

# Learning image convolutional representations and complete tags jointly

Yanbin Wu<sup>1</sup> · Hongbin Zhai<sup>1</sup> · Mengna Li<sup>1</sup> · Fan Cui<sup>2</sup> · Li Wang<sup>3</sup> · Nitin Patil<sup>4</sup>

Received: 23 December 2016 / Accepted: 4 October 2017 / Published online: 17 October 2017  
© The Natural Computing Applications Forum 2017

**Abstract** In this paper, we develop a novel image tag completion method. We propose to represent the images by using the convolutional neural network (CNN) and predict the complete tags from the convolutional representations. The prediction is performed by a linear predictive model, and the complete tags are also imposed to be consistent to the existing elements of the incomplete tag matrix. We propose to learn the CNN parameters, the complete tags, and the predictive model parameters jointly. The learning problem is modeled by a minimization problem of an objective function composed of a consistency term between the learned complete tag vectors and the existing incomplete tag matrix, a prediction error term, and the convolutional similarity regularization term, and a sparsity term of the complete tag vector. The minimization problem is solved by an augmented Lagrangian method. The experiments over some benchmark data sets show that our method outperforms the state-of-the-art image tag completion methods.

**Keywords** Image tag completion · Convolutional neural network · Augmented Lagrangian method · Sparsity regularization

## 1 Introduction

### 1.1 Background

Image tagging problem is becoming more and more important for both image retrieval and classification applications [5, 7, 23, 32, 33, 40, 41]. This problem is defined as assigning a set of textual tags to an targeted set of images. An example is the keyword-based image searching application. In this application, given a query keyword, we return a list of images whose tags are similar to the query words. In this application, the quality of the tags assigned to the images is critical for the performance of the retrieval. In the ideal case, the tags of an image should be accurate and complete. However, in the real-world applications, the tags of images are usually incomplete, and it is necessary to complete the tags of images. For example, when a social network user updates an image to his/her page, some tags may be provided. But the provided tags usually cannot cover all the semantic concepts of the images, and the tags are incomplete. In this scenario, it is necessary to obtain the complete list of tags of the image before we use it for image retrieval problem. The problem of completing the tags of images is called image tag completion [10, 14, 25, 44]. To solve this problem, many approaches have been proposed [6, 15, 17, 18, 39, 42]. The matrix completion is a popular method for the tag completion problem [2–4]. However, the performance of these works are not satisfied yet. Meanwhile, convolutional neural network (CNN) has been proved to be an effective

---

✉ Li Wang  
wang.li@radi.ac.cn

<sup>1</sup> College of Management Science and Engineering, Hebei University of Economics and Business, Shijiazhuang 050061, China  
<sup>2</sup> University Party and Government Office, China University of Mining and Technology (Beijing), Beijing 100083, China  
<sup>3</sup> State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China  
<sup>4</sup> Savitribai Phule Pune University, Pune, Maharashtra 411007, India

tool to represent images [8, 20, 21, 43]. However, surprisingly, CNN has not been applied to the problem of image tag completion. In this paper, we propose to use CNN as the representation method for the problem of image tag completion and develop a joint learning method of convolutional representations and complete tags of images.

## 1.2 Relevant works

Our work is an image tag completion method; thus, we introduce a few state-of-the-art image tag completion methods. Lin et al. [18] proposed a novel method for image tag completion by using the image-specific and tag-specific linear sparse reconstructions. The incomplete initial tagging matrix is reconstructed by the linear sparse reconstructions which optimally reconstructs images and the tags. The reconstruction is constrained by sparsity, image–image similarity, image–tag association, and tag–tag concurrence. Wu et al. [39] presented the image–tag relation by a tag matrix and completed the tags by learning an optimal tag matrix. The completion is constrained by both the observed tags and the visual similarity. A new algorithm was proposed to solve the argued optimization problem. Feng et al. [6] showed that the problem of tag completion can be solved as a noisy matrix recovery problem. The missing tags are recovered, and the noisy tags are de-emphasized jointly in this method. Moreover, a graph Laplacian component derived from visual features are also used to regularize the noisy matrix recovery. Lin et al. [17] developed a novel approach for image tag completion using dual-view linear sparse reconstructions. This method conducts tag completion from two views, which are image view and tag view, and uses available context information. The image–image correlations are used to linearly reconstruct the low-level image features and the tag vectors jointly. Moreover, the tag–tag correlations are also used to obtain the reconstruction of the tag vector from the initial incomplete tags. Both the image view- and tag view-reconstructed tag vectors are combined to predict the complete tags. Xia et al. [42] designed a regularized optimization framework for the tag completion problem, by jointly conducting nonnegative matrix factorization (NMF) and the holistic visual diversity minimization to complete the tag-image matrix. The NMF modeling map the tag-image matrix into a latent low-rank space and use the semantic relevance of the tags to complete the tags partially. Moreover, the holistic visual diversity is used to regularize the NMF to utilize the content-similar images. Li et al. [15] proposed a novel tag completion method by using the low-rank and error sparsity, the local reconstruction structure consistency, and the promote basis diversity. The tag matrix is decomposed as the combination

of a low-rank matrix and a sparse error matrix, and the local linear reconstruction is further applied to regularize the learn of the tag vectors.

## 1.3 Our contributions

In our paper, we propose a novel image tag completion method. This method is based on the convolutional representation of the images and the linear prediction of the tag assignment vectors. We first use a CNN model to represent the images to the convolutional vectors and then apply a linear predictive model over the convolutional representations to obtain the complete tag assignment vectors of the images. To learn the parameters of the CNN and the linear predictive model, we impose the learned tag assignment vectors to be consistent to the available elements of the incomplete tag vectors, minimizing the prediction errors of the linear predicative model over the image set. We also argue that the tag assignment vectors of the images with large convolutional similarities should be similar and minimize the distance between their tag assignment vectors weighted by their convolutional similarity. Finally, we apply the sparsity penalty to the tag assignment vectors and minimize the  $L_1$  norm of the tag assignment vectors. The learning problem is modeled as a minimization problem with regard to the CNN filters, the complete tag assignment matrix, and the predictive model parameters jointly. The objective function of the problem is the weighted combination of a the squared Frobenius norm distance between the tag assignment elements and available tag elements, a predictive error term, a convolutional similarity regularization term, and a sparsity term of tag assignment vectors. To solve this problem, we use the augmented Lagrangian method (ALM). The experimental results over some benchmark data sets show that the proposed convolutional representation-based tag completion method outperforms the state-of-the-art tag completion methods.

## 1.4 Paper organization

The rest parts of this paper are organized as follows. In Sect. 2, we introduce the proposed method. In Sect. 3, we conduct experiments to evaluate the proposed method. In Sect. 4, we give the conclusion of this paper with some future works.

# 2 Proposed method

## 2.1 Problem modeling

We suppose that we have a set of images, denoted as  $\mathcal{I} = \{I_1, \dots, I_n\}$ , where  $I_i$  is the  $i$ th image, and  $n$  is the total

number of the images. A set of candidate tags of these images are also given  $\mathcal{T} = \{T_1, \dots, T_m\}$ , where  $T_j$  is the  $j$ th tag and  $m$  is the number of tags. The problem of image tagging is to assign the tags in  $\mathcal{T}$  to the images in  $\mathcal{I}$ . To denote the assigning relation between the images and the tags, we define a matrix of assignment,  $T = [t_{ji}] \in \{1, 0\}^{m \times n}$ , and its  $(j, i)$ th entity  $t_{ji}$  is set to 1 if  $T_j$  is assigned to  $I_i$ , and 0 otherwise. In many applications, we already have a existing assignment matrix  $\widehat{T} = [\widehat{t}_{ji}] \in \{1, 0\}^{m \times n}$ , and its entities are of the same meaning as  $T$ , but it is incomplete, i.e., many of its entities are missing. To denote which entities are missing, we further define a binary matrix  $\Phi = [\phi_{ji}] \in \{1, 0\}^{m \times n}$ , where its  $(j, i)$ th entity is defined to indicate if  $\widehat{t}_{ji}$  is missing,

$$\phi_{ji} = \begin{cases} 1, & \text{if } \widehat{t}_{ji} \text{ is available,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Thus with these definitions, the problem of image tagging is transformed to the learning of a complete assignment matrix  $T$  from the images set  $\mathcal{I}$ , the incomplete assignment matrix  $\widehat{T}$ , and its corresponding missing entity indicator matrix  $\Phi$ . To this end, we propose to present an image to a convolutional representation by a convolutional network and then use a linear model to predict its assignment vector from its convolutional representation. The learning of parameters of the convolutional network and the complete assignment matrix are conducted jointly. In the following sections, we will introduce the model of predicting tag assignment vectors and the learning of the parameters.

### 2.1.1 Tagging model

To complete the tags of an image,  $I$ , we propose to learn its convolutional representation and the complete tag assignment vector jointly. Given the image  $I$  we use a sliding window to split the image to  $n_I$  overlapping small image patches,  $I \rightarrow \{\mathbf{x}_1, \dots, \mathbf{x}_{n_I}\}$ , where  $\mathbf{x}_i \in \mathbb{R}^d$  is the visual feature vector of the  $i$ th image patch. We further organize the feature vectors of the image patches as a matrix  $X = [\mathbf{x}_1, \dots, \mathbf{x}_{n_I}] \in \mathbb{R}^{d \times n_I}$ , where  $\mathbf{x}_i$  is the  $i$ th column of the matrix. To obtain the convolutional representation, we first perform the filtering and nonlinear transformation to the matrix  $X$ ,

$$G = g(W^T X) \in \mathbb{R}^{r \times n_I}, \text{ where } g(x) = \frac{1}{1 + \exp(-x)}, \quad (2)$$

$W = [\mathbf{w}_1, \dots, \mathbf{w}_r] \in \mathbb{R}^{d \times r}$  is the matrix of  $r$  filters, and  $\mathbf{w}_k \in \mathbb{R}^d$  is the  $k$ th filter vector.  $g(\cdot)$  is a element-wise nonlinear transformation function. In our experiments, we use the sigmoid function as the nonlinear transformation function. Then, we perform the row-wise max-pooling to the output matrix  $G$ ,

$$\mathbf{y} = \max(G) = [y_1, \dots, y_r]^T \in \mathbb{R}^r \quad (3)$$

where  $\max(\cdot)$  gives the row-wise maximum elements, and  $y_k$  is the maximum entity of the  $k$ th row of  $G$ ,

$$y_k = \max(G_{k,:}), \quad (4)$$

where  $G_{k,:}$  is the  $k$ th row of matrix  $G$ . The output vector of max-pooling,  $\mathbf{y}$ , is the convolutional representation vector of the image  $I$ . The tag completion problem of this image to learn a tag assignment vector  $\mathbf{t} = [t_1, \dots, t_m] \in \{1, 0\}^m$ , where  $t_j = 1$  indicates that the  $T_j$  is assigned to the image  $I$ .

*Remark* Each image can be assigned to multiple tags, not just one tag. Thus this is an multi-tag completion problem.

Direct learning of the binary vector is difficult; thus, we relax the learning of binary vector to the learning of a continues value vector,  $\mathbf{t} \in \mathbb{R}^m$ . The elements of the assignment vector can be treated as the scores of assigning the tags to this image. To learn the tag assignment vector  $\mathbf{t}$  of an image from its convolutional representation vector  $\mathbf{y}$ , we use a linear function to predict  $\mathbf{t}$  from  $\mathbf{y}$ ,

$$\mathbf{t} \leftarrow (U\mathbf{y} - \mathbf{b}) \quad (5)$$

where  $U \in \mathbb{R}^{m \times r}$  and  $\mathbf{b} \in \mathbb{R}^m$  are the parameters of the predictive model for the assignment vector. To learn the complete tag assignment matrix  $T$  for all the images in  $\mathcal{I}$ , we apply the convolutional representation and linear tag assignment prediction to all the images. The problem of the learning of the complete tag assignment matrix is considered jointly with the learning problem of the convolutional representation parameters, i.e., the filter matrix, and the linear predictive model parameters.

### 2.2 Learning problem

To jointly learn the complete tag assignment matrix  $T$ , the convolutional representation filter matrix  $W$ , and the linear predictive model  $U$  and  $\mathbf{b}$  jointly, we consider the following few problems to construct the joint objective function of the learning problem.

- For the  $i$ th training image  $I_i$ , its existing incomplete assignment vector is given as  $\widehat{\mathbf{t}}_i = [\widehat{t}_{1i}, \dots, \widehat{t}_{mi}]^T \in \{1, 0\}^m$ , which is the  $i$ th column of the matrix  $\widehat{T}$ . This vector is not complete, and the  $i$ th column of  $\Phi$  indicates the missing elements of  $\widehat{\mathbf{t}}_i$ ,  $\phi_i = [\phi_{1i}, \dots, \phi_{mi}]^T \in \{1, 0\}^m$ . If  $\phi_{ji} = 1$ ,  $\widehat{t}_{ji}$  is available, not missing. In this case, the learned tag assignment score  $t_{ji}$  should be as close to the available  $\widehat{t}_{ji}$ . To measure how close  $t_{ji}$  is to its corresponding available  $\widehat{t}_{ji}$ , we calculate the squared Frobenius norm distance between them,  $(t_{ji} - \widehat{t}_{ji})^2$ . The squared

Frobenius norm distance between  $t_{ji}$  and  $\hat{t}_{ji}$  weighted by the  $\phi_{ji}$  is minimized with regard to  $t_{ji}$ . The minimization is applied over all the images and tags,

$$\min_T \left\{ \sum_{i=1}^n \sum_{j=1}^m \phi_{ji} (t_{ji} - \hat{t}_{ji})^2 \right. \\ \left. = \sum_{i=1}^n \text{Tr} \left( (\mathbf{t}_i - \hat{\mathbf{t}}_i)^\top \text{diag}(\boldsymbol{\phi}_i) (\mathbf{t}_i - \hat{\mathbf{t}}_i) \right) \right\}, \tag{6}$$

where  $\text{Tr}(\cdot)$  is the trace of a matrix, and  $\text{diag}(\boldsymbol{\phi}_i)$  is a diagonal matrix with  $\boldsymbol{\phi}_i$  as its diagonal vector. By minimizing this problem, we hope the learned  $T$  can respect the incomplete tag assignment vector,  $\hat{T}$ .

- Since we use a linear model to predict the tag assignment vector from the convolution representation vectors, we show the prediction error can be as small as possible. Again, we use the squared Frobenius norm distance to measure the prediction error. For the  $i$ th training image, the squared Frobenius norm distance between its tag assignment vector,  $\mathbf{t}_i$ , and the output vector of the linear predictive model of its corresponding convolutional representation vector,  $\mathbf{y}_i = \max(g(W^\top X_i))$ , is minimized. The minimization problem is also applied over the training set,

$$\min_{T,U,\mathbf{b},W} \sum_{i=1}^n \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2. \tag{7}$$

By this way, we bridge the problems of learning complete tags and convolutional representations of the images.

- We also use the similarities between the visual features of different images to regularize the learning of the complete tag assignment vectors. For a images  $I_i$ , we seek its  $k$ -nearest neighbors to present its visually similar images. The set of  $k$ -nearest neighbors of  $I_i$  is denoted as  $\mathcal{N}_i$ . To measure the similarity between  $I_i$  and a neighboring image  $I_{i'} \in \mathcal{N}_i$  is measured by the normalized Gaussian kernel of the Frobenius norm distance between their convolutional representation vectors,

$$S_{i i'} = \begin{cases} \frac{\exp(-\gamma \|\mathbf{y}_i - \mathbf{y}_{i'}\|_F^2)}{\sum_{i'' \in \mathcal{N}_i} \exp(-\gamma \|\mathbf{y}_i - \mathbf{y}_{i''}\|_F^2)}, & \text{if } I_{i'} \in \mathcal{N}_i \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

If  $S_{i i'}$  is large, the  $I_i$  and  $I_{i'}$  are visually similar, we expect their complete tag assignment vectors,  $\mathbf{t}_i$  and  $\mathbf{t}_{i'}$  should also be close to each other. Thus, we minimize the squared Frobenius norm distance between  $\mathbf{t}_i$  and  $\mathbf{t}_{i'}$

weighted by  $S_{i i'}$ . The minimization is also applied over all pairs of training images,

$$\min_T \sum_{i,i'=1}^n S_{i i'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2. \tag{9}$$

- We also observed that for each image, it is usually assigned to only a few tags, and the complete tag assignment vector should be sparse. To encourage the sparsity of the tag assignment vector, we use a  $L_1$  norm term to regularize the tag assignment vector to force most of its elements to zero. The  $L_1$  norm of the tag assignment vector  $\mathbf{t}_i$ ,  $\|\mathbf{t}_i\|_1 = \sum_{j=1}^m |t_{ji}|$  is minimized, and the minimization problem is applied over all the training images,

$$\min_T \left\{ \sum_{i=1}^n \|\mathbf{t}_i\|_1 = \sum_{i=1}^n \sum_{j=1}^m |t_{ji}| \right\}. \tag{10}$$

The overall optimization problem of the learning is the weighted summation of the above objectives,

$$\min_{T,U,\mathbf{b},W} \left\{ \sum_{i=1}^n \text{Tr} \left( (\mathbf{t}_i - \hat{\mathbf{t}}_i)^\top \text{diag}(\boldsymbol{\phi}_i) (\mathbf{t}_i - \hat{\mathbf{t}}_i) \right) + \lambda_1 \sum_{i=1}^n \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 \right. \\ \left. + \lambda_2 \sum_{i,i'=1}^n S_{i i'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2 + \lambda_3 \sum_{i=1}^n \|\mathbf{t}_i\|_1 \right\} \tag{11}$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are the weights of different regularization terms of the objective, and their values are decided by cross-validation in our experiments. They play the roles of trade-off of different terms. Please note that  $\mathbf{y}_i$  is not listed as a variable in this problem. It is a function of the convolutional filter matrix  $W$ . Directly solving  $W$  of (11) is difficult, we propose to explicitly introduce  $\mathbf{y}_i, i = 1, \dots, n$  as variable vectors, and constrains to impose  $\mathbf{y}_i = \max(g(W^\top X_i))$ . The optimization problem of (11) is transformed to a constrained optimization problem,

$$\min_{T,U,\mathbf{b},W,Y} \left\{ \sum_{i=1}^n \text{Tr} \left( (\mathbf{t}_i - \hat{\mathbf{t}}_i)^\top \text{diag}(\boldsymbol{\phi}_i) (\mathbf{t}_i - \hat{\mathbf{t}}_i) \right) + \lambda_1 \sum_{i=1}^n \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 \right. \\ \left. + \lambda_2 \sum_{i,i'=1}^n S_{i i'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2 + \lambda_3 \sum_{i=1}^n \|\mathbf{t}_i\|_1 \right\} \\ \text{s.t. } \mathbf{y}_i = \max(g(W^\top X_i)), \forall i = 1, \dots, n, \tag{12}$$

where  $Y = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{R}^{m \times n}$  is the matrix of convolutional representations.

### 2.3 Problem optimization

To solve the problem in (12), we use the augmented Lagrangian method (ALM) [1, 9, 26, 28, 30, 31, 45–47].

The augmented Lagrangian function of the problem in (12) is given as follows,

$$\begin{aligned} \mathcal{L}(T, U, \mathbf{b}, W, Y, A) = & \sum_{i=1}^n Tr\left((\mathbf{t}_i - \widehat{\mathbf{t}}_i)^\top \text{diag}(\phi_i)(\mathbf{t}_i - \widehat{\mathbf{t}}_i)\right) \\ & + \lambda_1 \sum_{i=1}^n \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 \\ & + \lambda_2 \sum_{i,i'=1}^n S_{ii'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2 + \lambda_3 \sum_{i=1}^n \|\mathbf{t}_i\|_1 \\ & + \sum_{i=1}^n \alpha_i^\top (\mathbf{y}_i - \max(g(W^\top X_i))) \\ & + \frac{\beta}{2} \sum_{i=1}^n \|\mathbf{y}_i - \max(g(W^\top X_i))\|_F^2, \end{aligned} \tag{13}$$

where  $\alpha_i \in \mathbb{R}_+^m$  is the Lagrange multiplier vector for the constraint  $\mathbf{y}_i = \max(g(W^\top X_i))$ , and  $\beta$  is the positive penalty parameter. The optimization problem is transformed to the following minimization–maximization coupled problem,

$$\begin{aligned} \max_A \min_{T,U,\mathbf{b},W,Y} \mathcal{L}(T, U, \mathbf{b}, W, Y, A) \\ \text{s.t. } A \geq 0, \end{aligned} \tag{14}$$

where  $A = [\alpha_1, \dots, \alpha_n] \in \mathbb{R}_+^{m \times n}$  is the matrix of Lagrange multiplier vectors. To solve this coupled problem, we use the alternating direction method of multipliers (ADMM). This method updates the multiplier matrix  $A$  and the other variables,  $T, U, \mathbf{b}, W$  and  $Y$  alternately. In the following sections, we discuss how to optimize these variables, respectively.

### 2.3.1 Optimizing $T$

To optimize the complete tag assignment matrix  $T$ , we fix all the other variables, and the following sub-optimization problem is obtained from (14),

$$\begin{aligned} \min_T \left\{ \sum_{i=1}^n Tr\left((\mathbf{t}_i - \widehat{\mathbf{t}}_i)^\top \text{diag}(\phi_i)(\mathbf{t}_i - \widehat{\mathbf{t}}_i)\right) + \lambda_1 \sum_{i=1}^n \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 \right. \\ \left. + \lambda_2 \sum_{i,i'=1}^n S_{ii'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2 + \lambda_3 \sum_{i=1}^n \|\mathbf{t}_i\|_1 \right. \\ = \sum_{i=1}^n \left( Tr\left((\mathbf{t}_i - \widehat{\mathbf{t}}_i)^\top \text{diag}(\phi_i)(\mathbf{t}_i - \widehat{\mathbf{t}}_i)\right) + \lambda_1 \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 \right. \\ \left. + \lambda_2 \sum_{i'=1}^n S_{ii'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2 + \lambda_3 \|\mathbf{t}_i\|_1 \right) = \sum_{i=1}^n g(\mathbf{t}_i) \left. \right\}. \end{aligned} \tag{15}$$

From (15), we observe that the objective of (15) is a summation of a function  $g(\mathbf{t}_i)$  over independent tag assignment vectors  $\mathbf{t}_i$ . Thus the optimization of each  $\mathbf{t}_i$  is

independent from the optimization problems of other  $\mathbf{t}_{i'}$ ,  $i' \neq i$ . We can optimize  $\mathbf{t}_i$  by only optimizing  $g(\mathbf{t}_i)$ , and the following optimization problem is obtained,

$$\begin{aligned} \min_{\mathbf{t}_i} \left\{ g(\mathbf{t}_i) = Tr\left((\mathbf{t}_i - \widehat{\mathbf{t}}_i)^\top \text{diag}(\phi_i)(\mathbf{t}_i - \widehat{\mathbf{t}}_i)\right) + \lambda_1 \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 \right. \\ \left. + \lambda_2 \sum_{i'=1}^n S_{ii'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2 + \lambda_3 \|\mathbf{t}_i\|_1 \right\}. \end{aligned} \tag{16}$$

In  $g(\mathbf{t}_i)$ , the last term  $\|\mathbf{t}_i\|_1$  is non-smooth, and it is difficult to minimize it directly. To overcome this problem, we use the fix-point method. The last term  $\|\mathbf{t}_i\|_1$  can be rewritten as follows,

$$\|\mathbf{t}_i\|_1 = \sum_{j=1}^m |t_{ji}| = \sum_{j=1}^m \frac{t_{ji}^2}{|t_{ji}|} = \mathbf{t}_i^\top A(\mathbf{t}_i) \mathbf{t}_i, \tag{17}$$

where  $A(\mathbf{t}_i) = \text{diag}\left(\frac{1}{|t_{i1}|}, \dots, \frac{1}{|t_{im}|}\right)$ . It could be observed that if we fix the matrix  $A(\mathbf{t}_i)$ ,  $\mathbf{t}_i^\top A(\mathbf{t}_i) \mathbf{t}_i$  is a quadratic function of  $\mathbf{t}_i$ . In each iteration of an iterative algorithm, we firstly use the previous solution  $\widetilde{\mathbf{t}}_i$  to update the  $A(\widetilde{\mathbf{t}}_i)$  and then fix it to solve the minimization problem of (16). Meanwhile, we should also note that in the third term,  $\sum_{i'=1}^n S_{ii'} \|\mathbf{t}_i - \mathbf{t}_{i'}\|_F^2$ , we also need to know the tag assignment vectors of other images,  $\mathbf{t}_{i'}$ ,  $i' \neq i$ , to update  $\mathbf{t}_i$ . We also use the previous solutions of  $\mathbf{t}_{i'}$ ,  $\widetilde{\mathbf{t}}_{i'}$ , to regularize the learning of  $\mathbf{t}_i$ . The obtained objective is given as follows,

$$\begin{aligned} \widetilde{g}(\mathbf{t}_i) = Tr\left((\mathbf{t}_i - \widehat{\mathbf{t}}_i)^\top \text{diag}(\phi_i)(\mathbf{t}_i - \widehat{\mathbf{t}}_i)\right) \\ + \lambda_1 \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 + \lambda_2 \sum_{i'=1}^n S_{ii'} \|\mathbf{t}_i - \widetilde{\mathbf{t}}_{i'}\|_F^2 \\ + \lambda_3 \mathbf{t}_i^\top A(\widetilde{\mathbf{t}}_i) \mathbf{t}_i \end{aligned} \tag{18}$$

To minimize this objective, we set its derivative  $\nabla \widetilde{g}(\mathbf{t}_i)$  to zero to obtain the optimal solution  $\mathbf{t}_i^*$ ,

$$\begin{aligned} \nabla \widetilde{g}(\mathbf{t}_i) = 2\text{diag}(\phi_i)(\mathbf{t}_i - \widehat{\mathbf{t}}_i) + 2\lambda_1(\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})) \\ + 2\lambda_2 \sum_{i'=1}^n S_{ii'} (\mathbf{t}_i - \widetilde{\mathbf{t}}_{i'}) + 2\lambda_3 A(\widetilde{\mathbf{t}}_i) \mathbf{t}_i = 0, \\ \Rightarrow \left( \text{diag}(\phi_i) + \left( \lambda_1 + \lambda_2 \sum_{i'=1}^n S_{ii'} \right) I + \lambda_3 A(\widetilde{\mathbf{t}}_i) \right) \mathbf{t}_i \\ = \left( \text{diag}(\phi_i) \widehat{\mathbf{t}}_i + \lambda_1 (U\mathbf{y}_i - \mathbf{b}) + \lambda_2 \sum_{i'=1}^n S_{ii'} \widetilde{\mathbf{t}}_{i'} \right), \\ \Rightarrow \mathbf{t}_i^* = \left( \text{diag}(\phi_i) + \left( \lambda_1 + \lambda_2 \sum_{i'=1}^n S_{ii'} \right) I + \lambda_3 A(\widetilde{\mathbf{t}}_i) \right)^{-1} \\ \left( \text{diag}(\phi_i) \widehat{\mathbf{t}}_i + \lambda_1 (U\mathbf{y}_i - \mathbf{b}) + \lambda_2 \sum_{i'=1}^n S_{ii'} \widetilde{\mathbf{t}}_{i'} \right). \end{aligned} \tag{19}$$

### 2.3.2 Optimizing $U$ and $\mathbf{b}$

To optimize the linear predictive model parameter  $U$  and  $\mathbf{b}$ , we fix the other variables and obtain the following sub-optimization problem.

$$\min_{U, \mathbf{b}} \left\{ h(U, \mathbf{b}) = \lambda_1 \sum_{i=1}^n \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 \right\}, \tag{20}$$

where  $h(U, \mathbf{b})$  is the objective function of this sub-optimization problem. To minimize this objective function with regard to  $\mathbf{b}$ , we firstly set its derivative with regard to  $\mathbf{b}$  to zero to obtain the optimal solution  $\mathbf{b}^*$ ,

$$\begin{aligned} \nabla_{\mathbf{b}} h(U, \mathbf{b}) &= \lambda_1 \sum_{i=1}^n (\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})) = 0, \\ \Rightarrow \mathbf{b}^* &= \frac{1}{n} \sum_{i=1}^n (U\mathbf{y}_i - \mathbf{t}_i). \end{aligned} \tag{21}$$

We substitute the optimal solution  $\mathbf{b}^*$  to (20) to reduce the problem to only one variable  $U$ ,

$$\begin{aligned} \min_U \left\{ \tilde{h}(U) &= \lambda_1 \sum_{i=1}^n \left\| \mathbf{t}_i - \left( U\mathbf{y}_i - \frac{1}{n} \sum_{i'=1}^n (U\mathbf{y}_{i'} - \mathbf{t}_{i'}) \right) \right\|_F^2 \right. \\ &= \left. \lambda_1 \sum_{i=1}^n \left\| \left( \mathbf{t}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{t}_{i'} \right) - U \left( \mathbf{y}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{y}_{i'} \right) \right\|_F^2 \right\}. \end{aligned} \tag{22}$$

To obtain the optimal solution of  $U$ , we set the derivative of  $\tilde{h}(U)$  to zero,

$$\begin{aligned} \nabla_U \tilde{h}(U) &= 2\lambda_1 \sum_{i=1}^n \left( \left( \mathbf{t}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{t}_{i'} \right) - U \left( \mathbf{y}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{y}_{i'} \right) \right) = 0, \\ \Rightarrow U \sum_{i=1}^n \left( \mathbf{y}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{y}_{i'} \right) &= \sum_{i=1}^n \left( \mathbf{t}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{t}_{i'} \right), \\ \Rightarrow U^* &= \sum_{i=1}^n \left( \mathbf{t}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{t}_{i'} \right) \left( \sum_{i=1}^n \left( \mathbf{y}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{y}_{i'} \right) \right)^{-1}. \end{aligned} \tag{23}$$

### 2.3.3 Optimizing $W$

To solve the filter matrix  $W$ , we fix the other variables and obtain the following sub-optimization problem,

$$\begin{aligned} \min_W \left\{ \sum_{i=1}^n \alpha_i^\top (\mathbf{y}_i - \max(g(W^\top X_i))) + \frac{\beta}{2} \sum_{i=1}^n \|\mathbf{y}_i - \max(g(W^\top X_i))\|_F^2 \right. \\ = \sum_{k=1}^r \sum_{i=1}^n \left( \alpha_{ik} (y_{ik} - \max(g(\mathbf{w}_k^\top X_i))) + \frac{\beta}{2} (y_{ik} - \max(g(\mathbf{w}_k^\top X_i)))^2 \right) \\ \left. = \sum_{k=1}^r f(\mathbf{w}_k) \right\}. \end{aligned} \tag{24}$$

According to (24), we can see that the objective of (24) is actually a summation of the function  $f(\mathbf{w}_k)$  over independent filters  $\mathbf{w}_k, k = 1, \dots, r$ . So we can optimize the filters independently by solving  $r$  sub-optimization problems. The sub-optimization of the  $k$ th filter is given as follows,

$$\min_{\mathbf{w}_k} \left\{ f(\mathbf{w}_k) = \sum_{i=1}^n \left( \alpha_{ik} (y_{ik} - \max(g(\mathbf{w}_k^\top X_i))) + \frac{\beta}{2} (y_{ik} - \max(g(\mathbf{w}_k^\top X_i)))^2 \right) \right\}, \tag{25}$$

where  $\max(g(\mathbf{w}_k^\top X_i))$  is the output of the convolutional representation model with regard to the  $k$ th filter,  $\mathbf{w}_k$ . It can be rewritten as follows by introducing a maximum responding instance vector,  $\boldsymbol{\eta}_{ik}$ ,

$$\begin{aligned} \max(g(\mathbf{w}_k^\top X_i)) &= \max_{\boldsymbol{\eta}_{ik}} (g(\mathbf{w}_k^\top X_i \boldsymbol{\eta}_{ik})), \\ \text{s.t. } \boldsymbol{\eta}_{ik} &\in \{1, 0\}^{n_i}, \boldsymbol{\eta}_{ik} \mathbf{1} = \mathbf{1}. \end{aligned} \tag{26}$$

$\boldsymbol{\eta}_{ik}$  is a binary vector, and it has only one element as one, while all the other elements are set to zero. The position of this one-value element indicates the position of the instance in  $X_i$  which gives the maximum response. Substituting (26) to (25), we have

$$\begin{aligned} \min_{\mathbf{w}_k} \left\{ f(\mathbf{w}_k) &= \sum_{i=1}^n \left( \alpha_{ik} (y_{ik} - g(\mathbf{w}_k^\top X_i \boldsymbol{\eta}_{ik}^*)) + \frac{\beta}{2} (y_{ik} - g(\mathbf{w}_k^\top X_i \boldsymbol{\eta}_{ik}^*))^2 \right) \right\}, \\ \text{s.t. } \boldsymbol{\eta}_{ik}^* &= \arg \max_{\boldsymbol{\eta}_{ik} \in \{1, 0\}^{n_i}} g(\mathbf{w}_k^\top X_i \boldsymbol{\eta}_{ik}), \quad i=1, \dots, n. \end{aligned} \tag{27}$$

We also use the fix-point method to solve this problem. In each iteration of an iterative algorithm, we first fix  $\mathbf{w}_k$  as its previous solution,  $\tilde{\mathbf{w}}_k$ , to update  $\boldsymbol{\eta}_{ik}^*$ . Then we fix  $\boldsymbol{\eta}_{ik}^*$  to update  $\mathbf{w}_k$  by using the gradient descent method,

$$\mathbf{w}_k \leftarrow \mathbf{w}_k - \eta \nabla_{\mathbf{w}_k} f(\mathbf{w}_k), \tag{28}$$

where  $\nabla_{\mathbf{w}_k} f(\mathbf{w}_k)$  is the gradient function of  $f(\mathbf{w}_k)$ ,

$$\nabla_{\mathbf{w}_k} f(\mathbf{w}_k) = \sum_{i=1}^n \left( -\alpha_{ik} \nabla_{\mathbf{w}_k} g(\mathbf{w}_k^\top X_i \boldsymbol{\eta}_{ik}^*) X_i \boldsymbol{\eta}_{ik}^* - \beta (y_{ik} - g(\mathbf{w}_k^\top X_i \boldsymbol{\eta}_{ik}^*)) \nabla_{\mathbf{w}_k} g(\mathbf{w}_k^\top X_i \boldsymbol{\eta}_{ik}^*) X_i \boldsymbol{\eta}_{ik}^* \right), \tag{29}$$

$\nabla g(\cdot)$  is the gradient function of the nonlinear transformation function, and  $\eta$  is the descent step size.

### 2.3.4 Optimizing Y

To optimize the convolutional representation vectors of the images, we fix the other variables and obtain the following sub-optimization problem,

$$\min_Y \left\{ \sum_{i=1}^n \left( \lambda_1 \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 + \boldsymbol{\alpha}_i^\top (\mathbf{y}_i - \max(g(W^\top X_i))) \right) + \frac{\beta}{2} \|\mathbf{y}_i - \max(g(W^\top X_i))\|_F^2 \right\} = \sum_{i=1}^n l(\mathbf{y}_i) \tag{30}$$

We also observe the objective function of this problem is a linear combination of an independent function  $l(\mathbf{y}_i)$  over the training images. Thus we can consider the minimization of  $l(\mathbf{y}_i)$  with regard to  $\mathbf{y}_i$  independently. The minimization problem of  $l(\mathbf{y}_i)$  is given as follows,

$$\min_{\mathbf{y}_i} \left\{ l(\mathbf{y}_i) = \lambda_1 \|\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})\|_F^2 + \boldsymbol{\alpha}_i^\top (\mathbf{y}_i - \max(g(W^\top X_i))) + \frac{\beta}{2} \|\mathbf{y}_i - \max(g(W^\top X_i))\|_F^2 \right\}. \tag{31}$$

To solve this problem, we set the directive of  $l(\mathbf{y}_i)$  with regard to  $\mathbf{y}_i$  to zero and have the following optimal solution  $\mathbf{y}_i^*$

$$\begin{aligned} \nabla l(\mathbf{y}_i) &= -2\lambda_1 U^\top (\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})) + \boldsymbol{\alpha}_i + \beta (\mathbf{y}_i - \max(g(W^\top X_i))) = 0, \\ \Rightarrow -2\lambda_1 U^\top (\mathbf{t}_i - (U\mathbf{y}_i - \mathbf{b})) + \boldsymbol{\alpha}_i + \beta (\mathbf{y}_i - \max(g(W^\top X_i))) &= 0, \\ \Rightarrow (\beta I - 2\lambda_1 U^\top U) \mathbf{y}_i &= 2\lambda_1 U^\top (\mathbf{t}_i + \mathbf{b}) + \beta \max(g(W^\top X_i)) - \boldsymbol{\alpha}_i, \\ \Rightarrow \mathbf{y}_i^* &= (\beta I - 2\lambda_1 U^\top U)^{-1} (2\lambda_1 U^\top (\mathbf{t}_i + \mathbf{b}) + \beta \max(g(W^\top X_i)) - \boldsymbol{\alpha}_i). \end{aligned} \tag{32}$$

### 2.3.5 Optimizing A

To optimize the Lagrange multiplier matrix, we fix the other variables and obtain the following sub-optimization problem,

$$\max_A \left\{ s(A) = \sum_{i=1}^n \boldsymbol{\alpha}_i^\top (\mathbf{y}_i - \max(g(W^\top X_i))) = Tr(A^\top \Theta) \right\} \tag{33}$$

where  $\Theta = [(\mathbf{y}_1 - \max(g(W^\top X_1))), \dots, (\mathbf{y}_n - \max(g(W^\top X_n)))] \in R^{m \times n}$ . We use the gradient ascent method to update this matrix to maximize the objective function  $s(A)$ ,

$$A \leftarrow A + \varsigma \nabla s(A) \tag{34}$$

where  $\nabla s(A)$  is the gradient function of  $s(A)$ , which is defined as

$$\nabla s(A) = \Theta, \tag{35}$$

and  $\varsigma$  is the ascent step size.

## 2.4 Algorithm

Based on the optimization results of the above section, we design an iterative algorithm to learn these variables. The algorithm, named as Joint Image Convolutional Representation and Tag Completion (JICRTC) algorithm, is described in Algorithm 1.

**Algorithm 1** Iterative optimization algorithm of JICRTC.

Inputs :  $X_i, \hat{\mathbf{t}}_i, i = 1, \dots, n$ .  
 Initialization :  $\mathbf{t}_i, \mathbf{y}_i, i = 1, \dots, n, U, \mathbf{b}, W$ , and  $A$ .  
 Repeat :

1. Calculate the image pair-wise convolutional similarities for  $i, i' = 1, \dots, n$ :

$$S_{ii'} = \begin{cases} \frac{\exp(-\gamma \|\mathbf{y}_i - \mathbf{y}_{i'}\|_F^2)}{\sum_{i'' \in \mathcal{N}_i} \exp(-\gamma \|\mathbf{y}_i - \mathbf{y}_{i''}\|_F^2)}, & \text{if } i' \in \mathcal{N}_i \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

2. Update the complete tag assignment score vectors for  $i = 1, \dots, n$ :

$$\begin{aligned} \mathbf{t}_i &= \left( \text{diag}(\boldsymbol{\phi}_i) + \left( \lambda_1 + \lambda_2 \sum_{i'=1}^n S_{ii'} \right) I + \lambda_3 A(\mathbf{t}_i) \right)^{-1} \\ &\quad \left( \text{diag}(\boldsymbol{\phi}_i) \hat{\mathbf{t}}_i + \lambda_1 (U\mathbf{y}_i - \mathbf{b}) + \lambda_2 \sum_{i'=1}^n S_{ii'} \mathbf{t}_{i'} \right). \end{aligned} \tag{37}$$

3. Update the predictive model parameters,  $U$  and  $\mathbf{b}$ ,

$$\begin{aligned} U &= \sum_{i=1}^n \left( \mathbf{t}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{t}_{i'} \right) \left( \sum_{i=1}^n \left( \mathbf{y}_i - \frac{1}{n} \sum_{i'=1}^n \mathbf{y}_{i'} \right) \right)^{-1}, \text{ and} \\ \mathbf{b} &= \frac{1}{n} \sum_{i=1}^n (U\mathbf{y}_i - \mathbf{t}_i). \end{aligned} \tag{38}$$

4. Update the maximum responding instance indicators for  $i = 1, \dots, n$ :

$$\eta_{ik}^* = \arg \max_{\eta_{ik} \in \{1,0\}^n, \eta_{ik} \mathbf{1} = \mathbf{1}} g(\mathbf{w}_k^\top X_i \eta_{ik}), \quad k = 1, \dots, r. \quad (39)$$

5. Update the filters for  $k = 1, \dots, r$ :

$$\mathbf{w}_k = \mathbf{w}_k - \eta \left( \sum_{i=1}^n (-\alpha_{ik} \nabla_{\mathbf{w}_k} g(\mathbf{w}_k^\top X_i \eta_{ik}^*) X_i \eta_{ik}^* - \beta (y_{ik} - g(\mathbf{w}_k^\top X_i \eta_{ik}^*))) \nabla_{\mathbf{w}_k} g(\mathbf{w}_k^\top X_i \eta_{ik}^*) X_i \eta_{ik}^* \right). \quad (40)$$

6. Update the convolutional representations of the images:

$$\mathbf{y}_i = (\beta I - 2\lambda_1 U^\top U)^{-1} (2\lambda_1 U^\top (\mathbf{t}_i + \mathbf{b}) + \beta \max(g(W^\top X_i)) - \alpha_i). \quad (41)$$

7. Update the Lagrange multiplier matrix  $A$ :

$$A = A + \zeta ([(\mathbf{y}_1 - \max(g(W^\top X_1))), \dots, (\mathbf{y}_n - \max(g(W^\top X_n)))]). \quad (42)$$

Until convergence.

Outputs :  $\mathbf{t}_i, i = 1, \dots, n$ .

### 3 Experiments

In this section, we evaluate the proposed algorithm, JICRTC, over a few benchmark image data sets. The evaluation is based on two different computer vision problems, which are image annotation and image retrieval.

#### 3.1 Data sets and experimental setting

In the experiments, three data sets of images are used. The data sets are introduced as follows.

- Corel dataset is composed of 4993 images. The tag set is composed of 260 different tags. Each image has at most five tags.
- Labelme data set is composed 2900 images, and its tag set has 495 tags. Each image of this data sets has at most 48 tags.

**Table 1** Data set statistical information

Data sets	# images	# tags	Max # tags per image
Corel	4993	260	5
Labelme	2900	495	48
Flickr	1,000,000	>1000	76

- Flickr data set has 1,000,000 images. Its tag set has more than 1,000 tags. Each images is tagged by at most 76 tags.

The statistical information of the data sets is summarized in Table 1.

### 3.2 Image annotation experiments

#### 3.2.1 Experimental setting

We first evaluate our image tag completion method over the image annotation problem [27]. Given an image, and a set of candidate tags, the problem of image annotation is to predict its true complete list of tags relevant to the image. This is an special case of image tag completion, where each image has no existing tags and all the tags are missing. We should predict all its missing tags.

We use the fourfold cross-validation protocol to split the training/test subsets. Each data set is split to four subsets, and the four subsets are of the same size. We use each subset as the test set in turns, and use the remaining three subsets as the training sets. We first apply the proposed JICRTC method to the training set to learn the parameters of CNN and linear predictive model and apply the models to the test images to predict the tags. The image tags assignment scores are the outputs of the model, and the top-ranked tags are returned as the tags of the candidate image. We set the number of returned top-ranked images to 5 and 10, respectively, and report the precisions and recalls as the performance measures.

#### 3.2.2 Experimental results

We compare our method to the existing stat-of-the-art methods, including the methods proposed by Lin et al. [18], Wu et al. [39], Feng et al. [6], Lin et al. [17], and Li et al. [15]. The results are reported in Table 1 and Fig. 1. From the results reported in Fig. 1, the proposed method JICRTC outperforms the compared methods over all the three data sets on the four performance measures. This is a strong evidence of the advantage of the CNN model for the tag completion and annotation of images. Over the three data sets, it seems the compared methods obtain the best results over the Corel data set, and the worst results over the Flickr data set.

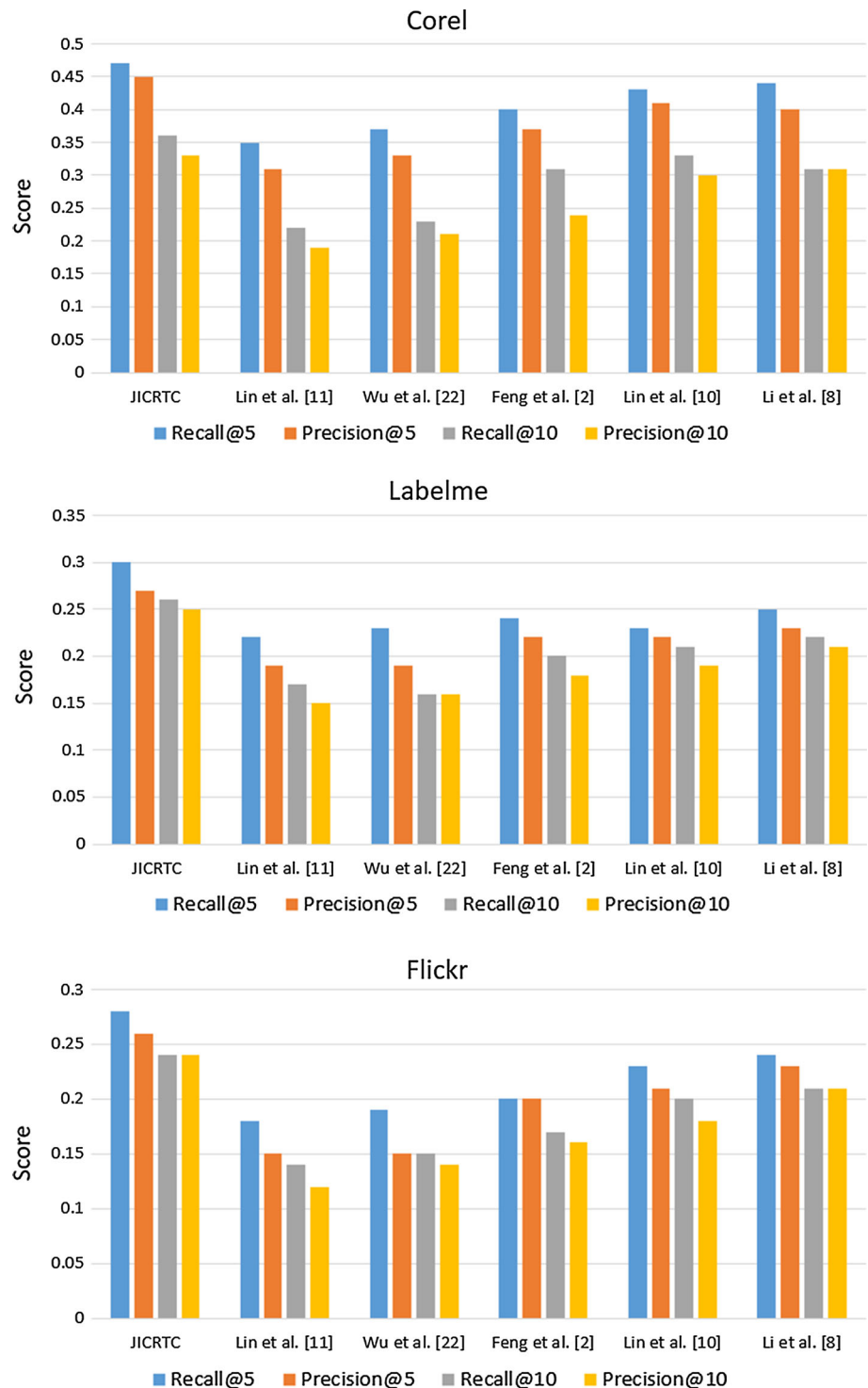
### 3.3 Image retrieval experiments

#### 3.3.1 Experimental setting

Then we evaluate the proposed method over the problem of tag-based image retrieval [16]. This problem uses tags as queries to retrieve the images from the database of images.



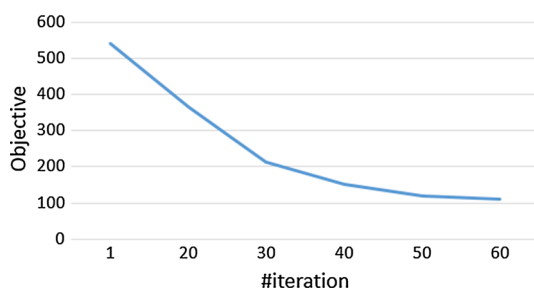
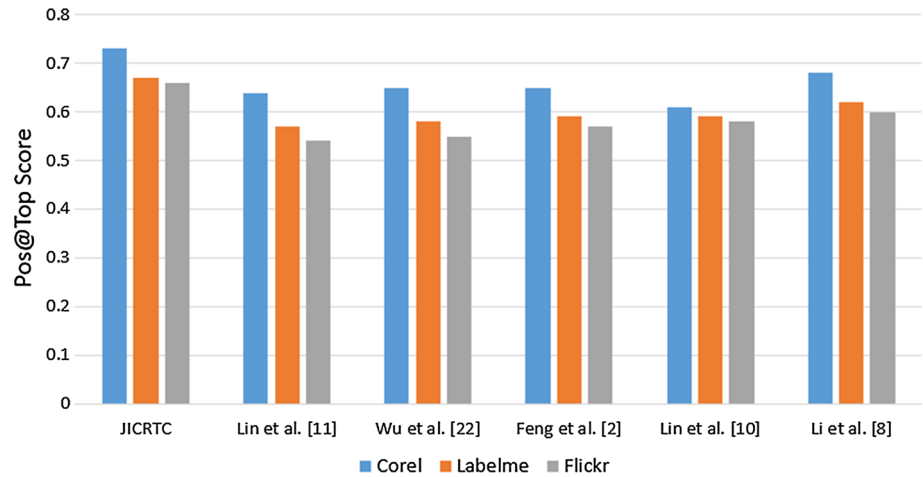
**Fig. 1** Figures of results of image annotation



In each data set of images, we remove some tags of the images to set up the image tag completion problem and then apply the image tag completion algorithm to complete the missing tags. Given a query tag, we return the images which has the same tag in the database as the retrieval results. We measure the retrieval performance by the positive at top (Pos@Top). The usage of this performance

measure is motivated by the works of Liang et al. [12, 16]. The works of Liang et al. [12, 16] show that the Pos@Top is a robust and parameter-free performance measure, which is suitable for most database retrieval problems. Following the works of Liang et al. [12, 16, 29], we adapt this performance measure to evaluate the results of the image retrieval problem in our experiments.

**Fig. 2** Figure of results of image retrieval experiments measured by Pos@Top



**Fig. 3** Convergence curve over Corel data set

### 3.3.2 Experimental results

The retrieval results of different methods are reported in Fig. 2. We can observe from this table that the proposed method JICRTC outperforms the other methods over all the three data sets. The Pos@Top values of JICRTC are higher than all the compared tag completion methods. This indicates that the convolutional-based tag completer works better than the other tag completer based on matrix completion, etc.

### 3.4 Convergence of the alternating gradient descent

We also plot the curve of the objective values with regard to increasing iterations for the alternating gradient descent algorithm. The curve of experiments over the Corel data set is shown in Fig. 3. According to Fig. 3, the algorithm converge after 40 iterations.

## 4 Conclusion and future works

In this paper, we proposed a novel image tag completion method. This method is based on the CNN model. We use the CNN model to represent the image and then predict the complete image tags from the CNN representations. The

complete tag assignment score vectors are also regularized by the visual similarities calculated from the CNN representations. We develop an iterative algorithm to learn the parameters of the CNN model, the linear predictive model, and the complete tags. The experiments of the problems of image annotation and image retrieval based on image tag completion over three benchmark data sets show the advantage of the proposed method.

In the future, we will extend our work of CNN model to other machine learning problems beside image tag completion, such as communication networks [48–51], human action recognition [19, 52–54], biometrics [35–38], medical imaging [11, 13, 22], computational mechanics [24, 34], and big data analysis [26, 28].

**Acknowledgements** This work was supported by the Natural Science Foundation of Hebei Province (D2015207008), Talent Training Project of Hebei Province (A201400215) and National High Technology Research and Development Program of China (863 Program No. 2014AA06A511), and the National Natural Science Foundation of China (41371358).

### Compliance with ethical standards

**Conflict of interest** The authors of this paper are stating no conflict of interest for the work described in this paper.

## References

1. Bhimani J, Yang J, Yang Z, Mi N, Xu Q, Awasthi M, Pandurangan R, Balakrishnan V (2016) Understanding performance of I/O intensive containerized applications for NVMe SSDs. In: 2016 IEEE 35th international performance computing and communications conference (IPCCC). IEEE, pp 1–8
2. Cai JF, Candès EJ, Shen Z (2010) A singular value thresholding algorithm for matrix completion. *SIAM J Optim* 20(4):1956–1982
3. Candès E, Recht B (2012) Exact matrix completion via convex optimization. *Commun ACM* 55(6):111–119
4. Candès EJ, Plan Y (2010) Matrix completion with noise. *Proc IEEE* 98(6):925–936

5. Charpenay V, Egyed-Zsigmond E, Kosch H (2016) Knowledge-driven reverse geo-tagging for annotated images. *Doc Numer* 19(1):83–102
6. Feng Z, Feng S, Jin R, Jain A (2014) Image tag completion by noisy matrix recovery. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8695(PART 7), 424–438
7. Fu J, Wu Y, Mei T, Wang J, Lu H, Rui Y (2016) Relaxing from vocabulary: robust weakly-supervised deep learning for vocabulary-free image tagging. In: *Proceedings of the IEEE international conference on computer vision*, 11–18-December-2015, pp 1985–1993. doi:[10.1109/ICCV.2015.230](https://doi.org/10.1109/ICCV.2015.230)
8. Gando G, Yamada T, Sato H, Oyama S, Kurihara M (2016) Fine-tuning deep convolutional neural networks for distinguishing illustrations from photographs. *Expert Syst Appl* 66:295–301
9. Gao H, Yang Z, Bhimani J, Wang T, Wang J, Sheng B, Mi N (2017) Autopath: harnessing parallel execution paths for efficient resource allocation in multi-stage big data frameworks. In: *26th international conference on computer communications*
10. Hou Y, Lin Z (2016) Image tag completion and refinement by subspace clustering and matrix completion. In: *2015 visual communications and image processing, VCIP 2015*, p 7457875
11. King DR, Li W, Squiers JJ, Mohan R, Sellke E, Mo W, Zhang X, Fan W, DiMaio JM, Thatcher JE (2015) Surgical wound debridement sequentially characterized in a porcine burn model with multispectral imaging. *Burns* 41(7):1478–1487
12. Li Q, Zhou X, Gu A, Li Z, Liang RZ (2016) Nuclear norm regularized convolutional max pos@top machine. *Neural Comput Appl* 1–10
13. Li W, Mo W, Zhang X, Squiers JJ, Lu Y, Sellke EW, Fan W, DiMaio JM, Thatcher JE (2015) Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging. *J Biomed Opt* 20(12):121305
14. Li X, Shen B, Liu BD, Zhang YJ (2016) A locality sensitive low-rank model for image tag completion. *IEEE Trans Multimed* 18(3):474–483
15. Li X, Zhang YJ, Shen B, Liu BD (2016) Low-rank image tag completion with dual reconstruction structure preserved. *Neurocomputing* 173:425–433
16. Liang RZ, Shi L, Wang H, Meng J, Wang JJY, Sun Q, Gu Y (2016) Optimizing top precision performance measure of content-based image retrieval by learning similarity function. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE
17. Lin Z, Ding G, Hu M, Lin Y, Sam Ge S (2014) Image tag completion via dual-view linear sparse reconstructions. *Comput Vis Image Underst* 124:42–60
18. Lin Z, Ding G, Hu M, Wang J, Ye X (2013) Image tag completion via image-specific and tag-specific linear sparse reconstructions. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, pp 1618–1625
19. Liu J, Chen C, Zhu Y, Liu W, Metaxas DN (2016) Video classification via weakly supervised sequence modeling. *Comput Vis Image Underst* 152:79–87
20. Lopes A, de Aguiar E, De Souza A, Oliveira-Santos T (2017) Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern Recognit* 61:610–628
21. Ma J, Wu F, Zhu J, Xu D, Kong D (2017) A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics* 73:221–230
22. Mo W, Mohan R, Li W, Zhang X, Sellke EW, Fan W, DiMaio JM, Thatcher JE (2015) The importance of illumination in a non-contact photoplethysmography imaging system for burn wound assessment. In: *SPIE BiOS*. International Society for Optics and Photonics, pp 93,030M–93,030M
23. Nie W, Liu A, Wang Z, Su Y (2016) Geo-location driven image tagging via cross-domain learning. *Multimed Syst* 22(4):395–404
24. Peng B, Liu Y, Zhou Y, Yang L, Zhang G, Liu Y (2015) Modeling nanoparticle targeting to a vascular surface in shear flow through diffusive particle dynamics. *Nanoscale Res Lett* 10(1):235
25. Qin Z, Li CG, Zhang H, Guo J (2016) Improving tag matrix completion for image annotation and retrieval. In: *2015 Visual communications and image processing, VCIP 2015*, p 7457871. doi:[10.1109/VCIP.2015.7457871](https://doi.org/10.1109/VCIP.2015.7457871)
26. Roemer J, Groman M, Yang Z, Wang Y, Tan CC, Mi N (2014) Improving virtual machine migration via deduplication. In: *2014 IEEE 11th international conference on mobile ad hoc and sensor systems (MASS)*. IEEE, pp 702–707
27. Russell BC, Torralba A, Murphy KP, Freeman WT (2008) LabelMe: a database and web-based tool for image annotation. *Int J Comput Vision* 77(1–3):157–173
28. Tai J, Liu D, Yang Z, Zhu X, Lo J, Mi N (2015) Improving flash resource utilization at minimal management cost in virtualized flash-based storage systems. *IEEE Trans Cloud Comput* 5(3):537–549
29. Wang J, Wang H, Zhou Y, McDonald N (2015) Multiple kernel multivariate performance learning using cutting plane algorithm. In: *2015 IEEE international conference on systems, man, and cybernetics (SMC)*, pp 1870–1875
30. Wang J, Wang T, Yang Z, Mao Y, Mi N, Sheng B (2017) Seina: A stealthy and effective internal attack in Hadoop systems. In: *2017 international conference on computing, networking and communications (ICNC)*. IEEE, pp 525–530
31. Wang J, Wang T, Yang Z, Mi N, Sheng B (2016) eSplash: Efficient speculation in large scale heterogeneous computing systems. In: *2016 IEEE 35th international performance computing and communications conference (IPCCC)*. IEEE, pp 1–8
32. Wang J, Zhou Y, Duan K, Wang J, Bensmail H (2015) Supervised cross-modal factor analysis for multiple model data classification. In: *2015 IEEE international conference on systems, man, and cybernetics (SMC)*, pp 1882–1888
33. Wang J, Zhou Y, Wang H, Yang X, Yang F, Peterson A (2015) Image tag completion by local learning. In: *International symposium on neural networks*. Springer, Berlin, pp 232–239
34. Wang S, Zhou Y, Tan J, Xu J, Yang J, Liu Y (2014) Computational modeling of magnetic nanoparticle targeting to stent surface under high gradient field. *Comput Mech* 53(3):403–412
35. Wang X, Guo R, Kambhamettu C (2015) Deeply-learned feature for age estimation. In: *2015 IEEE Winter conference on applications of computer vision*. IEEE, pp 534–541
36. Wang X, Kambhamettu C (2013) Gender classification of depth images based on shape and texture analysis. In: *2013 IEEE global conference on signal and information processing (GlobalSIP)*. IEEE, pp 1077–1080
37. Wang X, Kambhamettu C (2014) Leveraging appearance and geometry for kinship verification. In: *2014 IEEE international conference on image processing (ICIP)*. IEEE, pp 5017–5021
38. Wang X, Ly V, Lu G, Kambhamettu C (2013) Can we minimize the influence due to gender and race in age estimation? In: *2013 12th international conference on machine learning and applications (ICMLA)*, vol 2. IEEE, pp 309–314
39. Wu L, Jin R, Jain A (2013) Tag completion for image retrieval. *IEEE Trans Pattern Anal Mach Intell* 35(3):716–727
40. Wu Q, Boulanger P (2016) An unified image tagging system driven by image-click-ads framework. In: *Proceedings—2015 IEEE international symposium on multimedia, ISM 2015*, pp 369–372. doi:[10.1109/ISM.2015.12](https://doi.org/10.1109/ISM.2015.12)

41. Xia C, Hu J, Zhu Y, Naaman M (2015) What is new in our city? A framework for event extraction using social media posts. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, pp 16–32
42. Xia Z, Feng X, Peng J, Wu J, Fan J (2015) A regularized optimization framework for tag completion and image retrieval. *Neurocomputing* 147(1):500–508
43. Yang W, Chen Y, Liu Y, Zhong L, Qin G, Lu Z, Feng Q, Chen W (2017) Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. *Med Image Anal* 35:421–433
44. Yang X, Yang F (2016) Completing tags by local learning: a novel image tag completion method based on neighborhood tag vector predictor. *Neural Comput Appl* 27(8):2407–2416
45. Yang Z, Awasthi M, Ghosh M, Mi N (2016) A fresh perspective on total cost of ownership models for flash storage in datacenters. In: 2016 IEEE international conference on cloud computing technology and science (CloudCom). IEEE, pp 245–252
46. Yang Z, Tai J, Bhimani J, Wang J, Mi N, Sheng B (2016) GRem: dynamic SSD resource allocation in virtualized storage systems with heterogeneous IO workloads. In: 2016 IEEE 35th international performance computing and communications conference (IPCCC). IEEE, pp 1–8
47. Yang Z, Wang J, Evans D, Mi N (2016) Autoreplica: Automatic data replica manager in distributed caching and data processing systems. In: 2016 IEEE 35th international performance computing and communications conference (IPCCC). IEEE, pp 1–6
48. Zeng L, Chen H, Xiao Y (2011) Accountable administration and implementation in operating systems. In: 2011 IEEE global telecommunications conference-GLOBECOM 2011
49. Zeng L, Xiao Y, Chen H (2015) Accountable logging in operating systems. In: 2015 IEEE international conference on communications (ICC). IEEE, pp 7163–7167
50. Zeng L, Xiao Y, Chen H (2015) Auditing overhead, auditing adaptation, and benchmark evaluation in linux. *Secur Commun Netw* 8(18):3523–3534
51. Zeng L, Xiao Y, Chen H (2015) Linux auditing: overhead and adaptation. In: 2015 IEEE international conference on communications (ICC). IEEE, pp 7168–7173
52. Zhu Y, Tian Y, Mexatas D, Dollár P (2015) Semantic amodal segmentation. arXiv preprint [arXiv:1509.01329](https://arxiv.org/abs/1509.01329)
53. Zhu Y, Zhang S, Liu W, Metaxas DN (2014) Scalable histopathological image analysis via active learning. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 369–376
54. Zhu Y, Zhao X, Fu Y, Liu Y (2010) Sparse coding on local spatial-temporal volumes for human action recognition. In: Asian conference on computer vision. Springer, Berlin, pp 660–671