Review on “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”

Authors: Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille

# Short Summary

In this paper, the authors explore methods to improve the task of image segmentation with deep learning. This is explored through addressing three key issues with existing approaches: feature resolution, object scale, and the consequences of invariance. Their exploration begins by building upon well-known image classification networks, VGG-16 and ResNet-101, by adding atrous convolutions (convolution with upsampled filters) following the final layers responsible for decreasing image resolution. This technique allows the user to control effective Field of View by varying kernel size and the sampling rate. Experimenting with varying sampling rates, their results showed that smaller kernel sizes with larger sampling rates are competitive with the base network architectures but with a fraction of the parameters. With control over FoV and building on a previous work, the paper also theorizes that the issue of object scale can also be dealt with using atrous convolutions. They introduce a new method, entitled atrous spatial pyramid pooling, which extracts features using multiple parallel atrous convolutional layers at different sample rates. Experiment results showed that this technique yields an increase in performance (approx. 2-3% higher mean IOU) with larger sampling rates providing more benefit. Lastly, the paper also implements fully-connected Conditional Random Fields (CRFs) and shows that their inclusion in distinct architectures also yields a boost in performance both quantitatively and qualitatively. The final model, including various data augmentation strategies, achieves the top score on the on the PASCAL VOC 2012 benchmark with a mean IOU of 79.7.

# Main Contributions

* Experimented with filter upsampling (atrous convolutions) to address feature resolution in DCNNs
  + Discussed applicability and practical implementation using existing architectures
  + Quantified the effects of modifying the effective field of view through the filter size and sampling rate
  + Demonstrated that segmentation performance is comparable to base models with significantly fewer parameters (3.36 times faster and 20.5M parameters compared to 134.3M baseline)
* Introduced the Atrous Spatial Pyramid Pooling (ASPP) to address object scale concerns
  + Proposed two schemes with different sampling rates
  + Showed that ASPP improves performance over the baseline
  + Found that larger sampling rates may yield better performance
* Implemented Conditional Random Fields to improve boundary detection
  + Qualitatively and quantitatively showed that CRFs can improve image segmentation offering up to a 3-5% absolute increase in mean IOU.

# High-Level Evaluation of Paper

The paper makes novel contributions to image segmentation techniques with comprehensive research to support their work. The authors highlight issues with existing approaches and how these motivate the proposed improvements. Their final models, which combine and build upon existing techniques, sometimes exceed state-of-the-art performance across different datasets. Visual aids showing the segmentation in practice highlight these improvements. However, in their discussions, the authors unnecessarily repeat information and critically do not address the limitations associated with their proposals. Explanations about sampling techniques and sampling rate selection, which could provide valuable context, are also omitted. Inexplicably, one of the more important figures in the paper—the illustration detailing atrous convolution in 2D—shows only the result of the technique and does not actually describe its operation. Despite these few shortcomings, the novel techniques described in the paper, its accessibility and impressive results qualify it as valuable research.

# Discussion on Evaluation Methodology

The authors of this paper did an excellent job establishing a thorough evaluation methodology for their proposals. Performance is measured using mean pixel intersection-over-union (IOU) and applied to several well-established datasets performing a variety of distinct image segmentation tasks. Given the nature of the problem, these datasets are labelled at a pixel level. All major evaluations, summarized in tables, show both the iterative improvement over a baseline by applying each technique proposed in this paper and a comparison to other state-of-the-art models. Results are also presented with visual aids to demonstrate the practical benefits of their models.

# Possible Directions for Future Work

As suggested in the paper, future work should aim to improve object boundary detection and increase efficiency when dealing with higher resolution images.