International Journal of Advanced Technology and Innovative Research
www.ijatir.org
ISSN 2348–2370
Vol.08,Issue.17,
October-2016,
Pages:3292-3299

Learning to Rank Image Tags with Limited Training Examples

E. SRAVYA[1], S. RAJESHWAR[2], B. RAMYA[3]
[1]PG Scholar, Dept of SE, Arjun College of Technology and Sciences, Hyderabad, India, E-mail: sravyaeaga@gmail.com.
[2]HOD, Dept of CSE, Arjun College of Technology and Sciences, Hyderabad, India.
[3]Assistant Professor, Dept of CSE, Arjun College of Technology and Sciences, Hyderabad, India.

Abstract: Many social image search engines are based on keyword/tag matching. This is because tag-based image retrieval (TBIR) is not only efficient but also effective. The performance of TBIR is highly dependent on the availability and quality of manual tags. Recent studies have shown that manual tags are often unreliable and inconsistent. Most studies cast image annotation into a multi-label classification problem. The main shortcoming of this approach is that it requires a large number of training images with clean and complete annotations in order to learn a reliable model for tag prediction. We address this limitation by developing a novel approach that combines the strength of tag ranking with the power of matrix recovery. Instead of having to make a binary decision for each tag, our approach ranks tags in the descending order of their relevance to the given image, significantly simplifying the problem. In addition, the proposed method aggregates the prediction models for different tags into a matrix, and casts tag ranking into a matrix recovery problem. It introduces the matrix trace norm to explicitly control the model complexity so that a reliable prediction model can be learned for tag ranking even when the tag space is large and the number of training images is limited. Experiments on multiple well-known image datasets demonstrate the effectiveness of the proposed framework for tag ranking compared to the state-of-the-art approaches for image annotation and tag ranking.

Keywords: Automatic Image Annotation, Tag Ranking, Matrix Recovery, Low-Rank, Trace Norm.

I. INTRODUCTION

How to accurately retrieve images from enormous collections of digital photos has become an important research topic. Content-based image retrieval (CBIR) addresses this challenge by identifying the matched images based on their visual similarity to a query image. However due to the semantic gap between the low-level visual features used to represent images and the high-level semantic tags used to describe image content, limited performance is achieved by CBIR techniques. To address the limitation of CBIR, many algorithms have been developed for tag based image retrieval (TBIR) that represents images by manually assigned keywords/tags. It allows a user to present his/her information needs by textual information and find the relevant images based on the match between the textual query and the assigned image tags. Recent studies have shown that TBIR is usually more effective than CBIR in identifying the relevant images since it is time-consuming to manually label images, various algorithms have been developed for automatic image annotation in this work; we focus on the tag ranking approach for automatic image annotation. Instead of having to decide, for each tag, if it should be assigned to a given image, the tag ranking approach ranks tags in the descending order of their relevance to the given image.

By avoiding making binary decision for each tag, the tag ranking approach significantly simplifies the problem, leading to a better performance than the traditional classification based approaches for image annotation. In addition, studies have shown that tag ranking approaches are more robust to noisy and missing tags than the classification approaches. Although multiple algorithms have been developed for tag ranking; they tend to perform poorly when the number of training images is limited compared to the number of tags, a scenario often encountered in real world applications. In this work, we address this limitation by casting tag ranking into a matrix recovery problem. The key idea is to aggregate the prediction models for different tags into a matrix. Instead of learning each prediction model independently, we propose to learn all the prediction models simultaneously by exploring the theory of matrix recovery, where trace norm regularization is introduced to capture the dependence among different tags and to control the model complexity. We show, both theoretically and empirically, that with the introduction of trace norm regularize, a reliable prediction model can be learned for tag ranking even when the tag space is large and the number of training images is small.

II. LITERATURE SURVEY

In this section we review the related work on automatic image annotation and tag ranking.

A. Automatic Image Annotation

Automatic image annotation aims to find a subset of keywords/tags that describes the visual content of an image. It plays an important role in bridging the semantic gap between low-level features and high-level semantic content of images. Most automatic image annotation algorithms can be classified into three categories (i) generative models that model the joint distribution between tags and visual features,

(ii) discriminative models that view image annotation as a classification problem, and (iii) search based approaches. Below, we will briefly review approaches in each category. Both mixture models and topic models, two well-known approaches in generative model, have been successfully applied to automatic image annotation. In, a Gaussian mixture model is used to model the dependence between keywords and visual features. Since a large number of training examples are needed for estimating the joint probability distribution over both features and keywords, the generative models are unable to handle the challenge of large tag space with limited number of training images. Discriminative models, views image annotation as a multi-class classification problem, and learn one binary classification model for either one or multiple tags. A structured max-margin algorithm is developed in to exploit the dependence among tags. One problem with discriminative approaches for image annotation is imbalanced data distribution because each binary classifier is designed to distinguish image of one class from images of the other classes. It becomes more severe when the number of classes/tags is large

B. Tag Ranking

Tag ranking aims to learn a ranking function that puts relevant tags in front of the irrelevant ones. In the simplest form, it learns a scoring function that assigns larger values to the relevant tags than to those irrelevant ones. In the Authors develop a classification framework for tag ranking that computes tag scores for a test image based on the neighbor voting. It was extended in to the case where each image is represented by multiple sets of visual features. Liu et al. utilizes the Kernel Density Estimation (KDE) to calculate relevance scores for different tags, and performs a random walk to further improve the performance of tag ranking by exploring the correlation between tags. Similarly, Tang et al. proposed a two-stage graph-based relevance propagation approach. In a two-view tag weighting method is proposed to effectively exploit both the correlation among tags and the dependence between visual features and tags. In a max-margin riffled independence model is developed for tag ranking. As mentioned in the introduction section, most of the existing algorithms for tag ranking tend to perform poorly when the tag space is large and the number of training images is limited.

III. EXISTING AND PROPOSED SYSTEMS

A. Existing System

Most automatic image annotation algorithms can be classified into three categories (i) generative models that model the joint distribution between tags and visual features, (ii) discriminative models that view image annotation as a classification problem, and (iii) search based approaches. In one of the existing system, a Gaussian mixture model is used to model the dependence between keywords and visual features. In another system, kernel density estimation is applied to model the distribution of visual features and to estimate the conditional probability of keyword assignments given the visual features. Topic models annotate images as samples from a specific mixture of topics, which each topic is a joint distribution between image features and annotation keywords.

B. Proposed System

In this work, we have proposed a novel tag ranking scheme for automatic image annotation. We first present the proposed framework for tag ranking that is explicitly designed for a large tag space with a limited number of training images. The proposed scheme casts the tag ranking problem into a matrix recovery problem and introduces trace norm regularization to control the model complexity. Extensive experiments on image annotation and tag ranking have demonstrated that the proposed method significantly outperforms several state-of-the-art methods for image annotation especially when the number of training images is limited and when many of the assigned image tags are missing.

Advantages of Proposed System:

- The proposed scheme casts the tag ranking problem into a matrix recovery problem and introduces trace norm regularization to control the model complexity.
- Extensive experiments on image annotation and tag ranking have demonstrated that the proposed method significantly outperforms several state-of-the-art methods for image annotation especially when the number of training images is limited and when many of the assigned image tags are missing.

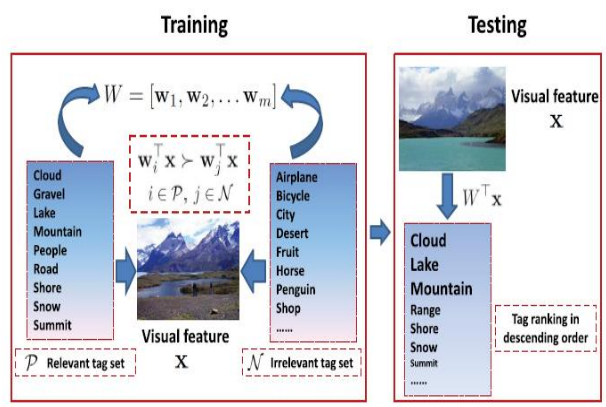


Fig.1. System Architecture

IV. EXPERIMENTAL RESULTS

In this section, we first describe our experimental setup, including image datasets, feature extraction, and evaluation measures. We then present three sets of experiments to verify the effectiveness of the proposed tag ranking approach, where the first experiment evaluates the performance of image annotation with limited training examples, the second experiment evaluates the performance of image annotation using training images with missing tags, and the last experiment examines the performance of the proposed algorithm for tag ranking. We finally evaluate the sensitivity of the proposed algorithm to parameter λ.

A. Image Datasets

To evaluate the proposed algorithm for image tagging, we conduct extensive experiments on five benchmark datasets for image annotation/tagging, including Corel5K, ESP Game, IAPRTC-12, Pascal VOC2007 and SUN Attribute. The first three image datasets are used to evaluate the performance of automatic image annotation, and the last two image datasets are used to evaluate tag ranking since a relevance score is provide for every assigned tag. Table I summarizes the statistics of the image datasets used in our study.

1. Corel5K: This dataset contains about 5, 000 images that are manually annotated with 1 to 5 keywords. The annotation vocabulary contains 260 keywords. A fixed set of 499 images are used as test and the rest images are used for training.

2. ESP Game: This dataset is obtained from an online game named ESP. We use a subset of around 20,000 images that are publicly available [12].

3. IAPRTC-12: This image collection is comprised of 19, 627 images, each accompanied with descriptions in multiple languages that were initially published for cross-lingual retrieval. Nouns are extracted from the textual descriptions to form the keyword assignments to images. We use the annotation results provided in [12].

TABLE I. Statistics for the Datasets Used in the Experiments. the Bottom Two Rows Are Given in the Format Mean/Maximum

	Corel5K	ESPGame	IAPRTC-12	Pascal VOC2007	SUNAttribute
No.of images	4,999	20,770	19,627	9,963	14,340
Vocabulary size	260	268	291	399	102
Tags per image	3.4/5	4.69/15	5.72/23	4.2/35	15.5/37
Image per tag	58.6/1,004	363/5,059	386/5,534	53/2,095	2,183/11,878

4. Pascal VOC2007: This dataset is comprised of 9, 963 images. We use the tags provided in that are collected from 758 workers using Amazon Mechanical Turk. As a result, for each image, we compute the relevance score for each assigned tag based on its votes from different workers. This relevance score will be used to evaluate ranking performance. On average, each image in this dataset is annotated by 4.2 tags from a vocabulary of 399 tags.

5. SUN Attribute: The SUN Attribute dataset contains 14,340 images and 102 scenes attributes spanning from materials, surface properties, lighting, functions and affordances, to spatial envelope properties. Similar to Pascal VOC2007, the annotated tags are collected from a large number of workers using the Amazon Mechanical Turk and therefore the votes from different workers can be used to compute the relevance score for different tags. For Corel5K, ESP Game and IAPRTC-12 image datasets, a bag-of-words model, based on densely sampled SIFT descriptors, is used to represent the visual content of images [8]. For Pascal VOC2007 dataset, we follow and extract three types of image features: Gist, color histogram, and bag-of-words histograms. For the SUN Attribute dataset, we follow and represent each image using four types of features: Gist, HOG2 × 2, a self-similarity, and geometric context color histogram. For simplicity, features provided in both Pascal VOC2007 and SUN Attribute datasets are directly combined by merging the feature vectors of each image. We subtract every element of each dimension of features by the mean of all elements in this dimension, and then divide by the standard variation of all elements in this dimension to normalize the features.

B. Evaluation Measures

Firstly, to evaluate the performance of automatic image annotation, we adopt the Average Precision (*AP@K*) and Average Recall (*AR@K*) as the evaluation metrics, which are defined as [4]:

$$AP@K = \frac{1}{n_t}\sum_{i=1}^{n_t}\frac{N_c(i)}{K} \tag{1}$$

$$AP@K = \frac{1}{n_t}\sum_{i=1}^{n_t}\frac{N_c(i)}{N_g(i)} \tag{2}$$

where K is the number of truncated tags, n_t is the number of test images, $N_c\ (i\ )$ is the number of correctly annotated tags for the i th test image, $N_g\ (i\ )$ is the number of tags assigned to the i th image. Both average precision and recall compares the automatically annotated image tags to the manually assigned ones. In addition, we use the Normalized Discounted Cumulative Gains at top K (*NDCG@K*) to measure the performance of different tag ranking approaches. It reflects how well a computer ranking agrees with the ideal (ground truth) ranking, with the emphasis on the accuracy of the top ranked items. It is defined as:

$$NDCG@K = \frac{1}{Z}\sum_{i=1}^{K}\frac{2^{rel\,(i)}-1}{log\,(1+i)} \tag{3}$$

where K is called truncation level, Z is the normalization constant to make sure the optimal ranking get the NDCG score of 1, and $rel(i\ )$ is the relevance score for the i –th ranked tag. Finally, for the proposed method, we set $\lambda = 1$ for all experiments except for the last one where we evaluate the impact of parameter λ.

C. Experimental (I): Automatic Image Annotation with Limited Number of Training Images

In the first experiment, we evaluate the annotation performance of the proposed image tagging method with limited training images. To this end, we randomly sample only 10% of images for training and use the remaining 90% for testing. Each experiment is repeated 10 times, each with a different splitting of training and testing data. We report the result based on the average over the trials. The following state-of-the-art approaches for image annotation are used as the baseline approaches in our evaluation:

- **Joint Equal Contribution Method (JEC) [5]:** It finds appropriate annotation words for a test image based on a k nearest neighbor classifier that used a combined distance measure derived from multiple sets of visual features.
- **Tag Propagation Method (Tag Prop) [8]:** It propagates the tag information from the labeled images to the unlabeled ones via a weighted nearest neighbor graph, where RBF kernel function is used for computing weights between images.
- **Multi-Class SVM Method (SVM):** It simply implements One-versus-All (OvA) SVM classifier for

each tag, and ranks the tags based on the output probability values.

- **Fast Image Tagging Method (Fast Tag):** It explores multi-view learning technique for multi-label learning. In particular, it defines two classifiers, one for each view of the data, and introduces a co-regularize in the objective function to enforce that the predictions based on different views are consistent for most training examples.
- **Efficient Multi-Label Ranking Method (MLR):** This approach explores the group lasso technique in multi-label ranking to effectively handle the missing class labels. It has been shown to outperform many multi-label learning algorithms.

The key parameter for Tag Prop is the number of nearest neighbors used to determine the nearest neighbor graph. We set it to be 200 as suggested by the original work [8]. For both SVM and MLR methods, linear function instead of RBF kernel function is adopted here for fair comparison. The optimal value for penalty parameter *C* in both methods is found by cross validation. Note that although Fast Tag method also adopts linear image feature classifiers, it incorporates non-linearity into the feature space as a preprocessing step. First, we show the comparison of average precision/recall for the first 5 returned tags for Corel5K dataset1 and the top 10 returned tags for both ESP Game and IAPRTC-12 datasets in Fig.2. It is not surprising to observe that with increasing number of returned tags, average precision declines while average recall improves. This is also called precision-recall trade-off, a phenomenon that is well known in information retrieval [4]. Second, we observe that our method significantly outperforms two nearest-neighbor based methods (JEC and Tag Prop) on the given datasets since the performance of nearest-neighbor based methods largely depend on the number of training samples. Specifically, at the truncation level of 4 (AP@4), we see our method yields around 5.6%, 4% and 8.76% improvement over Tag Prop on Corel5K, ESP Game and IAPRTC-12 dataset, respectively. In addition, the proposed method also outperforms multi-class SVM and Fast Tag algorithms, two classification based approaches, and MLR, a multi-label ranking approach. We attribute the success of the proposed approach to the special design of the proposed approach that nicely combines the ranking approach with trace norm regularization: it is the ranking approach that allows us to avoid making binary classification decision, and it is the trace norm regularization that makes our approach robust to the limited number of training examples.

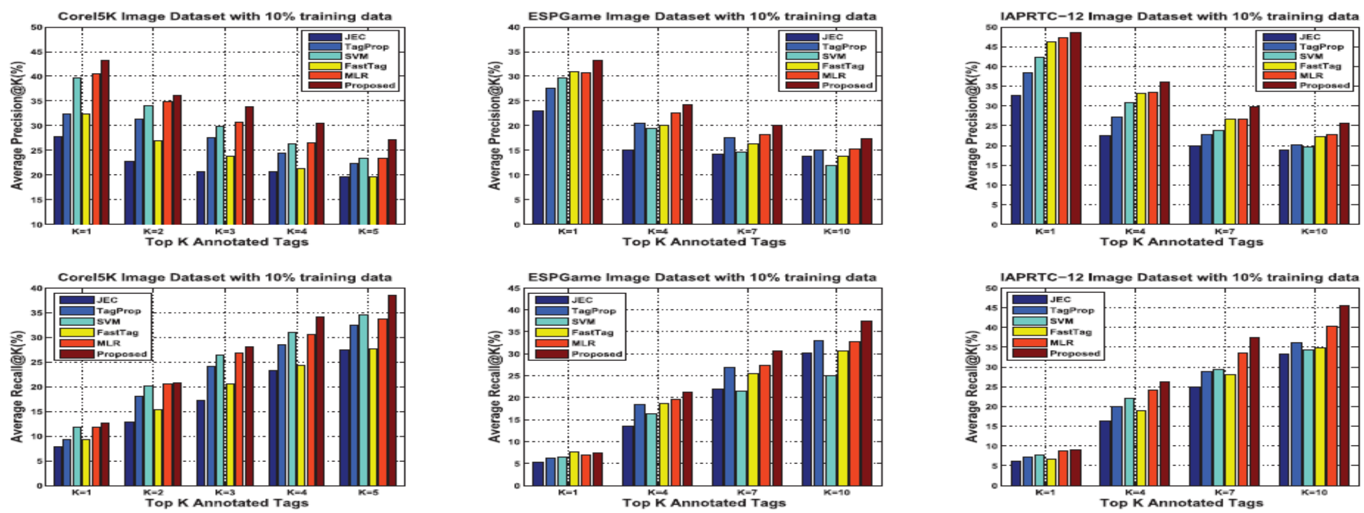


Fig.2. Average Precision and Recall for Automatic Image Annotation on Corel5K, ESP Game and IAPRTC-12 Datasets.

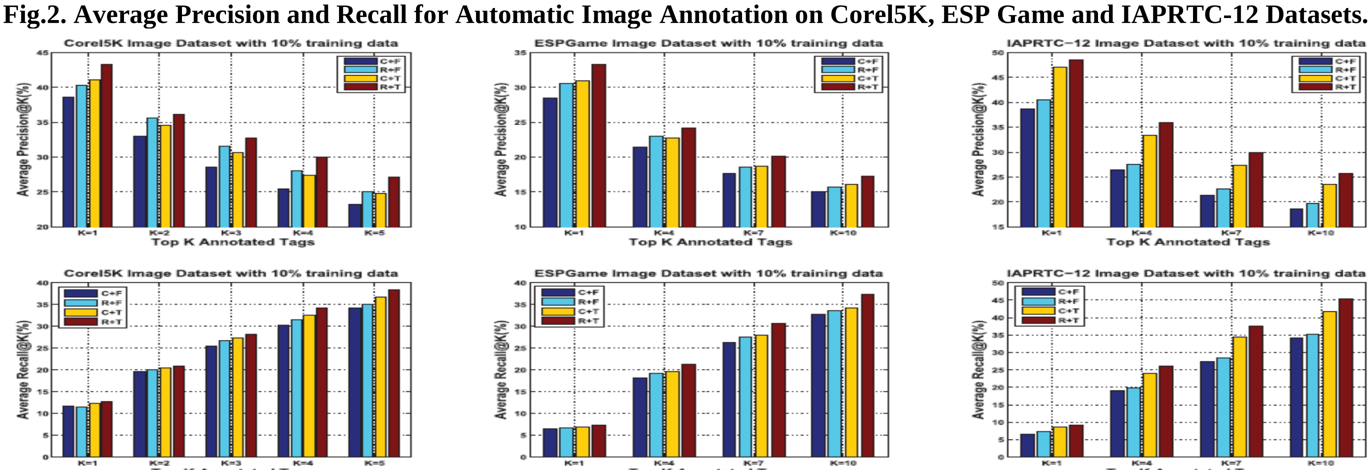


Fig.3. Evaluation of Different Loss Functions and Matrix Regularizes for Automatic Image Annotation.

To further investigate the advantages of the proposed approach, we evaluate the two components, i.e. ranking loss and trace norm regularization, separately. More specifically, we develop two baseline approaches, one replacing the ranking loss in the proposed framework with classification loss (**C+T**) and the other replacing trace norm with Frobenius norm for regularization (**R+F**). We also include the last baseline (**C+F**) that combines the classification loss with the Frobenius norm regularization. Following the naming convention here, we refer to the proposed approach as **R+T**. Figure 4 shows the prediction results that are based on 10% of images for training. We observe that the proposed framework outperforms the other three baselines, and the classification loss with Frobenius norm yields the worst performance among four approaches. Both observations indicate that the combination of ranking loss with trace norm regularization is important when the number of training images is limited. Finally, for the completeness of our experiment, we evaluate the performance of automatic image annotation by varying the number of training samples from 10% to 90%. Fig.4 summarizes the performance of *AP@*5 for three different datasets. We observe that annotation performance of all methods improves with increasing numbers of training images. We also observe that the improvement made by the proposed algorithm over the baseline methods reduces as the number of training images increase.

D. Experiment (II): Automatic Image Annotation with Incomplete Image Tags

In this experiment, we examine the performance of the proposed method when training image is partially annotated. To this end, similar to, we randomly select only 20%, 40%, and 60% of the assigned tags for training images. This setting allows us to test the sensitivity of the proposed method to the missing tags. Since the maximum number of annotated tags for the Corel5K dataset is 5, we only conduct the experiments on ESP Game and IAPRTC-12 datasets, where the maximum number of assigned tags are 15 and 23, respectively. The results of average precision

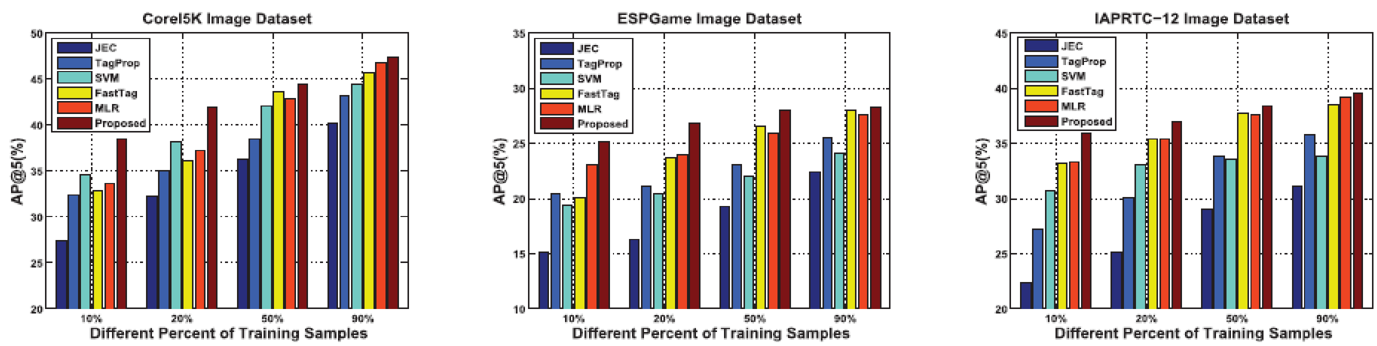


Fig.4. Average Precision at Rank 5 (*AP@*5) With Varied Numbers of Training Images for Datasets Corel5K, ESP Game, and IAPRTC-12.

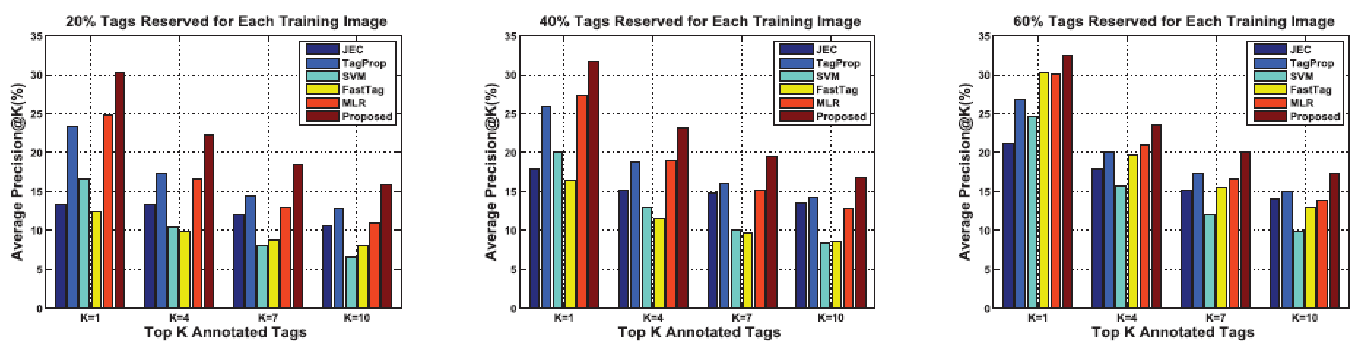


Fig.5. Performance Of Automatic Image Annotation On The ESP Game Dataset With Incomplete Image Tags, Where The Number Of Observed Tags Is Varied From 20%, 40% To 60%.

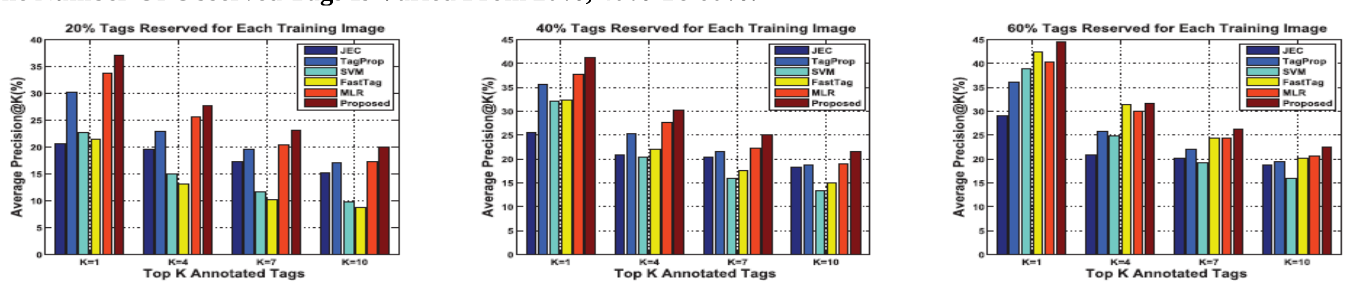


Fig.6. Performance Of Automatic Image Annotation On The IAPRTC-12 Dataset With Incomplete Image Tags, Where The Number of Observed Tags Is Varied From 20%, 40% To 60%.

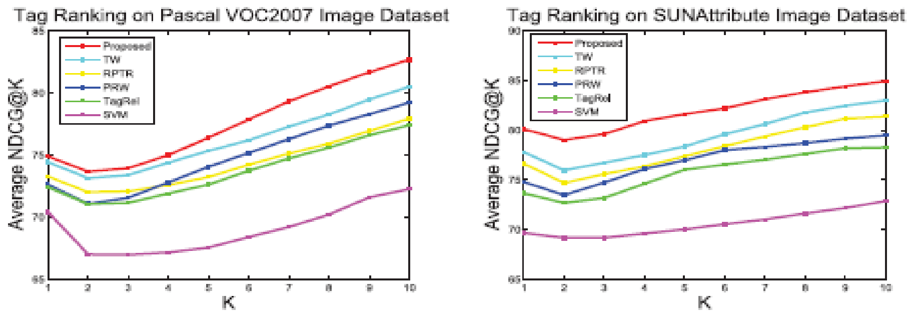
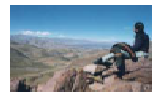
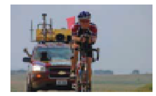


Fig.7. Performance of Tag Ranking, Measured By NDCG, For Dataset Pascal VOC2007 and SUN Attribute.

							
Ground Truth	**building door frame sky street window**	**adult child front house square woman**	**fog mountain roof stripe train**	**grandstand lawn roof round spectator stadium**	**hill landscape mountain rock woman**	**bike car cycling sky cyclist frame helmet jersey rack landscape roof short**	**boat green hill jacket lake orange mountain people range shore life sky summit**
JEC	*table front man house roof wall woman boy* **building** *chair*	*tourist* **woman child front** *wall classroom people room table building*	*front child classroom man sky table board car* **mountain** *wall*	*grass bush hill house* **lawn** *man people player short slope*	*sky* **landscape hill rock** *tree lake man* **mountain** *bay cliff*	**cyclist jersey helmet sky** *side road* **short bike** *sand* **cycling**	**mountain** *sea* **sky shore** *cloud man house* **lake** *rock tourist*
TagProp	*front wall table man rail house woman level* **building** *child*	*people wall tourist* **front woman** *table man side round classroom*	*man sky* **mountain** *front classroom wall child table tourist house*	*man short sky tree* **lawn** *house grass bush rock people*	*sky* **rock** *man* **landscape mountain** *jeep sea* **hill** *tree palm*	**cyclist sky helmet jersey short** *highway road side meadow* **bike**	**sea shore sky mountain lake** *cloud man house* **boat** *woman*
SVM	*rail level front house* **building** *table roof wall man* **sky**	*people side* **front** *wall tree* **woman** *tourist building sky* **house**	**mountain** *sky front man cloud wall child classroom table car*	*man short woman sky tree house grass wall bush people*	*jeep sky* **landscape mountain rock** *lake road* **hill** *man tree*	**sky** *side landscape highway* **car** *tree* **short** *road* **bike** *people*	**mountain lake** *sea* **sky** *cloud* **boat** *fountain* **hill** *house man*
FastTag	**sky** *front man stripe level tree people house rack pond*	*sky* **front** *man people rail* **house** *frame tree bedside centre*	*sky gate front man people* **train** *tree penguin portrait carpet*	*sky people front man tree* **lawn** *dirt short house tussock*	*sky front man tree house tussock kid people* **mountain** *formation*	**sky** *front tree man highway people* **short rack** *trouser pinnacle*	**sky** *man front lagoon tree* **shore** *fountain garden* **mountain** *cloth*
MLR	*house wall room cobblestone* **door** *front palm lamp tower bed*	**house** *tree landscape people jacket* **square** *building sweater shelf dog*	**mountain** *house palm flower landscape wall gate orange sky* **train**	*people house* **lawn round** *view green field building stand* **stadium**	**mountain** *grass* **landscape** *house shrub* **rock** *wall tussock slope snow*	**cyclist sky car** *tree* **jersey helmet cycling short** *sign sand*	**range shore** *field* **sky** *lagoon bay* **lake** *river tourist road*
Proposed	*wall* **building** *front* **window** *house* **door street** *room column balcony*	**front** *man* **house** *wall people* **woman child** *tourist room building*	**mountain** *sky* **train** *front cloud tourist door* **roof** *window wall*	**stadium lawn** *slope house* **grandstand** *road field tree player people*	**mountain landscape** *sky* **rock** *middle* **hill** *desert lake man cliff*	**cyclist jersey sky short** *road* **cycling helmet bike** *pole* **car**	**mountain sky lake range** *cloud* **shore summit** *sea stone* **hill**

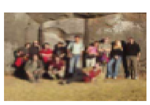
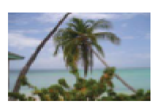
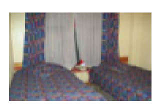
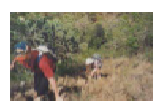
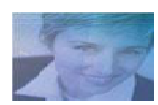
							
Ground Truth	**front group meadow stone tourist wall**	**boat man ocean sea water wave**	**balcony building sky fountain lamp palm square street**	**cloud leave palm plant sea sky**	**bed bedcover bedside curtain lamp room table wall**	**bush cactus man slope tree woman**	**face girl hair smile white woman**
JEC	*slope jersey man tree cyclist helmet cycling* **wall** *sky* **front**	**man** *green people grass picture photo family blue woman black*	**sky** *front table child grey tree wall classroom house man*	*man round tee shirt wall woman building child classroom jacket*	*front* **room** *fence gravel child cloud* **curtain** *door floor man*	**man** *forest* **grass bush** *cliff middle adult fence hill leave*	*website red* **woman** *black art box brown building church colors*
TagProp	*man jersey slope* **front** *tree* **wall** *mountain bush* **meadow** *people*	*airplane arrow red* **man** *blue sky* **water** *maga ine word* **ocean**	*front* **sky** *group people table man wall tree room house*	*man bench blanket wall table woman front child classroom* **sky**	*front child people* **wall room table** *sky fence man house*	**man** *hill grass forest* **tree** *pant sky* **bush** *rock middle*	*man cd red* **white** *black circle green logo* **face** *people*
SVM	*man* **front meadow wall** *cycling cyclist forest helmet tree bike*	**water man sea** *sand tree word blue car gray mountain*	*front group* **sky** *wall tree people table man house* **building**	*bench* **sky** *man wall people woman* **sea** *tree building tourist*	*front child* **wall room** *man people tourist fence gravel* **table**	*hill pant* **man** *sky* **tree** *grass* **bush** *forest meadow cliff*	*man logo red black cd* **face white** *circle green pink*
FastTag	*sky man tree bone people* **front** *shelter building pant house*	**water** *red sky* **ocean** *tree smile beak nose green cloud*	**sky** *front tree harbour man people house* **building** *pot* **fountain**	**sky** *jetty front man tree house corridor edge lagoon flagpole*	*sky tree table cloth flagpole* **wall** *man ravine* **room** *neck*	*sky* **man** *front* **tree** *people house grass bicycle lawn leave*	**hair face** *man* **smile** *nose* **woman girl** *circle glasses teeth*
MLR	*building trunk dog man* **tourist** *photo sky* **meadow** *sand paving*	**water ocean** *tree nose blue fire white roof cloud wing*	**building sky** *house tree tower* **palm street** *people sign man*	**sky** *woman house tree* **sea cloud** *salt lake orange tower*	**room bed wall** *tree house* **curtain** *night side painting* **bedside**	*grass* **tree** *leave sand rock dog* **bush** *house cliff trail*	**white** *pink square purple* **face** *bald nose man word blue*
Proposed	**front** *man* **wall** *people* **tourist** *helmet* **group** *woman photo* **meadow**	**ocean sea** *sky* **water** *sand blue red* **boat man** *tree*	**building sky lamp** *tree* **street** *front people tower* **palm** *grey*	**sky cloud** *man* **sea palm** *tree lake grass house shore*	**wall room curtain** *front* **bed** *painting* **bedside** *fence blanket* **table**	**tree bush man** *grass forest* **slope** *middle rock jungle path*	**white** *man* **hair** *black red* **face** *blue* **woman girl** *hat*

Fig.8. Examples of Test Images From Both The ESP Game And IAPRTC-12 Datasets With Top 10 Annotations Generated By Different Methods. The Correct Tags Are Highlighted By Bold Font Whereas The Incorrect Ones Are Highlighted By Italic Font.

for both datasets are reported in Fig.5 and Fig.6, respectively. It is not surprising to observe that annotation performance of all methods drops as the number of observed annotations decreases, indicating that the missing annotations could greatly affect the annotation performance. On the other hand, compared to the baseline methods, the proposed method is more resilient to the missing tags: on the ESP Game dataset,

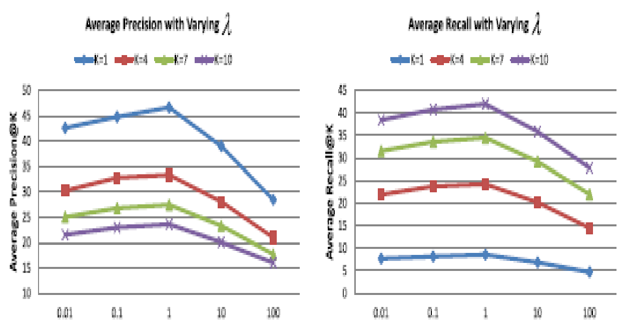


Fig.9. Average Precision and Recall Of The Proposed Method For The IAPRTC-12 Dataset With Varied Λ.

it only experiences a 1.41% drop in average precision when the number of observed tags decreases from 60% to 20%, while the other five baseline methods suffer from 4% to 8% loss for *AP@*4. This result indicates that the proposed method is more effective in handling missing tags. Fig.9 provides examples of annotations generated by different approach for the ESP Game and IAPRTC-12 datasets when only 20% of the assigned tags are observed for each training image. These examples further confirm the advantage of using the proposed approach for automatic image annotation when training images are equipped with incomplete tags.

E. Experimental (III): Tag Ranking

In this subsection, we evaluate the proposed algorithm for tag ranking. Given an image and a list of associated tags, the goal of tag ranking is to rank the tags according to their relevance to the image content. Both the Pascal VOC2007 and SUN Attribute datasets are used in this experiment since a relevance score is provided for each assigned tag. We randomly select 10% of images from each dataset for training, and use the remaining 90% for testing. We repeat the experiment 10 times and repeat the averaged NDCG. Using the votes collected from different workers, according to the settings, we create three levels of relevance score for each assigned tag: Most Relevant (score 4), Relevant (score 3) and Less Relevant (score 2). To make the problem challenging enough, for each image, we add three randomly sampled irrelevant tags (score 1) to the tag list. As a result, each tag list is comprised of labels with four relevance levels, ranging from the irrelevant category to the most relevant one. The following algorithms are used as the baselines in the evaluation of tag ranking. The first baseline uses the classification scores output from the one-vs-all SVM with linear function to rank tags. The second baseline, named Tag Rel, is based on the neighbor voting strategy for tag ranking, and the neighbor number is empirically set to 100. The third baseline, abbreviated as PRW, combines the probabilistic tag ranking approach with a random walk-based tag ranking approach, and we use the same parameter settings suggested by the origin work. The fourth baseline, named RPTR, is a relevance propagation tag ranking approach which combines both tag graph and image graph. The last baseline, which is known as TW, is a two-view tag weighting method that combines the local information both in tag space and visual space, and the trade-off hyper-parameters used in the algorithm is adopted as suggested by the origin work. Fig.7 reports NDCG values for the proposed algorithm and the four baseline methods on datasets Pascal VOC2007 and SUN Attribute. We can see that the proposed method significantly outperforms most baselines on both datasets. When evaluating with respect to the first five ranked tags (i.e. *NDCG@*5), we see our method yields about 9% improvement over SVM and 3.5% to 4% improvement over Tag Rel, RPTR and PRW on Pascal VOC2007 dataset. Furthermore, although our method only achieves around 2% improvement over TW, it is much more scalable than TW due to the fact that TW is essentially a transductive learning manner, which is not suitable for unseen test images. Similar improvements are observed on the SUN Attribute dataset. The experimental results prove that the proposed method is effective for tag ranking especially when the training samples are limited.

F. Experiment (IV): Sensitivity to Parameter λ

In this experiment, we examine the sensitivity of the proposed method to parameter λ using the dataset IAPRTC-12. In general, a larger λ will lead to a higher regularization capacity, and as a sequence, a larger bias and a smaller variance for the final solution. In order to understand how the parameter affects the annotation performance, we conduct the experiment by varying λ from 0.01 to 100 and measure average precision and recall for the learned annotation model, as shown in Fig.8. We observe that the proposed method yields the best performance when λ is around 1.

V. CONCLUSION

We have proposed a novel tag ranking scheme for automatic image annotation. The proposed scheme casts the tag ranking problem into a matrix recovery problem and introduces trace norm regularization to control the model complexity. A tag matrix completion method for image tagging and image retrieval we consider the image-tag relation as a tag matrix, and aim to optimize the tag matrix by minimizing the difference between tag based similarity and visual content based similarity. The proposed method falls into the category of semi-supervised learning in that both tagged images and untagged images are exploited to find the optimal tag matrix. Extensive experiments on image annotation and tag ranking have demonstrated that the proposed method significantly outperforms several state-of-the-art methods for image annotation especially when the number of training images is limited and when many of the assigned image tags are missing. In the future, we plan to apply the proposed framework to the image annotation problem when image tags are acquired by crowd sourcing that tend to be noisy and incomplete.

VI. REFERENCES

[1]Songhe Feng, Zheyun Feng, and Rong Jin, "Learning to Rank Image Tags With Limited Training Examples", EEE Transactions on Image Processing, Vol. 24, No. 4, April 2015.

[2]R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., vol. 40, no. 2, 2008, Art. ID 5.

[3]J. Wu, H. Shen, Y. Li, Z.-B. Xiao, M.-Y. Lu, and C.-L. Wang, "Learning a hybrid similarity measure for image retrieval," Pattern Recognit., vol. 46, no. 11, pp. 2927–2939, 2013.

[4]L. Wu, R. Jin, and A. K. Jain, "Tag completion for image retrieval," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 716–727, Mar. 2013.

[5]A. Makadia, V. Pavlovic, and S. Kumar, "Baselines for image annotation," Int. J. Comput. Vis., vol. 90, no. 1, pp. 88–105, 2010.

[6]J. Tang, R. Hong, S. Yan, T.-S. Chua, G.-J. Qi, and R. Jain, "Image annotation by kNN-sparse graph-based label propagation over noisily tagged web images," ACM Trans. Intell. Syst. Technol., vol. 2, no. 2, pp. 1–16, 2011.

[7]J. Tang, S. Yan, R. Hong, G.-J. Qi, and T.-S. Chua, "Inferring semantic concepts from community-contributed images and noisy tags," in Proc. 17th ACM Int. Conf. Multimedia, 2009, pp. 223–232.

[8]M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tag Prop: Discriminative metric learning in nearest neighbor models for image auto-annotation," in Proc. IEEE 12th Int. Conf. Comput. Vis., Sep./Oct. 2009, pp. 309–316.

[9]W. Liu and D. Tao, "Multi-view Hessian regularization for image annotation," IEEE Trans. Image Process., vol. 22, no. 7, pp. 2676–2687, Jul. 2013.

[10]S. Zhang, J. Huang, Y. Huang, Y. Yu, H. Li, and D. N. Metaxas, "Automatic image annotation using group sparsity," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 3312–3319.

[11]Y. Verma and C. V. Jawahar, "Image annotation using metric learning in semantic neighborhoods'," in Proc. 12th Eur. Conf. Comput. Vis., 2012, pp. 836–849.

[12]Z. Feng, R. Jin, and A. Jain, "Large-scale image annotation by efficient and robust kernel metric learning," in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 1609–1616.

Author's Profile:

E.Sravya, PG Scholar, Dept of SE, Arjun College of Technology and Sciences, Hyderabad, India,
E-mail: sravyaeaga@gmail.com.

S. Rajeshwar, HOD, Dept of CSE, Arjun College of Technology and Sciences, Hyderabad, India.

B. Ramya, Assistant Professor, Dept of CSE, Arjun College of Technology and Sciences, Hyderabad, India.