \sum_j w_{ij} = 1}{\min} \sum_{i,j} \left( w_{ij} + \frac{e_{ij}^2}{\beta} \right)^2 \quad (6)$$

The problem (6) can be written in the vector form since it is independent for different  $i$  [27].

$$\min_{w_i \geq 0, w_i^T \mathbf{1} = 1} \left\| w_i + \frac{h_i}{\beta} \right\|_2^2 \tag{7}$$

where  $H = E \odot E$

The associated Lagrangian function is:

$$\mathfrak{L}(w_i, c, m_i) = \frac{1}{2} \left\| w_i + \frac{h_i}{\beta} \right\|_2^2 - c(w_i^T \mathbf{1} - 1) - m_i^T w_i \tag{8}$$

where  $c$  and  $m_i$  are the Lagrangian multipliers associated to the boundary constraints on  $w_i$ .

Given the fact that  $m_{ij}w_{ij} = 0$  according to the KKT condition [42], we have:

$$w_{ij} = \left( c - \frac{h_{ij}}{\beta} \right)_+ \tag{9}$$

Finally, we update the Lagrangian multiplier  $c$  according to the constraint  $w_i^T \mathbf{1} = 1$  as follows:

$$\sum_{i=1}^N \left( c - \frac{h_{ij}}{\beta} \right) = 1 \Rightarrow c = \frac{1}{N} + \frac{1}{N\beta} \sum_{j=1}^N h_{ij} \tag{10}$$

**Step 2:** We can obtain the error matrix  $E$  by solving the following problem:

$$\min_E \|W^{1/2} \odot E\|_F^2 + \frac{\mu}{2} \|E - G\|_F^2 \text{ where } G = X - XZ + \frac{C_1}{\mu} \tag{11}$$

The problem (11) is equivalent to :

$$\sum_{i,j} \min_{e_{ij}} \left( e_{ij} - \frac{\mu g_{ij}}{\mu + 2w_{ij}} \right)^2 \tag{12}$$

Then, the optimal solution of each element  $e_{ij}$  is

$$e_{ij} = \frac{\mu g_{ij}}{\mu + 2w_{ij}} \tag{13}$$

**Step 3:** We can obtain the matrix  $J$  by solving the following problem:

$$\min_J \lambda \|J\|_F^2 + \frac{\mu}{2} \|Z - J + \frac{C_2}{\mu}\|_F^2 \tag{14}$$

The close-form of  $J$  can be obtained by setting the derivative of (14) w.r.t  $J$  to zero:

$$J^* = \frac{\mu G}{\mu + 2\lambda} \text{ where } G = Z + \frac{C_2}{\mu} \quad (15)$$

**Step 4:** The variable  $Z$  can be obtained by solving the following problem:

$$\begin{aligned} \min_Z \quad & \gamma \text{tr}(D^T Z) + \frac{\mu}{2} (\|M_1 - XZ\|_F^2 + \|Z - M_2\|_F^2) \\ \text{s.t.} \quad & Z \geq 0, \text{diag}(Z) = 0, Z\mathbf{1} = \mathbf{1} \end{aligned} \quad (16)$$

where  $M_1 = X - E + \frac{C_1}{\mu}$  and  $M_2 = J - \frac{C_2}{\mu}$

Considering the following unconstrained problem:

$$\text{argmin}_Z \quad \gamma \text{tr}(D^T Z) + \frac{\mu}{2} (\|M_1 - XZ\|_F^2 + \|Z - M_2\|_F^2) \quad (17)$$

The problem (17) has a closed-form solution obtained by setting its derivative equal to zero:

$$\widehat{Z} = (X^T X + I)^{-1} \left( X^T M_1 + M_2 - \frac{\gamma}{\mu} D \right) \quad (18)$$

Then, the optimal solution  $Z$  of the problem (16) can be obtained more efficiently by solving the following problem:

$$\min_{Z \geq 0, \text{diag}(Z)=0, Z\mathbf{1}=\mathbf{1}} \|Z - \widehat{Z}\|_F^2 \Leftrightarrow \min_{z_{ij} \geq 0, z_{ii}=0, \sum_i z_{ij}=1} (z_{ij} - \widehat{z}_{ij})^2 \quad (19)$$

We obtain the optimal solution for each row  $z_i$  as in problem (6):

$$z_i = (\eta_i I_f^T + \bar{z}_i)_+ \quad (20)$$

where  $I_f$  is a column vector whose elements are equal to one except the  $i$ -th is set equal to zero.  $\bar{z}_i$  is defined as:

$$\bar{z}_i = \begin{cases} \widehat{z}_{ij} & i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

$\eta_i$  is the Lagrangian multiplier which is calculated as:

$$\eta_i = \frac{1 + \bar{z}_i \mathbf{1}}{N - 1} \quad (22)$$

**Step 5:** We update the Lagrangian multipliers and the penalty parameter as follows, respectively:

$$C_1 = C_1 + \mu (X - XZ - E) \quad (23)$$

$$C_2 = C_2 \mu (Z - J) \quad (24)$$

$$\mu = \min(\mu_{max}, \mu \rho) \quad (25)$$

**Convergence and Computational Complexity.** In this section, we first analyze the computational complexity of the proposed representation model. Clearly, the most computationally-demanding step in the ADMM-based Optimization is the step 4 which includes matrix multiplication and matrix inverse operations. It costs  $O(N^3)$  for  $N \times N$  matrix. Fortunately, the term  $(X^T X + I)^{-1}$  can be pre-calculated before the iteration loop since it is independent from all variables and. The first two steps are efficiently calculated since they can be considered as element-wise operations. The third step mainly involves matrix addition operation. Hence, their computational complexities can be ignored compared to the fourth step.

### 3.2 The Proposed Hypergraph Construction Scheme

In this work, we assume that the representation vectors corresponding to two similar samples should be close since they can be similarly represented using remaining ones. More formally, we measure the similarity between two data samples as follows:

$$A(i, j) = z_i \cdot z_j \quad (26)$$

In terms of hypergraph, such information is very useful to characterize the incidence relations between hyperedges and their vertices:

$$h(v_i, e_j) = \begin{cases} A(i, j), & \text{if } z_{ij} \geq \theta \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

Here, we set  $\theta$  as the mean values of  $\{z_{ik}\}_{k=1}^N$ . According to this formulation, each vertex  $v_i$  is associated to hyperedge  $e_j$  based on whether it has prominently contributed in the representation of its centroid  $v_j$ . Moreover, for each centroid, the number of neighbors is adaptively selected. Hence, its distinctive neighborhood structure is well preserved.

### 3.3 The Hypergraph-Based Re-Ranking

In this work, we formulate the visual re-ranking problem as a transductive learning framework on the adaptive collaborative hypergraph model  $G = (V, E, \omega)$ :

$$\arg \min_f \{\Omega(f) + \mu R_{emp}(f)\} \quad (28)$$

where the vector  $f$  is constituted of the relevance scores to be learned.

Following the Zhou' works [44], the regularization term can be written as follows:

$$\Omega(f) = f^T (I - \Theta) f = f^T \left( I - D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2} \right) f \quad (29)$$



The empirical loss  $R_{emp}(f)$  guarantees that final ranking scores are close to the initial ones. It is defined as:

$$R_{emp}(f) = \|f - y\|^2 = \sum_{v_i \in V} (f(v_i) - y_i)^2 \quad (30)$$

Where the initial ranking vector  $y$  is uniformly defined as:

$$y_i = 1 - \frac{i}{N} \quad (31)$$

By substituting (29) and (30) into (28) and setting the derivative of (28) with respect to  $f$  to 0, we have

$$f(I - \Theta) + \mu(f - y) = 0 \Rightarrow f = \frac{\mu}{1 + \mu} (I - \frac{\Theta}{1 + \mu})^{-1} y \quad (32)$$

**Table 1.** Description of databases

Database	Description	No. of images
Landmark-30 [16]	30 one-concept locations queries	8923
Landmark-123 [16]	123 one-concept locations queries	36452
General-65 [18]	65 complex and multi-concept queries	20000
General-70 [18]	70 complex and multi-concept queries	30000

## 4 Experiments

### 4.1 Experimental Settings

In this section, we have conducted visual re-ranking experiments on four public databases designed within the MediaEval 2014 [16] and MediaEval 2016 [18] competitions and listed in Table 1. In particular, the MediaEval 2014 benchmark consists of information for 153 one-concept location queries (e.g., buildings, museums, roads, bridges, sites, monuments, etc) with about 300 photos per location [16]. The MediaEval 2016 benchmarks consists of 135 complex and general-purpose multi-concept queries (e.g., animals at zoo, sunset in the city, accordion player, etc)[18]. We choose those databases for the following reasons: (1) they are consisted of real-world images (i.e. images are initially retrieved from Flickr in response to a textual query) (2) they are publicly available and (3) annotations are carried out by experts [17].

We use the convolutional neural networks based descriptors to represent images of all databases for its impressive performance in image retrieval [43]. In all experiments, we followed the rules of the MediaEval competitions. Indeed,

in evaluation, a photo is considered to be relevant if it is a common photo representation of the query [16, 18]. Experiments were carried out for different cut-off points,  $X \in \{5, 10, 20, 30, 40, 50\}$ . For performance evaluation, we adopt the precision  $P@20$  as the official ranking for both MediaEval 2014 and MediaEval 2016 benchmarks was set to a cut-off of 20 images [16, 18]. For fair comparison, we conducted all experiments on the same platform, i.e., Matlab platform running on Windows7, with an Intel (R)-Core(TM) i7-4500U 3.40 GHz processor and 8 GB memory. Moreover, we manually tuned the parameters of all other methods to obtain their optimal results.

## 4.2 Performance Comparison with State-of-the-art Methods

This experiment is conducted in order to compare our method with other methods that achieved best performance during the MediaEval competitions. In this experiment, we select only those visual-based methods. Comparison results are reported in Table 2. First, it can be observed that our method achieves a consistent improvement over the Flickr baseline on all databases. For examples, at a cut-off point  $X = 20$ , the precision gains of ACR-HG over Flickr are 6.67%, 8.29%, 10.07% and 6.49% on Landmark-30, Landmark-123, General-65 and General-70 respectively. Second, our method almost always outperforms other methods on all databases. For example, on Landmark-123, the precision of our method is  $P@20 = 0.8894$  while other methods achieve 0.769 (TUW)[28], 0.7561 (SocSens) [31] and 0.748 (PeRCeiVe)[29]. On the General-70 database, which is a complex and general-purpose multi-concept database, we achieve a  $P@20 = 0.7921$  compared to  $P@20 = 0.5437$  achieved by the best team (LAPI) [6]. Our method, which not only models the complex and high-order relationships among visual samples via hypergraph but also capture the overall contextual information by the means of collaborative representation, achieves the best performance among the compared methods. This clearly demonstrates the validity of our method for visual re-ranking not only on for landmark image retrieval but also for multi-topic image retrieval.

**Table 2.** Performance comparison to state-of-the-art re-ranking methods.

Methods	P@20		Methods	P@20	
	Landmark-30	Landmark-123		General-70	General-65
Flickr	0.8333	0.8065	Flickr	0.6914	0.5531
PeRCeiVe [29]	0.866	0.748	UPMC [33]	0.631	0.5200
SocSens [31]	0.815	0.7561	LAPI [6]	0.6514	0.5437
TUW [28]	0.805	0.769	RECOD [13]	0.6821	0.5180
ACR-HG (ours)	<b>0.9</b>	<b>0.8894</b>	ACR-HG (ours)	<b>0.7921</b>	<b>0.618</b>

**Table 3.** Performance comparison to graph/hypergraph-based methods

Methods	P@20			
	Landmark30	Landmark-123	General-70	General-65
Flickr	0.8333	0.8065	0.6914	0.5531
VR [20]	0.8517	0.8314	0.74	0.5492
MR [5]	0.8251	0.8045	0.7293	0.5383
Knn-HG [2]	0.865	0.8537	0.7364	0.5461
SR-HG [36]	0.88	0.8541	0.6971	0.5531
CR-HG [41]	0.8883	0.8728	0.7564	0.5758
ACR-HG (ours)	<b>0.9</b>	<b>0.8894</b>	<b>0.7921</b>	<b>0.618</b>

### 4.3 Performance Comparison for Hypergraph Learning

In this experiment, we aim to validate the superiority of our hypergraph model over the conventional graph/hypergraph models. Results are showed in Table 3. From the results, the following observations can be drawn:

- Despite their ability in refining the initial retrieval results, graph-based re-ranking methods are almost outperformed by the hypergraph-based ones. This demonstrates that, in contrast to graph model, hypergraph model has an inherent ability to capture the local group information and latent high-order relationships among samples.
- The experimental results reveal also the good robustness and discriminative power of representation based hypergraph learning compared to neighborhood based hypergraph learning. On different databases, the representation based hypergraph ranking achieves the highest precision compared to hypergraph ranking based on neighborhood relationships. In particular, our method consistently and significantly achieves the best relevance improvement among other representation based hypergraph ranking.
- The adaptive collaborative representation has brought more robustness and discriminative power to the hypergraph than the collaborative representation. For instance, the precision gains of ACR-HG over the CR-HG are 1.17%, 1.66%, 3.57% and 4.22% on Landmark-30, Landmark-123, General-70 and General-65 respectively. One explanation is that the adaptive collaborative representation imposes a locality-preserving regularizer on the representation coefficients which enable to capture the global and local structures of data during the hypergraph learning.

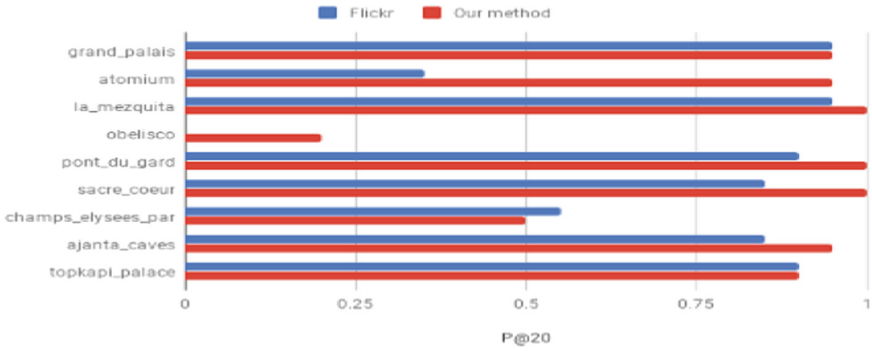


Fig. 1. Evolution curve of relevance for different landmark query topics

#### 4.4 Performance Evaluation per Topic Class

The aim of this experiment is to investigate the performance stability of our method for different query topics. Comparison results are presented in Figs. 1 and 2. We find that our method outperforms Flickr for almost all query topics. The experimental results also reveal that, the relevance of retrieval results is higher for landmarks queries compared to complex queries. One explanation is that, non-relevant images were likely to be arisen when the query is ambiguous or involve multiple topics. For example, the query ‘baby in stroller’ may give rise to images that contain an empty stroller. Another interesting observations, is that the retrieval performance is degraded for some queries (e.g. ‘baby in stroller’). This can be attributed to the fact that a high relevance score for a non-relevant image will be propagated to its visually similar neighbors since only the visual information is used for building the hypergraph.

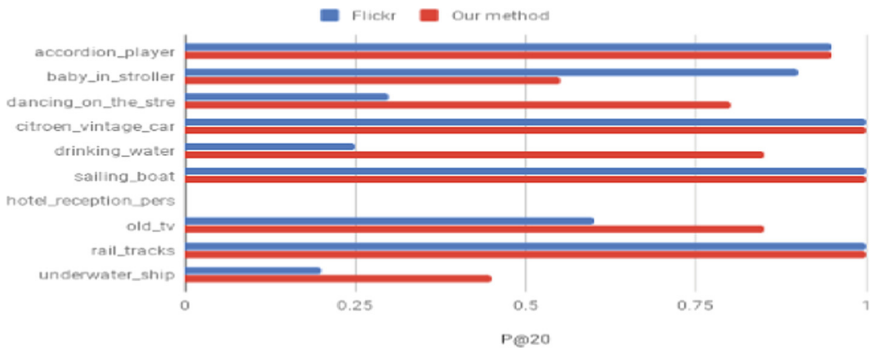


Fig. 2. Evolution curve of relevance for different general multi-concept query topics

## 5 Conclusion

In this paper, we proposed a novel hypergraph-based visual re-ranking method to enhance the performance of text-based image retrieval. At the core of our method is the data representation. Particularly, we proposed a novel representation technique called adaptive collaborative representation to build a more informative hypergraph. By constraining the self-representation term with an weighted matrix, the effect of those redundant and useless features can be adaptively minimized so that a more robust hypergraph can be constructed. In addition, our data representation technique has the advantage of simultaneously capturing both global and local structures of data during hypergraph learning by introducing a locality-preserving term. Based on the obtained representation matrix, we showed how to generate consistent hyperedge connections and hyperedge weights. Finally, a transductive learning is successfully performed upon the constructed hypergraph to learn the images' relevance scores. Experimental results performed on public MediaEval benchmarks demonstrate that our method achieves consistently superior results compared to state-of-the art re-ranking methods.

**Acknowledgements.** The research leading to these results has received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

## References

1. Boteanu, B., Mironică, I., Ionescu, B.: Hierarchical clustering pseudo-relevance feedback for social image search result diversification. In: Proceedings - International Workshop on Content-Based Multimedia Indexing (2015)
2. Bouhleb, N., Feki, G., Ben Ammar, A., Ben Amar, C.: A hypergraph-based reranking model for retrieving diverse social images. In: Felsberg, M., Heyden, A., Krüger, N. (eds.) CAIP 2017. LNCS, vol. 10424, pp. 279–291. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-64689-3\\_23](https://doi.org/10.1007/978-3-319-64689-3_23)
3. Bouhleb, N., Ksibi, A., Ben Ammar, A., Ben Amar, C.: Semantic-aware framework for mobile image search. In: International Conference on Intelligent Systems Design and Applications, ISDA, vol. 2016-June, pp. 479–484. IEEE (2016)
4. Cai, J., Zha, Z.J., Wang, M., Zhang, S., Tian, Q.: An attribute-assisted reranking model for web image search. *IEEE Trans. Image Process.* **24**(1), 261–272 (2015)
5. Cheng, X.Q., Du, P., Guo, J., Zhu, X., Chen, Y.: Ranking on data manifold with sink points. *IEEE Trans. Knowl. Data Eng.* **25**(1), 177–191 (2013)
6. Constantin, M.G., Boteanu, B., Ionescu, B.: LAPI at mediaeval 2016 predicting media interestingness task, October 2016
7. Feki, G., Fakhfakh, R., Ben Ammar, A., Ben Amar, C.: Knowledge structures: which one to use for the query disambiguation? In: 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 499–504, December 2015
8. Feki, G., Fakhfakh, R., Ben Ammar, A., Ben Amar, C.: Query disambiguation: user-centric approach. *J. Inform. Assur. Secur.* **11**, 144–156 (2016)

9. Feki, G., Fakhfakh, R., Bouhleb, N., Ben Ammar, A., Ben Amar, C.: REGIM @ 2016 retrieving diverse social images task. In: Working Notes Proceedings of the MediaEval 2016 Workshop, 20–21 October 2016, Hilversum, The Netherlands (2016)
10. Feki, G., Ksibi, A., Ben Ammar, A., Ben Amar, C.: Improving image search effectiveness by integrating contextual information. In: 2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI), pp. 149–154 (2013)
11. Feki, G., Ammar, A.B., Amar, C.B.: Adaptive semantic construction for diversity-based image retrieval. In: KDIR 2014 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, Rome, Italy, 21–24 October 2014, pp. 444–449 (2014)
12. Feki, G., Ammar, A.B., Amar, C.B.: Towards diverse visual suggestions on Flickr. In: Ninth International Conference on Machine Vision, ICMV 2016, Nice, France, 18–20 November 2016, p. 103411Z (2016)
13. Ferreira, C., et al.: Recod @ mediaeval 2016: Diverse social images retrieval, October 2016
14. Guedri, B., Zaied, M., Ben Amar, C.: Indexing and images retrieval by content. In: 2011 International Conference on High Performance Computing Simulation, pp. 369–375 (2011)
15. Hong, C., Zhu, J.: Hypergraph-based multi-example ranking with sparse representation for transductive learning image retrieval. *Neurocomputing* **101**, 94–103 (2013)
16. Ionescu, B., Popescu, A., Lupu, M., Gîncă, A.L., Boteanu, B., Müller, H.: Div150Cred: a social image retrieval result diversification with user tagging credibility dataset. In: Proceedings of the 6th ACM Multimedia Systems Conference, MMSys 2015, pp. 207–212. ACM, New York (2015)
17. Ionescu, B., Popescu, A., Radu, A.-L., Müller, H.: Result diversification in social image retrieval: a benchmarking framework. *Multimed. Tools Appl.* **75**(2), 1301–1331 (2014). <https://doi.org/10.1007/s11042-014-2369-4>
18. Ionescu, B., Zaharieva, M.: Retrieving diverse social images at MediaEval 2016: challenge, dataset and evaluation. In: Gravier, G., et al. (eds.) Working Notes Proceedings of the MediaEval 2016 Workshop, pp. 20–22 (2016)
19. Jing, P., Su, Y., Xu, C., Zhang, L.: HyperSSR: a hypergraph based semi-supervised ranking method for visual search reranking. *Neurocomputing* **274**, 50–57 (2018)
20. Jing, Y., Baluja, S.: VisualRank: applying PageRank to large-scale image search. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(11), 1877–1890 (2008)
21. Ksibi, A., Feki, G., Ben Ammar, A., Ben Amar, C.: Effective diversification for ambiguous queries in social image retrieval. In: Wilson, R., Hancock, E., Bors, A., Smith, W. (eds.) CAIP 2013. LNCS, vol. 8048, pp. 571–578. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-40246-3\\_71](https://doi.org/10.1007/978-3-642-40246-3_71)
22. Liu, Q., Sun, Y., Wang, C., Liu, T., Tao, D.: Elastic net hypergraph learning for image clustering and semi-supervised classification. *IEEE Trans. Image Process.* **26**(1), 452–463 (2017)
23. Liu, Y., Shao, J., Xiao, J., Wu, F., Zhuang, Y.: Hypergraph spectral hashing for image retrieval with heterogeneous social contexts. *Neurocomputing* **119**, 49–58 (2013)
24. Mei, T., Rui, Y., Li, S., Tian, Q.: Multimedia search reranking: a literature survey. *ACM Comput. Surv.* **46**(3), 1–38 (2014)
25. Mejdoub, M., Fonteles, L., BenAmar, C., Antonini, M.: Fast indexing method for image retrieval using tree-structured lattices. In: 2008 International Workshop on Content-Based Multimedia Indexing, pp. 365–372, June 2008

26. Mejdoub, M., Fonteles, L., Ben Amar, C., Antonini, M.: Embedded lattices tree: an efficient indexing scheme for content based retrieval on image databases. *J. Vis. Commun. Image Represent.* **20**(2), 145–156 (2009)
27. Nie, F., Wang, X., Jordan, M.I., Huang, H.: The constrained Laplacian rank algorithm for graph-based clustering. In: 30th AAAI Conference on Artificial Intelligence, AAAI 2016, no. 1, pp. 1969–1976 (2016)
28. Sabetghadam, S., Palotti, J.R.M., Rekabsaz, N., Lupu, M., Hanbury, A.: TUW @ mediaeval 2015 retrieving diverse social images task. In: Working Notes Proceedings of the MediaEval 2015 Workshop, 14–15 September 2015, Wurzen, Germany (2015)
29. Spampinato, C., Palazzo, S.: PeRCeiVe lab@UNICT at MediaEval 2014 diverse images: random forests for diversity-based clustering. In: MediaEval (2014)
30. Spyromitros-Xioufis, E., Papadopoulos, S., Ginsca, A.L., Popescu, A., Kompatsiaris, Y., Vlahavas, I.: Improving diversity in image search via supervised relevance scoring. In: ICMR 2015 - Proceedings of the 2015 ACM International Conference on Multimedia Retrieval, ICMR 2015, pp. 323–330. ACM, New York (2015)
31. Spyromitros-Xioufis, E., Papadopoulos, S., Kompatsiaris, I., Vlahavas, I.: SocialSensor: finding diverse images at mediaeval 2014, vol. 1263, October 2014
32. Tian, X., Yang, L., Wang, J., Wu, X., Hua, X.S.: Bayesian visual reranking. *IEEE Trans. Multimed.* **13**(4), 639–652 (2011)
33. Tollari, S.: UPMC at MediaEval 2016 retrieving diverse social images task. In: CEUR Workshop Proceedings, vol. 1739 (2016)
34. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3360–3367, June 2010
35. Wang, M., Li, H., Tao, D., Lu, K., Wu, X.: Multimodal graph-based reranking for web image search. *IEEE Trans. Image Process.* **21**(11), 4649–4661 (2012)
36. Wang, M., Liu, X., Wu, X.: Visual classification by  $\ell_1$ -hypergraph modeling. *IEEE Trans. Knowl. Data Eng.* **27**(9), 2564–2574 (2015)
37. Wang, Y., Lin, X., Wu, L., Zhang, W.: Effective multi-query expansions: robust landmark retrieval. In: MM 2015 - Proceedings of the 2015 ACM Multimedia Conference, MM 2015, pp. 79–88. ACM, New York (2015)
38. Wen, J., Fang, X., Xu, Y., Tian, C., Fei, L.: Low-rank representation with adaptive graph regularization. *Neural Netw.* **108**, 83–96 (2018)
39. Xu, B., Bu, J., Chen, C., Wang, C., Cai, D., He, X.: EMR: a scalable graph-based ranking model for content-based image retrieval. *IEEE Trans. Knowl. Data Eng.* **27**(1), 102–114 (2015)
40. Yang, J., Luo, L., Qian, J., Tai, Y., Zhang, F., Xu, Y.: Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(1), 156–171 (2017)
41. Zhang, L., Yang, M., Feng, X.: Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of the IEEE International Conference on Computer Vision, pp. 471–478 (2011)
42. Zheng, J., Yang, P., Chen, S., Shen, G., Wang, W.: Iterative re-constrained group sparse face recognition with adaptive weights learning. *Trans. Image Process.* **26**(5), 2408–2423 (2017)
43. Zheng, L., Yang, Y., Tian, Q.: SIFT meets CNN: a decade survey of instance retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(5), 1224–1244 (2018). <https://doi.org/10.1109/TPAMI.2017.2709749>

44. Zhou, D., Huang, J., Schölkopf, B.: Learning with hypergraphs: clustering, classification, and embedding. In: *Advances in Neural Information Processing Systems* 19, vol. 19, no. Figure 1, pp. 1601–1608 (2007)
45. Zhuang, L., Gao, H., Lin, Z., Ma, Y., Zhang, X., Yu, N.: Non-negative low rank and sparse graph for semi-supervised learning. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2328–2335 (2012)