

Multi Label Image Annotation for Weakly Labeled Images using Discriminative Dictionary

Neethu Lakshmi K M
M Tech Student
College of Engineering Kidangoor
Kerala, India

Nisha C A
Assistant Professor
College of Engineering Kidangoor
Kerala, India

Abstract— Image annotation is one of the classical problem in computer vision. With the popularity of online photo sharing websites automatic image annotation is an interesting service for users. Image annotation tells the objects in an image. This paper presents a multi label image annotation framework for weakly labeled training images by learn the dictionary embedded with semantic label. This achieves the discriminative representation of labels. For this first extract low level features and labels from image and construct a label graph. Then detect the label group by using the subgraph seeking problem. Learn multiple dictionaries to make distinction between the feature representation of each label. Fisher discriminative dictionary learning is used for that. Expand the semantic relationship between the visual words in the dictionary and labels based on the mutual information between them. Find the reconstruction coefficients based on the sparse reconstruction framework. In the test stage compute the reconstruction coefficients and score to the label list.

Keywords—Multi-Label Image Annotation, Discriminative Dictionary Learning, Inter Label Correlation Learning, Sparse Reconstruction.

I. INTRODUCTION

Nowadays digital images are easily accessible because of the emerging technologies in digital photography, networking etc. Photo and video hosting websites such as Flickr, Instagram, YouTube, Picasa are popular today. For example, Flickr, the photo and video sharing website reported that it was sharing more than 6 billion images and they give tags to describe the content of that images. For an effective retrieval mechanism to obtain the content of that data, automatic image annotation received much research interest. The different types of automatic image annotation are single labeling annotation, multi-labeling annotation and web based image annotation[1].

In the case of single labeling image annotation, low level visual features are extracted from the image and that is given to a conventional classifier which gives yes or no vote as its output. That output obtained as the semantic concept is used for image annotation. Machine learning tools used for this are support vector machine, artificial neural network and decision tree.

In multi-labeling annotation an image is annotated with multiple concept/categories. In the case of multi label

annotation it is important to understand the semantic context in large scale weakly labeled image set. The two groups in this are parametric and non-parametric. The steps in parametric annotation are feature extraction, dictionary learning, feature encoding, pooling and classifier training. Image annotation in non-parametric method consider as a sparse reconstruction problem.

In multi label sparse coding framework for automatic image annotation includes three components [2]. They are 1) feature extraction based on all patches from the training images 2) label sparse coding for feature extraction 3) Sparse coding for multi label data to proliferate the multi labels of the training images to the test image with reconstruction coefficients.

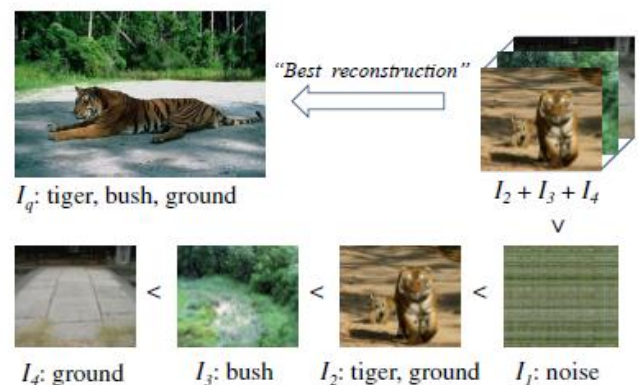


Figure 1 One to all sparse reconstruction

The two main limitation of existing image annotation technique are 1) discerning image representation 2) correlations between the co-occurrence labels. In this paper a dictionary learning method is used for embedding the semantic labels. The basic steps in this paper are, first extract the low level features and concepts/labels from the weakly labeled training images and compute a label matrix. Then identify the label groups from the computed matrix. Train dictionaries for each label group. After training dictionaries for each label group, the next step is dictionary label expansion. Then the above said sparse coding framework is used to create the reconstruction coefficients. With this values and dictionary labels the test image is annotated.

II. METHODOLOGY

In this paper the training phase consists of five steps. They are feature extraction, label discovery, label embedding dictionary learning, semantic label correlation learning and sparsity based reconstruction

A. Label Group Detection

We first construct the graph $O = \langle V, E \rangle$ and whose vertices are the labels from the training images. The label weighted matrix W associated with O is: $W_{ij} = 1$ if label i and label j never appear together in the same image and $W_{ij} = 0$ otherwise. We find subgraphs from this graph O [3][4]. Suppose we extracted G label groups and each group contains labels that form the subgraph. There are some constraints to remove the redundant groups, we merge two groups that have a large amount of overlap labels. This is repeated until no merging occurs. Computational efficiency drives to restrict the number of labels in each group.

B. Dictionary Learning

This part contains the dictionary learning by embedding the semantic label. We train the dictionary for each group in parallel because the label groups are independent with each other. The selection of dictionary is important for the success of sparse representation model. The KSVD dictionary learning algorithm has achieved great success in the case of image restoration, but they are not discriminative for image classification [5]. In this paper we follow the Fisher Discrimination Dictionary learning framework is used [6]. In FDDL we obtain both the representation residual and the representation coefficients and these are discriminative for the query. In this dictionary learning we obtain dictionaries for each group by minimizing the intra class scatter matrix and maximizing the inter class scatter matrix.

Steps in dictionary learning

1. Input: Training data
2. Initialize dictionary D
3. Update the reconstruction coefficients X
4. Update dictionary D with X fixed
5. Repeat step 2 until maximum number of iterations reached or the objective function values in iterations are close enough
6. Output: Coefficients X and dictionary D

Discriminative Dictionary Learning

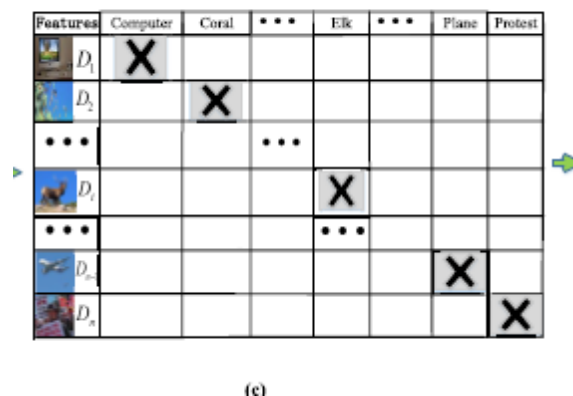


Figure 2. Discriminative dictionary representation

C. Semantic Label Correlation Learning

The association between co-occurrence labels should be transferred to compute the relatedness between labels and the visual words. A trace regularization based multi task learning framework is adopted to compute the correlation based on the reconstruction coefficients and we get the correlation parameter.

Before the dictionary label expansion the visual word is related only with one label. After the expansion the visual word is related with multiple labels. The inter-label correlation matrix helps us to understand the interdependency between concepts and co-occurrence of labels in the same image [7].

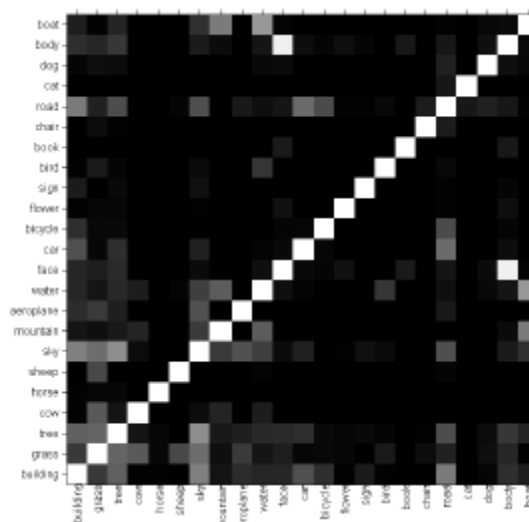


Figure 3. The inter label correlation matrix

We can construct a label correlation matrix from the training image labels. Suppose two labels p and q , we calculate the conditional probabilities $P(p/q)$ and $P(q/p)$, the harmonic mean and finally define the correlation matrix. The brighter blocks show the stronger correlation between labels. In the above figure face and body shows the strong correlation and a weak correlation between sky and bird.

D. Image Reconstruction

The corresponding dictionaries are selected by computing the discriminating score. After selecting the corresponding dictionary, the image annotation could be done based on the reconstruction coefficients obtained from the dictionary learning. The relevant label groups will give the non zero reconstruction coefficients

F. Label Propagation

At the time of testing, we extract the visual low level features from the test image and calculate the reconstruction coefficients and transfer the dictionary label in to the query image.



III. DATASET

The three commonly used image annotation datasets NUS-WIDE-LITE, Corel 5K, IAPR-TC12 are used in this proposed approach. In the case of NUS-WIDE-LITE dataset, the dataset is divided into training and testing part and the training part is used to construct the dictionary [8].

IV. COMPUTATIONAL COMPLEXITY

This proposed model consists of one offline training part and online testing part. The complexity of the training part is based on the dictionary learning and updating the dictionary items. At the time of testing the computational complexity is based on the group sparse reconstruction.

V. CONCLUSION

To bridge the gap between training and testing data in the case of weakly supervised setting this paper propose dictionary representation by embedding the semantic label. The main contribution of this paper lies in the dictionary representation and the inter label correlation among multiple labels and obtain a solution for inconsistent label combination. This method also has some limitation. They are size of the dictionary for each label, incomplete training labels and unbalanced training data distribution.

REFERENCES

- [1] Dengsheng Zhang, Md. Monirul Islam, Guojun Lu, "A review on automatic image annotation techniques" Gippsland School of Information Technology, Monash University, Churchill, Vic. 3842, Australia.
- [2] C. Wang, S. Yan, L. Zhang, and H.-J. Zhang, "Multi-label sparse coding for automatic image annotation" Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 1643–1650.
- [3] X. Chen, X.-T. Yuan, Q. Chen, S. Yan, and T.-S. Chua, "Multilabel visual classification with label exclusive context," in Proc. ICCV, Nov. 2011, pp. 834–841.
- [4] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," Pattern Recognit., vol. 42, no. 2, pp. 218–228, Feb. 2009.
- [5] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in Proc. IEEE Conf. CVPR, Jun. 2011, pp. 1697–1704.
- [6] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse Representation Based Fisher Discrimination Dictionary Learning for Image Classification," Int. J. Comput. Vis., vol. 109, no. 3, pp. 209–232, 2014.
- [7] Xiangyang Xue¹, Wei Zhang¹, Jie Zhang¹, Bin Wu¹, Jianping Fan², Yao Lu¹, "Correlative Multi-Label Multi-Instance Image Annotation".
- [8] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "NUS-WIDE: A real-world Web image database from National University of Singapore," in Proc. ACM CIVR, 2009, p. 48.