Rahul Thapa
Paper: Deep Residual Learning for Image Recognition

Summary

In this paper, the authors talk about a very deep network with shortcuts that they designed for image recognition on popular datasets such as ImageNet and COCO. The authors acknowledge the fact that a deeper neural network is difficult to train. They present an approach of explicitly reformulating the layers as learning residual functions with reference to the layer inputs instead of learning unreferenced functions. With this approach, the authors are able to train a very deep neural network over 1000 layers without the problem of overfitting and vanishing gradient, prevalent in deep neural networks.

The papers present recent evidence about network depth being of crucial importance to the performance of the network. However, on training a very deep neural network, a problem very persistent is vanishing/exploding gradient problem. This hampers convergence from the very beginning. This problem can be effectively tackled using normalization; however, it still cannot solve the problem of degradation. The authors describe the degradation problem as, with the increase in the network depth, accuracy gets saturated and then degrades rapidly. Unexpectedly, the authors found out that such degradation is not caused by overfitting. In this paper, the authors address the degradation problem by introducing a deep residual learning framework. In this approach, the authors let the stacked layers fit a residual mapping. The formulation of this network can be realized by a feedforward neural network with shortcut connections i.e. skipping one or more layers. In this paper, the shortcut connections simply perform identity mapping, and their outputs are added to the outputs of the stacked layer. The unique feature of identity mapping is it does not add extra parameters to the network and does not increase the network complexity. This allows the author to fairly compare between plain and residual networks that simultaneously have the same number of parameters, depth, width, and computation cost. The network that the authors used in this paper have fewer filters and lower complexity than VGG nets, previously developed in dealing with the same dataset.

There are two approaches authors used to introduce the shortcuts in the plain network are: First, the shortcuts perform identity mapping with extra zero entries padded for increasing dimensions. This approach introduces no extra parameters. Second, the projection shortcut is used to match dimensions. The identity shortcuts can be directly used when the input and the outputs are of the same dimension. We only use the above two options when the dimension increases.

The results show that the deeper 34-layer plain net has a higher validation error than the shallower 18-layer plain net. The authors observed degradation problem to be the reason for this increase in a validation error. However, in the case of the residual network, the validation error decreased when increasing the size of the network from 18 to 34-layers. This indicated that the degradation problem is well addressed in the setting and we manage to obtain accuracy gains from increased depth.

Strengths

- The Paper is Organized very nicely with multiple headings and subheadings explaining a certain aspect of the experiment/network.
- The paper has multiple figures and tables that help understand their result better and help understand how good the resnet is in comparison to other approaches.

- For those who want to learn more about their experiment and network architecture, there is a dedicated appendix section.

Weaknesses

- There isn't any concluding paragraph in the paper, which made it difficult to understand what exactly they achieved while skimming the paper for the first time. It might not be in favor of paper as many researchers tend to skim through the paper to see if the paper interests them in the beginning.
- The introduction is long and some of the things he says in the introduction, he repeats in other sections too.
- It is hard to understand the math behind how the residual network works. They could have explained it a bit more clearly.

Confusion

- What is Identity shortcut and how is it different than projection shortcut.
- How exactly does this math works: $F(x) := H(x) – x$? What are $F(x)$ and $H(x)$ in the equation?

Discussion

- What is dehydration problem and how does resnet deals with it?
- What exactly happens when you have shortcuts in the network? How does it deal with the problem of overfitting even for a significantly deeper network being trained on not so big sized data?
- The author mentions that deep residual networks may be used for applications that object detection. What other applications do you think deep resnet will be able to achieve a good result?