Resnet Architecture

Problem: In theory, deeper neural networks should result in better performances. However, with more layers, the problem of vanishing/exploding gradients arise. While it is addressed largely in normalized initialization and intermediate normalization layers, it still exists with an increase of layers. Another problem was degradation, where an increase of layers causes degrading in training accuracy from back-propagation.

Solution: The proposed architecture residual net includes a shortcut through every two layers (stride of 2). The shortcut is an identity function that carries x to be added to F(x) after the two layers. In the case of vanishing/exploding gradients, the addition allows for direct propagation back to any desired layer, and prevents the problems caused from going through the weighted layers. In the case of degradation, after the network reaches an optimal solution, no matter how many more layers it needs to go through, the optimal solution can be passed down through the shortcut which prevents degrading of training accuracy. Also proven through tests.

Interpretation: The resnet Architecture can be thought of as a simple game of telephone, where the information passed through will drive further away from the original with more people. And the resnet acts as a shortcut where the information is directly passed on from person one to person three, increasing the speed and accuracy.

Advantages: The resnet architecture solves the problems of vanishing/exploding gradients and degradation during training of deep neural networks.

Achievements: 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation in ILSVRC & COCO 2015 competitions.

Alexnet Architecture

Architecture: Some of Alexnet's key modifications were data augmentation, dropout regularization, Rectified Linear units, and MaxPooling. Data augmentation makes sight changes to the data such as cropping, brightness, color to improve robustness to changes of data between the training and validation data. Dropout is used to prevent overfitting and reduce complex co-adaptation of neurons. ReLu replaced sigmoid nonlinearity to prevent vanishing gradients since a half of the function was an identity function. MaxPooling replaced average pooling to allow more shift invariants in features.

Advantages: Prevents overfitting using data augmentation and dropout. Prevents vanishing gradient with Relu function. And allows more variation in features with MaxPooling.

Disadvantages: Much more complex and memory consuming. Needed two GPUs to run at the time since Alexnet required more memory space. Also needed the two GPUs to synchronize through scripts.

Achievements: ImageNet LSVRC-2010 contest top-1 and top-5 error rates. ILSVRC-2012 competition top-5 test error rate.