

Semi-Supervised Learning on Bi-Relational Graph for Image Annotation

Hien Duy Pham¹, Kye-Hyeon Kim², and Seungjin Choi^{1,2}

¹ Division of IT Convergence Engineering

² Department of Computer Science and Engineering

Pohang University of Science and Technology

77 Cheongam-ro, Nam-gu, Pohang 790-784, Korea

Email: {hienpham1311, fenrir, seungjin}@postech.ac.kr

Abstract—We present a semi-supervised learning algorithm based on local and global consistency, working on a bi-relational graph of images and labels. By incorporating two types of different entities (images and labels) in a single graph, label propagation can exploit label correlations for measuring the relevance between unannotated images and labels, leading to a significant improvement in performance. In our propagation process, images belonging to the same label (or labels belonging to the same image) are not treated equally: our method allows that those images (or labels) have different weights in the label propagation process according to their semantic reliability to the label (or to the image), so that can achieve further improvement in the image annotation performance, compared to the existing work using a bi-relational graph. We apply our method to two benchmark multi-label image datasets, and obtain some encouraging experimental results.

I. INTRODUCTION

Advances in digital imagery has led to a sharp increase in the prevalence of digital images. As a result, there is an increasing requirement on designing image retrieval systems in order to create, manage and query image databases effectively. One important task in image retrieval is *image annotation*, a process of assigning relevant keywords or captions to images. Since manual annotation by human is a costly and tedious procedure, automatic image annotation has attracted more and more attention recently.

However, automatic image annotation is a difficult topic in computer vision because of the *semantic gap* between low-level image features and high-level semantic words. Extensive work has been performed to bridge the semantic gap problem (see [1] and references therein), following three different approaches: generative models, discriminative models and graph-based models.

Generative models compute a joint distribution over image features and labels, which leads to a conditional probability over labels given image features by using Bayes' rule. *Topic models* are well-known generative models, assuming that each image can exhibit multiple components or topics. Some examples of topic models are PLSA-based methods [2] and LDA-based methods [3]–[6]. A major drawback of generative models is a huge computational cost for parameter estimation in the learning process.

Discriminative models consider each annotation keyword as a class, and cast an image annotation task to a multi-class classification problem [7], [8]. By directly modeling

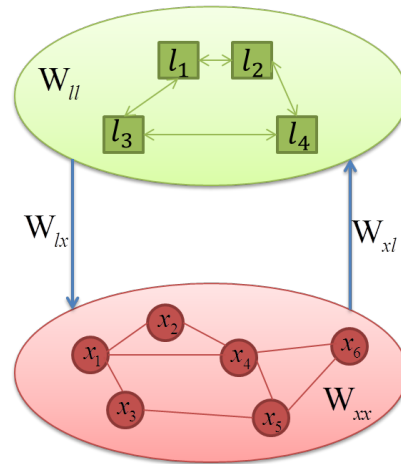


Fig. 1. A bi-relational graph (BG) of images and labels. Two subgraphs are modeled for representing correlations among images (W_{xx}) and labels (W_{ll}). The connections between annotated images and their labels (W_{xl} and W_{lx}) transfer the information between two groups, where the weights are set to a constant value in existing work. Our method allows different weights from/to images and labels, according to the actual semantic closeness of each image-label pair.

the conditional distribution of image features given annotation words, these models cannot exploit label correlations that are important for improving the overall classification performance.

Graph-based models formulate an image annotation task as *semi-supervised learning* [9], which can take advantage of a large number of unlabeled images. *Label propagation* is usually applied for transferring label information from annotated images to unannotated images over a graph. Traditional graph models connect both annotated and unannotated images according to their similarities. Each node corresponds to each image, and a pairwise similarity between images is assigned to each edge. However, traditional models do not consider label correlations [10], [11], or they just consider label correlations for simple refinement processes [12]. Some recent work proposed to incorporate label correlations into graph weights [13] or to exploit label correlations for subspace learning [14], [15] on a graph of images.

Recently, Wang et al. [16] proposed a *bi-relational graph* (BG, Fig. 1) of images and annotation words, and then developed the *random walk with restart* (RWR) label propagation process, which can directly measure the semantic closeness

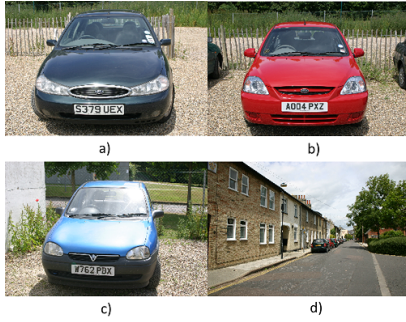


Fig. 2. An example that images belonging to the same label “car” have different contributions to the semantic closeness between the images and the label (W_{xl}). Compared to the lower-right image, the first three images are more representative for the label “car”, so that they should have larger weights to the label.



Fig. 3. An example that labels (“face”, “body”, and “book”) belonging to the same image have different contribution to the semantic closeness between the labels and the image (W_{lx}). According to the ground-truth in the right panel, “face” should have larger weights to the image than the weights from “body” and “book”.

between an annotation word and an image. While the work achieved a significant improvement in the classification performance, one limitation is that the weights between images and labels (W_{xl} and W_{lx} in Fig. 1) are simply fixed to a constant value. That is, images belonging to the same label (or labels belonging to the same image) contribute equally to the semantic closeness, while they actually have different contributions. Fig. 2 shows an example: Every image belongs to the class “car”, but intuitively, Fig. 2(a)-(c) should have larger weights to the class than Fig. 2(d). A similar example is shown in Fig. 3. The labels “face” and “body” should have more contribution than “book” for labeling the image.

In this paper, we propose a new BG model for allowing different contributions from/to images and labels, and then develop the label propagation algorithm on the proposed BG model. Below we summarize our contributions:

- Our BG model maintains different weights from/to images and labels, according to the actual semantic closeness of each image-label pair. The similarity between a label and an image is asymmetric, from which we can consider the relative reliability of nodes in the label propagation process.
- We extend the local and global consistency method [17] to a multi-entity data graph, by formulating a joint regularization framework and then deriving the corresponding RWR algorithm.
- Compared to the previous method of Wang et al. [16], our model does not need to learn the label-to-label re-

lationship iteratively. Also, experimental results show that our method outperforms the previous method on two multi-label image databases.

II. PROBLEM FORMULATION

Suppose that we have N images $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ and L labels $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$. We define \mathcal{C}_k as a set of images belonging to the label l_k , i.e., $i \in \mathcal{C}_k$ if \mathbf{x}_i belongs to l_k . Assuming that the first N_L images in \mathcal{X} are (partially) annotated, our goal is to predict labels for the remaining $N - N_L$ unannotated images.

A. Bi-relational Graph Model

A BG consists of two subgraphs of image nodes and label nodes and the intermediate connections between those subgraphs:

- \mathcal{G}_X captures the pairwise similarity between images.
- \mathcal{G}_L captures the correlation between labels.
- \mathcal{G}_B is the bipartite graph between \mathcal{X} and \mathcal{L} , representing the label assignment of the first N_L annotated images in \mathcal{X} .

Then the affinity of the whole BG is decomposed into 4 sub-matrices:

$$\mathbf{W}_G = \begin{bmatrix} \mathbf{W}_{xx} & \mathbf{W}_{xl} \\ \mathbf{W}_{lx} & \mathbf{W}_{ll} \end{bmatrix}, \quad (1)$$

where W_{xx} and W_{ll} are the similarity matrix of images in \mathcal{G}_X and correlation matrix of labels in \mathcal{G}_L respectively. W_{xl} and W_{lx} are defined on \mathcal{G}_B , representing the relation between the first N_L annotated images and their labels.

The similarity between images is defined as the cosine similarity:

$$W_{xx}(i, j) = \frac{\mathbf{x}_i^\top \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}, \quad (2)$$

where $W_{xx}(i, j)$ denotes the (i, j) -th entry in W_{xx} . The label correlation matrix W_{ll} is calculated as

$$W_{ll}(h, k) = \frac{\mathbf{y}_h^\top \mathbf{y}_k}{\|\mathbf{y}_h\|}, \quad (3)$$

where $\mathbf{y}_h \in \{0, 1\}^N$ is a binary vector such that $y_h(i) = 1$ for all $i \in \mathcal{C}_h$. Note that in the work of Wang et al. [16], W_{ll} is initialized as a symmetric matrix and then becomes asymmetric by learning, whereas in our model, W_{ll} is asymmetric from the beginning, and needs not to be updated.

Now we present how to construct W_{lx} and W_{xl} . In [16], $W_{xl}(i, k) = W_{lx}(k, i) = \beta$ for all $i \in \mathcal{C}_k$, given a constant $0 \leq \beta \leq 1$ (Fig. 4(a)). In other words, the similarity between a label and an image is always equal and symmetric. However, images belonging to the same label (or labels belonging to the same image) may have different contribution to the semantic closeness, as we have shown in Fig. 2 and 3.

Fig. 4(b) shows our proposed model for W_{lx} and W_{xl} : for each label l_k and the images $\{\mathbf{x}_i \mid i \in \mathcal{C}_k\}$,

$$W_{lx}(k, i) = \frac{\sum_{j \in \mathcal{C}_k} W_{xx}(i, j)}{\sum_{i' \in \mathcal{C}_k} \sum_{j \in \mathcal{C}_k} W_{xx}(i', j)}, \quad (4)$$

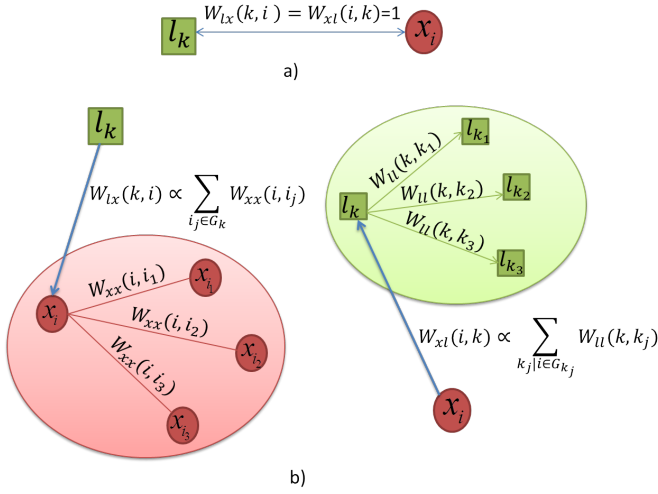


Fig. 4. The weight between an image and a label, defined in (a) the work of Wang et al. [16] and (b) our proposed BG.

and $W_{lx}(k, i) = 0$ for all $i \notin \mathcal{C}_k$. The normalization factor $\sum_{i' \in \mathcal{C}_k} \sum_{j \in \mathcal{C}_k} W_{xx}(i', j)$ enforces $\sum_i W_{lx}(k, i) = 1$ for all k . Eq. (4) implies that the label l_k has more effect on the image x_i if more neighboring images of x_i in \mathcal{G}_X also belongs to the label l_k . For example, in Fig. 3, the human face is dominant in the image. Thus, the image would be connected to other human face images in \mathcal{G}_X with large similarities $W_{xx}(i, j)$. Since those human face images would also tend to have the “face” label, Eq. (4) can enforce a large weight from the label “face” to the image in Fig. 3. On the other hand, the human body is not dominant in the image. Thus, there are only a few (or no) human body images among the neighbors of the image, leading to a small weight from the label “body” to the image.

Similarly, we define the image-to-label similarity W_{xl} as the following: for each annotated image x_i ($i = 1, \dots, N_L$) and its labels $\{l_k \mid i \in \mathcal{C}_k\}$,

$$W_{xl}(i, k) = \frac{\sum_{h \mid i \in \mathcal{C}_h} W_{ll}(k, h)}{\sum_{k' \mid i \in \mathcal{C}_{k'}} \sum_{h \mid i \in \mathcal{C}_h} W_{ll}(k', h)}, \quad (5)$$

and $W_{xl}(i, k) = 0$ if $i \notin \mathcal{C}_k$ or if x_i is unannotated. The normalization factor $\sum_{k' \mid i \in \mathcal{C}_{k'}} \sum_{h \mid i \in \mathcal{C}_h} W_{ll}(k', h)$ enforces $\sum_k W_{xl}(i, k) = 1$ for all annotated images $i = 1, \dots, N_L$.

B. Semi-supervised Learning on Bi-relational Graph

Now we present a novel label propagation method for image annotation with our BG model. Based on the local and global consistency (LGC) algorithm [17], which is a representative label propagation method for *single-label* semi-supervised classification (i.e., classes are assumed to be mutually exclusive), we extend the method to *multi-label* classification problems for image annotation (i.e., classes are often correlated). Fig. 5 clearly shows why our extension is necessary for image annotation.

We develop the joint regularization framework for the image subgraph \mathcal{G}_X , the label subgraph \mathcal{G}_L , and the bipartite graph \mathcal{G}_B . First, we consider the *image subgraph* \mathcal{G}_X . The cost

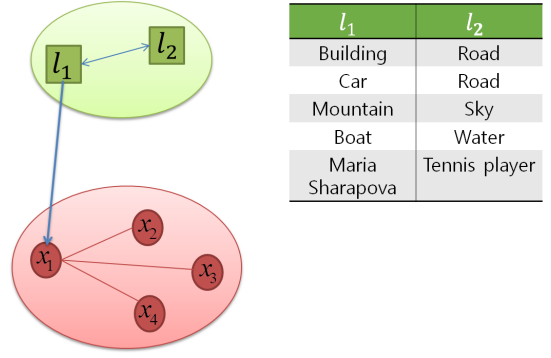


Fig. 5. Example of label propagation in our BG graph. When the label l_2 is missing in the image x_1 and its connected images, our method can transfer the label l_2 to x_1 through the high correlation from l_1 to l_2 , whereas traditional label propagation cannot. Some example pairs of labels are also shown, where l_2 exists with high probability if l_1 is given.

function associated with \mathcal{X} is defined as

$$Q(\mathbf{F}_X) = \sum_{i, j \in \mathcal{X}} W_{xx}(i, j) \left\| \frac{\mathbf{f}_X(i)}{\sqrt{D_{xx}(i, i)}} - \frac{\mathbf{f}_X(j)}{\sqrt{D_{xx}(j, j)}} \right\|^2 + \mu \sum_{i \in \mathcal{X}} \left\| \mathbf{f}_X(i) - \mathbf{f}_X^{(0)}(i) \right\|^2, \quad (6)$$

where

- $\mathbf{f}_X(i) \in \mathbb{R}^{1 \times L}$ is the i -th row of $\mathbf{F}_X \in \mathbb{R}^{N \times L}$, denoting the *relevance scores* of the image x_i for each label;
- $\mathbf{f}_X^{(0)}(i) = [W_{lx}(1, i), \dots, W_{lx}(L, i)]$ denotes *prior relevance scores*, which can be seen as the conditional probability of the image x_i given each label l_k without considering image-to-image and label-to-label correlations;
- $\mu > 0$ is the regularization parameter;
- D_{xx} is the diagonal matrix, whose (i, i) -th element is the sum of the i -th row of \mathbf{W}_{xx} , i.e., $D_{xx}(i, i) = \sum_j W_{xx}(i, j)$.

In Eq. (6), the first term of the right-hand side is the smoothness constraint, which means that the relevance score should not differ too much between nearby points (“*local consistency*”). The second term enforces the score should not differ too much from the initial score (“*global consistency*”).

Similarly, we define the cost function associated with the *label subgraph* \mathcal{G}_L as

$$Q(\mathbf{F}_L) = \sum_{i, j \in \mathcal{L}} W_{ll}(i, j) \left\| \frac{\mathbf{f}_L(i)}{\sqrt{D_{ll}(i, i)}} - \frac{\mathbf{f}_L(j)}{\sqrt{D_{ll}(j, j)}} \right\|^2 + \mu \sum_{i \in \mathcal{L}} \left\| \mathbf{f}_L(i) - \mathbf{f}_L^{(0)}(i) \right\|^2, \quad (7)$$

where $\mathbf{f}_L(i) \in \mathbb{R}^{1 \times L}$ (the i -th row of $\mathbf{F}_L \in \mathbb{R}^{L \times L}$) denotes the relevance between the label l_k and the other labels, and $\mathbf{f}_L^{(0)}(i) \in \{0, 1\}^{1 \times L}$ denotes the prior relevance whose i -th element is 1 and the other elements are 0.

Finally, we define two cost functions associated with the bipartite subgraph \mathcal{G}_B as

$$Q(\mathbf{F}_{XL}) = \sum_{j \in \mathcal{L}} \sum_{i \in \mathcal{C}_j} W_{xl}(i, j) \|\mathbf{f}_X(i) - \mathbf{f}_L(j)\|^2, \quad (8)$$

$$Q(\mathbf{F}_{LX}) = \sum_{j \in \mathcal{L}} \sum_{i \in \mathcal{C}_j} W_{lx}(i, j) \|\mathbf{f}_X(i) - \mathbf{f}_L(j)\|^2, \quad (9)$$

where we omit \mathbf{D}_{xl} and \mathbf{D}_{lx} since in Eq. (4) and (5) the sum of any nonzero row of \mathbf{W}_{xl} and \mathbf{W}_{lx} are always 1, i.e., $D_{xl}(i, i) = D_{xl}(j, j) = D_{lx}(i, i) = D_{lx}(j, j) = 1$ for all i, j . These two cost functions play a role of the smoothness constraints between two subgraphs \mathcal{G}_X and \mathcal{G}_L to avoid large differences in the relevance scores between \mathcal{G}_X and \mathcal{G}_L .

From Eq. (6)-(9), we derive the cost function over the whole BG:

$$Q(\mathbf{F}) = \alpha Q(\mathbf{F}_X) + \beta Q(\mathbf{F}_{XL}) + \gamma Q(\mathbf{F}_{LX}) + \delta Q(\mathbf{F}_L), \quad (10)$$

where $\mathbf{F} = [\mathbf{F}_X; \mathbf{F}_L] \in \mathbb{R}^{(N+L) \times L}$. Denoting $\mathbf{W} = \begin{bmatrix} \alpha \mathbf{W}_{xx} & \beta \mathbf{W}_{xl} \\ \gamma \mathbf{W}_{lx} & \delta \mathbf{W}_{ll} \end{bmatrix}$ and $\mathbf{D} = \begin{bmatrix} \alpha \mathbf{D}_{xx} & \beta \mathbf{D}_{xl} \\ \gamma \mathbf{D}_{lx} & \delta \mathbf{D}_{ll} \end{bmatrix}$, Eq. (10) can be rewritten as

$$Q(\mathbf{F}) = \sum_{i,j=1}^{N+L} W(i, j) \left\| \frac{\mathbf{f}(i)}{\sqrt{D(i, i)}} - \frac{\mathbf{f}(j)}{\sqrt{D(j, j)}} \right\|^2 + \mu \sum_{i=1}^{N+L} \left\| \mathbf{f}(i) - \mathbf{f}^{(0)}(i) \right\|^2, \quad (11)$$

where $\mathbf{F} = [\mathbf{f}_X(1); \dots; \mathbf{f}_X(N); \mathbf{f}_L(1); \dots; \mathbf{f}_L(L)]$ is the relevance score between $N+L$ nodes with L labels. Setting the derivative of $Q(\mathbf{F})$ to be zero, we have

$$\left. \frac{\partial Q}{\partial \mathbf{F}} \right|_{\mathbf{F}=\mathbf{F}^*} = \mathbf{F}^* - \mathbf{T}\mathbf{F}^* + \mu(\mathbf{F}^* - \mathbf{F}^{(0)}) = \mathbf{0}, \quad (12)$$

where

$$\mathbf{T} = \frac{1}{2}(\mathbf{S} + \mathbf{S}^\top) \quad (13)$$

and

$$\mathbf{S} = \begin{bmatrix} \alpha \mathbf{D}_{xx}^{-\frac{1}{2}} \mathbf{W}_{xx} \mathbf{D}_{xx}^{-\frac{1}{2}} & \beta \mathbf{D}_{xl}^{-\frac{1}{2}} \mathbf{W}_{xl} \mathbf{D}_{xl}^{-\frac{1}{2}} \\ \gamma \mathbf{D}_{lx}^{-\frac{1}{2}} \mathbf{W}_{lx} \mathbf{D}_{lx}^{-\frac{1}{2}} & \delta \mathbf{D}_{ll}^{-\frac{1}{2}} \mathbf{W}_{ll} \mathbf{D}_{ll}^{-\frac{1}{2}} \end{bmatrix}. \quad (14)$$

$\mathbf{D}_{xl,x} \in \mathbb{R}^{N \times N}$ and $\mathbf{D}_{xl,l} \in \mathbb{R}^{L \times L}$ are diagonal matrices, whose (i, i) -th elements are the sum of the i -th row and the i -th column of \mathbf{W}_{xl} , respectively. Similarly, $\mathbf{D}_{lx,x} \in \mathbb{R}^{L \times L}$ and $\mathbf{D}_{lx,l} \in \mathbb{R}^{N \times N}$ are diagonal matrices, whose (i, i) -th elements are the sum of the i -th row and the i -th column of \mathbf{W}_{lx} , respectively.

For the initial relevance $\mathbf{F}^{(0)} \in \mathbb{R}^{(N+L) \times L}$, we concatenate $\mathbf{F}_X^{(0)}$ and $\mathbf{F}_L^{(0)}$ with different weights as

$$\mathbf{F}^{(0)} = \begin{bmatrix} \lambda \mathbf{F}_X^{(0)} \\ (1 - \lambda) \mathbf{F}_L^{(0)} \end{bmatrix} \quad (15)$$

for a constant $0 \leq \lambda \leq 1$, which allows the relative importance of nodes belonging to each label l_k differently during the label propagation process, such that (1) $\lambda W_{lx}(k, i)$ for the i -th image node, and (2) $1 - \lambda$ for the label node l_k . The sum of the

importances (i.e., the sum of the k -th column of $\mathbf{F}^{(0)}$) is always 1.

Solving Eq. (12), we have the following closed-form solution:

$$\mathbf{F}^* = \frac{\mu}{1 + \mu} \left(\mathbf{I} - \frac{1}{1 + \mu} \mathbf{T} \right)^{-1} \mathbf{F}^{(0)}, \quad (16)$$

where \mathbf{I} denotes the $(N+L) \times (N+L)$ identity matrix. The i -th row vector $\mathbf{f}^*(i)$ measures the relevance between the i -th image x_i and each label optimally. Using the adaptive decision boundary method [15], one can refine the labels of N_L annotated images as well as assign labels for the remaining $N - N_L$ unannotated images.

III. EXPERIMENTS

We evaluate our proposed method, named *local and global consistency on bi-relational graph* (LGC-BG), on two well-known benchmark datasets for image annotation, namely *Microsoft Research Cambridge* image dataset (MSRC) and *LabelMe* image dataset [18].

We compare our methods to the iterative BG (I-BG) method [16] that performs label propagation on a BG in which the weights between annotated images and their labels are equal to a constant β .

For our approaches, we reports the results of two variants of our method:

- 1) LGC-BG1 does not consider the relative importance of image nodes during label propagation ($\lambda = 0$ in Eq. (15));
- 2) LGC-BG2 considers the relative importance of each node.

A. Data Description and Setup

MSRC dataset contains 591 images annotated by 23 labels. Each image has around 3 labels on average. As suggested by the official description of the dataset, we eliminated the ‘‘horse’’ label because only a few images belong to this label.

LabelMe dataset contains 2,687 images annotated by 200 labels. In our experiments, we used 8 categories from the dataset: ‘‘coast’’, ‘‘forest’’, ‘‘highway’’, ‘‘inside city’’, ‘‘mountain’’, ‘‘open country’’, ‘‘street’’ and ‘‘tall building’’.

For both datasets, we extracted global image descriptor features (GIST) [18] from each image. Then each image is represented by a 512-dimensional feature vector.

The parameter μ controls the trade-off between the local and the global consistency, which is usually set to a small value. For all of our experiments, we set $\mu = 0.01$. For LGC-BG2, we set $\lambda = 0.5$. For the other parameters $\alpha, \beta, \gamma, \delta$, we found the optimal values within $[0, 1]$ for each experiment.

B. Results

We evaluate our proposed methods and compare their annotation performance to the I-BG methods. For comparing the performance of image annotation, micro-averaged and macro-averaged versions of precision and F1 measures (Micro P, Macro P, Micro F1, Macro F1, respectively) were measured

TABLE I. IMAGE ANNOTATION PERFORMANCE (WITH 1 STANDARD DEVIATION) ON MSRC AND LABELME DATASETS

Dataset	Metric	I-BG	LGC-BG1	LGC-BG2
MSRC	Macro P	0.297±0.03	0.594±0.07	0.609±0.06
	Macro F1	0.371±0.02	0.585±0.05	0.613±0.05
	Micro P	0.302±0.04	0.564±0.07	0.628±0.07
	Micro F1	0.406±0.03	0.597±0.06	0.626±0.05
LabelMe	Micro P	0.210±0.05	0.503±0.09	0.506±0.08
	Micro F1	0.311±0.04	0.532±0.07	0.537±0.05



Fig. 6. Some examples of image annotation results on MSRC dataset. Boldface words are missing labels in the ground-truth annotation set but found by our method.

by 5-fold cross validation. For LabelMe dataset, we only measured Micro P and Micro F1 for comparison. In LabelMe dataset, some classes contain only one image, so we cannot compute the averaged precision and F1 score for each class, but only over the entire dataset.

Table I summarizes the results. LGC-BG methods consistently outperforms the I-BG approaches, clearly showing the effectiveness of our joint regularization framework. Moreover, I-BG iteratively updates the label correlation matrix for improving the overall performance, whereas our model simply uses the initial label correlation matrix and does not need to update it. The superiority of LGC-BG2 over LGC-BG1 shows the advantage of considering the relative reliability of image nodes in the propagation process. Fig. 6 and 7 show some examples of image annotation results obtained by our LGC-BG2 method.

IV. CONCLUSIONS

Bi-relational graph is a novel graph-based semi-supervised learning model to solve the image annotation problem (in particular) as well as multi-label classification problems (in general). By incorporating two types of different entities (data and labels) in a single graph, label propagation can exploit label correlations for multi-label classification. In this paper, we analyzed a major limitation of the existing BG-based method, and then overcome such a limitation using a novel BG model. We also developed a more general version of the local and global consistency method by investigating the joint regularization framework on our BG model. Experimental results clearly showed the significant improvement of our method compared to the previous approach. In future work, we will study the strategy for learning parameters in the joint regularization framework to achieve the best overall classification performance. We will also study how to fuse multiple BGs that are constructed from multiple types of features (e.g., SIFT, GIST, and color histogram) extracted from image data.



Fig. 7. Some examples of image annotation results on LabelMe dataset. Boldface words are missing labels in the ground-truth set but found by our method. Bold-italic words in brackets are given in the ground-truth set but cannot be found by our method, since those labels correspond to minor components in the images.

ACKNOWLEDGMENTS

This work was supported by National Research Foundation (NRF) of Korea (NRF-2013R1A2A2A01067464) and KEIT Machine Learning Research Center.

REFERENCES

- [1] F. Wang, "A survey on automatic image annotation and trends of the new age," *Procedia Engineering*, vol. 23, pp. 435–438, 2011.
- [2] F. Monary and D. Gatica-Perez, "PLSA-based image auto-annotation: Constraining the latent space," in *Proceedings of the ACM International Conference on Multimedia (MM)*, New York, New York, USA, 2004.
- [3] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *Journal of Machine Learning Research*, vol. 3, pp. 1107–1135, 2003.
- [4] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, Toronto, Canada, 2003.
- [5] C. Wang, D. M. Blei, and L. Fei-Fei, "Simultaneous image classification and annotation," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009.
- [6] D. Putthividhya, H. T. Attias, and S. S. Nagarajan, "Topic regression multi-modal latent Dirichlet allocation for image annotation," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, USA, 2010.
- [7] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.
- [8] G. Carneiro and N. Vasconcelos, "Formulating semantic image annotation as a supervised learning problem," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, 2005.
- [9] X. Zhu, "Semi-supervised learning literature survey," University of Wisconsin-Madison, Tech. Rep. Computer Sciences TR-1530, 2008.

- [10] H. Tong, J. He, M. Li, W.-Y. Ma, H.-J. Zhang, and C. Zhang, "Manifold-ranking-based keyword propagation for image retrieval," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–10, 2006.
- [11] P. Jia, N. Zhao, S. Hao, and J. Jiang, "Automatic image annotation by semi-supervised manifold kernel density estimation," *Information Sciences*, 2013, accepted.
- [12] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," *Pattern Recognition*, vol. 42, no. 2, pp. 218–228, 2009.
- [13] G. Chen, Y. Song, F. Wang, and C. Zhang, "Semi-supervised multi-label learning by solving a sylvester equation," in *Proceedings of the SIAM International Conference on Data Mining (SDM)*, Atlanta, GA, USA, 2008.
- [14] Z.-J. Zha, T. Mei, J. Wang, Z. Wang, and X.-S. Hua, "Graph-based semi-supervised learning with multi-label," in *Proceedings of IEEE International Conference on Multimedia and Expo*, Hannover, Germany, 2008.
- [15] H. Wang, H. Huang, and C. Ding, "Image annotation using multi-label correlated Green's function," in *Proceedings of the International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.
- [16] —, "Image annotation using bi-relational graph of images and semantic labels," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, 2011.
- [17] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 16. MIT Press, 2004, pp. 321–328.
- [18] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.