Branch&Rank: Non-Linear Object Detection

Alain D. Lehmann[1]
lehmann@vision.ee.ethz.ch

Peter V. Gehler[1,2]
pgehler@mpi-inf.mpg.de

Luc Van Gool[1,3]
vangool@vision.ee.ethz.ch

[1] Computer Vision Laboratory
ETH Zurich, Switzerland

[2] Computer Vision & Multim. Comp. Lab
MPI for Informatics, Germany

[3] ESAT-PSI/IBBT
KU Leuven, Belgium

Abstract

Branch&rank extends the idea of branch&bound [2]: ranking improves efficiency and detects object with often less than 100 classifier evaluations. We thereby enable the use of expensive classifiers (*e.g*. non-linear SVMs with RBF-χ^2 kernels) without cascade-like approximations. This is a crucial advance in object detection since strong classifiers are beneficial to properly model the object intra-class variations [4]. Fig. 1 demonstrate the algorithmic properties using the VOC'07 dataset [1].

1 Motivation

Object detection is challenging because of two problems. First, objects may appear anywhere in an image and need to be localised; this leads to a large-scale search problem. Second, objects exhibit strong appearance variations that are best captured with strong classifiers [4]. Unfortunately, such classifiers are expensive to evaluate, which makes exhaustive search intractable. One can either try to reduce the cost of a classifier, *e.g*. cascades [5], or to reduce the number of classifier evaluations, *e.g*. branch&bound [2]. Branch&rank follows the ideas of branch&bound but abandons the notion of bounds as tight bounds are unavailable in practice [3, 4]. In other words, we adopt the best-first search and "branch" but do not "bound"; we *rank*. We thereby integrate the idea of scoring sets into the training. Intuitively speaking we aim to "learn the bound".

2 Ranking

The core of our detector is a ranking function f. It governs the adaptive search space partitioning (*c.f*. [2, 3]) and supersedes the notion of bounds. Its goal is to prioritise hypothesis sets Λ^+ that *contain an object* over sets Λ that do not, *i.e*., $f(\Lambda^+) > f(\Lambda)$. A function that fulfills this condition would yield detections in logarithmic time. Of course, we cannot expect this in practice and every incorrect ranking increases the number of iterations till detection. The main assumption of branch&rank is that it is possible to decide if a set (*i.e*. a sub-image *c.f*. Fig. 2) contains an object or not. This is exactly the problem that image classification [1] is addressing with much success .

3 Multi-Task

An essential aspect of branch&rank is that sets are already integrated in the training. The ranking function can thus leverage information of the sets. Our setup accounts for the set size (*c.f*. Fig. 3), which makes the system effective: we thereby separate image classification from object recognition, yet combines them in a single, structured SVM formulation. This grouping reduces intra-task variance and often led to a 10% increase of accuracy.

4 Conclusion

In summary, the concept of ranking has the following benefits: (1) The ranking condition combines model estimation and acceleration of the test-time problem in a *joint objective*: improving the ranking function (classifiers) leads to better *and more efficient* object detection. (2) The ranking condition is flexible; it allows for *arbitrary* (ranking-)classifiers since no bound has to be derived. (3) Branch&rank is efficient and enables the use of strong and costly classifiers (like non-linear SVMs) without the need for cascade-like approximations.

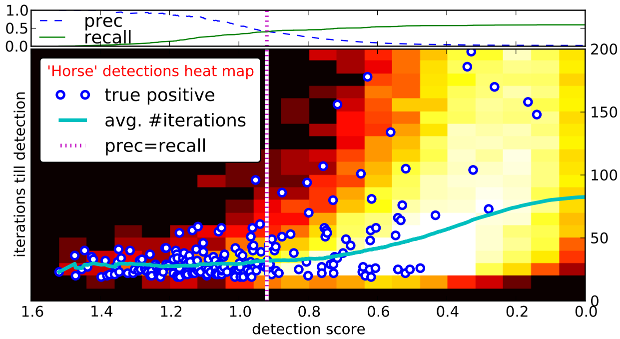

Figure 1: Branch&rank localises objects with often less than 100 classifier evaluations. This efficiency enables strong but more costly classifiers. We use non-linear SVMs with RBF-χ^2 kernel throughout the entire detection process; we do not use any approximations like *e.g*. cascades.

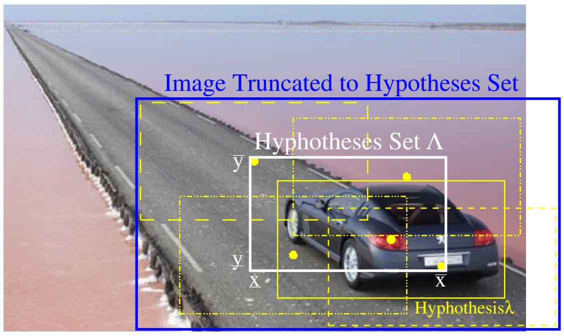

Figure 2: The sub-image which covers all bounding boxes of a hypothesis set is the (visual) input to our ranking function.

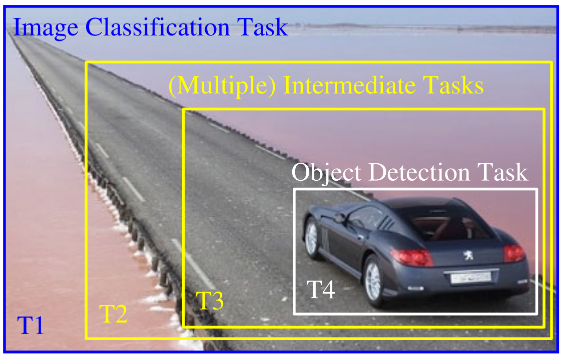

Figure 3: The hypothesis set *size* provides additional information to the ranking function: large sets relate to image classification as they cover the entire image and objects are *not* centered in it; bounding boxes of smaller sets frame objects more tightly as usual in detection. We distinguish yet combine these tasks and allow for a smooth transition between them.

[1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results.

[2] C. H. Lampert, M. B. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *CVPR*, 2008.

[3] A. Lehmann, B. Leibe, and L. Van Gool. Fast prism: Branch and bound hough transform for object class detection. *IJCV*, 94(2):175–197, 2011.

[4] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple kernels for object detection. In *ICCV*, 2009.

[5] P. A. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.