Int J Comput Vis (2012) 99:257–258
DOI 10.1007/s11263-012-0530-y

Guest Editorial: Special Issue on Structured Prediction and Inference

Matthew B. Blaschko · Christoph H. Lampert

Received: 16 April 2012 / Accepted: 19 April 2012 / Published online: 1 May 2012
© Springer Science+Business Media, LLC 2012

1 Overview of this Special Issue

Computer vision has been profoundly influenced by machine learning in the past two decades. Canonical papers in the field, such as Turk and Pentland (1991), have applied statistical methods to visual data to achieve results that are both compelling and accurate. Classification algorithms, such as support vector machines, are now commonly applied to discriminatively train vision systems. Such systems empirically minimize risk functionals that measure the expected classification error rate given an i.i.d. sampling assumption (Vapnik 1998; Schölkopf and Smola 2002). While many intermediate goals in computer vision can be formulated as classification, regression, or dimensionality reduction—the settings most commonly addressed in the machine learning literature—more appropriate statistical methods are needed to train vision systems that make collections of related predictions, such as in segmentation, parts based models, or scene layout analysis. In these settings, the independence assumptions made by binary classifiers no longer hold, and independent prediction may not be computationally or statistically feasible.

Structured output prediction (Bakır et al. 2007; Nowozin and Lampert 2011) is the task of predicting related variables given some input, such as image or video data. Study of structured output prediction in the machine learning field has led to a number of techniques being proposed to better match the training algorithm to the prediction setting, without making implicit independence assumptions such as those made by training components of a prediction architecture with an unmodified binary classification algorithm. Instead resulting algorithms can be used for discriminative training of graphical models taking into account dependencies between output variables (Taskar et al. 2004; Tsochantaridis et al. 2004). These methods have found wide application in the computer vision literature in the last decade.

Structured output prediction is complicated by one factor in particular, that the output space tends to be exponential in size. This requires several innovations to make training feasible in practice. While generic techniques such as cutting plane optimization (Tsochantaridis et al. 2004) have been proposed in the machine learning literature, the application of structured output prediction techniques to new problems is colored by the interplay between learning and inference. The specific characteristics of a problem domain dictate the resulting methods that enable tractable and accurate solutions.

Vision is key application area of structured prediction, and statistical methods can benefit from properly encoding spatial constraints and prior probabilities. This special issue consists of four articles that illustrate central issues of representation, the encoding of interdependencies, optimization, and statistical efficiencies.

The paper "Discriminative Appearance Models for Pictorial Structures" provides an interesting study of the relationship between representation and discriminative training for a pictorial structures model (Andriluka et al. 2012). Discriminative training is combined with a pictorial structures model, including a kinematic prior, to achieve state-of-the-

M.B. Blaschko (✉)
École Centrale Paris, Grande Voie des Vignes,
92295 Châtenay-Malabry, France
e-mail: matthew.blaschko@inria.fr

C.H. Lampert
Institute of Science and Technology Austria, Am Campus 1,
3400 Klosterneuburg, Austria
e-mail: chl@ist.ac.at

art results on people detection, upper body pose estimation, and full body pose estimation.

Taxonomic representations have been of particular interest in the vision community as they promise both improvements in semantic interpretability and statistical efficiency. "On Taxonomies for Multi-class Image Categorization" provides an up-to-date study on the application of taxonomies to visual categorization (Binder et al. 2012). This work makes use of taxonomies in a discriminative learning framework and illuminates the interplay between taxonomic feature map and loss functions in this structured prediction setting.

Inference in arbitrary graphical models is NP hard, and practical approaches consequently make simplifying assumptions about the form of the graphical model. "Fast Structured Prediction using Large Margin Sigmoid Belief Networks" describes one such family of graphical models that exhibits several favorable computational and statistical properties (Miao and Rao 2012). Fast inference is achieved using a branch-and-bound strategy, with results demonstrated for scene recognition, optical character recognition, and large scale image annotation.

"On Learning Conditional Random Fields for Stereo: Exploring Model Structures and Approximate Inference" shows the applicability of structured prediction methods to depth from stereo, and the connection between learning and approximate inference (Pal et al. 2012). They employ sparse variational message passing to achieve substantial increases in efficiency, while learning from newly available stereo datasets with complete ground truth annotation.

We believe these papers provide an interesting cross-section of structured prediction in computer vision. The central issues in statistical learning and inference applied to visual data are apparent, and we hope this will inspire future work in this direction. We would like to take this opportunity to especially thank the reviewers for their contributions to the special issue.

References

Andriluka, M., Roth, S., & Schiele, B. (2012). Discriminative appearance models for pictorial structures. *International Journal of Computer Vision*. doi:10.1007/s11263-011-0498-z.

Bakır, G. H., Hofmann, T., Schölkopf, B., Smola, A. J., Taskar, B., & Vishwanathan, S. V. N. (2007). *Predicting structured data*. Cambridge: MIT Press.

Binder, A., Müller, K. R., & Kawanabe, M. (2012). On taxonomies for multi-class image categorization. *International Journal of Computer Vision*. doi:10.1007/s11263-010-0417-8.

Miao, X., & Rao, R. (2012). Fast structured prediction using large margin sigmoid belief networks. *International Journal of Computer Vision*. doi:10.1007/s11263-011-0423-5.

Nowozin, S., & Lampert, C. H. (2011). Structured learning and prediction in computer vision. *Foundations and Trends in Computer Graphics and Vision*, *6*, 185–365.

Pal, C., Weinman, J., Tran, L., & Scharstein, D. (2012). On learning conditional random fields for stereo. *International Journal of Computer Vision*. doi:10.1007/s11263-010-0385-z.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Cambridge: MIT Press.

Taskar, B., Guestrin, C., & Koller, D. (2004). Max-margin Markov networks. In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (p. 16). Cambridge: MIT Press.

Tsochantaridis, I., Hofmann, T., Joachims, T., & Altun, Y. (2004). Support vector machine learning for interdependent and structured output spaces. In *International conference on machine learning (ICML)* (pp. 104–112).

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86. doi:10.1162/jocn.1991.3.1.71.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley-Interscience.