# Deep Residual Learning for Image Recognition

Ravi Raju

12-15-2018

## 1 Motivation

This paper introduces the notion of deep residual networks. Residual networks are those that have connections from previous layers into later layers of the neural network. This is motivated from the fact that the authors had done some preliminary work and found that a degradation existed when more layers were added to the network. This is clearly seen in Figure 1.

Even when identity layers were added performance dropped while the norm of the gradients were not ill-behaved so it does not seem to indicate exploding/vanishing gradients.

## 2 Solution

The solution introduced was to use shortcut connections to skip layers to deeper layers in the network. The formal idea presented here is that we would like to learn some representation, $H(x)$. Learning $H(x)$ is difficult so we reformulate the problem into $F(x) := H(x) - x$. The logic here is very simple: if the identity mapping is desired, it is easier to push the stacked layers, the residual block, to zero rather than learning the identity block.

## 3 Results

The results, for its time, are impressive. No other network architecture is able to compete with it. One interesting note is that their 1202 layer network performs similarily to one at 100 layers suggesting that there is some upper bound when it comes to adding depth to a network.

## 4 Discussion/Takeaway

There is not so much left to discuss other than the fact that it is annoying that how backprop is changed. I would assume I guess it's not the different since it was not mentioned detail. They should have tried to see if the inception model from Google could be incorporated into the residual block.