Invited Applications Paper

Discriminative Latent Variable Models for Object Detection

Pedro Felzenszwalb PFF@CS.UCHICAGO.EDU
University of Chicago

Ross Girshick RGB@CS.UCHICAGO.EDU
University of Chicago

David McAllester MCALLESTER@TTI-C.ORG
Toyota Technological Institute at Chicago

Deva Ramanan DRAMANAN@ICS.UCI.EDU
UC Irvine

(To be presented by Deva Ramanan). In this talk, I will discuss recent work by colleagues and myself on discriminative latent-variable models for object detection. Object recognition is one of the fundamental challenges of computer vision. We specifically consider the task of localizing and detecting instances of a generic object category, such as people or cars, in cluttered real-word images. Recent benchmark competitions such as the PASCAL Visual Object Challenge suggest our method is the state-of-the-art system for such tasks. This success, combined with publically-available code that runs orders of magnitude faster than comparable approaches, has turned our system into a standard baseline for contemporary research on object recognition (Felzenszwalb et al., 2008; 2009).

This talk will focus on the machine learning aspects of our approach. Our system is trained with a latent variable extension of support vector machines that we call a latent SVM. The formulation is equivalent to the MI-SVM framework for multiple instance learning. Latent variables provide a formalism for modeling structured variation in object appearance due to deformation, viewpoint, and other factors. The resulting learning problem is no longer convex, but admits a coordinate descent algorithm that exploits a 'semi-convex' property. Notable aspects of our system involve (a) weakly-supervised learning, in which hidden latent structure is automatically inferred; (b) out-of-core learning algorithms for learning from large-scale datasets that do not fit in memory; and (c) efficient algorithms for searching over latent variables.

Most of this talk represents joint work with Pedro Felzenszwalb, Ross Girshick, and David McAllester. I will conclude by highlighting various extensions from our groups including: generic object grammars that model variable object structure (Felzenszwalb & McAllester, 2010), efficient cascade implementations that result in real-time detection performance (Felzenszwalb et al., 2010), multilinear encodings that exploit the spatial structure inherent in image features (Pirsiavash et al., 2009), and latent-variable models that infer photometric as well as geometric properties of objects (Yang et al., 2010).

References

Felzenszwalb, P. and McAllester, D. Object detection grammars. Technical Report TR-2010-02, University of Chicago, 2010.

Felzenszwalb, P., McAllester, D., and Ramanan, D. A discriminatively trained, multiscale, deformable part model. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2008.

Felzenszwalb, P., Girshick, R., McAllester, D., and Ramanan, D. Object detection with discriminatively trained part-based models. In *IEEE Pattern Analysis and Machine Learning (PAMI)*, 2009.

Felzenszwalb, P., Girshick, R., and McAllester, D. Cascade object detection with deformable part models. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2010.

Pirsiavash, H., Ramanan, D., and Fowlkes, C. Bilinear classifiers for visual recognition. In *Neural Info. Proc. Systems (NIPS)*, 2009.

Yang, Y., Hallman, S., Ramanan, D., and Fowlkes, C. Layered object detection for multi-class segmentation. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2010.