SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

December 2015, Vol. 58 122104:1–122104:15
doi: 10.1007/s11432-015-5486-4

Image retrieval based on multi-concept detector and semantic correlation

XU HaiJiao[1,2], HUANG ChangQin[2,4*], PAN Peng[1], ZHAO GanSen[2], XU ChunYan[3], LU YanSheng[1], CHEN Deng[1] & WU JiYi[4]

[1]*School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan* 430074*, China;*
[2]*Research Center for Information Services and Software Technology, South China Normal University, Guangzhou* 510631*, China;*
[3]*Electrical and Computer Engineering, National University of Singapore,* 21 *Lower Kent Ridge Road* 119077*, Singapore;*
[4]*E-Service Research Center, Zhejiang University, Hangzhou* 310027*, China*

Received May 28, 2015; accepted October 22, 2015; published online November 12, 2015

Abstract With the rapid development of future network, there has been an explosive growth in multimedia data such as web images. Hence, an efficient image retrieval engine is necessary. Previous studies concentrate on the single concept image retrieval, which has limited practical usability. In practice, users always employ an Internet image retrieval system with multi-concept queries, but, the related existing approaches are often ineffective because the only combination of single-concept query techniques is adopted. At present semantic concept based multi-concept image retrieval is becoming an urgent issue to be solved. In this paper, a novel Multi-Concept image Retrieval Model (MCRM) based on the multi-concept detector is proposed, which takes a multi-concept as a whole and directly learns each multi-concept from the rearranged multi-concept training set. After the corresponding retrieval algorithm is presented, and the log-likelihood function of predictions is maximized by the gradient descent approach. Besides, semantic correlations among single-concepts and multi-concepts are employed to improve the retrieval performance, in which the semantic correlation probability is estimated with three correlation measures, and the visual evidence is expressed by Bayes theorem, estimated by Support Vector Machine (SVM). Experimental results on Corel and IAPR data sets show that the approach outperforms the state-of-the-arts. Furthermore, the model is beneficial for multi-concept retrieval and difficult retrieval with few relevant images.

Keywords multi-concept image retrieval, semantic correlation, probability estimation, concept learning, visual evidence

Citation Xu H J, Huang C Q, Pan P, et al. Image retrieval based on multi-concept detector and semantic correlation. Sci China Inf Sci, 2015, 58: 122104(15), doi: 10.1007/s11432-015-5486-4

1 Introduction

With the rapid development of the society and scientific technologies, it can be foreseen that network technologies and network contents will play a more important and involved role in human life, and based

* Corresponding author (email: cqhuang@zju.edu.cn)

Figure 1 (Color online) Training examples for conventional single-concept detectors and the proposed multi-concept detector. (a) Concept ⟨water⟩; (b) concept ⟨boat⟩; (c) concept ⟨harbor⟩; (d) multi-concept ⟨water, boat, harbor⟩.

on the profoundly developed network in the future, multimedia data will continue to grow explosively. There is no doubt that people need to learn more about big data and much more related research will be conducted. Big data is an enormous dataset that may contain rich media data, most of which belong to non-structured data (such as images). Current information collection techniques are no longer restricted to only using text or attribute keywords to characterize the target object. Moreover, multimedia data such as images are used to give a visual representation of the object, in order to minimize the loss of information. Therefore, an efficient image retrieval engine is much needed to help users search and browse interesting images from a large image dataset, and semantic retrieval should be further considered.

During multi-concept image retrieval, the paradigm allows users to provide multiple desired concepts and retrieve relevant images[1)] containing all the target concepts. Some solutions use concept detectors to automatically detect and recognize image concepts in the un-annotated image set [1]. The concepts may be about a scene (such as ⟨harbor⟩) or the nature (such as ⟨sunset⟩). The essential idea of concept detection is to model the associations between images and concepts.

In practice, users always use the image retrieval system with multi-concept queries (such as a scene query ⟨water, boat, harbor⟩). To deal with this kind of retrieval, the existing approaches [2,3] perform multi-concept retrieval by combining single-concept detectors. These approaches may be ineffective in some cases. In Figure 1, three single-concepts ⟨water⟩, ⟨boat⟩ and ⟨harbor⟩ in the visual feature space are considered, denoted in blue, red and yellow, respectively. The overlapping orange area denotes a multi-concept ⟨water, boat, harbor⟩ in the visual feature space and it is considered that it has a special visual appearance which may be difficult to distinguish solely by conventional single-concept detectors. The conventional single-concept detectors and multi-concept detectors can complement each other and their combination may lead to better performance of multi-concept retrieval, which is our research motivation.

Besides, many concept-based retrieval approaches [1,3,4] do not explicitly consider the semantic correlations among concepts to simplify the complexity of the proposed models. In other words, these approaches are based on the assumption that concepts are independent of each other. However, this assumption cannot be satisfied in some cases. For example, the single-concept ⟨sunset⟩ usually appears with ⟨cloud⟩ or ⟨horizon⟩ in the images, and a multi-concept ⟨water, boat⟩ is likely to co-occur with a single-concept ⟨harbor⟩ or a multi-concept ⟨city, harbor⟩ in an image. To solve the above-mentioned problems, a probabilistic model is proposed and the semantic correlations between single-concepts and multi-concepts are incorporated into the probabilistic model.

The remainder of the paper is organized as follows. Section 2 briefly reviews the related work. Section 3 presents our probabilistic model and the retrieval algorithm, and then in Section 4, the image relevance estimation is described. Experimental results and analyses are reported in Section 5. Finally, the conclusion is given in Section 6.

2 Related work

Two major conventional image retrieval paradigms are content-based image retrieval (CBIR) [5] and text-based image retrieval. CBIR is neither intuitive nor user-friendly. Hence, the concept-based image retrieval [6,7] is proposed and has received increasing attention. These approaches employ concept detectors to automatically distinguish semantic concepts in the un-annotated images. In general, the existing concept detection can be classified into three categories: the generative model, the discriminative model and the nearest neighbor model.

1) Images are considered relevant for a query Q when they contain all single-concepts $w_i \in Q$.

The generative model usually learns the joint probability $p(w_i, I)$ and calculates the prediction $p(w_i|I)$ by using Bayes rules, given the retrieval concept w_i and the un-annotated image I. A pioneer work was Cross Media Relevance Model (CMRM) [8], which utilizes the co-occurrences of w_i and all visual blobs b_i of I to generate the joint probability $p(w_i, b_1, \ldots, b_m)$. CMRM considers multi-concept image retrieval including single-concept, 2-concept[2)], 3-concept and 4-concept, which is performed through the combination of single-concept detectors. Inspired by the discrete CMRM model, Continuous-space Relevance Model (CRM) [9] was proposed and the efficiency of the CRM detector for multi-concept retrieval has been greatly improved by about 51% compared with CMRM. In [10], a probabilistic generative approach called Multiple Bernoulli Relevance Model (MBRM) was proposed, which is based on CRM. Benefiting from the assumption of the multiple Bernoulli distribution, MBRM improves about 15% over CRM for single-concept retrieval. Similar to CMRM, Cross Media Translation Table (CMTT) [11] also assumes that an image region can be represented by using a discrete vocabulary of b_i. It employs the G-means algorithm [12] to automatically determine the optimal number m of b_i. Its single-concept detectors achieve a 45% relative improvement in the accuracy of concept detection, compared with previous methods. The previous approaches mainly focus on image representation and the link $p(w_i, I)$ learning. A few studies (such as CMRM) consider the multi-concept retrieval through the single-concept detector techniques.

Topic models which originate from text mining are also generative models, and have been widely applied in the image-related problems [13]. In the data generation process, these approaches do not use training images but use hidden topics (aspects) z_k as latent variables which link concepts w_i and images I. Latent Dirichlet Allocation (LDA) [14] and Probabilistic Latent Semantic Analysis (PLSA) [15] are two classical examples of the topic model approaches. Correspondence Latent Dirichlet Allocation (Corr-LDA) [16] extends the basic LDA model to learn the joint correlations. Topic model detectors utilize hidden topics z_k to link w_i and I, which may improve the accuracy. However, the correlations between concepts and multi-concept detectors are not considered.

Discriminative models directly estimate the posterior probability $p(w_i|I)$ or learn a map from I to w_i, such as Support Vector Machine (SVM) [17] and Passive-Aggressive Model for Image Retrieval (PAMIR) [2]. PAMIR introduces a learning procedure, where a criterion related to the multi-concept retrieval performance is optimized, and achieves better multi-concept retrieval performance. Different from the existing methods, Group Sparsity approach (GS) [1] investigates the properties of visual features. Ref. [18] proposed a joint convex optimization formulation which minimizes ranking errors while simultaneously conducting feature selection. Owing to the continuous efforts in concept detection, these discriminative model detectors show competitive retrieval performance in single-concept retrieval. However, multi-concept retrieval is performed through the combination of single-concept retrieval techniques and the associations between concepts are not used for concept detection.

Compared with many parametric models mentioned above, the nearest neighbor models are more attractive as a simple yet powerful alternative, such as ranking-oriented nearest-neighbor method [19] and the utilization of the annotation propagations over a similarity graph of the annotated and un-annotated images [20]. In [21], a greedy label transfer algorithm was presented to transfer annotations from visual neighbors. Ref. [3] proposed a new nearest neighbor approach, named TagProp. TagProp detectors produce the relevance estimates of a concept for the images by adopting a weighted combination of a concept presence among visual neighbors. This single-concept detector shows the state-of-the-art retrieval performance because weight learning is integrated in the prediction model.

Many concept-based retrieval methods neglect the semantic links among concepts. To address this problem, WordNet-based approaches [22] were proposed to refine the concept detection results. However, WordNet ontology is too small and cannot deal with the concepts that do not exist in their lexicons (such as ⟨balcony⟩ and ⟨frozen⟩), which may limit its application in the cases of large vocabularies, such as Corel and IAPR containing hundreds of concepts. Besides, it calculates word similarities instead of the correlations between visual concepts. Ref. [23] applied the Google semantic distance to yield better data results. In [24], Google distances are used to find out the most relevant information in the top retrieval results, which can achieve higher retrieval accuracy.

2) The multi-concept with length n is called the n-concept.

Generally, given a single-concept query Q, the conventional approaches employ single-concept detectors to calculate whether an image I is relevant for Q. For a multi-concept query Q, the prevailing approaches are still based on the single-concept detector techniques. Conventional single-concept detectors can effectively detect single-concepts in images while multi-concept detectors can effectively detect multi-concept scenes. Their combination may lead to the improved performance of multi-concept retrieval.

3 Multi-Concept image Retrieval Model (MCRM)

The proposed probabilistic model focuses on the multi-concept image retrieval task, so it is called Multi-Concept Retrieval Model (MCRM for brevity). Let $\mathcal{T} = \{I_1, \ldots, I_T\}$ be a set of training images and $\mathcal{Y} = \{w_1, \ldots, w_V\}$ be a vocabulary of V semantic single-concepts. The training set $\{(I_1, Y_1), \ldots, (I_T, Y_T)\}$ consists of pairs of images and their corresponding concept annotation sets, with each $Y_i \subseteq \mathcal{Y}$. Each semantic multi-concept $W_i^{(n)} = \{w_1, \ldots, w_i, \ldots, w_n\}$ is an element of the power set of $\mathcal{Y}$, i.e., $W_i^{(n)} \in 2^{\mathcal{Y}}$ or $W_i^{(n)} \subseteq \mathcal{Y}$, where $n \geqslant 1$ is the length $|W_i^{(n)}|$ of $W_i^{(n)}$. If $n = 1$, $W_i^{(n)}$ becomes exactly a conventional single-concept w_i. Given a test set $\mathcal{S}$ of un-annotated images and a retrieval multi-concept $Q \subseteq \mathcal{Y}$, the goal of multi-concept image retrieval is to search for the most relevant images $I \in \mathcal{S}$ that contain all target single-concepts $w_i \in Q$.

For consistence, the additional mathematical notations in the proposed model are defined as follows:

- N denotes the total size of the test set, namely $N = |\mathcal{S}|$;
- q denotes the length of the retrieval concept Q, namely $q = |Q|$ $(Q \subseteq \mathcal{Y})$;
- K denotes the number of the returned images for the query Q $(K \leqslant N)$;
- $\mathcal{W}^{(n)}$ denotes an n-concept set. The multi-concept $W_i^{(n)}$ with length n is called an n-concept, i.e., $\mathcal{W}^{(n)} = \{W_i^{(n)} \mid |W_i^{(n)}| = n\}$;
- $\mathcal{Y}^q$ denotes the multi-concept vocabulary for the retrieval concept Q and it is a subset of $\bigcup_{n=1}^{q} \mathcal{W}^{(n)} \subseteq 2^{\mathcal{Y}}$ $(1 \leqslant n \leqslant q)$;
- $\mathcal{R}(Q) = \{W_i^{(n)} \mid W_i^{(n)} \in \mathcal{Y}^q, p(W_i^{(n)}|Q) > 0\}$ is a subset of $\mathcal{Y}^q$, which includes all semantic neighbors of the retrieval concept $Q \in \mathcal{Y}^q$ in the semantic space;
- $\mathcal{R}_{\mathrm{RC}}(Q) \subseteq \mathcal{R}(Q)$ denotes the retrieval context which is a subset of $\mathcal{R}(Q)$;
- K_{RC} denotes the number of elements in the set $\mathcal{R}_{\mathrm{RC}}(Q)$, namely $K_{\mathrm{RC}} = |\mathcal{R}_{\mathrm{RC}}(Q)|$;
- $\mathrm{Nr}(W_i^{(n)})$ denotes the total number of occurrences of the multi-concept $W_i^{(n)} \in \mathcal{Y}^q$ in the training set $\mathcal{T}$, namely the multi-concept frequency;
- $\mathrm{Nr}(W_i^{(n)}, W_i^{(m)})$ denotes the total co-occurrences of both the multi-concepts $W_i^{(n)}, W_i^{(m)} \in \mathcal{Y}^q$ in the training set $\mathcal{T}$.

3.1 The framework of MCRM

Given a retrieval multi-concept $Q = \{w_1, \ldots, w_i, \ldots, w_q\}$, previous approaches are mainly based on fusion techniques of the single-concept detectors, such as the product fusion [3] or the addition fusion [2].

A general framework of the proposed MCRM is shown in Figure 2. In the solid box, q single-concepts $w_i \in Q$ are denoted as solid circles. The corresponding conventional single-concept detectors SD_i are denoted as solid octagons, which produce the visual evidence $p(I|w_i)$ of w_i in $I \in \mathcal{S}$. In the dashed box, Q and its K_{RC} related multi-concepts $W_i^{(n)} \in \mathcal{Y}^q$ are denoted as dashed circles, and they form a retrieval context $\mathcal{R}_{\mathrm{RC}}(Q)$. Q and $\forall W_i^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q)$ are linked by edges (dashed arrow) weighted according to semantic correlations $\varpi_i = p(W_i^{(n)}|Q)$. Its corresponding weighted multi-concept detectors MD_i are denoted as dashed octagons, which model the maps from $W_i^{(n)}$ to I and produce the visual evidence $\nu_i = p(I|W_i^{(n)})$ of $W_i^{(n)}$ in I. The semantic correlation ϖ_i can be seen as the weight of MD_i. I and the ranked result set $O = \{I(1), \ldots, I(K)\}$ are represented as ellipses. As can be seen from Figure 2, three technical points are considered and shown as follows.

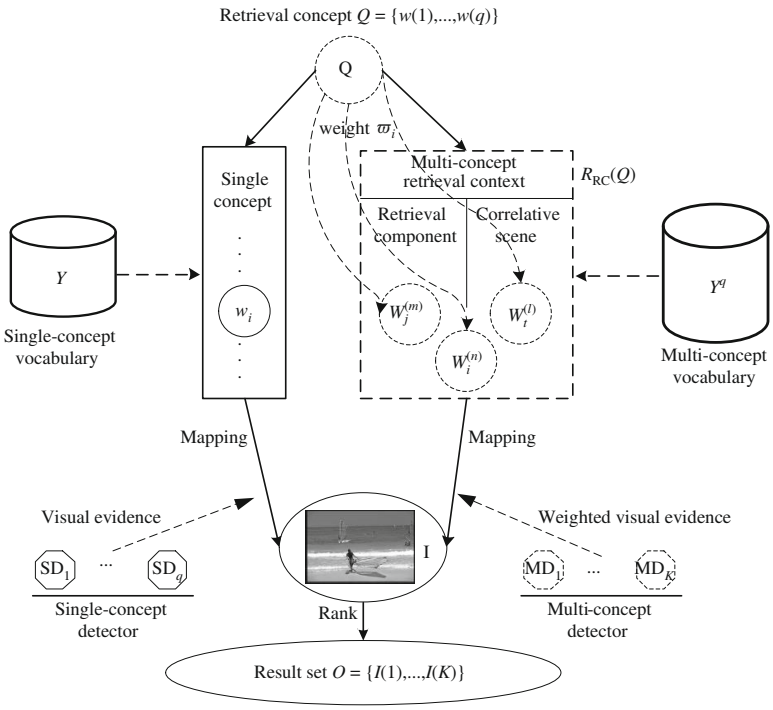

Figure 2 The framework of the MCRM model.

3.1.1 *Multi-concept vocabulary production*

Given a Q and the original $\mathcal{Y}$, MCRM produces a fixed multi-concept vocabulary $\mathcal{Y}^q$ for all q-concept queries with length q in an off-line process. To avoid the meaningless concept permutation (such as ⟨tree, tail, terrace⟩), MCRM selects meaningful n-concepts $W_i^{(n)} \in 2^{\mathcal{Y}}$ to generate the vocabulary $\mathcal{Y}^q$ by the following co-occurrence rule over the training set $\mathcal{T}$:

$$\mathrm{Nr}(W_i^{(n)}) \geqslant c \quad (1 \leqslant n \leqslant q). \tag{1}$$

If $q = 1$ (i.e., a single-concept query Q), $\mathcal{Y}^q$ becomes exactly a single-concept vocabulary $\mathcal{Y}$ (i.e., $\mathcal{Y}^1 = \mathcal{Y}$). If the size of the set $\mathcal{Y}^q$ is very large, the co-occurrence count c in (1) is adjusted to reduce the computational cost. In this way, the multi-concept vocabulary $\mathcal{Y}^q$ is generated. The corresponding generation algorithm is summarized in Algorithm 1.

Algorithm 1 The multi-concept vocabulary generation algorithm

Require: Training set $\mathcal{T}$ with single-concept vocabulary $\mathcal{Y}$ and multi-concept query Q;
Ensure: The multi-concept vocabulary $\mathcal{Y}^q$;
  Initialize the vocabulary, i.e., $\mathcal{Y}^q(1) \Leftarrow \mathcal{Y}$;
  for $n = 2$ to q **do**
    Calculate the n-concept set $\mathcal{W}^{(n)}$ according to (1);
    $\mathcal{Y}^q(n) \Leftarrow \mathcal{Y}^q(n-1) \cup \mathcal{W}^{(n)}$;
  end for
  $\mathcal{Y}^q \Leftarrow \mathcal{Y}^q \cup Q$;
  return $\mathcal{Y}^q$;

3.1.2 *Retrieval context production*

A retrieval concept Q is augmented into two parts: retrieval components $\mathrm{Com}(Q)$ and correlative scene concepts $\mathrm{Csc}(Q)$, which are closely associated with Q and can be seen as the retrieval context. First, the semantic neighbor set $\mathcal{R}(Q) = \{W_i^{(n)} \mid W_i^{(n)} \in \mathcal{Y}^q, p(W_i^{(n)}|Q) > 0\}$ is produced as the candidate concept pool where $p(W_i^{(n)}|Q)$ is the semantic correlation probability between two concepts $W_i^{(n)}$ and Q. Second, all the retrieval components $c_j \in \mathrm{Com}(Q) = \{W_j^{(m)} \mid W_j^{(m)} \in \mathcal{R}(Q), W_j^{(m)} \subseteq Q\}$ can be taken as

the elements of the retrieval context $\mathcal{R}_{\mathrm{RC}}(Q)$. Clearly, $Q \in \mathrm{Com}(Q)$. Last, top t , the most correlative concepts from the set $\{W_t^{(l)} \mid W_t^{(l)} \in \mathcal{R}(Q), W_t^{(l)} \notin \mathrm{Com}(Q)\}$ are chosen into the set $\mathrm{Csc}(Q)$. It is worth noting that each multi-concept $\forall x \in \mathcal{R}_{\mathrm{RC}}(Q)$ is regarded as its own semantic neighbor and $p(x|x) = 1$. In this way, with the two sets $\mathrm{Com}(Q)$ and $\mathrm{Csc}(Q)$, the retrieval context $\mathcal{R}_{\mathrm{RC}}(Q) = \mathrm{Com}(Q) \cup \mathrm{Csc}(Q)$ is produced, which consists of K_{RC} elements.

3.1.3 *Multi-concept map*

The multi-concept map models the link between Q and I, and outputs the image relevance estimate $p(I|Q)$:

$$e_M(I,Q) = \sum_{i=1}^{K_{\mathrm{RC}}} \nu_i \varpi_i = \sum_{i=1}^{K_{\mathrm{RC}}} p\left(I|W_i^{(n)}\right) p\left(W_i^{(n)}|Q\right) \quad \left(W_i^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q)\right), \tag{2}$$

$$e_S(I,Q) = \prod_{i=1}^{q} p(I|w_i), \tag{3}$$

$$p(I|Q) = \lambda_1 e_M(I,Q) + \lambda_2 e_S(I,Q) \quad (0 \leqslant \lambda_1, \lambda_2 \leqslant 1), \tag{4}$$

where the quantity $e_M(I,Q)$ and $e_S(I,Q)$ denote the relevance estimates produced by the multi-concept detectors and the single-concept detectors respectively, λ_1 and λ_2 are the parameters of MCRM to be estimated and s.t. $\lambda_1 + \lambda_2 = 1$. They determine the trade-off between multi-concept detectors and single-concept detectors. The probability $p(I|W_i^{(n)})$ can be seen as visual evidence ν_i of $W_i^{(n)}$ in I. The semantic correlation probability $p(W_i^{(n)}|Q)$ can be regarded as the weight ϖ_i of visual evidence $\nu_i = p(I|W_i^{(n)})$. If $q = 1$, namely, the case of single-concept retrieval, Eq. (4) can tackle it as well. The estimations of semantic correlation $p(W_i^{(n)}|Q)$ and visual evidence $p(I|W_i^{(n)})$ or $p(I|w_i)$ will be presented in Subsections 4.1 and 4.2.

Given Q, if there is strong visual evidence for Q and other highly correlative concepts $W_i^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q)$, the terms $p(I|W_i^{(n)})p(W_i^{(n)}|Q)$ become large and the relevance $e_M(I,Q)$ also gets large. Conversely, if the detectors cannot find strong visual evidence for Q and other correlative concepts $W_i^{(n)}$, $e_M(I,Q)$ gets small.

Compared with previous methods, there are three major differences: (1) given a complex scene query, multi-concept detectors as well as single-concept detectors are concerned with concept detection instead of sole single-concept detectors; (2) the semantic links among the concepts are considered; (3) MCRM makes an assumption that the link between Q and I can be modeled as a map probability $p(I|Q)$ from Q to I.

3.1.4 *The algorithm of multi-concept retrieval*

To effectively perform multi-concept retrieval, the algorithm of multi-concept retrieval is proposed in Algorithm 2. First, a multi-concept vocabulary $\mathcal{Y}^q$ is generated in this algorithm. Second, the retrieval component set $\mathrm{Com}(Q)$ and the correlative scene concept set $\mathrm{Csc}(Q)$ are generated, which constitute the retrieval context $\mathcal{R}_{\mathrm{RC}}(Q)$. Third, for each test image $I \in S$, the relevance probability $p(I|Q)$ is obtained by fusing multi-concept and single-concept detector estimates. It is easy to discern that the time complexity and the space complexity are $O(N)$ and $O(1)$ respectively. Last, the heap sort over all probabilities $p(I|Q)$ is performed and the retrieved image rank $\{I(1), I(2), \ldots, I(K)\}$ is returned. The time complexity and the space complexity of the heap sort are $O(N \log N)$ and $O(1)$ respectively and hence the time complexity and the space complexity of the retrieval algorithm are also $O(N \log N)$ and $O(1)$ respectively.

3.2 Parameter estimation

To find the optimal parameters $\Lambda = (\lambda_1, \lambda_2)$, the log-likelihood function of the predictions over the training set $\mathcal{T}$ is maximized. $y_{Qi} \in \{0, 1\}$ is used to denote the absence/presence of the retrieval multi-

Algorithm 2 The multi-concept retrieval algorithm

Require: Training set $\mathcal{T}$ with single-concept vocabulary $\mathcal{Y}$, Test set $\mathcal{S}$ and multi-concept query Q;
Ensure: The ranked list of images $\{I(1), \ldots, I(K)\}$;
Generate multi-concept vocabulary $\mathcal{Y}^q$ using Algorithm 1;
Calculate semantic neighbor set $\mathcal{R}(Q) = \{W_i^{(n)} \mid W_i^{(n)} \in \mathcal{Y}^q, p(W_i^{(n)}|Q) > 0\}$
Calculate retrieval component set $\mathrm{Com}(Q) = \{W_j^{(m)} \mid W_j^{(m)} \in \mathcal{R}(Q), W_j^{(m)} \subseteq Q\}$
Select top t, the most correlative concepts from the set $\{W_t^{(l)} \mid W_t^{(l)} \in \mathcal{R}(Q), W_t^{(l)} \notin \mathrm{Com}(Q)\}$ as the set $\mathrm{Csc}(Q)$ and obtain the retrieval context: $\mathcal{R}_{\mathrm{RC}}(Q) \Leftarrow \mathrm{Com}(Q) \cup \mathrm{Csc}(Q)$
for each $I \in S$ **do**
  Calculate relevance estimate $e_M(I, Q)$ of I and Q produced by multi-concept detector according to (2);
  Calculate relevance estimate $e_S(I, Q)$ of I and Q produced by single-concept detector according to (3);
  Fuse multi-concept and single-concept detector estimate and obtain final relevance probability $p(I|Q)$ of I and Q according to (4);
end for
Perform the heap sort in a descending order over all probabilities $p(I|Q)$ for obtaining top K images $I \in \mathcal{S}$;
return the path $\{I(1), I(2), \ldots, I(K)\}$ which stands for the retrieved image rank;

concept Q for the training image $I_i \in \mathcal{T}$ and the prediction $p(y_{Qi})$ is given by

$$p(y_{Qi} = 1) = p(I_i|Q), \tag{5}$$

$$p(y_{Qi} = 0) = 1 - p(I_i|Q), \tag{6}$$

$$p(y_{Qi}) = p(I_i|Q)^{y_{Qi}}(1 - p(I_i|Q))^{1-y_{Qi}}. \tag{7}$$

The log-likelihood function of the retrieval concept Q can be written as

$$\mathcal{L}_Q = \sum_{i=1}^{T} n_{Qi} \log p(y_{Qi}), \tag{8}$$

where T denotes the total size of the training set, and n_{Qi} is a cost which considers the imbalance of the number $N+$ of the positive examples and the number $N-$ of negative examples for the retrieval concept Q. $n_{Qi} = 1/N+$ if $y_{Qi} = 1$ and $n_{Qi} = 1/N-$ otherwise. Substitute (4) and (7) into (8), and then the following log-likelihood function is obtained:

$$\mathcal{L}_Q = \sum_{i=1}^{T} n_{Qi} \log\{(\Lambda[e_M(I, Q), e_S(I, Q)]^T)^{y_{Qi}}(1 - \Lambda[e_M(I, Q), e_S(I, Q)]^T)^{1-y_{Qi}}\}. \tag{9}$$

The log-likelihood function $\mathcal{L}_Q$ is maximized by the gradient descent approach [25]. The gradient of (9) is

$$\frac{\partial \mathcal{L}_Q}{\partial \lambda_k} = \sum_{i=1}^{T} \frac{n_{Qi}(\frac{\partial \Lambda}{\partial \lambda_k}[e_M(I, Q), e_S(I, Q)]^T)^{y_{Qi}}(-\frac{\partial \Lambda}{\partial \lambda_k}[e_M(I, Q), e_S(I, Q)]^T)^{1-y_{Qi}}}{(\Lambda[e_M(I, Q), e_S(I, Q)]^T)^{y_{Qi}}(1 - \Lambda[e_M(I, Q), e_S(I, Q)]^T)^{1-y_{Qi}}}, \tag{10}$$

where $k \in \{1, 2\}$.

4 The estimation of semantic correlation and visual evidence

4.1 The estimation of semantic correlation probability

The semantic correlation probability $p(W_i^{(n)}|Q)$ is calculated with three correlation measures between the multi-concept $x = W_i^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q)$ and the retrieval concept $y = Q$, denoted by NGD, CO1 and CO2, respectively. According to these three measures, the probability $p(W_i^{(n)}|Q)$ is called as $p_{\mathrm{NGD}}(W_i^{(n)}|Q)$, $p_{\mathrm{CO1}}(W_i^{(n)}|Q)$ and $p_{\mathrm{CO2}}(W_i^{(n)}|Q)$ respectively. Our correlation measures can calculate the conventional correlations between single-concepts and among single-concepts and multi-concepts.

4.1.1 *Google correlation*

The local image corpus is used to calculate the normalized Google distance in order to reflect the properties of the local image dataset. Given two concepts x and y, the Google distance $\mathrm{Dist}(x, y)$ is defined as follows [24]:

$$\mathrm{Dist}(x, y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x, y)}{\log N_G - \min\{\log f(x), \log f(y)\}}, \tag{11}$$

where $f(x)$ and $f(y)$ denote the number of images containing x and y respectively, $f(x, y)$ denotes the number of images containing x and y simultaneously, and N_G denotes the total number of all images.

The semantic correlation between two concepts $x = W_i^{(n)}$ and $y = Q$ is then calculated as the negative exponentiation of $\mathrm{Dist}(x, y)$:

$$p_{\mathrm{NGD}}\left(W_i^{(n)}|Q\right) = \exp\left(-\frac{\mathrm{Dist}(W_i^{(n)}, Q)}{\delta}\right), \tag{12}$$

where δ is a distance smoothing parameter.

4.1.2 *Co-occurrence correlation*

Co-occurrence in a linguistic sense can be interpreted as an indicator of semantic correlation, and it assumes the interdependency of two concepts in a document. Two co-occurrence correlation measures (such as CO1 and CO2) between the concepts $x = W_i^{(n)}$ and $y = Q$ over the training data are defined as follows:

$$p_{\mathrm{CO1}}\left(W_i^{(n)}|Q\right) = \frac{\mathrm{Nr}(W_i^{(n)}, Q)}{\mathrm{Nr}(W_i^{(n)})}, \tag{13}$$

$$p_{\mathrm{CO2}}\left(W_i^{(n)}|Q\right) = \frac{2 \times \mathrm{Nr}(W_i^{(n)}, Q)}{\mathrm{Nr}(W_i^{(n)}) + \mathrm{Nr}(Q)}. \tag{14}$$

The correlation probability between the concept Q and itself is one, e.g., $p(Q|Q) = 1$. In order to keep the probabilistic attribute of the semantic correlation, $p(W_i^{(n)}|Q)$ are normalized as follows:

$$p\left(W_i^{(n)}|Q\right) = \begin{cases} \frac{p(W_i^{(n)}|Q)}{\sum_{j=1}^{K_{\mathrm{RC}}} p(W_j^{(n)}|Q)}, & \text{if } W_i^{(n)}, W_j^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q), \\ 0, & \text{elsewhere.} \end{cases} \tag{15}$$

The multi-concepts $W_i^{(n)}$ with a high correlation have high weights $\varpi_i = p(W_i^{(n)}|Q)$ of the corresponding multi-concept detectors. Of all the multi-concepts $W_i^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q)$, Q has the highest correlation, which ensures that its corresponding detector $p(I|Q)$ plays a key role in the detection of Q.

4.2 The estimation of visual evidence

The probability $p(I|W_i^{(n)})$ can be seen as visual evidence of the multi-concept $W_i^{(n)}$ in the image I, and by Bayes' theorem, it can be expressed in the form:

$$p\left(I|W_i^{(n)}\right) = \begin{cases} \frac{p(W_i^{(n)}|I)p(I)}{p(W_i^{(n)})}, & \text{if } \mathrm{Nr}(W_i^{(n)}) > 0, \\ 0, & \text{elsewhere,} \end{cases} \tag{16}$$

where the value of $p(I|W_i^{(n)})$ is set to zero if the multi-concept $W_i^{(n)}$ has no occurrence in the training set. The denominator $p(W_i^{(n)})$ in (16) can be interpreted as the prior probability for the multi-concept $W_i^{(n)}$ following the Bernoulli distribution, like the previous MBRM model [10]. Through Bayesian treatment with the beta conjugate prior probability distribution, it can be estimated as follows [25]:

$$p\left(W_i^{(n)}\right) = \frac{\mathrm{Nr}(W_i^{(n)}) + b}{T + a}, \tag{17}$$

where a and b act as the smoothing parameters and T is the total size of the training set.

The probability $p(I|W_i^{(n)})$ in (16) can be estimated by a concept detector, such as SVM, Naive Bayesian, Random Forest, etc. In this work, SVM is adopted because it is an effective detector and can deal with high-dimensional visual features. However, SVM is a decision machine and provides the decision score $f(W_i^{(n)}, I)$ rather than the posterior probability. Therefore, the decision score is mapped to the probability by the logistic sigmoid function $\sigma(\eta)$ [25], and Eq. (16) can be rewritten as follows:

$$p\left(I|W_i^{(n)}\right) = \begin{cases} \sigma\left(\frac{f(W_i^{(n)},I)p(I)}{p(W_i^{(n)})}\right), & \text{if } \mathrm{Nr}(W_i^{(n)}) > 0, \\ 0, & \text{elsewhere.} \end{cases} \tag{18}$$

The quantity $p(I)$ in (18) can be interpreted as the prior probability for the image I following the uniform distribution, i.e., $p(I)$ is a constant. The estimation of the probability $p(I|w_i)$ of the single-concept w_i in the image I is the same as that of $p(I|W_i^{(n)})$.

For each concept, a two-class SVM detector is trained based on a one-versus-the-rest approach. Given a $w_i \in Q$, the conventional single-concept detector is used which considers images $I \in \mathcal{T}$ annotated with w_i as positive samples and the rest as negative samples.

For a scene multi-concept $W_i^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q)$, the original single-concept training data are rearranged. The positive sample set $\mathrm{Po}(W_i^{(n)})$ and the negative sample set $\mathrm{Ne}(W_i^{(n)})$ are constructed as follows:

$$\begin{aligned} \mathrm{Po}\left(W_i^{(n)}\right) &= \left\{I|W_i^{(n)} \subseteq A(I)\right\}, \\ \mathrm{Ne}\left(W_i^{(n)}\right) &= \left\{I|I \notin \mathrm{Po}(W_i^{(n)})\right\}, \end{aligned} \tag{19}$$

where $A(I)$ is the annotation concept set for the training images $I \in \mathcal{T}$. In addition, considering the imbalance between the size $N+$ of the positive sample set $\mathrm{Po}(W_i^{(n)})$ and the size $N-$ of the negative sample set $\mathrm{Ne}(W_i^{(n)})$, the weights $1/N+$ and $1/N-$ are given for positive samples and negative samples for $W_i^{(n)}$ training, respectively. In this way, with these two sets $\mathrm{Po}(W_i^{(n)})$ and $\mathrm{Ne}(W_i^{(n)})$, each two-class SVM is trained as the multi-concept detector and the learned detector can output $p(W_i^{(n)}|I)$.

5 Experiments and analysis

5.1 Datasets

The proposed MCRM is evaluated on two public image datasets, namely Corel [26] and IAPR TC-12 [27]. The Corel dataset contains about 5000 images. Each image is manually annotated with 1–5 concepts from a vocabulary $\mathcal{Y}$ consisting of 260 semantic concepts. The IAPR dataset consists of about 20000 still images and $\mathcal{Y}$ contains 291 concepts.

Note that all test images have no semantic annotations. $\mathcal{Y}$ contains a few hundred semantic concepts, and about 75% of the semantic concepts have frequencies less than the average concept frequency.

5.2 Experimental setup

5.2.1 *Visual features*

In the experiments, eleven visual features are used, including four SIFT features for dense SIFT and Harris SIFT, a Gist feature, and six color features for RGB, LAB and HSV. They are the same as the work of [3] and are publicly available for download[3)].

LIBSVM software [17] is used for all of our SVM experiments. To compute the distance between two visual features, the HI measure [28] is employed for the SIFT features and the RBF measure [29] for the rest {Gist, RGB, LAB, HSV}. Following the previous work [3], the mean of all distances is employed for the SVM.

3) http://lear.inrialpes.fr/people/guillaumin/data.php.

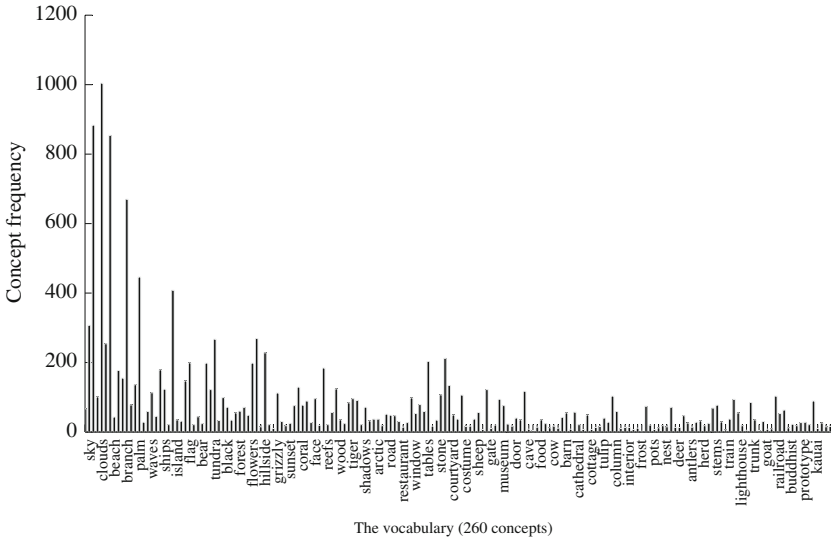

Figure 3 Unbalanced concept distribution on Corel 5K [26].

5.2.2 *Baseline methods*

Two conventional baseline methods are considered: a random method[4)] and an SVM method. The baselines use product fusion and addition fusion to obtain $p(Q|I)$. It is found that product fusion gives better results and hence product fusion is adopted for the baselines.

5.2.3 *Test query set*

Following [3], on the Corel dataset, the same subset is used, which consists of 179 $w_i \in \mathcal{Y}$ that appear at least twice in the test set. Besides, 2062 multi-concept queries with at least one relevant image in the test set [3] are considered, i.e., 967 2-concepts, 873 3-concepts and 222 4-concepts. In this way, the test query set $\mathcal{Q}^T$ is constructed. The corresponding vocabularies $\mathcal{Y}^1 = \mathcal{Y}$, $\mathcal{Y}^2$, $\mathcal{Y}^3$ and $\mathcal{Y}^4$ contain a total of 4105 concepts. The parameters c, δ, a and b are set to 2, 0.1, 300 and 200, respectively. In order to set the size K_{RC} of the retrieval context $\mathcal{R}_{\mathrm{RC}}(Q)$, 10-fold cross-validation is performed. By observing a range of $K_{\mathrm{RC}}([1, 30])$ during validation, $K_{\mathrm{RC}} = 15$ is used as a default choice as it shows good performance.

For experiments on the IAPR set, besides all 291 $w_i \in \mathcal{Y}$, 3900 multi-concept queries are randomly selected, which consist of 1300 2-concepts, 1300 3-concepts and 1300 4-concepts. The vocabularies $\mathcal{Y}^1 = \mathcal{Y}$, $\mathcal{Y}^2$, $\mathcal{Y}^3$ and $\mathcal{Y}^4$ contain 6565 multi-concepts in all. The parameters are set to be $c = 30$, $\delta = 0.1$, $a = 400$ and $b = 200$.

For full comparability, the recall curves, the precision-recall (PR) curves and the mean average precisions (MAP) over concepts [30] are taken as performance measures. The higher the MAP score is, the better the retrieval performance will be.

5.3 Experimental results and analysis

5.3.1 *Frequent and rare concept experiment*

Most concepts from $\mathcal{Y}$ have frequencies less than the average concept frequency. For clearness, all concept frequencies on the training images are shown in Figure 3. This highly unbalanced concept distribution is often found in engineering applications and may impact the retrieval performance. Concept detectors always have good accuracy on the frequent concepts[5)] but very poor accuracy on the rare concepts[6)] [31].

First, the proposed MCRM[7)] is compared with two baseline methods for the frequent and rare concepts

4) A random method means randomly retrieving the images.

5) The frequent concepts mean the concepts with high frequency appearing in the training set.

6) The rare concepts mean the concepts with low frequency appearing in the training set.

7) Our MCRM is denoted with three correlation measures as NGD+MCRM, CO1+MCRM and CO2+MCRM, respectively.

Table 1 Retrieval performance comparison (MAP Scores %) for frequent and rare concepts on Corel

Experiment	Frequent 1-concept	Rare 1-concept	Frequent 2-concept	Rare 2-concept
Random Baseline	8.8	1.2	8.7	1.1
SVM Baseline	59.0	32.7	36.0	15.7
NGD+MCRM	59.0	32.7	41.1	36.0
CO1+MCRM	60.8	38.2	41.3	**36.6**
CO2+MCRM	**61.0**	**39.6**	**41.5**	34.7

Table 2 Multi-concept retrieval performance (MAP Scores %) with different semantic measures on Corel

Experiment	1-concept	2-concept	3-concept	4-concept
Random Baseline	3.7	1.6	1.6	1.8
SVM Baseline	44.5	32.0	31.8	33.7
NGD+MCRM	47.0	40.1	**39.2**	**40.9**
CO1+MCRM	47.9	**40.8**	**39.2**	40.7
CO2+MCRM	**48.6**	**40.8**	**39.2**	40.6

on the Corel dataset. The 50 most frequent and 50 most rare single-concepts $w_i \in \mathcal{Q}^T$ are respectively taken as the frequent and rare 1-concept query sets, and each frequent and each rare 2-concept $W_i^{(2)} \in \mathcal{Q}^T$ contains such frequent and rare concepts w_i separately. Table 1 shows the MAP scores and bold number means the highest MAP score (%). From Table 1, it can be observed that all approaches achieve high MAP scores of about 60% for the frequent single-concept retrieval, while the MAP scores of the rare single-concept are much lower. For example, the conventional SVM get the MAP score of 59.0% for the frequent single-concepts, while for the rare single-concepts the score sharply declines by 45%. Similar results have also been found on multi-concept retrieval. For example, for the rare 2-concept retrieval, the SVM baseline score sharply declines by 56% compared with frequent 2-concept. It can be observed that better scores tend to occur on the frequent concepts.

Another interesting thing is that MCRM achieves remarkable improvements of about average 128% on 2-concept and about average 13% on single-concept for the rare concept queries, respectively. A rare concept Q may be difficult to be detected, but it may have correlative frequent concepts $W_i^{(n)} \in \mathcal{R}_{\mathrm{RC}}(Q)$, which can be detected efficiently. Thus, this concept Q can be efficiently detected through its correlative frequent concepts. This is one of the reasons why our approach can achieve the improved performance.

From Column 2 and 4 in Table 1, it can be seen that our method also obtains high retrieval performance on frequent concepts. The MCRM improvements are observed, about 2.2% on single-concept, 14.7% on 2-concept.

5.3.2 *Multi-concept retrieval experiment*

Next the retrieval performance is compared for all 2241 queries $\mathcal{Q}^T$ containing 2062 multi-concepts and 179 single-concepts over the Corel set. The MAP scores are presented in Table 2. As can be seen from Table 2, the retrieval performances of the proposed MCRM are superior to the baselines. From Column 2 in Table 2, it can be seen that our MCRM averagely improves about 7.5% than the SVM baseline in the conventional single-concept retrieval.

Since our approach utilizes semantic correlations between concepts and does not use multi-concept detectors in single-concept retrieval, this performance improvement may be caused by using semantic correlation. For n-concept queries ($n \geqslant 2$), the SVM baseline yields inferior retrieval ranking, while our approach achieves marked improvements of about 27% on 2-concept queries, 23% on 3-concept ones and 21% on 4-concept ones on average. It can be easily noted that the MAP scores of multi-concept retrieval are much lower than those of single-concept retrieval. It is not surprising to consider the difficulty of recognizing multiple concepts in an image. The conventional SVM method only uses single-concept detectors for multi-concept scene retrieval, while our approach, besides concept correlations, combines multi-concept detectors and conventional single-concept detectors. Hence, significant improvements for multi-concept retrieval have been obtained.

Table 3 Comparison of the proposed MCRM and the previous approaches in terms of MAP on dataset Corel

Experiment	All Query	Difficult	Easy	Multi-concept	Single-concept
Random Baseline	1.8	1.2	4.2	1.6	3.7
CMRM [8]	19.2	15.8	34.0	18.6	25.8
CMTT [11]	19.8	17.2	31.3	19.3	26.4
PLSA [15]	20.7	16.7	38.0	19.7	31.7
PAMIR [2]	26.3	22.4	43.3	25.7	34.0
GS [1]	27	22.3	47.4	25.5	44.0
SVM Baseline	29.5	23.7	54.6	28.2	44.5
TagProp [3]	36	32	55	35	46
NGD+MCRM	40.2	36.3	57.2	39.8	47.0
CO1+MCRM	40.7	36.7	**58.3**	**40.1**	47.9
CO2+MCRM	**40.8**	**36.8**	57.9	**40.1**	**48.6**

5.3.3 *Difficult/easy and multi-concept/single-concept retrieval experiment*

Table 3 presents the performances of the proposed MCRM and the previous retrieval approaches. The set $\mathcal{Q}^T$ of all 2241 test queries is divided into two groups: (1) a "difficult" set consisting of 1820 queries with only one or two relevant images in the test set $\mathcal{S}$ and an "easy" set consisting of 421 queries with 3 or more relevant images in the set $\mathcal{S}$; (2) a multi-concept set consisting of 2062 n-concept queries ($n \geqslant 2$) and a single-concept set consisting of 179 single-concept queries. This division follows [2,3]. According to the split, the corresponding MAP scores (%) are reported separately. It can be observed from Table 3 that our proposed approach outperforms the previous methods and its advantage shows more obvious for the multi-concept queries as well as the "difficult" queries. Compared with SVM and TagProp, the two best alternative methods, our mean improvements are 41.8% or 14.3% on multi-concept queries and 54.4% or 14.4% on "difficult" queries, respectively.

The Group Sparsity approach (GS) gains a competitive MAP score of 44% on single-concept queries and outperforms CMRM, CMTT, PLSA and PAMIR. However, on multi-concept queries as well as the "difficult" queries, this new approach is close to PAMIR and inferior to our model. These two aforementioned queries (i.e., multi-concept queries and "difficult" queries) are harder queries since the relevant test images account for 1.76 and 1.19 per query on average respectively, which should be compared with 9.36 and 7.48 per query on average for the single-concept queries and the "easy" queries, respectively. Clearly, the advantage of our multi-concept retrieval model is confirmed on these two queries, getting the average scores of 40.0% and 36.6%. A conventional single-concept detector can effectively detect a single-concept in an image, while a multi-concept detector can effectively detect a multi-concept scene with characteristic visual appearance. A combination of them can improve the performance of multi-concept retrieval. Besides, MCRM improves the performance on rare concepts by taking the relationships among concepts into consideration. There are two reasons behind the improvements. For all 2241 queries $\mathcal{Q}^T$, CO2+MCRM, CO1+MCRM and NGD+MCRM get the average MAP scores of 40.8%, 40.7% and 40.2% respectively. They improve about 38.3%, 38.0%, 36.3% over the SVM baseline respectively, and about 13.3%, 13.1%, 11.7% over TagProp respectively.

5.3.4 *The recall curves and precision recall curves*

In Figure 4, the recall curves are shown in 4(a) and the precision recall curves are given in 4(b) for the SVM baseline, TagProp, which are the two best alternative approaches, and our NGD+MCRM, CO1+MCRM and CO2+MCRM. As can be seen from Figure 4(a), when the number of the returned image examples gets large, the number of retrieved relevant images becomes larger. When the returned images increase to about 20% of the size of the test set, the 90% relevant images are retrieved by MCRM while SVM and TagProp respectively retrieve 64% and 82% relevant images. For all 2241 queries containing 90%+ multi-concepts queries, MCRM shows higher recall at each level (see Figure 4(a)).

As can be seen from Figure 4(b), MCRM has higher precision than SVM and TagProp at every level of recall. The two best previous methods do not explicitly consider the correlations between concepts, or

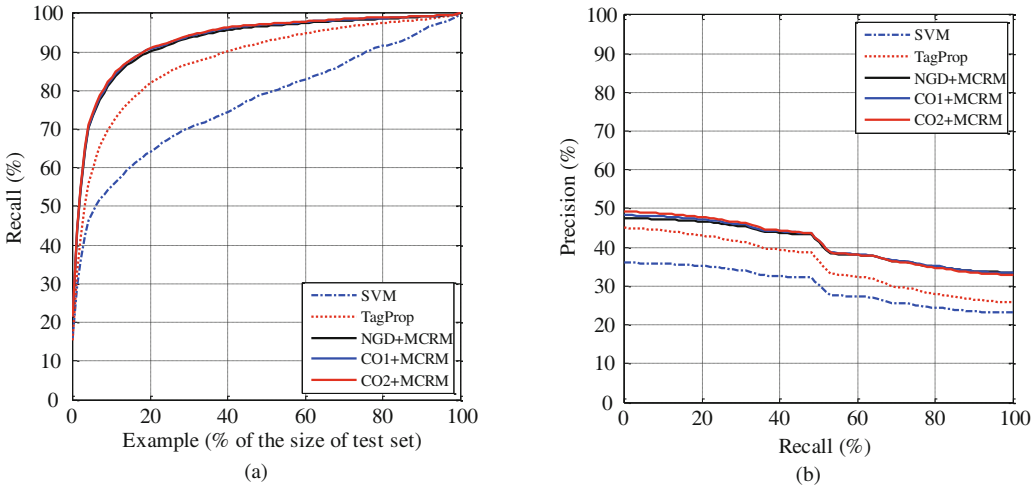

Figure 4 Plots of the recall curves (a) and the precision recall curves (b) over all 2241 queries.

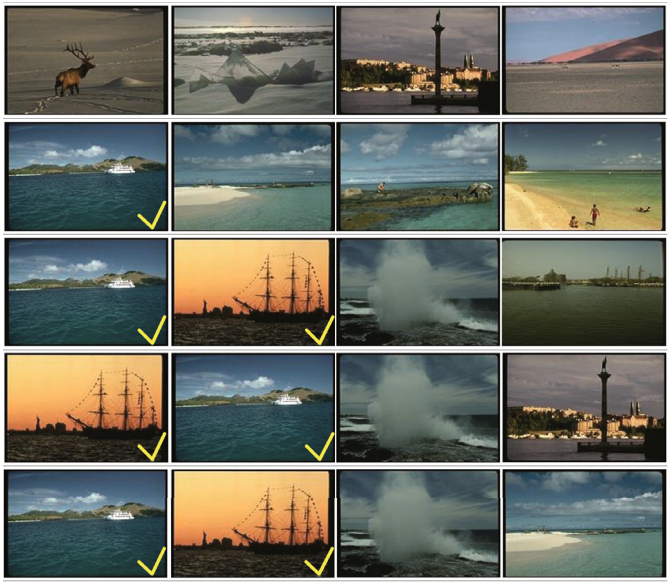

Figure 5 An example: a multi-concept retrieval ⟨sky, water, ship⟩ with two relevant images in the test set.

combine the multi-concept detectors to effectively detect a multi-concept scene for multi-concept retrieval.

5.3.5 *An illustration of multi-concept retrieval on Corel*

The advantages of our multi-concept approaches are illustrated in Figure 5. The images retrieved by SVM, TagProp, NGD+MCRM, CO1+MCRM and CO2+MCRM are shown in the first to the fifth rows, respectively. As can be seen from Figure 5, the SVM and TagProp detectors find no and one relevant image simultaneously containing three concepts ⟨sky⟩, ⟨water⟩ and ⟨ship⟩, respectively, while MCRM succeeds in retrieving both the two relevant images in the top four positions.

For a Q = ⟨sky, water, ship⟩ which represents a meaningful scene, the SVM detector distinguishes the concept ⟨sky⟩, yet it does not effectively distinguish the scene concept ⟨sky, water, ship⟩. It cannot find relevant images for Q. Maybe the SVM detector is confused by the similar visual content such as snow scenes and water scenes. The TagProp single-concept detectors favor the frequent concept ⟨sky⟩ and ⟨water⟩ (respectively with concept frequency 1004 and 883) at the expense of the rare concept ⟨ship⟩ (only with concept frequency 21). In this example, only MCRM multi-concept detectors succeed in retrieving both the two relevant images in the top four positions. The reason may be that our detectors improve the performance on the rare single-concept ⟨ship⟩ using semantic correlations. On the concept ⟨ship⟩,

Table 4 Comparison of MCRM and the previous approaches in terms of MAP on IAPR 20K

Experiment	All Query	Difficult	Easy	Multi-concept	Single-concept
Random Baseline	0.9	0.7	1.2	0.8	2.6
SVM Baseline	31.6	22.8	44.4	31.2	36.6
TagProp [3]	30.5	25.6	37.7	30.1	**39.9**
NGD+MCRM	**37.6**	32.4	**45.2**	**37.6**	37.0
CO1+MCRM	37.2	32.5	44.1	37.2	37.4
CO2+MCRM	37.2	**32.6**	44.0	37.2	37.2

the average score of 83% is achieved by MCRM while the scores of 76% and 71% are respectively got by SVM and TagProp. Furthermore, the multi-concept scene ⟨sky, water, ship⟩ is effectively distinguished by a combination of multi-concept detectors and single-concept detectors (the score 100% is obtained).

5.3.6 *Difficult/easy and multi-concept/single-concept retrieval experiment on IAPR*

Last, the experimental results performed on the dataset IAPR are reported. This dataset contains more images with varying appearances and assorted aspects of the contemporary life, and the concepts are extracted from free-flowing text captions. Hence, this dataset is closer to real world cases. The average number of concepts of each test image is 5.6 which should be compared with 3.5 for the dataset Corel. Recognizing a multi-concept on these complex images is a challenging task.

Similarly to [3], all 4191 queries are divided into: (1) 2492 difficult ones and 1699 easy ones; (2) 3900 multi-concept and 291 single-concept ones. According to the split, the corresponding MAP scores (%) are listed separately. Our multi-concept image retrieval method is compared with the above two best previous methods (i.e., SVM and TagProp) and the performances are shown in Table 4. As can be observed from Table 4, for almost all queries our model outperforms SVM and TagProp and gains about 18% and 22% improvements over all 4191 queries respectively. Especially for multi-concept queries (Column 5) and "difficult" queries (Column 3), MCRM yields average MAP scores of 37.3% and 32.5% and obtains significant improvements of about 20% or 24% on multi-concept queries and 43% or 27% on "difficult" queries respectively, compared with SVM and TagProp. For single-concept queries, TagProp uses the weights based on the single-concept distance and uses the metric learning method to learn the weights, showing the best performance in single-concept retrieval. However, for multi-concept retrieval, TagProp performs a scene retrieval solely by single-concept detectors, which does not work well and is close to the SVM baseline.

6 Conclusion

In this paper, a novel probabilistic model has been proposed for multi-concept retrieval. The proposed multi-concept detectors in the model can effectively identify a multi-concept in the image. In contrary to previous single-concept models, both single-concept detectors and multi-concept detectors are incorporated in the multi-concept recognizing procedure. A group of correlative weighted detectors are involved in such a multi-concept recognition through retrieval context. Both of them are important and their combination can yield better retrieval performances for the multi-concept retrieval, as shown by our experiments. To better capture the correlations among concepts, further research will be conducted to explore the connections among the concepts such as semantic hierarchy or high-order correlations. Besides, more effective training methods will be exploited in our future work.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant Nos. 61370229, 61370178, 61272067), National Key Technology R&D Program (Grant No. 2013BAH72B01), MOE-China Mobile Research Fund (Grant No. MCM20130651), the Natural Science Foundation of GDP (Grant No. S2013010015178), and Science-Technology Project of GDED (Grant No. 2012KJCX0037).

References

1 Zhang S, Huang J, Li H, et al. Automatic image annotation and retrieval using group sparsity. IEEE Trans Syst Man Cybern Part B-Cybern, 2012, 42: 838–849
2 Grangier D, Bengio S. A discriminative kernel-based approach to rank images from text queries. IEEE Trans Patt Anal Mach Intell, 2008, 30: 1371–1384
3 Guillaumin M, Mensink T, Verbeek J, et al. Tagprop: discriminative metric learning in nearest neighbor models for image auto-annotation. In: Proceedings of 12th International Conference on Computer Vision, Kyoto, 2009. 309–316
4 Chen M, Zheng A, Weinberger K. Fast image tagging. In: Proceedings of 30th International Conference on Machine Learning, Atlanta, 2013. 1274–1282
5 Xu C, Wang T, Gao J, et al. An ordered-patch-based image classification approach on the image grassmannian manifold. IEEE Trans Neural Netw Learn Syst, 2014, 25: 728–737
6 Truong B Q, Sun A X, Bhowmick S S. CASIS: a system for concept-aware social image search. In: Proceedings of 21st International World Wide Web Conference, Lyon, 2012. 425–428
7 Gao Y, Wang M, Zha Z J, et al. Visual-textual joint relevance learning for tag-based social image search. IEEE Trans Image Process, 2013, 22: 363–376
8 Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models. In: Proceedings of 26th Annual International ACM SIGIR Conference, Toronto, 2003. 119–126
9 Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures. In: Thrun S, Saul L K, Schölkopf B, eds. Advances in Neural Information Processing Systems 16. Cambridge: MIT Press, 2003. 553–560
10 Feng S, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington, 2004. 995–1002
11 Pan J Y, Yang H J, Duygulu P, et al. Automatic image captioning. In: Proceedings of IEEE International Conference on Multimedia and Expo, Taipei, 2004. 1987–1990
12 Debatty T, Michiardi P, Mees W, et al. Determining the k in k-means with MapReduce. In: Proceedings of EDBT/ICDT 2014 Joint Conference, Athens, 2014. 19–28
13 Nguyen C T, Kaothanthong N, Tokuyama T, et al. A feature-word-topic model for image annotation and retrieval. ACM Trans Web, 2013, 7: 12–35
14 Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation. J Mach Learn Res, 2003, 3: 993–1022
15 Monay F, Gatica-Perez D. Plsa-based image auto-annotation: constraining the latent space. In: Proceedings of 12th Annual ACM International Conference on Multimedia, New York, 2004. 348–351
16 Blei D M, Jordan M I. Modeling annotated data. In: Proceedings of 26th Annual International ACM SIGIR Conference, Toronto, 2003. 127–134
17 Chang C C, Lin C J. Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol, 2011, 2: 27
18 Lai H J, Pan Y, Tang Y, et al. Fsmrank: feature selection algorithm for learning to rank. IEEE Trans Neural Netw Learn Syst, 2013, 24: 940–952
19 Cui C, Ma J, Lian T, et al. Ranking-oriented nearest-neighbor based method for automatic image annotation. In: Proceedings of 36th International ACM SIGIR Conference, Dublin, 2013. 957–960
20 Liu J, Li M, Liu Q, et al. Image annotation via graph learning. Patt Recognit, 2009, 42: 218–228
21 Makadia A, Pavlovic V, Kumar S. Baselines for image annotation. Int J Comput Vision, 2010, 90: 88–105
22 Jin Y, Khan L, Wang L, et al. Image annotations by combining multiple evidence & wordnet. In: Proceedings of 13th Annual ACM International Conference on Multimedia, Hilton, 2005. 706–715
23 Cilibrasi R L, Vitanyi P M. The google similarity distance. IEEE Trans Knowl Data Eng, 2007, 19: 370–383
24 Chen P I, Lin S J, Chu Y C. Using google latent semantic distance to extract the most relevant information. Expert Syst Appl, 2011, 38: 7349–7358
25 Bishop C M. Pattern Recognition and Machine Learning. New York: Springer, 2006
26 Duygulu P, Barnard K, Freitas J F, et al. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In: Proceedings of 7th European Conference on Computer Vision, Copenhagen, 2002. 97–112
27 Grubinger M, Clough P, Müller H, et al. The IAPR TC-12 Benchmark: a new evaluation resource for visual information systems. In: Proceedings of International Conference on Language Resources and Evaluation, Genoa, 2006. 13–23
28 Maji S, Berg A C, Malik J. Classification using intersection kernel support vector machines is efficient. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, 2008. 1–8
29 Das S R, Panigrahi P K, Das K, et al. Improving RBF kernel function of support vector machine using particle swarm optimization. Int J Adv Comput Res, 2012, 2: 130–135
30 Manning C D, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge: Cambridge University Press, 2008
31 Ganganwar V. An overview of classification algorithms for imbalanced datasets. Int J Emerg Technol Adv Eng, 2012, 2: 42–47