Object Proposal by Multi-branch Hierarchical Segmentation

Chaoyang Wang[1,3], Long Zhao[2,3], Shuang Liang[2], Liqing Zhang[1], Jinyuan Jia[2], Yichen Wei[3]
[1]Shanghai Jiao Tong University. [2]Tongji University. [3]Microsoft Research.

Segmentation based object proposal methods [1, 2, 5, 6, 8] have become an important step in modern object detection paradigm. Among those, hierarchical segmentation is favorable for its ability to capture objects of all scales and has fast implementation such as hierarchical greedy merging [6, 8]. However, single-way hierarchical greedy merging is fundamentally flawed in that the errors in early steps of greedy merging cannot be corrected and accumulate. In this work, we propose a novel multi-branch hierarchical segmentation approach that alleviates such problems by learning multiple merging strategies in each step in a complementary manner, such that errors in one merging strategy could be corrected by the others. This approach turns the original hierarchical greedy merging's sequential evolution structure into a tree-like structure. As illustrated in Fig. 1, different merging strategies are tried throughout the greedy merging process, which we call *branching*. Objects, especially large and complex ones will get a better chance to be detected in one of those tree branches.

To make our multi-branched hierarchical segmentation effective, we address two important problems: how to train complementary merging strategies capable of fixing mistakes made by others, and how to organize branching since arbitrary branching causes exponentially large search space. We handle those issues as below:

Learning complementary merging strategies. For each branch, we model the merging strategy as a binary classifier by using its decision value as merge score. For the branches starting from the same parent branch, their merging strategies are trained to be complementary in a sequential manner, which makes each of them capable of fixing some mistakes made by their predecessors. We use linear SVM with weighted loss to train those linear classifiers sequentially: at each time a classifier is trained, the loss weight is increased by a portion if its corresponding training sample is wrongly classified. This sequential training with weighted loss is very similar to boosting, except that our goal is not to combine the classifiers into a strong one, but to obtain a set of complementary ones to enrich the chances of finding objects missed in each other.

Multi-staged branching. Using the complementary merging strategies (classifiers), we branch the searching into multiple directions. New branches start only when the classifier's decision scores of all pairs of adjacent segments are below zero. This setting splits one single greedy merging process into multiple stages. Merging strategies in each stages are different.

Since the total number of branches at the last stage is exponential to the stage number, it appears that this multi-staged branching would generate too many proposals. But as shown in the paper, if we use a small branch degree (2 in our experiment) and a large merging pace (merges 50% pairs of segments) in each stage, the proposal number can be kept relatively small. We propose to control the pace of each stage by setting a target miss rate for each merging strategy. This target miss rate is achieved by searching a bias threshold for each classifier on a separate training set.

In [7], merging strategies are also modeled as linear classifiers and organized in cascaded stages. Compared to their work, ours emphasizes on the complementary property between multi-branches, rather than high boundary recall which results in a single sequence of cascaded classifiers with a lot of stages.

Our work is mostly related to [8], as both use hierarchical image segmentation and multiple merging strategies. Our approach differs in that it is built on two more general principles: multi-staged branching and automatic learning of complementary merging strategies. As shown in experiments, our method is more effective as it finds a smaller number of proposal with higher qualities.

In this paper, extensive comparison to previous object proposal methods indicates that our approach achieves the state-of-the-art results in terms of object proposal evaluation protocols. Moreover, in order to investigate how well these methods perform for real-world object detection tasks, we test all compared object proposal methods using the state-of-the-art R-CNN detector [4] on PASCAL VOC2007 [3]. To the best of our knowledge, we are among the first to give this end to end comparison. As a result, our approach achieves the best mAP rate comparing to previous methods.

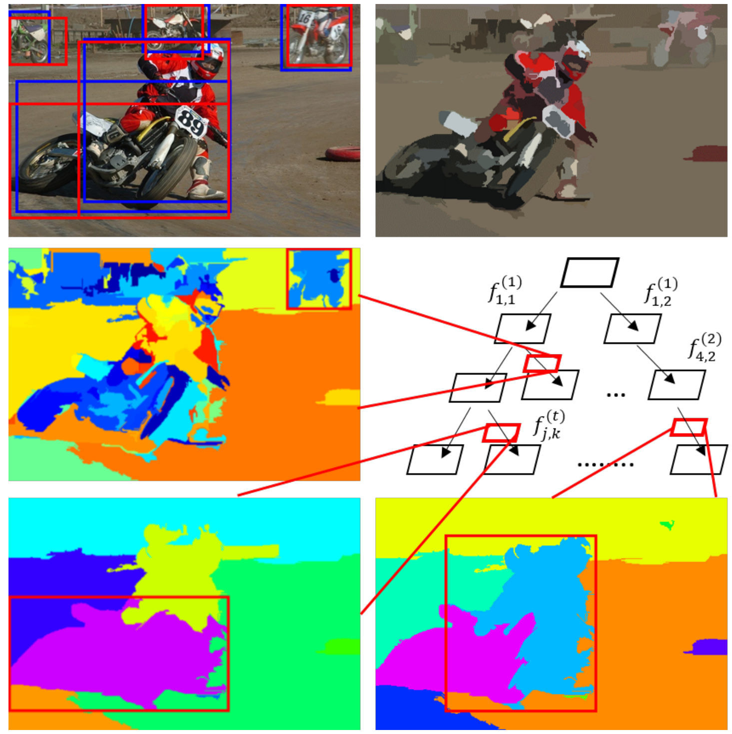

Figure 1: Illustration of our approach. Top left: ground truth bounding box annotation (blue), our proposal windows with highest IoU score (red). Top right: initial oversegmentation rendered with mean color. Mid right: the tree structure of multi-branch hierarchical image segmentation. Bottom & mid left: three objects successfully detected at different branches.

[1] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *CVPR*, 2014.

[2] J. Carreira and C. Sminchisescu. CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(7):1312–1328, 2012.

[3] M. Everingham, L. J. V. Gool, C. K. I. Williams, J. M. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision (IJCV)*, 88(2):303–338, 2010.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

[5] P. Krähenbühl and V. Koltun. Geodesic object proposals. In *ECCV*, 2014.

[6] P. Rantalankila, J. Kannala, and E. Rahtu. Generating object segmentation proposals using global and local search. In *CVPR*, 2014.

[7] Z. Ren and G. Shakhnarovich. Image segmentation by cascaded region agglomeration. In *CVPR*, 2013.

[8] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *International Journal of Computer Vision (IJCV)*, 104(2):154–171, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.