*Deep Residual Learning for Image Recognition:*

Deeper neural networks are difficult to train. During training a deeper network, one of the obstacle is vanishing/exploding gradients. This problem is largely normalized by normalized initialization and intermediate normalized layers .When deeper layer are able to start converging, a degradation problem has been exposed. Unexpectedly, this problem is not caused by overfitting.This problem is addressed by using a deep residual learning framework.

In residual learning, instead of hoping each few stacked layers directly fit a desired underlying mapping, we explicitly let these layers fit a residual mapping.

Formally, denoting the desired underlying mapping as H(x), we let the stacked nonlinear layers fit another mapping of F(x) := H(x)−x. The original mapping is recast into F(x)+x. We hypothesize that it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. To the extreme, if an identity mapping were optimal, it would be easier to push the residual to zero than to fit an identity mapping by a stack of nonlinear layers. The formulation of F(x) +x can be realized by feed forward neural networks with “shortcut connections”.

Using residual we found:

: 1) Extremely deep residual nets are easy to optimize, but the counterpart “plain” nets (that simply stack layers) exhibit higher training error when the depth increases; 2) Deep residual nets can easily enjoy accuracy gains from greatly increased depth, producing results substantially better than previous networks.

These shortcut connections are of two types:

> Identity connection y = F(x, {WI}) + x.

>Projection or Convolution connection y = F(x, {Wi}) + Wsx

(We actually perform a convolution step with weight matrix as Ws so as to make dimension of Wsx = F(x, {Wi})

The dimensions of x and F must be equal when doing identity mapping if dimensions changes we have two options:

(A) The shortcut still performs identity mapping, with extra zero entries padded for increasing dimensions. This option introduces no extra parameter; (B) The projection shortcut in is used to match dimensions (done by 1×1 convolutions).

We majorly use identity connections when dimensions are same and go with option B when dimensions changes.

Through building various models on Imagenet ,CIFAR10, it is found residual networks results in good yield .Increasing the depth decreases the training error and also the validation error.