DIRAC Training - Report

LUO Jie (Roger)
IDIAP Research Institute
Centre du Parc, Rue Marconi 19, Martigny, Switzerland
jluo@idiap.ch

September 18, 2008

Object localization is one of the core technicals in DIRAC's application scenarios. In recent years object recognition/localization have become a major focus of computer vision and have shown substantial progress. However, they still can not give robust performance under complex and unexpected environment, which is the topic DIRAC interested in. Current trend is to use multiple cues instead of single cue, and use machine learning algorithms to learn the best combination of these cues. These methods are expected to achieve much robust performance than single cue methods.

The BIWI group are actively doing research on object localization. The idea of my visiting is to learn the state-of-art approaches on object feature representation, then combine them with my learning algorithm. During my three month visiting, I first worked under the supervision Dr. Bastian Leibe in the first two weeks. After Dr. Leibe left BIWI, I continued to work with Prof. Vittorio Ferrari. I will briefly summarize what have been done in the following part of this report.

1. In the first two weeks, I learned the state-of-art approaches on object localization, mainly on feature representations. I studied features, such as HOG [Dalal and Triggs CVPR04], color [Gevers et. al.], self-similarity [Shechtman and Irani CVPR07] and bag-of-words, whose code are public available, and re-implemented part of them for my specific requirement. After that I reproduced the experiments on standard benchmark PASCAL VOC 2006/2007 dataset using SVM classifiers with common sliding windows approach, and achieved similar results as those reported in original papers.

2. Object localization is commonly performed using sliding windows classifiers. The classifiers trained a discriminant function and then scan over different locations, usually at multiple scales and windows sizes.

And the classifiers predict that one object is present in sub-windows with high confidence. People usually trained the classifiers using cropped images with the object presented, and background images like the classification task. However, it is not clear if this way of training the classifier is also optimal for the localization task. We proposed a method that employs a discriminative learning procedure. Its learning phase directly aimed at achieving a higher area under the ROC curve (AUC), which is the standard measure for evaluating object localization. This is done by modifying the cost function of a SVM classifier using the *Wilcoxon-Mann-Whiteny* statistic [Cortes and Mohri NIPS03]. Our classifiers took positive and negative examples as training examples, and could be trained efficiently using optimization toolboxes. We applied the methods for training both the classifiers to learn the single cue and classifier for combining multiple cues. It is shown to be competitive with the large margin approach and attain higher AUC over the training examples and unseen examples.

3. I studies different learning distance functions [Bar-Hillel et. al JMLR, Hertz et. al ICML04, Weinberger et. al NIPS05, Goldberger et. al NIPS06] during my six weeks visit to Prof. Weinshall, sponsored by the DIRAC training program. Learning a distance function is to learn a similarity measure for the data from a given collection of pair of similar/dissimilar points that preserves the distance relation among the training data. In recent years, many studies have demonstrated that a learned distance function can significantly improve the performance in classification, clustering and retrieval tasks. Its advantage over the canonical distance functions are that it learns a similarity measure that embeds domain specific knowledge. To the best of my knowledge, although many approaches on learning a distance function have been proposed, only few of them have been applied on object recognition and localization scenario. During my visit, I tested learning distance function using positive and negative constrains with a KNN classifiers. The idea for using equivalent constrains instead of labels is simple. The assumption is that some categories could not be easily separated using a linear classifier. For example, a car and a sheep could both be white, and the shape of an apple and tomato may look similar from certain view points. Preliminary experiments showed that our simple nearest neighbor classifier using a learned distance measure outperform a linear svm classifier in most of the tasks on PASCAL VOC dataset. The main disadvantage of our method is that its computational complexity is higher than a linear svm classifier. And it is not obvious how to improve it by using efficient search algorithm [Lampert et. al. CVPR08].

In the future, we will try to improve the efficiency and performance of our methods, and possibly write a conference submission if the results are promis-

ing. Ideally, I also plan to extend the approach to multi-modal information, which would be my Ph.D topic. This work holds promise to be relevant for the goals in DIRAC-WP4, and will pave the way to develop methods for learning distance functions on audio-visual data, a topic of great relevance for DIRAC.

Acknowledge

I am grateful to Prof. Vittorio Ferrari for supervising my work and Prof. Bastian Leibe for the useful introduction to the state-of-art object recognition/localization algorithms. I would also like to acknowledge Dagan Eshar, Andreas Ess, Gabriele Fanelli, Alain Lehmann and Stefano Pellegrini for their fruitful discussions. Also many thanks to all staffs of BIWI, ETHZ for their friendliness and support.