Review of “Deep Residual Learning for Image Recognition”

Riyad Bin Rafiq

1. Paper summary

The practice of the deep Convolutional Neural Network (CNN) started when AlexNet won the ILSVRC in 2012. But there was a limitation before then because very deep-layer networks performed worse as depth increased. So the paper [1] addressed the problem and proposed a new technique called residual learning for training very deep CNN architecture without any performance degradation. The paper applied 152-layered CNN on the ImageNet dataset and an ensemble of these residual nets achieved 3.57% error on the test set. The authors experimented with the layer size ranging from 18 to 1000. They also applied the network to different datasets including CIFAR-10, COCO etc. that proved the generalization power of the architecture across different datasets and tasks.

2. Contribution

2.1 Very deep-layered networks with less complexity

The paper introduced very deep-layer CNN architectures with great performance accuracy. In the paper, the authors introduced 34, 50, 101 and 152-layer ResNet and applied the model on different datasets. They also showed how the proposed deep models performed compared to deep plain CNN models and state-of-the-art models. Moreover, the ResNet architectures have less parameters than VGG although the 152-layer Resnet is 8 times deeper than VGG. In my opinion, this is one of the best contributions in the field of computer vision.

2.2 Residual learning to solve degradation problem

Very deep networks converge in terms of accuracy, when more layers are added to the network. The paper introduced a new approach named residual learning by shortcut connections that removes the degradation problem in very deep networks. In the paper, they applied residual learning to every few stacked layers that introduce neither extra-parameter nor computation complexity. If the dimensions of the input and the output of residual function don’t match, the authors proposed three options for this particular problem using shortcut connections. Moreover, to develop more deep ResNet architecture (such as 50,101, etc), the authors also described how they applied residual function using a stack of 3 layers instead of 2.

3. Critique

3.1 What is the reason for the degradation problem?

Deep CNN networks performed poorly as more layers were added in the network. The authors mentioned this issue as a degradation problem. At first, I thought that this problem occurred due to the vanishing gradients. But in the paper, the authors stated that the optimization difficulty was unlikely to be caused by vanishing gradients. They also mentioned an example where deep plain networks with batch normalization didn't exhibit any vanishing gradients. So the reason for the performance degradation is still unknown. The paper didn't tell any probable reasons for this. I think the reason for this optimization difficulty may be discovered in this research as it has solved the problem.

3.2 Layers of residual function

In the paper, it was stated that the residual function was flexible and the authors implemented only two or three layers in the function. But a question emerges as to how the model performs if more layers are added in the function. Also, the authors said that there was no advantage in applying one layer in the function. But no result was given regarding this. I think a comparison study could be provided regarding the number of layers of the residual function and by supplying this, the role of residual learning could be more clarified.

3.3 Lack of proof for overfitting with 1000 layers

The authors developed a very deep network (1202-layer network) on the CIFAR-10 dataset. The method didn't exhibit any optimization difficulty and the training error was low. The test error was high and they said that it was because of overfitting. But this was not proved as the research didn't apply any regularization technique such as max out or dropout in training the deep ResNet. If dropout was applied and the test error became low, then we could mention that performance degraded due to overfitting. Otherwise, some new question emerges for these very deep networks. What were the reasons for low test accuracy? In my opinion, the paper overlooked this scenario.

Reference

1. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 770–778.