Paper Review and Notes For DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs

Jin Hyoung Joo
hyoungjoo.j@gmail.com

Abstract

This paper [1]

1. Key Points

1.1. Proposed Methods for Segmentation

The challenges of image segmentation and the authors' proposed method of handling each problem can be seen as below.

- Image segmentation needs to predict the output with reduced feature resolutions. The authors used *atrous convolutions* (convolution with upsampled filters) to help with this problem.
- Image segmentation needs to segment objects at multiple scales. The authors used *atrous spatial pyramid pooling (ASPP)* to help with this problem.
- Deep convolution networks reduce the localization accuracy due to its invariance. The authors used *fully connected Conditional Random Fields (CRF)s* to help with this problem.

1.2. Advantages of DeepLab

DeepLab has the following advantages.

- DeepLab can operate at fast speeds.
- DeepLab achieves high accuracy.
- DeepLab has a simple structure.

2. Technical Details

2.1. Atrous Convolutions

Reduced feature resolutions are due to repeated combinations of pooling and downsampling. Therefore to solve this problem, the last few pooling layers are removed and instead the filters are upsampled in the subsequent convolution layers. The original image dimensions are recovered using atrous convolutions and bilinear interpolation. Compared to regular convolution with larger filters, atrous convolution allows us to effectively enlarge the field of view of filters without increasing the number of parameters or the amount of computation.

2.2. Atrous Spatial Pyramid Pooling

One method of dealing with objects at multiple scales is to compute feature maps at multiple scales and aggregating the output feature maps. This approach increases the accuracy but it computationally expensive. Instead, DeepLab uses multiple parallel atrous convolutions with different sampling rates.

2.3. Fully Connected CRFs

The model's ability to capture fine details are boosted by employing a fully-connected Conditional Random Field. This method efficiently captures the fine edge details while also catering for long range dependencies.

3. Further Research

References

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 1