Week 5 paper summary

Sina Mehdinia

04/30/2021

Summary

He et al. [1] introduce residual learning deep learning method (ResNets) for image recognition. Going deeper with neural networks (adding more depth) rises a problem which is not from overfitting but a degradation when accuracy gets saturated. The core idea of ResNet is identity shortcut connections that skip one or more layers. The output of shortcut connections are simply added to the output of stacked layers. The authors show that the performance of the ResNet models are better (lower training and testing error with increasing depth) than plain architecture (simply stacking layers) and other famous networks such as VGG. The results are tested on ImageNet as well as some other datasets. It is important that the shortcut connections do not add any complexity and parameters to the model. There is not dropout in its implementation. The residual block has two convolutional layers with the same number of output channels. Each convolutional layer is followed by a batch normalization layer and a ReLU activation function. Then, the input is directly added before the final ReLU activation function. The output of the convolutional layers has to be of the same shape as the input, so zero padding or projection shortcut can be performed. Deeper ResNets are a bit different than 34 layer version to account for training time. There are stack of three layers instead of two. The middle convolution layer is a bottleneck with two 1×1 layer before and after it to first reduce and then restore the dimensions. The ResNet 152 is still less complex than VGG nets.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.