Learning to Propose Objects

Philipp Krähenbühl[1], Vladlen Koltun[2]
[1]UC Berkeley [2]Intel Labs

Abstract. We present an approach for highly accurate bottom-up object segmentation. Given an image, the approach rapidly generates a set of regions that delineate candidate objects in the image. The key idea is to train an ensemble of figure-ground segmentation models. The ensemble is trained jointly, enabling individual models to specialize and complement each other. We reduce ensemble training to a sequence of uncapacitated facility location problems and show that highly accurate segmentation ensembles can be trained by combinatorial optimization. The training procedure jointly optimizes the size of the ensemble, its composition, and the parameters of incorporated models, all for the same objective. The ensembles operate on elementary image features, enabling rapid image analysis. Extensive experiments demonstrate that the presented approach outperforms prior object proposal algorithms by a significant margin, while having the lowest running time. The trained ensembles generalize across datasets, indicating that the presented approach is capable of learning a generally applicable model of bottom-up segmentation.

Introduction. Object proposal algorithms aim to identify a small set of regions such that each object in the image is approximately delineated by at least one proposed region. Object proposals can be computed bottom-up, based only on low-level boundary detection and category-independent grouping [1, 3, 6, 8]. They are used as a starting point for both object detection and semantic segmentation, and have become a standard first step in state-of-the-art image analysis pipelines [2, 4, 5, 8].

To support diverse image parsing tasks, object proposal algorithms must have a number of characteristics. They need to provide region proposals with informative shape for semantic segmentation and instance segmentation [2, 4, 5, 7]. They must have high recall, producing corresponding regions for as many genuine objects as possible. They must generate a manageable number of proposals to limit unnecessary workload. And they must be fast to support high-performance image parsing [4, 8].

In this paper, we present an object proposal algorithm that has all of these characteristics. The key idea is to optimize an ensemble of figure-ground segmentation models. Given a new image, the algorithm simply applies each model and outputs all of the produced foreground segments. The algorithm is fast since each model is highly efficient and operates on elementary image features. Proposals produced by a trained ensemble are shown in Figure 1.

The presented approach optimizes a diverse ensemble of segmentation models globally during training. The training objective is the accuracy of the generated proposal set balanced by its size. We show that the training objective can be expressed in terms of the uncapacitated facility location problem and optimized by combinatorial techniques. The training jointly optimizes the size of the ensemble, its composition, and the parameters of the incorporated models, all for the same objective. The number of generated proposals can be controlled at training time and there is no need for test-time ranking.

Results. We conduct extensive experiments on the Pascal VOC2012 dataset and the recent Microsoft COCO dataset, comparing the performance of the presented approach to state-of-the-art object proposal algorithms. We evaluate both region proposal accuracy and bounding box proposal accuracy. In region proposal accuracy, our approach outperforms prior methods by a wide margin, while having the lowest running time. For example, the approach achieves 94% recall on the VOC 2012 dataset as measured by detailed shape overlap: the highest ever reported. Our approach also yields the highest bounding box proposal accuracy simply by taking the bounding boxes of the proposed regions.

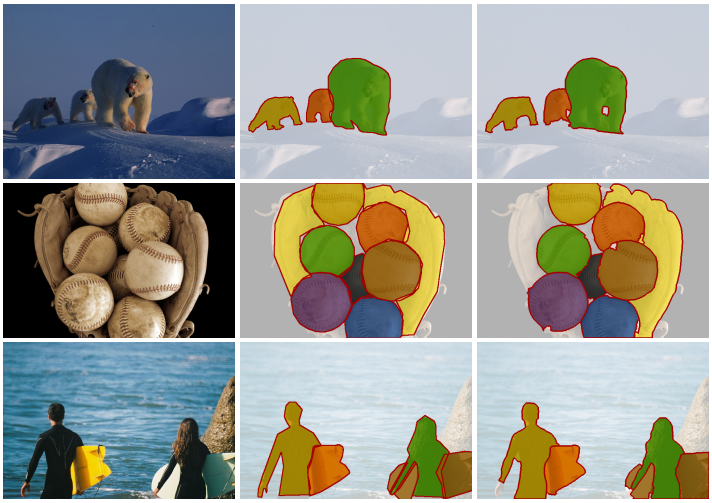

Figure 1: Object proposals for three images from the Microsoft COCO dataset. From left to right: input images, ground-truth instance segmentations, region proposals generated by the presented approach. Note the accurate instance proposals in the top and middle rows, despite color and texture similarity across instances. In the bottom row, the trained ensemble correctly identifies the white surfboard as a single object with three connected components.

We have also trained models on the entire VOC 2012 segmentation dataset and then evaluated them on COCO. Models trained on COCO and models trained on VOC perform similarly. This strongly suggests that our approach is capable of learning a general model of bottom-up object segmentation, biased neither to a specific dataset nor to specific object classes.

[1] João Carreira and Cristian Sminchisescu. CPMC: automatic object segmentation using constrained parametric min-cuts. *PAMI*, 34(7), 2012.

[2] João Carreira, Rui Caseiro, Jorge Batista, and Cristian Sminchisescu. Free-form region description with second-order pooling. *PAMI*, 2015. To appear.

[3] Ian Endres and Derek Hoiem. Category-independent object proposals with diverse ranking. *PAMI*, 36(2), 2014.

[4] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

[5] Bharath Hariharan, Pablo Arbeláez, Ross B. Girshick, and Jitendra Malik. Simultaneous detection and segmentation. In *ECCV*, 2014.

[6] Philipp Krähenbühl and Vladlen Koltun. Geodesic object proposals. In *ECCV*, 2014.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014.

[8] Jasper R. R. Uijlings, Koen E. A. van de Sande, Theo Gevers, and Arnold W. M. Smeulders. Selective search for object recognition. *IJCV*, 104(2), 2013.

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