

Image region label refinement using spatial position relation graph

Jing Zhang^{*}, Zhenkun Wang, Yakun Mu, Zhe Wang

Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China

ARTICLE INFO

Article history:

Received 12 April 2018

Received in revised form 4 December 2018

Accepted 8 December 2018

Available online 13 December 2018

MSC:

00–01

99–00

Keywords:

Image region annotation

Label refinement

Spatial position relation graph

Random-walking

Graph matching

ABSTRACT

With the exponential growth of massive image data, automatic image annotation is becoming more important in image management and retrieval. Traditional image region annotation methods, through machine learning and low-level visual features, typically yield incorrect annotation results owing to the influence of the Semantic Gap. We herein propose a novel label refinement method for improving the image region annotation results. A spatial position relation graph with co-occurrence relations and spatial position relations among labels is proposed to analyze the latent semantic correlations among image region labels. Moreover, an incremental iterative random-walking algorithm is proposed to reconstruct the region relation graph for detecting non-dependable regions whose labels do not fit the semantic context of an image. Subsequently, a graph matching algorithm with semantic correlation and spatial relation analysis is proposed for non-dependable region label completion. Experiments on Corel5K demonstrate that our proposed spatial-position-relation-graph-based label refinement method can achieve good performance for image region label refinement.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

With the development of advanced digital capturing technologies and social networks, the quantity of images has increased significantly and the requirements of retrieval have become more complex [1]. Representing images by semantic labels instead of low-level visual features is important in image management and retrieval [2,3]. Traditional automatic image annotation methods still struggle in bridging the semantic gap between the high-level semantics and low-level features by machine learning [4,5], and most of them regard that the labels in an image are mutually independent and ignore the relations between semantics and position [6,7]. Hence, some inconceivable annotation results may be obtained, as shown in Fig. 1, in which “sea” is to be located above “building”, and “water” is wrongly marked as “cloud”.

The semantic information of an image is typically extremely complex and can be described by a group of labels. These labels are strongly correlated in terms of semantics and positions. For example, the regions of “sea” typically appears with the regions of “sand”, and they are located in the area above the image. However, because the low-level features of “sea” and “sky” regions are highly similar, our machine-learning-based annotation methods would sometimes annotate the region of “sky” as “sea”, thus causing the “sea” region to be located above the “building” region in the image,

as illustrated in Fig. 1. Obviously, these annotation results deviate from the objective principle.

Upon careful study, we found that some label pairs, such as “sky” and “cloud”, appear together frequently in an image. Further, the other label pairs, such as “train” and “sea”, rarely appear in the same image. In addition, some label pairs always present fixed spatial position relations including topological relations and directional relations. For example, the “building” region is always in proximity to the “plant” region, and the “sky” region is typically located above the “building” and “sea” regions in an image. Hence, the spatial position relation between labels in images can be applied effectively for label refinement, where the aim is to detect erroneous annotations and predict more suitable labels to replace them [8–10]. Although many image label refinement methods have been proposed in the last decade [11–25], a complete spatial semantic correlation model based on images has not been studied systematically and constructed.

We herein propose a novel image region label refinement framework. First, a spatial position relation graph (SPRG) is constructed by analyzing the relationships of labels between semantics and region positions, in which the spatial position relations and semantic correlations among labels are fully defined and learned. Subsequently, incremental iterative random-walking is applied in reconstructing a region relation graph (RRG) to achieve the dependable value of each label. Subsequently, non-dependable labels will be replaced by more suitable labels obtained by our proposed label completion algorithm.

The original contributions of this study are illustrated as follows.

^{*} Corresponding author.

E-mail addresses: jingzhang@ecust.edu.cn (J. Zhang), wangzhe@ecust.edu.cn (Z. Wang).



Fig. 1. Examples of wrong annotation results in the images.

- A novel SPRG that considers both the co-occurrence relations and spatial position relations (including topological relations and directional relations) is constructed, which can represent the latent semantic correlation among the labels in an image effectively.
- Incremental iterative random-walking algorithm is proposed to reconstruct the RRG, which iteratively achieves dependable regions by analyzing the dependable value of each annotation result.
- A graph matching method with semantic correlation and spatial position relationship analysis based on SPRG is proposed to predict the labels of non-dependable regions for label completion.

The remainder of this paper is organized as follows. Section 2 reviews the related work. In Section 3, we describe the framework of automatic image region annotation with label refinement. Subsequently, the model of SPRG is introduced in Section 4, and label refinement method based on SPRG is described in Section 5. Experimental results and analysis are presented in Section 6, followed by the conclusions and future work in Section 7.

2. Related work

As image region annotation results based on machine learning are not adequate, researchers have proposed some image label refinement methods and attempt to further improve the accuracy of annotation results [26]. Primarily, three types of methods are used in image label refinement: label refinement with low-level visual feature, label refinement based on ontology-based semantic network, and label refinement by label semantic correlation analysis.

Some researchers analyzed the relations between labels and visual features of an image, such as color and texture, to obtain the relationships between visual descriptors and labels for image region label refinement. Li et al. [11] proposed a deep matrix factorization algorithm that reveals the latent relation between image visual feature descriptors and label concepts embedded in a latent subspace. Tiberio et al. [12] proposed and thoroughly evaluated the application of nearest-neighbor methods for tag refinement. Kuo et al. [13] proposed a framework to augment each image with the relevant semantic (visual and textual) features using graphs among images. A label completion framework was proposed by Hou et al. [14], in which the low rank representation method based on subspace clustering was used to complete the labels. To refine the labels, they used matrix completion with the inductive matrix completion model. Uricchio et al. proposed a label propagation framework based on kernel canonical correlation analysis that builds a latent semantic space where the correlation of visual and textual features is preserved into a semantic embedding [15].

In addition, some researchers proposed refining labels by semantic ontology, such as WordNet [16,17], OWL [18], ORN [19], and ImageNet. Ullah et al. [17] extended a semantic ontology

method to extract semantic related terms of labels from training image datasets and enhanced their correlation by WordNet and ConceptNet. Im et al. [20] proposed a semi-automatic image annotation system based on the semantic network by bridging the semantic links between the labels. Linked data such as DBpedia were exploited to connect the image labels with a property value through capturing the contexts of the image labels.

Limited by the domain of a specific ontology, it is difficult to adapt label refinement methods based on ontology to various situations. Hence, constructing a new semantic network by extracting the relationships in images between labels is proposed and has attracted increasing attention. Zhong et al. [21] presented a relevance feedback algorithm based on multi-view non-negative matrix factorization to improve the annotation performance. Tian et al. [22] built a joint framework with the relation of the “label set” to the image and the internal correlation in the label set, and a heuristic algorithm based on the nearest neighbor was used to annotate image collections in real environments. Pobar et al. [7] proposed an annotation refinement procedure that used the knowledge about the co-occurrence frequency of objects to improve the annotation precision by detecting and abandoning labels that did not fit the context of other whole image labels.

Considering that low-level features and semantic networks are all useful for label refinement, some researches proposed approaches to automatically refine image labels based on fusing label semantics and visual similarity [23–25]. For example, Wang et al. [23] realized tag refinement from the perspective of topic modeling and presented a graphical model regularized latent Dirichlet allocation. Tang et al. [24] propose a tri-clustered tensor completion framework to collaboratively explore these multi-category information to improve the performance of social image tag refinement. Sang et al. [25] proposed a method of ranking based on multi-correlation tensor factorization to model the ternary relations jointly among users, images, and tags.

3. Automatic image region annotation with label refinement

Our proposed label refinement method is based on the automatic image region annotation method proposed by [27]. The framework of automatic image region annotation with proposed refinement method is illustrated in Fig. 2. It primarily includes two parts: image region annotation and label refinement.

We adopt the texture-enhanced JSEG algorithm to segment the image to semantic regions, and subsequently extended the bag-of-words [28–30] model with visual context information to represent the content of the regions. Subsequently, a multi-classifier learned by a maximal figure-of-merit [31–33] algorithm was used to predict the labels of the regions.

Based on the image region annotation results of [27], we propose an SPRG model and label refinement method based on the SPRG for wrong label revision, whose framework is illustrated in Fig. 3.

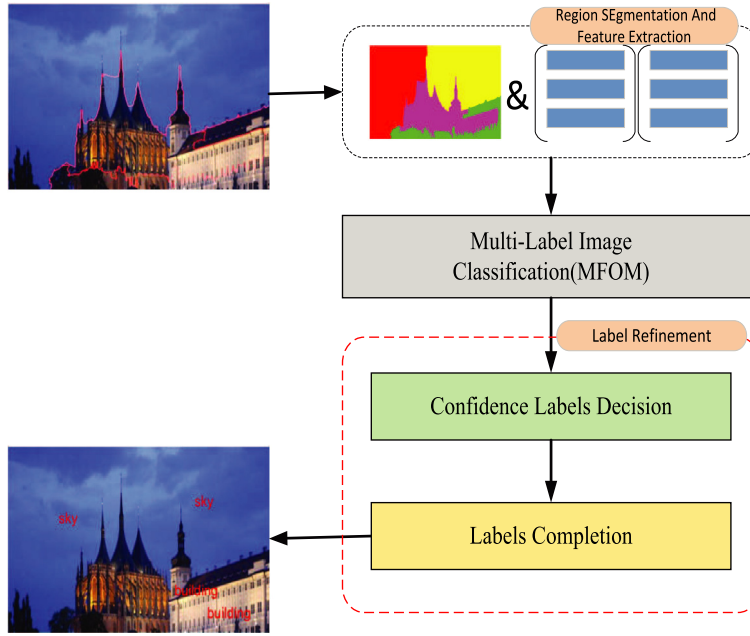


Fig. 2. Framework of automatic image region annotation with label refinement.

In the training part, we constructed an SPRG for all the labels, in which labels are represented by nodes, and the semantic correlations and spatial positional relations between labels are represented by edges and weights. During testing, we completed label refinement through two steps. First, we constructed a RRG for every testing image and adopted our proposed incremental iterative random-walking algorithm to obtain non-dependable label regions. Subsequently, a graph matching method is proposed to predict the labels of non-dependable regions.

4. Spatial position relation graph

Typically, an image can be described by complex semantic information, and different regions apply different semantic concepts. These concepts always implicit all types of latent relations, including semantic correlation and position correlation. Traditional image annotation methods do not consider these complex relations between labels in an image, which produces unsatisfactory results. Hence, we propose a SPRG by analyzing the co-occurrence relation and spatial position relation between different semantic labels of the image region to improve the image annotation results in this study.

In our SPRG model, $\Phi = \{l_1, l_2, \dots, l_N\}$ is denoted as the label set of training images, and N is the size of the label set. $\mathbf{I}^{\text{train}} = \{I_1, I_2, \dots, I_M\}$ represents the training image set, and M is the size of the training image set. Specifically, $I_k \rightarrow L_k, L_k = \{l_1, l_2, \dots, l_S\}$, L_k denotes the set of labels in image I_k , and S is the size of L_k .

Definition 1. Image region semantic relation can be modeled by the following SPRG:

$$G = \langle V, E, W \rangle \quad (1)$$

where V is the set of vertices in the graph G , which denotes a collection of semantic labels of the training images and $V = \{v_1, v_2, \dots, v_N\}$. E is the set of edges in the graph G , which denotes a collection of spatial position relations between labels in the training images, $V \times V \rightarrow E$. W is the set of values of the edges attributes, which represents the spatial position relation between labels. The attributes matrix W is defined as follows:

$$W = \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix} \quad (2)$$

where W_1 denotes the co-occurrence relation matrix, W_2 denotes the topological relation matrix, and W_3, W_4 denotes the directional relation matrix.

In the SPRG, we set vertex v_i as label l_i , and e_{ij} as the edge between vertices v_i and v_j , $e_{ij} \in E$. Edge e_{ij} and a set of attributes $w(e_{ij})$ denote the co-occurrence relation and spatial position between labels l_i and l_j , $w(e_{ij}) \in W$. The attributes $w(e_{ij}) = \{w_{1,ij}, w_{2,ij}, w_{3,ij}, w_{4,ij}\}$ of the edge e_{ij} are composed of two parts: global spatial co-occurrence measurement and relative position relation measurement, where $w_{1,ij}$ denotes the co-occurrence relation, $w_{1,ij} \in W_1$; $w_{2,ij}$ denotes the topological relation, $w_{2,ij} \in W_2$; $w_{3,ij}$ and $w_{4,ij}$ denote the direction relation, $w_{3,ij} \in W_3$, $w_{4,ij} \in W_4$.

The global spatial co-occurrence relation is computed by the proportion of co-occurrence semantic label pairs. In image I_k , we define $u_{\text{coo}}^k(i, j)$ as whether the label l_i and label l_j appear simultaneously, where

$$u_{\text{coo}}^k(i, j) = \begin{cases} 1, & \text{if } l_i \text{ and } l_j \text{ are co-occurrence} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The spatial co-occurrence relation $w_{1,ij}$ between label l_i and l_j is defined as follows:

$$w_{1,ij} = \frac{1}{M} \sum_{k=1}^M \mu_{\text{bor}}^k(i, j) \quad (4)$$

The relative position relations include topological relations and direction relations of image region labels; they can reflect the layout of the image region and contain implicitly deeper semantic relationships of the labels.

The topological relations include *disjointing* and *bordering* [34, 35], which is illustrated in Fig. 4(a).

We define matrix $M_{\text{bor}}^k \in \mathbf{R}^{S \times S}$ to represent the adjacent relation of image I_k , and each element $\mu_{\text{bor}}^k(i, j)$ is calculated by the following:

$$\mu_{\text{bor}}^k(i, j) = \begin{cases} 1, & \text{if } r_i \text{ is bordering to } r_j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where r_i and r_j represent the corresponding regions of labels l_i and l_j .

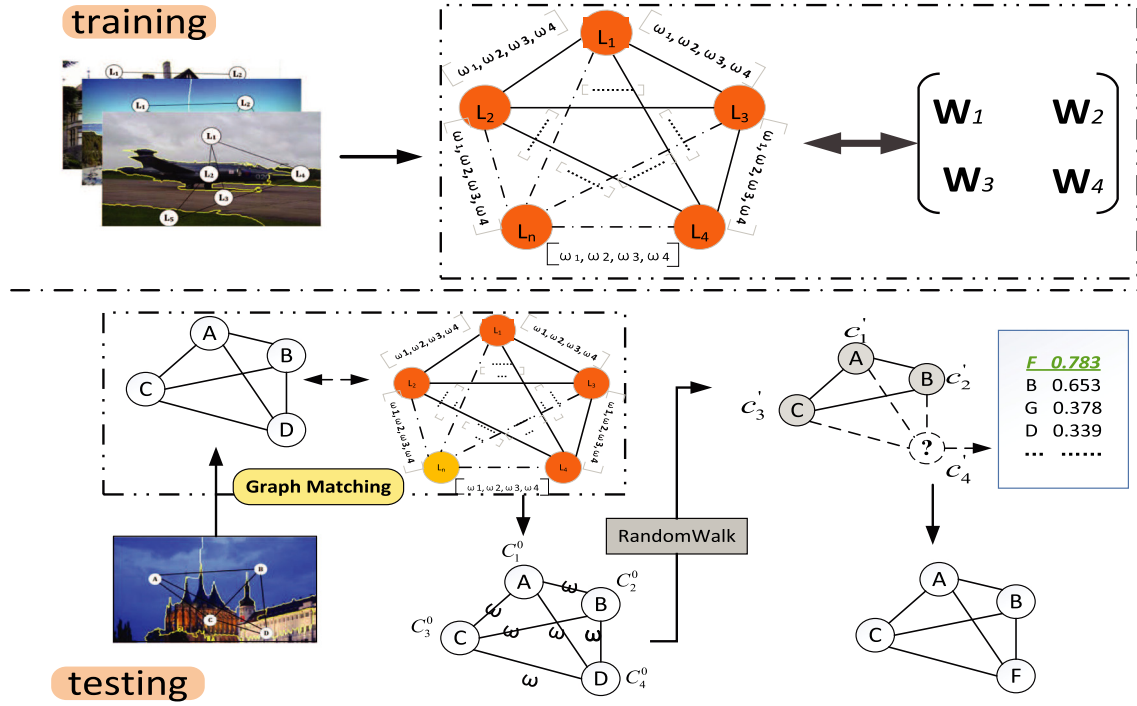


Fig. 3. Framework of our proposed label refinement method based on SPRG.

The value of topological relation $w_{2,ij}$ of the spatial position relation between region labels l_i and l_j in the SPRG is calculated by the following:

$$w_{2,ij} = \frac{1}{M} \sum_{k=1}^M \mu_{bor}^k(i, j) \quad (6)$$

The direction relations include *above*, *below*, *left*, and *right*. Considering the characteristics of natural scene images, *left* and *right* typically do not influence the image classification. Hence, the directional relations include *above*, *below*, and *beside*, which are shown in Fig. 4(b).

Subsequently, the directional relation of regions r_i and r_j in image I_k can be defined as $\mu_{abo}^k(r_i, r_j)$, $\mu_{bel}^k(r_i, r_j)$ and $\mu_{bes}^k(r_i, r_j)$, which can be computed by the following:

$$\mu_{abo}^k(r_i, r_j) = \begin{cases} 1, & \text{if } \frac{\pi}{6} < \theta_{ij} < \frac{5\pi}{6} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$\mu_{bel}^k(r_i, r_j) = \begin{cases} 1, & \text{if } -\frac{5\pi}{6} < \theta_{ij} < -\frac{\pi}{6} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$\mu_{bes}^k(r_i, r_j) = \begin{cases} 1, & \text{if } -\frac{\pi}{6} \leq \theta_{ij} \leq \frac{\pi}{6} \\ & \text{or } \frac{5\pi}{6} \leq \theta_{ij} \leq \pi \\ & \text{or } -\pi \leq \theta_{ij} \leq -\frac{5\pi}{6} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where θ_{ij} denotes the angle between the horizontal axis and the line joining the centers of two object regions in image I_k . We use matrix $M_{abo}^k \in \mathbf{R}^{S \times S}$ to represent the relation that the label l_i is above l_j in image I_k , and the $\mu_{abo}^k(i, j)$ ($\mu_{abo}^k(i, j) \in M_{abo}^k$) is obtained by the following:

$$\mu_{abo}^k(i, j) = \mu_{abo}^k(r_i, r_j) \quad (10)$$

where r_i and r_j represent the regions of label l_i and l_j , respectively, in image I_k .

Similarly, for image I_k , $M_{bel}^k \in \mathbf{R}^{S \times S}$ represents the relation that the label l_i is located below l_j . $M_{bes}^k \in \mathbf{R}^{S \times S}$ means the relation that the label l_i is next to l_j . Also, $\mu_{bel}^k(i, j) = \mu_{bel}^k(r_i, r_j)$, $\mu_{bes}^k(i, j) = \mu_{bes}^k(r_i, r_j)$.

While M_{abo}^k is used to represent that the label l_i is above l_j for one image, the attribute $w_{3,ij}$ of position relation between region label l_i and region l_j in the SPRG is determined by the following:

$$w_{3,ij} = \frac{1}{M} \sum_{k=1}^M \mu_{abo}^k(i, j) \quad (11)$$

M_{abo}^k describes that the label l_i is above l_j , which is not equal to M_{bel}^k . Therefore, it describes that the label l_j is above l_i . $w_{3,ji} \neq w_{3,ij}$, and

$$w_{3,ji} = \frac{1}{M} \sum_{k=1}^M \mu_{bel}^k(i, j) \quad (12)$$

While M_{bes}^k is used to represent that the label l_i is beside l_j for one image, the attribute $w_{4,ij}$ of the position relation between labels l_i and l_j in the SPRG is determined by the following:

$$w_{4,ij} = \frac{1}{M} \sum_{k=1}^M \mu_{bes}^k(i, j) \quad (13)$$

5. Label refinement based on SPRG

Definitions used in this section:

Dependable label: the label that fit the context of the scene in an image.

Dependable value: the quantitative value of the label fitting the context.

Non-dependable label: the label that cannot fit the context of the scene in an image.

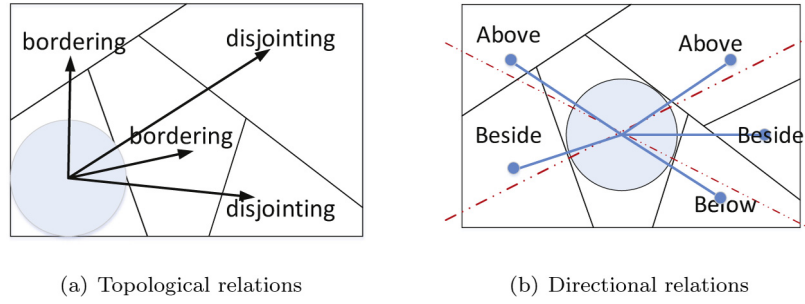


Fig. 4. Relative position relations between regions in an image.

Non-dependable label region: the region corresponding to the non-dependable label.

We herein propose a novel label refinement method based on the SPRG, which includes random-walking-based dependable label decision and graph-matching-based label completion. During the procedure of dependable label decision, we propose an incremental iterative random-walking algorithm to enhance or weaken the dependable value of each region label, thus allowing for the dependable label to achieve a higher dependable value and the non-dependable label to achieve a lower dependable value. Hence, we can effectively obtain the potential error labels. During label completion, we predict the highest probability label for non-dependable regions by analyzing the position and semantic correlation with dependable labels through the SPRG.

5.1. Incremental iterative random-walking-based dependable label decision

The random-walking algorithm is applied widely in machine learning and information retrieval [36–39], and is crucial in image retrieval and image annotation. We herein propose an incremental iterative random-walking algorithm to improve the reliability of the region label by reconstructing an RRG.

We defined $\Gamma_{test} = \{I_1, I_2, \dots, I_A\}$ as the testing image set with the original labels obtained by an automatic image region annotation algorithm in [27], where A denotes the quantity of images in the testing set. Specifically, $I_\tau \rightarrow L_\tau, L_\tau = \{l_1^\tau, l_2^\tau, \dots, l_Y^\tau\}$, L_τ denotes the set of labels in image I_τ , and Y is the size of L_τ .

Definition 2. RRG G_τ of image I_τ is defined as follows:

$$G_\tau = \{V_\tau, E_\tau, w_\tau\} \quad (14)$$

where V_τ is the set of vertices in the graph G_τ , which denotes a collection of labels of image I_τ . e_{ij}^τ ($e_{ij}^\tau \in E_\tau$) denotes the correlation between label l_i^τ and l_j^τ , $V_\tau \times V_\tau \rightarrow E_\tau$. w_τ is a set of edge attribute values, and its initial value is set as empty. G_τ is a complete graph, which is a subgraph of G , namely $G_\tau \subset G$.

Subsequently, the attribute $w_{1,ij}$ of e_{ij} in the G describes the semantic co-occurrence relation of the region label is assigned to the corresponding edge e_{ij}^τ of G_τ through graph matching.

$$w_{ij}^\tau = w_{1,ij} \quad (15)$$

To acquire dependable labels, an incremental iterative random-walking algorithm is proposed and applied on G_τ . Next, we will introduce the incremental iterative random-walking algorithm in detail.

We define the dependable label set as $\Psi = \{l_1^\tau, l_2^\tau, \dots, l_p^\tau\}$, in which P is the size of the dependable label set. Further, the dependable value set that corresponds to the dependable label set is denoted as $C = \{c_1, c_2, \dots, c_p\}$. The initial dependable values of the labels in L_τ were set using two methods. The first one considers

the relative position relation (RPR) between the label pairs, which is introduced in scheme 1. The other one considers the absolute position relation (APR) between the label pairs with the image layout, which is introduced in scheme 2.

Scheme 1. Initial value setting by RPR

RPR reflects the topological relations and directional relations between two labels in the SPRG. We can use the RPR to prove that a label fits the image context (or not) in the testing image. Hence, considering the RPR, the initial dependable value c_i^0 of label l_i^τ is computed as follows:

$$c_i^0 = \frac{1}{f_i} \sum_{j=1}^{f_i} \eta(i, j) \quad (16)$$

where

$$\eta(i, j) = \{\beta \times w_{2,ij} + (1 - \beta) \times [w_{4,ij} \times \mu_{bes}^\tau(i, j) + w_{3,ij} \times \mu_{abo}^\tau(i, j) + w_{3,ji} \times \mu_{bel}^\tau(i, j)]\} \quad (17)$$

where β is a parameter that determines the effects of the adjacent relation and the direction relation, and f_i is the in-degree of the corresponding vertex.

Scheme 2. Initial value setting by absolute position relation

APR typically represents the labels' layout of an image, which implies that some labels appear in some specific regions with high frequency. For example, "sky" always appears at the top of an image and "car" often appears at the bottom of an image. Hence, we can use the APR to verify whether a label fits the image layout.

We define the APR of labels by the region layout of an image. We divide the image into D grid regions (for example $D = 25$) on average and obtain the probabilities of the labels falling in each grid, which is shown in Fig. 5. $\mu_{fal}^k(i, j)$ denotes that label l_i falls in the grid g_j (or not) in image I_k , and it is computed by the following:

$$\mu_{fal}^k(i, j) = \begin{cases} 1, & \text{Count}(\text{pixel}(i, j)) \geq \varepsilon \times \frac{B}{D} \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where $\text{Count}(\text{pixel}(i, j))$ denotes the number of pixels that are included in grid g_j , and its label is defined as l_i . B denotes the total number of pixels in image I_k . ε is a parameter.

We use $M_{APR} \in \mathbf{R}^{N \times D}$ to represent the APR, whose element $w_{APR}(i, j)$ denotes the relation between label l_i and grid g_j , which is calculated as follows:

$$w_{APR}(i, j) = \frac{1}{M} \sum_{k=1}^M \mu_{fal}^k(i, j) \quad (19)$$

Considering the APR, the initial dependable value c_i^0 of label l_i^τ is computed by the following equation.

$$c_i^0 = \frac{\sum_{j=1}^D w_{APR}(i, j) \times \mu_{fal}^\tau(i, j)}{\sum_{j=1}^D \mu_{fal}^\tau(i, j)} \quad (20)$$



Fig. 5. Grid segmentation of image.

We set the label whose initial dependable value is the maximal as the initial dependable label, as shown in Fig. 6(a).

Algorithm 1 Incremental iterative random-walking algorithm

Input: SPRG: $G = \langle V, E, W \rangle$; $L_\tau = \{l_1^\tau, l_2^\tau, \dots, l_Y^\tau\}$; Initial dependable values set $C = \{c_1^0, c_2^0, \dots, c_Y^0\}$. Initial dependable label l_1 with dependable value c_1^0 .
Output: Final dependable values set C_{final}
1: $i = 1$; initialize a Graph G_τ with only a vertex l_1 and $\Psi^1 = \{l_1\}$; $C = \{c_1^0\}$
2: **while** $i < Y$ **do**
3: $i = i + 1$;
4: Choose label l_i with initial dependable value c_i^0 by Algorithm 2.
5: Put l_i into G_τ and update the weights of edges using Eq. (15).
6: $\Psi^i = \Psi^{i-1} \cup \{l_i\}$; $C = C \cup \{c_i^0\}$
7: Use the random-walking algorithm to update the dependable values by Algorithm 3, $C = C'$.
8: **end while**
9: $C_{final} = C$

Algorithm 2 Maximal dependable value label choosing algorithm

Input: SPRG: $G = \langle V, E, W \rangle$; $L_\tau = \{l_1^\tau, l_2^\tau, \dots, l_Y^\tau\}$; Initial dependable values set $C = \{c_1^0, c_2^0, \dots, c_Y^0\}$. Initial dependable label l_1 with dependable value c_1^0 . Ψ^{i-1}
Output: A label l_i with initial dependable value c_i^0 .
1: Initialize $L = L_\tau - \Psi^{i-1}$ and $Z = Y - i + 1$ as the size of L .
2: **for** $j = 1 : Z$ **do**
3: **for** $k=1:i-1$ **do**
4: Get the RPR value $\eta(k, j)$ between l_j and dependable label l_k using Eq. (17).
5: **end for**
6: $\eta(j) = \sum_{k=1}^{i-1} \eta(k, j)$
7: **end for**
8: Assigned label l_j with the maximum RPR value $\eta(j)$ to l_i and c_i^0 .

After obtaining the initial dependable label, we add other labels to the dependable set successively by the incremental iterative random-walking algorithm, which is illustrated in Algorithm 1. After achieving a new dependable label l_i , we add it to the primary dependable label set $\Psi^{i-1} = \{l_1, l_2, \dots, l_{i-1}\}$, and update the dependable label set as $\Psi^i = \{l_1, l_2, \dots, l_i\}$. The label dependable value corresponding to the dependable label is also updated by the random-walking algorithm. Subsequently, we can obtain the new set of label dependable value $\psi^i = \{c_1^{i-2}, c_2^{i-2}, \dots, c_i^1\}$, as shown in Fig. 6(b).

Given an RRG with P vertices, we subsequently use $c_h(i)$ to denote the dependable value of vertex i at the h – th iteration. Hence, the dependable values of all vertices in the graph at iteration h

form a column vector $\mathbf{c}_h \equiv [c_h(i)]_{P \times 1}$. We define $\mathbf{P}_t \in \mathbf{R}^{P \times P}$ as a transfer matrix, and each element $p_t(ij)$ of the transfer matrix represents the transfer probability from vertex i to vertex j ; the transfer probability is computed as follows:

$$p_t(ij) = \frac{w_{ij}^\tau}{\sum_h w_{ih}^\tau} \quad (21)$$

where e_{ij} represents the relation of co-occurrence between vertex i and vertex j (as shown in Eq. (15)).

The random-walking process is formulated as follows:

$$c_h(j) = \delta \sum_i c_{h-1}(i) p_t(ij) + (1 - \delta) c_j \quad (22)$$

where c_j represents the primary dependable value of node j ; δ is a transferring weight parameter ranging from 0 to 1. We can improve the dependable value of the correct annotation results, and simultaneously decrease the dependable value of the noise annotation results.

Algorithm 3 Random-walking algorithm on RRG

Input: RRG of $I_\tau: G_{fi} = \{V_\tau, E_\tau, w_\tau\}$; Dependable values set $C = \{c_1, c_2, \dots, c_P\}$. Transfer matrix \mathbf{P}_t
Output: C'
1: Initialize $h = 1$;
2: **for** $j=1:P$ **do**
3: $c_1(j) = c_j$
4: **end for**
5: **while** $\max\{|c_h(j) - c_{h-1}(j)|, j = 1, 2, \dots, P\} > \lambda ||h < 2$ **do**
6: $h = h + 1$;
7: **for** $j = 1 : P$ **do**
8: Update $c_1(j)$ using Eq. (22)
9: **end for**
10: **end while**
11: $C' = \{c_h(1), c_h(2), \dots, c_h(P)\}$;

We iteratively add the label that matches the dependable label set the most by the random-walking algorithm and update the dependable value until all labels in the RRG are added into the dependable set, as shown in Fig. 6(c) to (d). Subsequently, the labels with dependable values greater than the dependable threshold value σ are obtained, and the labels with dependable values less than the dependable threshold are discarded. Therefore, we can obtain the final label dependable set Ψ , the corresponding labels dependable value set C , and the non-dependable label set $\Omega = L_\tau - \Psi$, whose size is $q = N - P$.

5.2. Graph-matching-based label completion

We herein propose a novel label completion algorithm by graph matching that predicts the label for the non-dependable region label set Ω obtained from Section 5.1. The basis of the prediction is the relative spatial position relation and semantic relation between each non-dependable label l_i and every dependable label l_j .

The candidate label set of the prediction are set as $\Upsilon = \Phi = \{l_1, l_2, \dots, l_N\}$.

First, we obtain the relative spatial position relation between each non-dependable label l_i and each dependable label l_j , including the adjacent relation $\mu_{bor}^\tau(i, j)$ and direction position relation $\mu_{bes}^\tau(i, j)$, $\mu_{abo}^\tau(i, j)$ and $\mu_{bel}^\tau(i, j)$ by the algorithm proposed in Section 4.

Subsequently, we use candidate label l_k to replace the non-dependable label l_i and obtain a new graph $G_k = \{V_k, E_k\}$, where $V_k = \Psi \cup l_k$. For the candidate label set Φ , we can obtain $G' = \{G_1, G_2, \dots, G_N\}$. Subsequently, we match G' with the SPRG to

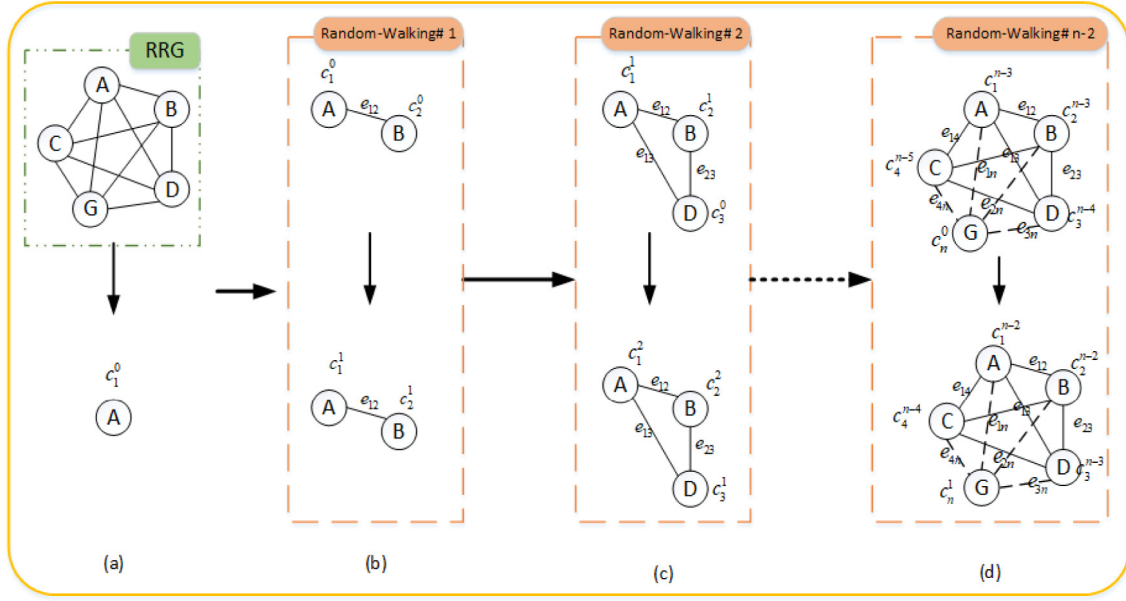


Fig. 6. Dependable label achieved by incremental iterative random-walking algorithm.

obtain the matching matrix $\mathbf{X} \in \mathbf{R}^{N \times P}$. $\chi(k, j)$ is defined as the matching degree between candidate label l_k and dependable label l_j , which is computed as follows:

$$\chi(k, j) = \psi(j) \eta(k, j) \quad (23)$$

where $\psi(j)$ denotes the dependable value of dependable label l_j .

Moreover, the integration matching degree of the candidate label l_k to the dependable label l_i is computed by the following:

$$p_{match}(k) = \sum_{j=1}^P \chi(k, j) \quad (24)$$

We chose the label with the highest integration matching degree as the result of label completion l_{final} , and used it to replace the non-dependable label l_i . The l_{final} is formulated as follows:

$$l_{final} = l_{\arg \max_k p_{match}(k)} \quad (25)$$

6. Experiments

Experiments were conducted on the subset of standard dataset Corel5K and web dataset Deriantart3K for verifying the performance of our proposed label refinement model. The experimental datasets and results are discussed as follows.

6.1. Dataset and evaluation metric

Corel5K is widely used in image annotation and retrieval, which contains 50 themes with 100 images for each theme, and each image includes an average of 3.14 labels. Considering the variety of the images, we chose 1500 images from Corel5K with three categories: *Building*, *Vehicle*, and *Seaside*. The web dataset Deriantart3K obtains 3000 images from the image-sharing site Deriantart. Table 1 illustrates the information of all the image datasets.

To evaluate the performance of the proposed algorithms more accurately, we used five different evaluation metrics: revised label precision, incorrect labels recall, mean average precision, mean average recall, and mF1, which are defined below.

Table 1

Image datasets.

Dataset	Training set	Testing set	Labels
Corel	1000	500	"sky", "plant", "sea", "sand", "rock", "road", "house", "building", "sun", "car", "plane", "boat", "land"
Deriantart3K 2000		1000	"sky", "plant", "water", "mountain", "grass", "sand", "rocks", "ground", "road", "castle", "chapel", "house", "building", "snow", "car", "train", "track", "plane", "runway", "smoke", "cloud"

Revised labels precision (P_{RL}) is a specialized evaluation metric that measures the efficiency of the proposed method on incorrect label refining, which is defined as follows.

$$P_{RL} = \frac{1}{\Lambda} \sum_{\tau=1}^{\Lambda} \frac{|T_{\tau} \cap M_{\tau} \cap \Omega_{\tau}|}{|\Omega_{\tau}|} \quad (26)$$

where Ω_{τ} denotes the non-dependable label set in image I_{τ} ; Λ is the size of the testing image set; T_{τ} is the label set of the ground truth; M_{τ} is the label set of the annotation result of image I_{τ} .

Incorrect label recall (R_{IL}) measures the efficiency of the proposed incremental iterative random-walking algorithm in detecting the confidence labels, which is defined as follows:

$$R_{IL} = \frac{1}{\Lambda} \sum_{\tau=1}^{\Lambda} \frac{|A_{\tau} \cap \bar{T}_{\tau} \cap \Omega_{\tau}|}{|A_{\tau} \cap \bar{T}_{\tau}|} \quad (27)$$

where A_{τ} is the label set of the original image region annotation results of image I_{τ} .

Mean average precision (mPr) is a typical evaluation metric that is used widely in image annotation. mPr is the mean of the average precision of all testing images, which is defined as follows:

$$mPr = \frac{1}{\Lambda} \sum_{\tau=1}^{\Lambda} Pr^{\tau} \quad (28)$$

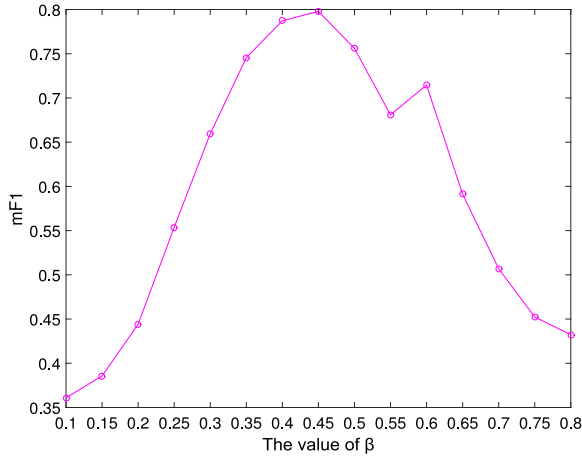
Fig. 7. Experiments with the effects of β .

Table 2
Experimental results on two datasets with P_{RL} and R_{IL} .

DataSet	P_{RL}	R_{IL}
Corel	63.28	92.28
Deriantart3K	81.37	87.82

For a given image I_τ , the average precision (Pr^τ) can be calculated as follows:

$$Pr^\tau = \frac{|T_\tau \cap M_\tau|}{|T_\tau|} \quad (29)$$

Similar to mPr , the mean average recall (mRe) is the mean of the average recall of all testing images, which is defined as follows:

$$mRe = \frac{1}{A} \sum_{\tau=1}^A Re^\tau \quad (30)$$

where

$$Re^\tau = \frac{|T_\tau \cap M_\tau|}{|M_\tau|} \quad (31)$$

mF_1 is the harmonic mean of mPr and mRe , and can be interpreted as a weighted average of precision and recall. For image I_τ , the F_1^τ is defined as follows:

$$F_1^\tau = \frac{2 \times Pr^\tau \times Re^\tau}{Pr^\tau + Re^\tau} \quad (32)$$

mF_1 is defined as follows:

$$mF_1 = \frac{1}{A} \sum_{\tau=1}^A F_1^\tau \quad (33)$$

6.2. Experiments with the influences of β

The effect of the β value on the average harmonic ratio: From the data in the table, we can conclude that the average harmonic value in the set of empirical values is the maximum at $\beta = 0.45$; therefore, we set the β value to 0.45 in the algorithm (see Fig. 7).

6.3. Experiments on various image datasets

For validating the effectiveness of our proposed method, we conducted experiments on two image datasets, and the experimental results are shown in Table 2.

The experimental results indicate that our proposed label refinement method obtains an average of 90.05% R_{IL} , thus implying

Table 3

Experimental results on baseline of [27] and our proposed label refinement method.

DataSet	SubDataset	Method	mPr	mRe	mF_1
Corel5k	Building	Baseline	79.82	81.66	79.07
		Our	86.11	83.49	83.30
	Seaside	Baseline	73.64	72.46	71.62
		Our	77.26	70.98	73.58
	Vehicle	Baseline	87.00	90.58	87.14
		Our	89.52	89.75	89.20
Deriantart3K	Building	Baseline	43.45	58.86	48.97
		Our	67.02	45.06	52.86
	Scenery	Baseline	42.16	68.77	53.95
		Our	61.19	70.62	63.62
	Vehicle	Baseline	48.07	63.40	53.61
		Our	57.89	65.67	59.97

Table 4

Experimental results with RPR or APR.

Dataset	Methods	mPr	mRe	mF_1
Corel	Baseline	78.73	79.76	77.70
	LR with RPR	82.64	81.15	79.79
	LR with APR	83.25	79.74	80.59
Deriantart3K	Baseline	44.59	63.68	52.18
	LR with RPR	60.01	68.36	63.68
	LR with APR	67.03	64.45	65.34

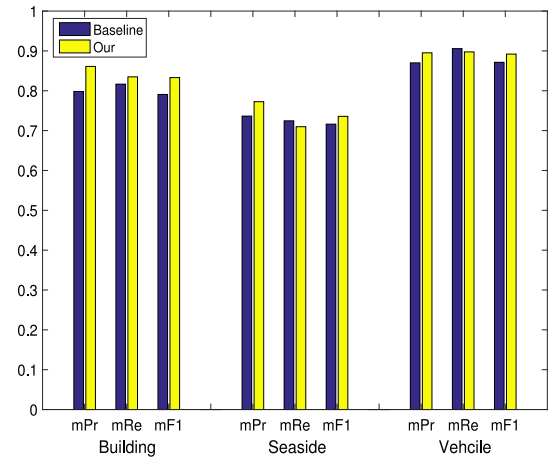


Fig. 8. Experimental results on Corel5k with baseline of [27] and our proposed label refinement method.

that our incremental iterative random-walking algorithm is highly effective in obtaining most of the wrong annotation labels. Regarding new label prediction for the non-dependable label regions, our proposed label completion method can acquire a 72.33% P_{RL} on average, thus implying that most of the wrong labels are corrected by our graph matching algorithm.


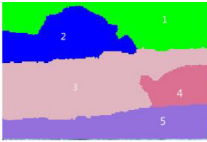

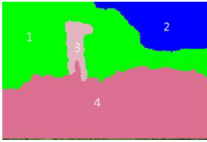

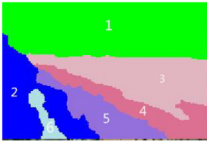

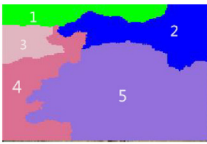



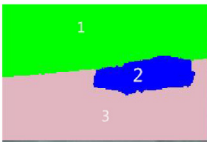
6.4. Contrast experiments

To prove that our proposed label refinement method can improve the accuracy of image region annotation effectively, we completed a series of experiments on three datasets of Corel5k, *Building*, *Seaside*, and *Vehicle*, and compared their performance with the baseline [27]. Corresponding, we performed a similar experiment on the three subsets of Deriantart3K, *Building*, *Scenery*, and *Vehicle*. The experimental results are shown in Table 3, Figs. 8 and 9.

The experimental results indicate that our proposed label refinement method improves the baseline of [27]. mPr and mF_1 improve 6.29% and 4.23%, respectively, on image set *Building*. With

Table 5

Examples of annotation results on various image datasets with different methods.

Image	Region	Annotation		
		GT	Baseline	Ours
		1: sky	sky	sky
		2: plant	plant	plant
		3: building	building	building
		4: plant	plant	plant
		5: water	sky	water
		1: building	building	building
		2: sky	sky	sky
		3: building	plant	building
		4: plant	plant	plant
		1: sky	sea	sky
		2: plant	building	building
		3: sea	sea	sea
		4: sand	plant	sand
		5: sand	sand	sand
		6: plant	plant	plant
		1: sky	sea	sky
		2: building	building	building
		3: sea	sea	sand
		4: sand	sand	sand
		5: sand	sand	sand
		1: sky	sea	sky
		2: plant	plant	plant
		3: land	land	land
		1: plant	plant	plant
		2: car	car	car
		3: land	sky	land

the effect of some rare labels, mRe only improves 1.83% on dataset *Building* by using our method. For the dataset *Seaside*, complex scenes and many different labels exist, thereby complicating the annotation and refinement. Nevertheless, we still achieved 77.26% mPr , and a 3.62% increase compared with previous work [27]. The experimental results on the dataset *Vehicle* achieve 89.52% mPr and 89.20% mF_1 , which correspond to improvements of 2.52% and 2.06% compared with previous work [27]. The experiments above demonstrate that our proposed method is highly effective on various image datasets, and can improve the image region annotation results by our proposed label refinement method.

6.5. Experiments with different initial confidence value choices

For validating the effectiveness of our proposed, i.e., two types of initial dependable value schemes, we conducted experiments on three datasets and the experimental results are shown in Table 4.

From our experimental results, we found that our proposed label refinement methods with different initial dependable value choices (RPR or APR) achieve various performances on the three image datasets. In most cases, the APR model achieves obviously better performance than the RPR model, thus verifying that using the APR to choose the initial dependable value is more effective on mPr . Particularly, based on the experiments on the Deriantart3K dataset, the APR model achieves a 67.03% mPr , which improves by 7.02% compared with the RPR model. However, based on the experiments on two datasets, the RPR model indicates a 2.66% increase in mRe compared with the APR model, implying that the RPR model can improve the image region annotation recall rate effectively; further, it is probably more suitable for the image sets with simple scenes and few labels. Meanwhile, the APR model improves the image region annotation accuracy rate effectively,

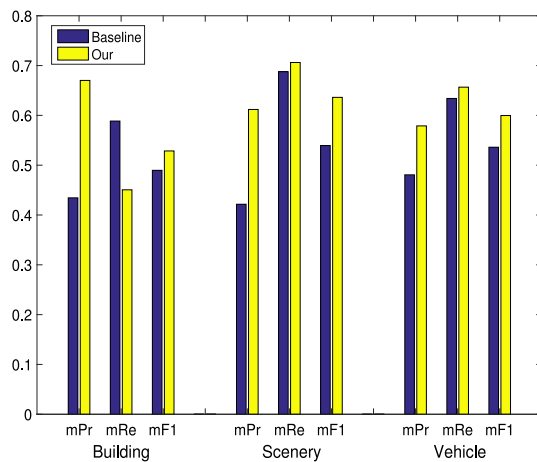


Fig. 9. Experimental results on Deriantart3K with baseline of [27] and our proposed label refinement method.

and demonstrates superiority on most image sets with complex scenes and many labels.

To illustrate the results visually, we present some images and their annotation results in Table 5. We highlight the cases where our algorithm is better than the baseline in red. We also highlight the unsuccessful cases with bold font in the table. We found that our method achieves more precise annotation results than the baseline algorithm [27], and can effectively distinguish some feature similar regions by region semantic analysis and spatial position analysis, such as “sky” and “sea”, “sky” and “water”, and “plant” and “building”.

7. Conclusion and future work

We herein propose a new framework for the label refinement of automatic image region annotation, in which an SPRG with co-occurrence relation and spatial position relation was proposed to reveal the semantic correlation among image region labels. Subsequently, an incremental iterative random-walking algorithm and graph matching method were proposed for dependable label decision and non-dependable region label prediction. Experiments on Corel5k illustrated that our proposed image region label refinement method could improve the accuracy of the image region annotation. For future work, we will add visual features relevance analysis to our framework for the further research on image label refinement.

Acknowledgments

This research has been supported by the National Nature Science Foundation of China (Grant 61402174 and Grant 61672227).

References

- [1] D.S. Zhang, M.M. Islam, G. Lu, A review on automatic image annotation techniques, *Pattern Recognat.* 45 (2012) 346–362.
- [2] L. Feng, B. Bhanu, Semantic concept co-occurrence patterns for image annotation and retrieval, *IEEE PAMI* 38 (4) (2016).
- [3] J. Zhang, S. Feng, D. Li, et al., Image retrieval using the extended salient region, *Inf. Sci.* 399 (2017) 154–182.
- [4] H. Zhang, A. Berg, M. Maire, et al., SVM-KNN: Discriminative nearest neighbor classification for visual category recognition, in: *Proc. CVPR*, Vol. 2 (2006) 2126–2136.
- [5] C.B. Yang, M. Dong, J. Hua, Region-based image annotation using asymmetrical Support Vector Machine-based multipleinstance learning, in: *Proc. CVPR*, 2006.

- [6] H. Ma, J. Zhu, M.R.T. Lyu, et al., Bridging the semantic gap between image contents and tags, *IEEE TMM* 12 (5) (2010) 462–473.
- [7] M. Pobar, M. Ivasis-Kos, Automatic image annotation refinement, in: *Proc. MIPRO*, 2016.
- [8] L. Chen, D. Xu, I.W. Tsang, et al., Tag-based image retrieval improved by augmented features and group-based refinement, *IEEE TMM* 14 (4) (2012) 1057–1067.
- [9] C. Wang, F. Jing, L. Zhang, et al., Image annotation refinement using random walk with restarts, in: *Proc. ACM Multimedia*, 2006.
- [10] H. Xu, J. Wang, X.S. Hua, et al., Tag refinement by regularized LDA, in: *Proc. ACM Multimedia*, 2009.
- [11] Z. Li, J. Tang, Weakly-supervised deep matrix factorization for social image understanding, *IEEE TIP* 26 (1) (2016) 276–288.
- [12] T. Uricchio, L. Ballan, M. Bertini, et al., An evaluation of nearest-neighbor methods for tag refinement, in: *Proc. IEEE ICME* (2013) 1–6.
- [13] Y.H. Kuo, W. Cheng, H. Lin, et al., Unsupervised semantic feature discovery for image object retrieval and tag refinement, *IEEE TMM* 14 (4) (2012) 1079–1090.
- [14] Y. Hou, Z. Lin, Image tag completion and refinement by subspace clustering and matrix completion, in: *Proc. CVPR*, 2015.
- [15] T. Uricchio, L. Ballan, L. Seidenari, et al., Automatic image annotation via label transfer in the semantic space, *Pattern Recognit.* 71 (2017) 144–157.
- [16] G. Mesnil, A. Bordes, J. Weston, et al., Learning semantic representations of objects and their parts, *Mach. Learn.* 94 (2) (2014) 281–301.
- [17] R. Ullah, A.B. Said, M.Q. Saleem, et al., Semantic ontology for annotated images, in: *International Conference on Computer and Information Sciences*, 2016, pp. 408–413.
- [18] Web Ontology Working Group, OWL Web Ontology Language Overview, W3C Candidate Recommendation, World Wide Web Consortium, 18 2003.
- [19] N. Chen, Q.Y. Zhou, V.K. Prasanna, Understanding web images by object relation network, *Creating and Using Links between Data Objects*, 2012.
- [20] D.H. Im, G.D. Park, Linked tag: image annotation using semantic relationships between image tags, *Multimed. Tools Appl.* 74 (2015) 2273–2287.
- [21] F. Zhong, L. Ma, Image annotation using multi-view non-negative matrix factorization and semantic co-occurrence, in: *IEEE Region 10 Conference (TENCON)- Proceedings of the International Conference*, 2016.
- [22] F. Tian, X. Shen, Learning label set relevance for search based image annotation, in: *International Conference on Virtual Reality and Visualization*, 2014.
- [23] J. Wang, J. Zou, H. Xu, et al., Image tag refinement by regularized latent dirichlet allocation, *CVIU* 12 (4) (2014) 61–70.
- [24] J. Tang, X. Shu, G. Qi, et al., Tri-Clustered tensor completion for Social-Aware image tag refinement, *IEEE PAMI*, 2016.
- [25] J. Sang, C. Xu, J. Liu, User-Aware image tag refinement via ternary semantic analysis, *IEEE TMM* 14 (3) (2012) 883–895.
- [26] X. Li, T. Uricchio, L. Ballan, et al., Socializing the semantic gap: a comparative survey on image tag assignment, refinement, and retrieval, *ACM Comput. Surv.* 49 (1) (2016).
- [27] J. Zhang, Y. Gao, S. Feng, et al., Automatic image region annotation through segmentation based visual semantic analysis and discriminative classification, in: *Proc. ICASSP* (2016) pp. 1956–1960.
- [28] Z. Lu, L. Wang, Learning descriptive visual representation for image classification and annotation, *Pattern Recognit.* 48 (2) (2015) 498–508.
- [29] J. Zhang, Y. Zhao, D. Li, et al., A novel image annotation model based on content representation with multi-layer segmentation, *Neural Comput. Appl.* 26 (6) (2015) 1407–1422.
- [30] J. Han, M. Xu, X. Li, et al., Interactive object-based image retrieval and annotation on ipad, *Multimed. Tools Appl.* 72 (3) (2014) 2275–2297.
- [31] S. Gao, W. Wu, C.H. Lee, et al., A maximal figure-of-merit learning approach to text categorization, in: *Proc. ACM SIGIR*, July 28 – August 1, 2003.
- [32] J. Zhang, D. Li, Y. Zhao, et al., Representation of image content based on Rol-BoW, *J. Vis. Commun. Image Represent.* 36 (2015) 37–49.
- [33] K. Li, Z. Huang, Y. Cheng, et al., A maximal figure-of-merit learning approach to maximizing mean average precision with deep neural network based classifiers, in: *Proc. ICASSP* (2014) pp. 4503–4507.
- [34] S. Aksoy, C. Tusk, K. Koperski, et al., Scene modeling and image mining with a visual grammar, *Front. Remote Sensing Inf. Proc.* (2003) 35–62.
- [35] C. Ri, M. Yao, Bayesian network based semantic image classification with attributed relational graph, *Multimed. Tools Appl.* 74 (13) (2015) 4965–4986.
- [36] H. Wang, H.g. Huang, C. Ding, Image annotation using Bi-Relational Graph of images and semantic labels, in: *Proc. CVPR*, 2012.
- [37] W.H. Hsu, L.S. Kennedy, S. Chang, Video search reranking through random walk over Document-Level context graph, in: *Proc. ACM Multimedia*, 2007.
- [38] Y. Jing, S. Baluja, Visual Rank: Applying pageRank to Large-Scale image search, in: *IEEE PAMI*, 2008.
- [39] D. Liu, X. Hua, L. Yang, Tag ranking, in: *International World Wide Web Conference Committee*, 2009.