International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 12, December 2017

Recognition, Description and Localisation of Images Using WS-MRF-SIBP

[1] P.Subashree Kasi Thangam
[1] Assistant Professor, Department of Computer Sciennce and Engineering,HolyCross Engineering College, Vagaikulam.

Abstract:-- **This paper proposed a WS-MRF-SIBP model to learn the weakly labeled images. The object, attribute and background appearances, object – attribute association and their locations from realistic weakly labeled images including multiple objects with cluttered background are learned from the images. Then a novel weakly supervised Bayesian model is formulated to learn and exploit spatial coherence and factor co-occurrence. Once learned from weakly labeled data, this model performs various tasks including semantic segmentation, image description and image query.**

Keywords:-- **Weakly supervised learning, object-attribute association, semantic segmentation, non-parametric Bayesian model, Indian Buffet Process**

1. INTRODUCTION

The human visual system generate rich description of the scene content in image. Such description has nouns and adjectives, based on the objects and their associated attributes respectively. Example : “A red flowers in a tree”. The humans can effortlessly trace each object in the scene. This ability is the key objectives of computer vision research.

The research include number of fundamental computer vision problems such as,
1. Recognizing objects in the scene (Object annotation)
2. Describing the objects using their attributes (Attribute prediction and association)
3. Localizing and delineating the objects (Object detection and semantic segmentation)

In Conventional Supervised Approach, the images are strongly labeled with object bounding boxes or segmentation masks, and associated attributes. From this, object detectors and attributes classifiers are learned. This is independent learning approach.

Given a new image,
1. The learned object detectors are first applied to find object locations
2. Then the attribute classifiers are applied to produce the object descriptions.
This method has some limitations :

1. This is not scalable due to the lack of fully labeled training data, inspite of the large number of object classes, distinguishable to humans, attributes to describe them and object – attribute combinations.
2. Only closely related tasks are tackled in a single model.
Indian Buffet Process (IBP) explain the multiple factors that simultaneously co-exist to define the appearance of a particular image or super pixel. The multiple factors can be object, its particular texture and color attributes. This is an infinite factor model. So, it can automatically discover and model latent factors not defined by the provided training data labels, corresponding to latent object/attributes and structured background ‘stuff’. Eg : Sky, Road

The Conventional IBP is unsupervised model. It is a flat model. So, it is applied to either super pixels or images, but not both. So, it can’t be directly applied to the proposed system.

The Standard IBP is unable to exploit cues critical for segmentation and object-attribute association by modeling the correlation of factors within and across super pixels in each image. The within-super pixel correlation captures the co-occurrence relations. Eg : Cars are metal and bananas are yellow. The across-super pixel correlation describes that neighbouring super pixels have similar labels.

An object recognition system finds objects in the real world from an image of the world, using object models which are known a priori. This task is surprisingly difficult. Humans perform object recognition effortlessly and instantaneously. Algorithmic description of this task for implementation on machines has been very difficult. There are different steps in object recognition and introduce some techniques that have been used for object recognition in many applications. Different types of recognition tasks that a vision system may need to perform. It analyze the complexity of these tasks and present approaches useful in different phases of the recognition task. The object recognition problem can be defined as a labeling problem based on models of known objects. Formally, given an image containing one or more objects of interest (and background) and a set of labels corresponding to a set of models known to the system, the system should assign correct labels to regions, or a set of regions, in the image. The object recognition problem is closely tied to the segmentation problem: without at least a partial recognition of objects, segmentation cannot be done, and without segmentation, object recognition is not possible.

The object recognition system has the following components to perform the task :

- **Model database** - The model database contains all the models known to the system. The information in the model database depends on the approach used for the recognition. It can vary from a qualitative or functional description to precise geometric surface information. A feature is some attribute of the object that is considered important in describing and recognizing the object in relation to other objects. Size, color, and shape are some commonly used features.

- **Feature detector**- The feature detector applies operators to images and identifies locations of features that help in forming object hypotheses. The features used by a system depend on the types of objects to be recognized and the organization of the model database.

- **Hypothesis**-The hypothesizer assigns likelihoods to objects present in the scene. This step is used to reduce the search space for the recognizer using certain features.

- **Hypothesis verifier**-The verifier then uses object models to verify the hypotheses and refines the likelihood of objects. The system then selects the object with the highest likelihood, based on all the evidence, as the correct object.

The objects are described using elementary components consisting of segments, facets and text. However, FORTRAN 3D does not ensure the automatic elimination of hidden parts (to be defined by the user). The description is done in the object's own coordinate system. The object, thus described, then undergoes linear transformations (translation, rotation, scaling etc) defined by the user in order to position it. A key goal is to describe objects and to learn from descriptions. Two objects with the same name (e.g., “car”) may have differences in materials or shapes, and able to recognize and comment on those differences. Further, that encounter new types of objects. Even though that can’t name them, that is to be able to say something about them. Finally, learn about new objects quickly, sometimes purely from a textual description**.**

A method for localising multi colored objects in images using MNS signatures. The idea is that an object of interest is likely to appear in the part of the image with the highest population of neighbourhoods with different colors, similar to the colors of the object. The object representation is computed from one or more example image regions. The proposed algorithm was tested on localising objects in two video sequences showing sport events. The results obtained were acceptable, given the appearance variations of the sought objects and the relatively poor quality of the test sequences. A comparison with a well known localisation method called histogram back projection was also performed. In both experiments, the MNS localisation success rate was higher than that obtained using back projection. It require all the image neighbourhoods with colors similar to the sought object lie in a relatively small image region. A template is placed over selected image pixels and the location with most neighbourhoods similar to the object inside the template region is returned as the object’s expected position.

In this paper, all the objects, attributes and their relationships, and the locations of objects are learned. Using the weakly labeled data, and once the objects are learned, the proposed model performs several tasks such as semantic segmentation, image description and image query.

The organization of the rest of this paper is as follows. A short review on the related work is provided in Section II and the overview of the proposed scheme is described in Section III. Feature Extraction is described in Section IV. In Sections V, the Object Prediction, Object Attribute Association and Object Localisation are introduced. Semantic Segmentation and Image Annotation are described in Section VI. Finally, the conclusion of the paper is given in Section VII.

II RELATED WORK

A. Learning object-attribute associations

Attributes describe objects, people, clothing, scenes, faces and video events. But, most studies, learn and guess object and attribute models separately. Some recent studies learned the object-attribute association explicitly. In existing system, only train and test on clear data is performed. That is, the images having a single dominant object, assuming object-attribute association is known at training. And they allocate one attribute per object exactly.

B.Weakly Supervised Semantic Segmentation

Most existing semantic segmentation models are fully supervised. They need pixel-level labels. The Convolutional Neural Networks work either in a fully supervised fashion or a weakly supervised fashion. But, these methods need a large-scale annotated dataset to train or pre-train a deep CNN model for feature representation. Eg : ImageNet. In two-or multi-class co-segmentation, the task is to segment shared objects from a set of images. The co-segmenttaion doesn't need image labels, and it assumes common objects across multiple training images. By using these methods, a model is learned to segment unseen and unlabelled test images, rather than segmenting training images as in co-segmentation. All existing system focus only on object labels (nouns).

C.Weakly Supervised Learning

The existing weakly supervised learning (WSL) methods are dominated by discriminative models. The several existing models are Conditional Random Field (CRF), Label Propagation, Clustering Models and Discriminative Multi-Instance Learning (MIL) models. These methods agree many high performance recognition and annotation.

III PROPOSED MODEL

The proposed non-parametric Bayesian model describe the images from weak image-level object and attribute tagging. In this model, each image is divided into super pixels. Given a set of images, the super pixels shared by all the images with a particular label, and the super pixels corresponding to unannotated background are learned. The appearance of objects, attributes and their associations are learned. This model outperforms weakly supervised alternatives. It performs variety of tasks including semantic segmentation, automatic image annotation and retrieval based on object-attribute associations.

The Procedure of the proposed model is given as follows:
Input : A given image.
Output : Final annotated image.
Steps :

1) Low Level Feature Extraction.
2) Clustering of Feature Space using EM Algorithm.
3) Object Prediction.
4) Object Attribute Association.
5) Localizing the Objects and Semantic Segmentation
6) Annotation of Images.

Object prediction is used to extract features from an image. There are variations like angle, size, perspective and illumination. The same objects look very different to a machine to its presented with a different perspectives. Object-attribute association is used for image retrieval. Humans describe images they tend to use combinations of nouns and adjectives, corresponding to objects and their associated attributes. This is achieved by introducing a novel weakly supervised non-parametric Bayesian model. This model generalizes the non-parametric IBP. It defines a probability distribution over equivalence classes of binary matrices with finite rows and unbounded columns. It represent objects using a potentially infinite array of features. Weakly supervised latent dirichlet allocation approach used for object localisation. Localize each class of objects independently from other classes. In Multi-instance multi-label, three types of cues for object localization. Combine these three cues for WSOL which employed a conditional random field (CRF). Simultaneously locating objects in images and learning their appearance using only weak labels indicating presence/absence of the object. Weakly supervised dual clustering is used for semantic segmentation. WSDC task to cluster super pixels and assign a suitable label to each cluster. Discriminative clustering outputs to be consistent with the outputs of spectral clustering. Then impose weakly supervised constraints during dual clustering process which can assign labels to cluster. Image annotation has three tasks. First task is free annotation. The second task is to annotate the images using object names. The third task is to annotate the images using object locations.

The architecture diagram of the proposed system is given as,

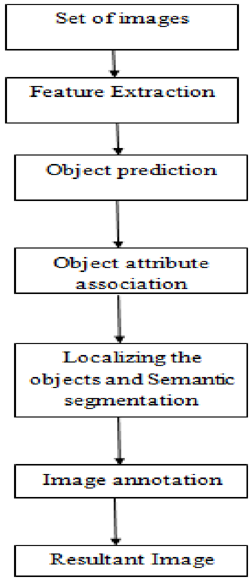

Fig. 1 Architecture diagram for proposed model

IV. FEATURE EXTRACTION

Each image is divided into super pixels. From each image, low level features are extracted. The color features are extracted by extracting the first four moments of three channels of CIE Luv color space. Then the Gabor texture feature is extracted, by using six scales and six orientations of Gabor transformation to extract their means and standard deviations. And the two normalized edge direction histogram feature, to represent the shape of the images.

The low-level feature vectors are clustered using mixture models that model the data by a number of Gaussian distributions. A cluster corresponds to a set of distributions, one for each dimension of the dataset. Each distribution is described in terms of mean and standard deviation. A probabilistic approach to assigning feature vectors to clusters is used.

For 1-D datasets, a mixture is a set of c Gaussian probability distributions, representing c clusters. The parameters of a mixture model are determined by the expectation maximization (EM) algorithm.

The EM algorithm is used to estimate the maximum likelihood L of θ given a set of features $\{x1, . . . , xN \}$. The algorithm results in a set of distributions, a vector of pairs of means μ and standard deviations σ, each of which corresponds to a feature, and outputs the size of the cluster (the number of vectors that belong to the cluster). The vector of means μ of the distributions for every feature represents the centroid of the cluster. The clusters resulting from the EM algorithm are considered as patterns extracted from the image database

V OBJECT PREDICTION, OBJECT ATTRIBUTE ASSOCIATION AND OBJECT LOCALISATION

Object prediction is used to extract features from an image. There are variations like angle, size, perspective and illumination. The same objects look very different to a machine to its presented with different perspectives.

Attribute centric approach is used for object prediction. It focus learning object attributes which can be semantic or not. Semantic attributes describe parts, shapes and materials. Object prediction allows us to describe objects and to identify them based on textual descriptions.

First select features that can predict attributes within object class and use only those to train attribute classifier. Two objects with the same name may have differences in materials or shape. Semantic attributes have advantage that they can be used to verbally describe new types of objects and to learn from textual description.

Object-attribute association is used for image retrieval. Humans describe images they tend to use combinations of nouns and adjectives, corresponding to objects and their associated attributes.

This is achieved by introducing a novel weakly supervised non-parametric Bayesian model. This model generalizes the non-parametric IBP(Indian Buffet Process). It defines a probability distribution over equivalence classes of binary matrices with finite rows and unbounded columns. It represent objects using a potentially infinite array of features.

Weakly supervised latent dirichlet allocation approach used for object localisation. Localize each class of objects independently from other classes. In Multi-instance multi-label, three types of cues for object localization.

- Saliency
- Intra-class
- Inter-class

Saliency- A region containing an object should look different from the majority of (background) regions.

Intra-class- A region containing an object should look similar to the regions containing the same class of objects in other images.

Inter-class- A region should look dissimilar to any regions.

Combine these three cues for WSOL which employed a conditional random field (CRF). Simultaneously locating objects in images and learning their appearance using only weak labels indicating presence/absence of the object.

VI SEMANTIC SEGMENTATION AND IMAGE ANNOTATION

Weakly supervised dual clustering is used for semantic segmentation. WSDC task to cluster super pixels and assign a suitable label to each cluster. It perform two clustering

- Spectral clustering
- Discriminative clustering

A spectral clustering is defined over the super pixels of all images to group the visually similar ones together.

A discriminative clustering outputs to be consistent with the outputs of spectral clustering. Then impose weakly supervised constraints during dual clustering process which can assign labels to cluster.

Image annotation has three tasks.

- Free annotation
- Annotation given object names
- Annotation given locations

Free annotation, where no constraint is given to a test image. There are a variable number of objects per image in aPascal, quantitatively evaluating free annotation is not straightforward. Therefore, evaluate only the most confident object and its associated top attributes in each image, although more could be described.

Annotation given object names, where named but not located objects are given. Here use the object's DPM detector to find the most confident bounding box. Then predict attributes for that box. Here, annotation accuracy is the same as attribute accuracy.

Annotation given locations, where object locations are given in the form of bounding boxes, and the attributes are predicted. Here simply predict attributes inside each bounding box. This becomes the conventional attribute prediction task for describing an object.

VII CONCLUSION

This is an effective model for weakly supervised learning of objects, attributes, and their locations and associations. Learning object attribute association from weak image-level labels is non-trivial but critical for learning from 'natural' data, and scaling to many classes and attributes. This is achieved by the first time through a novel weakly-supervised IBP model that simultaneously disambiguates super pixel annotation correspondence, and learns the appearance of each annotation and super pixel level annotation correlation. This results show that on a variety of tasks, this model often performs comparably to strongly supervised alternatives that are significantly more costly to supervise, and is consistently better than weakly supervised alternatives.

REFERENCES

[1] Aishwarya Agrawal, Jiasen Lu, Stanislaw Antol, Margaret Mitchell, C.Lawrence Zitnick, DhruvBatra, Devi Parikh (2015) 'VQA; Visual Question Answering',ICCV.

[2] BenjatSiddiquie, RogerioS.Feris, Larry S.Davis (2011) 'Image Ranking and Retrieval based on Multi-Attribute Queries', CVPR.

[3] Mohammad Rastegari, Ali Diba, Devi Parikh.(2013), 'Multi-attribute Queries: To Merge or Not to Merge?',CVPR.

[4] Oriol Vinyals, Alexander Toshev, SamyBengio, DumitruErhan. (2015) 'Show and tell: A neural image caption generator',CVPR.

[5] Pedro O. Pinheiro Ronan Collobert (2015) 'From Image-level to pixel- level labeling with convolutional networks',CVPR.

[6] Ronghang Hu, HuazheXu, Marcus Rohrbach,JiashiFeng, Kate Saenko, Trevor Darrell. (2016),"Natural language object retrieval", CVPR.

[7] Shuo Wang, JungseockJoo, Yizhou Wang, and Song-Chun Zhu(2013) 'Weakly supervised learning for attribute localizationin outdoor scene', CVPR.

[8] S.Zheng, M.M.Cheng, J.Warrell, P.Sturgess, V.Vineet, C.Rother and P.Torr,(2014) "Dense Semantic image segmentation with objects and attributes,"CVPR.