Mapping Auto-context Decision Forests to Deep ConvNets for Semantic Segmentation

David L. Richmond[*1]
Dagmar Kainmueller[*1]
Michael Y. Yang[2]
Eugene W. Myers[1]
Carsten Rother[2]

[1] Max Planck Institute of Molecular Cell
Biology and Genetics
Dresden, DE

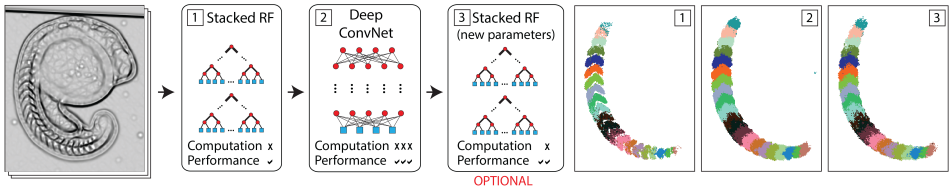[2] Technical University of Dresden
Dresden, DE

Figure 1: Our method and exemplary results for semantic segmentation of developing zebrafish. (1) A stacked Random Forest is trained, and then (2) mapped to a deep ConvNet and further trained by back-propagation. (3) Optionally, the ConvNet is mapped back to a stacked Random Forest with updated parameters.

In this paper, we propose a mapping from the Auto-context model to a deep Convolutional Neural Network (ConvNet), bridging the gap between these two models, and helping address the challenge of training ConvNets with limited training data.

Auto-context (AC) is a simple model for semantic segmentation that has proven powerful for, e.g., body-pose, facade, and brain segmentation. However, AC is limited by the fact that each classifier is trained greedily. ConvNets on the other hand, benefit from end-to-end training, and have demonstrated remarkable performance when large training data sets are available.

Here we demonstrate that AC can be mapped directly to a ConvNet, and thereby trained end-to-end. This mapping can be seen as an intelligent initialization of the ConvNet, enabling training of large models on limited data. We also describe an approximate mapping of our sparse, deep ConvNet back to a stacked Random Forest with updated parameters, for more computationally efficient evaluation. See Figure 1 for our proposed workflow.

Other works have explored the space between Random Forests and ConvNets [1, 2, 3]. In [2, 3], the authors map a single Random Forest to a shallow Neural Network model. We extend this work to Auto-context and deep ConvNets, and apply it to semantic segmentation.

The mapping that we propose leads to an interesting new ConvNet architecture that, to the best of our knowledge, hasn't previously been explored. One feature of the architecture is that it has large, sparse convolutional kernels. It also avoids max pooling and strided convolutions, which tend to lead to coarse outputs in semantic segmentation. Another result of the initialization from a stack of classifiers, is that specific layers in the ConvNet are directly interpretable as intermediate predictions of the net.

We experimentally verify that the mapping outperforms stacked Random Forests for two different applications: Kinect-based body part labeling from depth images, and somite segmentation in microscopy images of developing zebrafish. By directly visualizing the intermediate prediction layers, we observe that the ConvNet learns to smooth the intermediate predictions, a strategy that was previously developed to improve the performance of stacked classifiers.

[1] Peter Kontschieder, Madalina Fiterau, Antonio Criminisi, and Samuel Rota Bulo. Deep neural decision forests. In *ICCV*, 2015.

[2] Ishwar Sethi. Entropy nets: from decision trees to neural networks. *Proceedings of the IEEE*, 78 (10):1605–1613, 1990.

[3] Johannes Welbl. Casting random forests as artificial neural networks (and profiting from it). In *GCPR*, 2014.