RESNET SUMMARY

Literature preceding the “Deep Residual Learning for Image Recognition” paper showed that deeper Convolutional Neural Networks performed better than shallow networks on large datasets such as the ImageNet dataset. In this paper, the authors Kaiming He et al discuss the most prominent issue faced in training deeper networks, the vanishing gradients problem. To overcome the difficulties in training deeper networks, the authors introduce a novel approach, called residual learning, where the key idea is to reformulate network layers in such a way such that they learn residual functions, i.e. the difference between the desired output and the input to the layer, instead of learning the desired output entirely.

Residual Blocks:

The authors discuss how various experiments performed by them showed that as we increase network depth, the accuracy first saturates, and then starts degrading, which counter-intuitively, is not a result of overfitting, as it is seen that with the addition of layers, even the training error gets higher; this is indicative of the issue in optimization. These findings, hence, proved to be the motivation behind the introduction of skip connections - identity mappings that skip one or more layers in order to let information flow unchanged, preventing the degradation in accuracy.

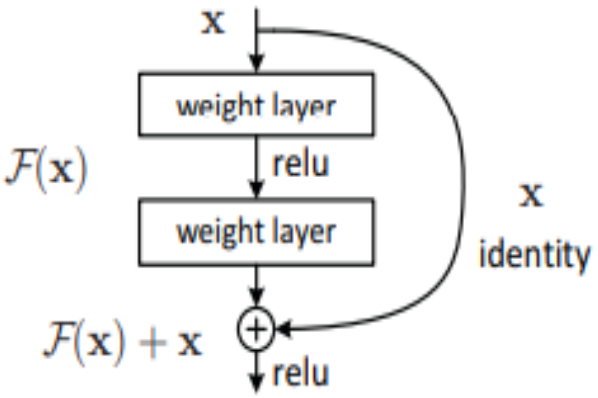

Figure 2. Residual learning: a building block.

Consider the above diagram of a residual block (a single block of layers with a skip connection):

Let x be the input to the above block of layers, and H(x) be the desired mapping learned by the block,

The original mapping H(x) is recast into F(x) + x, where F(x) is known as the residual mapping.

The skip connections allow for the models to learn the residual mapping F(x) = H(x) – x, with the hypotheses that it is easier to learn this residual mapping, than the original mapping.

Consequently, if the identity mapping (x) is optimal, the optimization would simply tend the weights to zero, therefore tending F(x) towards zero.

Architectures Proposed:

The authors designed several residual network (Resnet) architectures based on the proposed residual learning framework. Key architectural details include:

Basic Building Block:

Each residual block consists of two or three stacked layers, with shortcut connections performing identity mapping. The output of each block is the sum of the block's layers' output and the block's input, helping to avoid degradation as the network grows deeper.

Bottleneck Architecture:

For deeper networks, the authors employed a "bottleneck" design to reduce computational complexity. Each bottleneck block has three layers: a 1×1 convolution to reduce dimensions, a 3×3 convolution, and another 1×1 convolution to restore dimensions.

The paper tested networks with various depths: 18, 34, 50, 101, and 152 layers. These networks are much deeper than previously common architectures, such as VGG, which had up to 19 layers.

Notable Results:

The 152-layer ResNet achieved a top-5 error rate of 3.57%, outperforming both VGG-16 and GoogLeNet. This result won the 1st place on the ILSVRC 2015 classification task.

Various Resnets also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.