# Machine Learning for Visual Scene Understanding

Machines that understand images of natural scenes are a long-lasting dream of artificial intelligence research. Today, it is a driving force for research on high-level computer vision, where statistical machine learning techniques are used to study questions such as *What objects are visible in an image?* (Object Categorization and Image Retrieval), and *Where exactly are they?* (Object Localization and Image Segmentation).

**Object Categorization** Visual data allows several natural characterizations, for example by color, texture, or shape. Each of these gives rise to a family of positive definite kernel functions. In [6] we demonstrated that Linear Program Boosting can be used to learn the trade-offs between different features, whereas, contrary to dominant belief in the community, Multiple Kernel Learning (MKL) is not able to improve classification accuracy over simple baselines for such features. In [5] we introduced a way to link the final classifier system more directly to image content, resulting in an *end-to-end* learning system that integrates kernel design transparently into the learning task.

For the common situation that only a few or even no training examples are available, we introduced a kernel-based clustering technique in [2], showing that groups of semantically similar objects can also be found in an unsupervised way. In [8] we introduced the first method for visual object categorization based not on training examples of the classes of interest, but based on a description in terms of high-level attributes, which are learned from a set of disjoint classes.

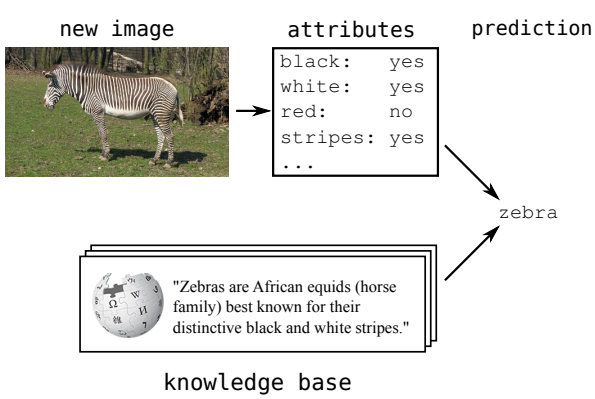


Figure 0.1: Attribute-based classification [8]. An image is categorized based on its semantic properties, not based on its similarity to reference or training images.

**Image Retrieval** Modern object classification systems rely on databases that contain several millions of images. In [4, 1] we show that the Jensen-Bregman LogDet diverge allows efficient nearest-neighbor image retrieval while at the same time allowing for queries with respect to a more natural distance between images than the Euclidean one.

**Object Localization** In object detection we are interested in identifying the location of an object in an image by means of a bounding box. In [3], we used recent structured prediction techniques to construct the first end-to-end learning system for this task, where previous techniques worked in a suboptimal two stage setting. We developed a branch-and-bound technique for finding the rectangular region in an image that maximizes a score function, which we subsequently extended to the case of multiple images [9], and to cascaded classifiers [10].

**Semantic Image Segmentation** Semantic image segmentation is used to obtain pixel-accurate prediction of object positions in an image. Modern techniques combine local image cues with domain specific image priors, typically second-order gradient statistics. In [11] we showed how higher-order statistics or constraints can be integrated, using the example of enforcing a path-connected foreground region.

Our related work [7] started as a theoretical investigation of submodular optimization and eventually led to interesting insights for computer vision. It extends the class of models for which the efficient GraphCut algorithm can be applied from modular to submodular functions, thereby enabling the use of more realistic image priors that encode the effect of diminishing returns.

Figure 0.2: Segmentation by CooperativeCuts [7]. Modular image priors tends to suppress thin elongated structures during segmentation (middle). With a submodular image prior such structures are preserved (right).

# Publications

## Journal Articles


[1] S Sra. The Stein divergence and metric on positive definite matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 2012. 1

[2] T Tuytelaars, CH Lampert, MB Blaschko, and W Buntine. Unsupervised object discovery: A comparison. *International Journal of Computer Vision*, 88(2):284–302, 6 2010. 1


## Articles in Conference Proceedings


[3] MB Blaschko and CH Lampert. Learning to localize objects with structured output regression. In D A Forsyth, P HS Torr, and A Zisserman, editors, *10th European Conference on Computer Vision*, pages 2–15, Marseille, France, 10 2008. Springer. Best Student Paper Award. 1

[4] A Cherian, S Sra, A Banerjee, and N Papanikolopoulos. Efficient similarity search for covariance matrices via the Jensen-Bregman logdet divergence. In DN Metaxas, L Quan, A Sanfeliu, and LJ Van Gool, editors, *13th International Conference on Computer Vision (ICCV 2011)*, pages 2399–2406, Barcelona, Spain, 2011. IEEE. 1

[5] PV Gehler and S Nowozin. Let the kernel figure it out: Principled learning of pre-processing for kernel classifiers. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2836–2843, Miami Beach, FL, USA, 6 2009. IEEE. 1

[6] PV Gehler and S Nowozin. On feature combination for multiclass object classification. In *Twelfth IEEE International Conference on Computer Vision*, pages 221–228, Kyoto, Japan, 10 2009. IEEE Computer Society. 1

[7] S Jegelka and J Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, pages 1897–1904, Colorado Springs, CO, USA, 6 2011. IEEE. 1

[8] C Lampert. Learning to detect unseen object classes by between-class attribute transfer. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, Miami Beach, FL, USA, 6 2009. 1

[9] CH Lampert. Detecting objects in large image collections and videos by efficient subimage retrieval. In *Twelfth IEEE International Conference on Computer Vision*, pages 987–994, Kyoto, Japan, 10 2009. IEEE Computer Society. 1

[10] CH Lampert. An efficient divide-and-conquer cascade for nonlinear object detection. In *23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pages 1022–1029, San Francisco, CA, USA, 6 2010. IEEE. 1

[11] S Nowozin and CH Lampert. Global connectivity potentials for random field models. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 818–825, Miami Beach, FL, USA, 6 2009. IEEE. 1