Object Recognition by Ranking Figure-Ground Hypotheses

Joao Carreira, Fuxin Li, Cristian Sminchisescu
Computer Vision and Machine Learning Group, INS, University of Bonn

We focus on designing methods that can segment images into a number of objects and label each image pixel with the object class to which it belongs (people, animals, bottles, etc). We pursue a segmentation-based front-end to infer plausible regions of support for feature extraction. Differently from existing approaches that dominantly use hierarchical segmentations to fully tile an image with multiple uniform regions [1, 2] - 'superpixels' - we compute a set of binary (figure/ground) partitionings, obtained using constrained parametric max flow procedures. This creates a flat family of figure/ground solutions at multiple scales, later assembled into a complete image interpretation using consistency rules. Our segment hypotheses are non-uniform, large in scope, and in practice overlap with entire image objects, or their major parts. This allows reasoning over a larger context and adds the possibility of exploiting Gestalt cues in order to discard implausible segments, re-rank the remaining ones, and build rich representations for processing by the next-level recognition module.

Our visual object-class recognition is based on continuous value ranking, based on estimates of spatial overlap of each segment hypothesis with putative classes. We differ from existing approaches not only in our assumption that good object segments can be obtained in a feed-forward fashion, as opposed to methods employing shape priors to help segmenting the objects [3], but also in framing recognition as a regression problem. Instead of focusing on a one-vs-all winning margin that may not respect the qualitative ordering inside the non-maximum (non-winning) set, our learning method produces a globally consistent ranking with close ties to segment quality, hence to the extent entire object or part segment hypotheses spatially overlap with the ground truth. This approach handles naturally diverse problems such as image classification, object detection and semantic segmentation, and attains encouraging results in a number of challenging datasets including Caltech-101 and ETHZ-Shape. This methodology ranked first in the PASCAL VOC 2009 Object Class Segmentation and Labeling Challenge. Foreground/background image segmentation results are illustrated in fig. 1 whereas the full recognition pipeline is depicted in fig. 2.

Topic: visual processing and pattern recognition **Preference: poster**

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. *CVPR*, 2009.

[2] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE PAMI*, 24(5):603–619, 2002.

[3] M. Kumar, P. Torr, and A. Zisserman. Obj cut. In *CVPR*, pages 18–25, Washington, DC, USA, 2005. IEEE Computer Society.

Figure 1: Some foreground/background image segmentation results. This segmentation front end has been used in our PASCAL VOC09 recognition framework.

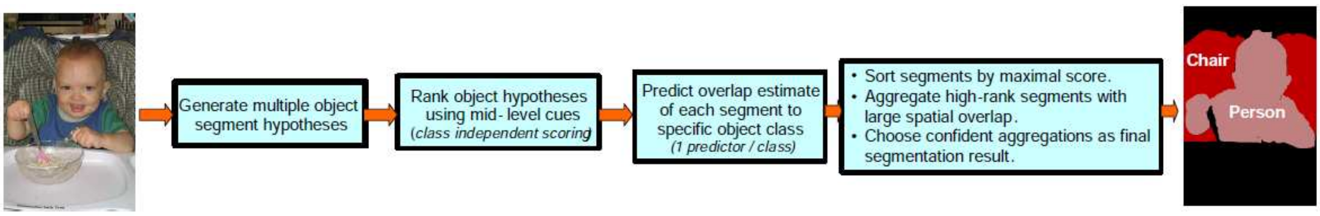

Figure 2: Recognition pipeline used in our PASCAL VOC 2009 entry. An image is segmented into multiple figure-ground segments. Ranking functions on different classes score each segment on each class. Then top-scoring overlapping segments are selected for combination. The final segmentation mask is combined from these segments, and its score is a weighted average of segment scores. In this example, our system is shown to correctly segment and recognize people (plotted in pink) and chairs (plotted in red), but the method is also capable of recognizing objects as diverse as cows, horses, bottles or motorbikes.