Abstract:

We introduce a purely feed-forward architecture for semantic segmentation. We map small image elements (superpixels) to rich feature representations extracted from a sequence of nested regions of increasing extent. These regions are obtained by "zooming out" from the superpixel all the way to scene-level resolution. This approach exploits statistical structure in the image and in the label space without setting up explicit structured prediction mechanisms, and thus avoids complex and expensive inference. Instead, superpixels are classified by a feedforward multilayer network. Our architecture achieves new state of the art performance in semantic segmentation, obtaining 64.4% average accuracy on the PASCAL VOC 2012 test set. Joint work with Mohammadreza Mostajabi and Payman Yadollahpour.