IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-3, Issue-3 E-ISSN: 2347-2693

A Review on: Visual Recognition Through Object Bank

Amrit Kumar Sharma
Department of Comp Sc & Engineering
Sikkim Manipal Institute Of Technology
Majitar, India
aks.kal6135@gmail.com
www.ijcseonline.org

Received: Feb/09/2015 Revised: Feb/22/2015 Accepted: Mar/10/2015 Published: Mar/31/2015

***Abstract*—** This report consists of a literature review of papers dealing with visual recognition using different techniques. Several papers that brought contribution to this field are summarized, analysed and compared. Different papers uses different moreover similar concepts for image/object recognition and their work brought average results in this field. By using the novel concept of Object Bank (OB) very good progress over image/object recognition has been done over recent years. Here we are stipulating the concept of Object Bank for high level visual recognition by using different Support Vector Machine (SVM) classifiers.

Keywords— Object Bank, Image recognition, Image representation, SVM, Semantic information, Feature extraction, Maximum Entropy.

I. INTRODUCTION

In the field of computer vision, image/object recognition has been one of the most challenging domains [1]. A great amount of research is being conducted on developing robust image representation.

Among the image representations widely done so far, most of them are low level image representations focusing on describing images by using some variant of image gradients, textures or colors. There exists a discrepancy between these low level image representations and the high level image recognition goals, which is called 'Semantic gap'. One way to bridge the semantic gap is by deploying increasingly sophisticated models, such as the probabilistic grammar model [3], compositional random fields and probabilistic models. While these approaches are based on much statistical formulation, good learning and inference are still extremely difficult. One notable development in image representation is the work that built image representations using intermediate 'attributes' [2][9]. It was much successful in recognition task due to the introduction of 'attribute', a high-level semantically meaningful representation. In attribute-based methods for object recognition, an object is represented by using visual attributes. Attribute-based methods have showed great potential in image classification. Image representations based on either attribute or spatial location have demonstrated potential in visual recognition tasks, which reminds us how human interprets an image. Objects are essential components to interpret an image. As human, we start to learn numerous objects from our childhood and memorize the appearance of learned objects. Their appearances are then used to effectively describe our visual world. Therefore, we can make out that object appearance and their spatial locations could be very useful for representing and recognizing images. In this paper, we are dealing with task of visual recognition using object bank (OB), a novel high level image feature to represent complex real-world image by collecting the responses of many object detectors at different spatial locations in the image.

.

II. LITERATURE REVIEW

Felzenszwalb et al.[1] describe an object detection system based on mixtures of multi-scale deformable part models. These models are trained using a discriminative procedure that only requires bounding boxes for the objects in a set of images. Their system is able to represent highly variable object classes and achieves state-of-the-art results in the PASCAL object detection challenges. They have made a robust object detector to produce responses of the object indicating the probability of its appearance at each pixel in an image. In deformable part models their value had not been demonstrated on difficult benchmarks such as the PASCAL datasets. Their system relies on new methods for discriminative training with partially labelled data. They have combined a margin sensitive approach for data-mining hard negative examples with a formalism called latent SVM. Ferrari, V.,&Zisserman.[2] presented a probabilistic generative model of visual attributes, together with an efficient learning algorithm. Attributes are visual qualities of objects, such as 'red', 'striped', or 'spotted'. These visual attributes are important for understanding object appearance and for describing objects to other people. The model sees attributes as patterns of image segments, repeatedly sharing some characteristic properties. Attributes with general appearance are taken into account, such as the pattern of alternation of any two colours which is characteristic for stripes. To enable learning from unsegmented training

© 2015, IJCSE All Rights Reserved

images, the model is learnt discriminatively, by optimizing a likelihood ratio. Automatic learning and recognition of attributes can complement category-level recognition and therefore improve the degree to which machines perceive visual objects. They proposed a probabilistic generative model of visual attributes, and a procedure for learning its parameters from real-world images. When presented with a novel image, the method infers whether it contains the learnt attribute and determines the region it covers.

Zhu et. al. [3] described an unsupervised method for learning a probabilistic grammar of an object from a set of training examples. Their approach is invariant to the scale and rotation of the object and the model of a hybrid object class where specific object, its position, scale or pose is not known. They have combined probability grammars with Markov RandomFields (MRF's) based on rigorous statistical formulation which resulted in structured models and have great representational power. It is a highly sophisticated model involving lots of computation. The results are obtained by learning the probability grammars from training datasets and evaluating them on the test datasets.

Somprasertsri et. Al.,[5] proposed an approach to product feature extraction using a maximum entropy model. Maximum entropy is a probability distribution estimation technique. The underlying principle of maximum entropy is that without external knowledge, one should prefer distributions that are uniform. Maximum entropy thresholding is based on the maximization of the information measure between object and background.

Felzenszwalb et. Al.,[8] applied algorithmic tricks to achieve speedups. This paper defines a root parts and n additional parts, which are searched over one by one, and thresholds are applied at each stage to prune away useless hypothesis. The thresholds are learned by looking at statistics of partial detection scores over positive examples. They used fresh positive training examples for selecting thresholds, separate from the examples used to train the models, and conducted evaluations on the PASCAL 2007 dataset. Testing on the 2007 dataset ensured that the statistics for the threshold training and test data were the same.

Kittikhun et. al.[7] proposed the maximum entropy-based image segmentation approach to segment a gray-scale face image. The approach performs with the Maximum Entropy Thresholding value (MET) of 2D image. The result obtained using presented maximum entropy-based thresholding approach was quite promising. It could separate robustly face image from background image better than Otsu approach. The comparison of this study with some other existing face segmentation approach such as the Centre of mass and the Iterative approach shows that they are having same performance. The threshold value of the proposed method is also at par performance of the two early approaches. However, the thresholdvalue from MET approach is dramatically different from the Otsu method. This is due to the boundary of segmentation image is more obvious than Otsu method. Once the segmentation of face boundary has been accomplished, then the subsequent steps in the pipeline of face recognition system are face detection and face extraction process. In this preprocessing state, several processes such as filtering, removing the noise are to be done before segmentation. Torresani et. Al.[9] introduced a new descriptor for images which allows the construction of efficient and compact classifiers with good accuracy on object category recognition. The descriptor is the output of a large number of weakly trained object category classifiers on the image. The trained categories are selected from the ontology of visual concepts, but the intention is not to encode an explicit decomposition of the scene. Characteristics can be combined to represent visual classes unrelated to the constituent categories semantic meanings. A compact descriptor is learned from a set of pre-trained concept classifiers. By using the training data from web image search in a novel way to train "category-like" classifiers, the descriptor is essentially given access to knowledge about what humans consider "similar" when they search for images. The classemes representation underscores the compactness of a feature representation for large scale visual tasks. It is shown that knowledge is effectively encoded in the classemes vector, and when this vector is quantized to below 200 bytes per image, gives competitive object category recognition performance.

Hossein Mobahi et. al.[14] presented an algorithm for segmentation of natural images that harnesses the principle of minimum description length (MDL). Their method is based on observations that a homogeneously textured region of a natural image can be well modeled by a Gaussian distribution and the region boundary can be effectively coded by an adaptive chain code. The optimal segmentation of an image is the one that gives the shortest coding length for encoding all textures and boundaries in the image, and is obtained via an agglomerative clustering process applied to a hierarchy of decreasing window sizes as multi-scale texture features. The optimal segmentation also provides an accurate estimate of the overall coding length and hence the true entropy of the image. They tested their algorithm on the publicly available Berkeley Segmentation Dataset. It achieves state-of-the-art segmentation results compared to other existing methods. Michael Stark et al.,[16] introduced a new dataset for fine-grained object categorization containing cluttered scenes with fully visible cars of different models. Two different object models are proposed. The first is based on bank of part detectors called Part Bank which is based on Object Bank paper from Fei-Fei Li. Basically at train and test times, they evaluate many part detectors at different scales inside a small window where they know a car to exist, and concatenate fully these detection maps into a single feature vector. Then they train linear SVMs (they say 'a' linear SVM) based on these

© 2015, IJCSE All Rights Reserved

feature vectors which tell which fine-grained category it really is. The second is simply based on DPM, by providing the fine-grained object label for each training example, and treating different fine-grained object categories as separate 'components' in the DPM.

Tim Althoff et. Al.,[15] proposed an image representation, called Detection Bank, based on the detection images from a large number of windowed object detectors where an image is represented by different statistics derived from these detections. This representation is extended to video by aggregating the key frame level image representations through mean and max pooling. It is showed that it captures complementary information to state-of-the-art representations such as Spatial Pyramid Matching and Object Bank. These descriptors combined with the Detection Bank representation significantly outperform any of the representations alone on TRECVID MED 2011 data.

Li-Jia Li et. Al. [17] introduced the concept of object bank (OB), a high-level image representation encoding object appearance and spatial location information in images. OB represents an image based on its response to a large number of pre-trained object detectors, or 'object filters', blind to the testing dataset and visual recognition task. It significantly outperforms traditional low level image representations in image classification on various benchmark image datasets by using simple, off-the-shelf classification algorithms such as linear SVM and logistic regression. The semantic information is obtained by running object detectors over multiple scales of images to capture the possibility of objects appear in the images. A spatial pyramid structure is applied to the response map representing the possibility of objects in an image to summarize the spatial statistics of objects.

Hao su et. Al.[18] proposed to use objects as attributes of scenes for scene classification. They represented images by collecting their responses to a large number of object detectors, or object filters. Such representation carried high-level semantic information rather than low-level image feature information, making it more suitable for high-level visual recognition tasks. Using very simple, off-the-shelf classifiers such as SVM, they showed that this object-level image representation can be used effectively for high-level visual tasks such as scene classification. The results were superior to reported state-of-the-art performance on a number of standard datasets. In this paper, they have only used very weak spatial information in the image representation through the spatial pyramid representation.

Ahmed Bassiouny et. al.[19] introduced a novel approach towards scene recognition using semantic segmentation maps as image representation. Given a set of images and a list of possible categories for each image, they assigned a category from that list to each image. Their approach is based on representing an image by its semantic segmentation map, which is a mapping from each pixel to a pre-defined set of labels. Among similar high-level approaches, this approach has the capability of not only representing what semantic labels the scene contains, but also their shapes, sizes and locations.

.

III. RESEARCH GAPS

Lot of research works has been carried out on visual recognition from a given image. These recognitions are based on image features. Most of the works used some variants of colors or textures for image classification but could not develop a robust image classifier.

Well performing object detectors have been introduced by [1], as well as the geometric context classifiers ('stuff' detectors) of [4]. These object recognition approaches the advancement of new state-of-the-art object detection and classification algorithms [12]. Visual attributes based research for image recognition [10][2][9] has achieved substantial progress recently. These approaches focus on single object classification based on visual attributes. These Cannot be useful for scene classification and Images rich in colours, textures and structure, they pose a considerable challenge for the classification task. The pre-defined concepts used in [6],[10] and [11] are not necessarily directly related to visual pattern in the images, e.g. 'eats fish' in [6], 'carnival' [10] and 'able-minded' in [11]. Different than these approaches, OB representation encodes semantic and spatial information of objects universally applicable for high level visual recognition tasks. While the classmes representation proposed by [9] underscores the compactness of a feature representation for large scale visual tasks it cannot be useful for large images or scene, for example, home photo retrieval, or object indexing of surveillance footage.

IV. CONCLUSION

Visual recognition with the help of objects can provide better information of the images. Over many techniques discussed above, the concept of object bank is unique and result oriented. So we choose the concept of object bank using different SVM classifiers for visual recognition task since object appearance and their spatial locations could be very useful for representing and recognizing images. Visual recognition using Object Bank can bridge the difference between low level image representation and the high level visual recognition tasks and provides better clarity and distinctibility in visual recognition tasks.

REFERENCES

[1]. Felzenszwalb, P., Girshick, R., McAllester, D., & Ramanan, "Object detection with discriminatively trained part based models". Journal of Artificial Intelligence Research, 29, 2007.

© 2015, IJCSE All Rights Reserved

[2]. Ferrari, V.,&Zisserman. "Learning visual attributes". In NIPS. ,2007.
[3]. Zhu, L., Chen, Y., & Yuille,. "Unsupervised learning of a probabilistic grammar for object detection and parsing". Advances in neural information processing systems, 19, 1617. 2007.
[4]. Fei-Fei, L., Fergus,R.,&Torralba, A. "Recognizing and learning object categories", 2007.
[5]. Somprasertsri, G.; Lalitrojwong, P., "A maximum entropy model for product feature extraction in online customer reviews," Cybernetics and Intelligent Systems, 2008 IEEE Conference on , vol., no., pp.575,580, 21-24 Sept. 2008.
[6]. Kittikhun Meethongjan, Dzulkifli Mohamad "Maximum Entropy-based Thresholding algorithm for Face image segmentation".,2009.
[7]. Lampert, C. H., Nickisch, H., & Harmeling, S. "Learning to detect unseen by between-class attribute transfer". In CVPR, 2009.
[8]. Felzenszwalb, Girschick, McAllester, "Cascade Object Detection with Deformable Part Models".,2009.
[9]. Torresani, L., Szummer, M., & Fitzgibbon, A. "Efficient object category recognition using classemes". In ECCV.,2010.
[10]. Farhadi, A., Endres, I., & Hoiem, D. "Attribute-centric recognition for cross-category generalization". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, (pp. 2352–2359). New York: IEEE.
[11]. Torresani, L., Szummer, M., & Fitzgibbon, A. "Efficient object category recognition using classemes". In ECCV.,2010.
[12]. Song, Z., Chen, Q., Huang, Z., Hua, Y., & Yan, S. "Contextualizing object detection and classification". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).,2011.
[13]. Dixit, M., Rasiwasia, N., & Vasconcelos, N. "Adapted Gaussian models for image classification". In CVPR, 2011.
[14]. Hossein Mobahi, Shankar R. Rao, Allen Y. Yang, Shankar S. Sastry, Yi Ma. "Segmentation of Natural Images by Texture and Boundary Compression". IJCV 2011.
[15]. Tim Althoff, Hyun Oh Song, Trevor Darrell. "Detection Bank: An Object Detection Based Video Representation for Multimedia Event Recognition"., 2011.
[16]. Michael Stark, Robert Patrick, James roch:"Fine-Grained Categorization for 3D Scene understanding" (BMVC 2012).
[17]. Li-Jia Li · Hao Su · Yongwhan Lim · Li Fei-Fei. "Object Bank: An Object-Level Image Representation for High-Level Visual Recognition". International Journal of Computer Vision, 2013.(pp. 630-660).
[18]. Hao su, Li-Jia Li, Yongwhan Lim, Li Fei-Fei:"Objects as Attributes for Scene Classification". ICCV, 2013.
[19]. Hetal J. Vala, Astha Baxi.: "A Review on Otsu Image Segmentation Algorithm": IJARCET, Volume 2, Issue 2, February 2013.
[20]. Ahmed Bassiouny, Motaz El-Saban:"Semantic Segmentation As Image Representation For Scene Recognition". Microsoft Advanced Technology Labs, Cairo, Egypt.
[21]. en.wikipedia.org/wiki/Support_vector_machine.
[22]. Digital Image Processing 2nd Edition by Gonzalez and Woods Pearson Publications.

© 2015, IJCSE All Rights Reserved