Theory of Learning with Few Examples and Object Localization

Erik Rodner, Joachim Denzler

Chair for Computer Vision, Friedrich Schiller University of Jena
07743, Jena, Ernst-Abbe-Platz 2, Germany
{rodner,denzler}@informatik.uni-jena.de

1 Introduction

Visual object localization and categorization is still a big challenge for current research and gets even more difficult when confronted with few training examples. Therefore we will present a Bayesian concept to enhance state-of-the-art machine learning techniques even when dealing with just a single view of an object category. Furthermore an object localization approach is presented, which can serve as a baseline for researchers within the area of object localization.

Recent work has shown impressive progress in the field of visual object recognition. Despite these results, nearly all those techniques depend on a large set of training examples. This shows the still existing gap between the capabilities of human and machine vision. The question arises which additional information source is used by the human vision system to solve this ill-posed task. An intuitive answer is the interclass paradigm, which emphasize the importance to transfer knowledge from similar object categories (support classes) previously learned. With this basic concept new machine learning techniques have been developed specialized on learning object models from small training sets [1].

2 Bayesian Framework for Learning with Few Examples

Nearly all machine learning techniques are based implicitly or explicitly on estimating classifier parameters θ_i of a specific class i using a set of training examples T^i. This estimation technique is mostly done with simple maximum likelihood (ML) estimation assuming an uniform prior distribution on θ. In the case of very few training examples this optimization problem becomes ill-posed and the estimation becomes impossible without further regularization.

Fig. 1. Some promising results and some failure cases of our detection algorithm: airplane, motorbike, bicycle, bottle, cat, sofa

Dagstuhl Event Proceedings 08422
Computer Vision in Camera Networks for Analyzing Complex Dynamic Natural Scenes
http://drops.dagstuhl.de/opus/volltexte/2009/1861

A typical example is the estimation of a covariance matrix $\mathbf{C}$ of a Gaussian distribution which is often regularized by adding a constant λ to the diagonal. It is well-known in statistics that this trick can be justified using the Bayesian concept of maximum-a-posteriori estimation with a Wishart prior distribution. Therefore the idea underlying all solutions of the "few examples"-problem is to learn hyperparameters of a suitable prior distribution $p(\theta)$ from support classes $\mathcal{S}$ to regularize the estimation of model parameters of a new class γ:

$$\hat{\theta}_i = \mathrm{argmax}_\theta \; p(T^i \mid \theta) \cdot p(\theta \mid T^{\mathcal{S}}).$$

Recent progress has shown that this simple principle can be used to enhance the performance of decision tree approaches significantly [2].

3 Object Localization and Semantic Segmentation

Visual recognition tasks can be categorized in three main directions: classification (label the whole image), detection or localization (results in bounding boxes containing instances of the object category) and semantic segmentation which performs a pixel-wise labeling. Within the context of the PASCAL Visual Object Challenge (VOC), we implemented a sliding-window approach for object localization and used this additionally to enhance the performance of a current approach to semantic segmentation [3]. As a classifier for each window, we used Gentle-Boost with randomized forests as weak classifiers. The feature pool consists of shape (histogram of oriented gradients) and color features to cope with the large intra-class variablity of the VOC datasets. The advantage of these features is the efficient computation using integral histograms. The training process builds a cascade of these classifier for each category individually. This enables us to incorporate misclassified random background patches into the negative training set. Combination of hypotheses is finally based on connected component analysis of a suitable graph of overlapping bounding boxes.

For semantic segmentation we used the recently published method of Shotton et al. [3] using our localization results as a detection based prior as proposed in the paper. Furthermore an image-level prior had been incorporated using the method of Hegazy and Denzler [4]. Examples and detailed quantitative results can be found in the PASCAL VOC 2008 evaluation.

References

1. Fei-Fei, L.: Knowledge transfer in learning to recognize visual objects classes. In: ICDL. (2006) 1–8
2. Rodner, E., Denzler, J.: Learning with few examples using a constrained gaussian prior on randomized trees. In: Vision, Modelling, and Visualization Workshop. (2008) 104–112
3. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: CVPR. (2008) 1–8
4. Hegazy, D., Denzler, J.: Boosting local colored features for generic object recognition. Pattern Recognition and Image Understanding **18** (2008) 323–327