Saathi

6/8 June 2023

Topic: Encoder-decoder with atrous separable convolutions for semantic image segmentation

Abstract:
Spatial pyramid pooling module or encode-decoder structure are used in deep neural networks for semantic segmentation task. The former networks are able to encode multi scale contextual information by probing the incoming features with filters or pooling operations at multiple rates and multiple effective fields of view, while the latter networks can capture sharper object boundaries by gradually recovering the spatial information. In this work, we propose to combine the advantages from both methods. Specifically, our proposed model, DeepLabv3+ extends DeepLabv3 by adding a simple yet effective decoder module to refine the segmentation results especially along object boundaries. We further explore the Xception model and apply depthwise separable convolution to both Atrous Spatial Pyramid pooling and decoder modules, resulting in a faster and stronger encoder-decoder network. We demonstrate the effectiveness of the proposed model on PASCAL VOC 2012

Date __/__/__ Saathi

and cityscapes datasets, achieving the test set performance of 89.0% and 82.1% without any post-processing. Our paper is accompanied with a publically available reference implementation of proposed models in Tensorflow.

Conclusion

Our proposed model "Deeplabv3+" employs the encoder decoder structure where Deeplab3 is used to encode the rich contextual information and a simple yet effective decoder module is adopted to recover the object boundaries. One could also apply the atrous convolution to extract the encoder features at an arbitrary resolution, depending on the available computation resources. We also explore the Xception model and atrous separable convolution to make the proposed model faster and stronger. Finally, our experimental results show that the proposed model sets a new state of the art performance on PASCAL VOC 2012 and cityscapes datasets.