GoogLe Net

Introduction

GoogLe Net is a deep CNN architecture, codenamed Inception, which set the SOTA for ILSVRC14. It is a 22 layer deep network with its architectural designs based on the Hebbian principle.

GoogLe Net was designed in such a way that the depth and width of the network was increased without increasing the computational budget. No new data sources were used for training. The network uses 12X less parameters as compared to AlexNet, proposed in 2012.

The Network-in-Network approach proposed by the Network in Network paper is used in the architecture. When applied to convolutional networks, it acts like a 1x1 convolutional layer followed by a ReLU activation layer. This method has dual purpose in GoogLe Net: first one as defined above and second one being that it acts as a dimension reducing module. This allows for increase not only in depth but also in width.

The brute force approach of increasing performance of a deep neural network is to increase its size, not only the depth but also the width of the network. But, this approach has two major drawbacks: one being that it increases the number of parameters thus the enlarged model is more prone to overfitting, the other being that it increases the use of computational power and resources.

Architecture

The architecture of the GoogLe Net consists of some repeated units known as the "Inception Module". Before the first Inception Module, the input image goes through 7x7, 3x3, 1x1, 3x3 convolutional layers in that order, followed by a 3x3 max pool layer. After this, the result then passes through 9 Inception Modules which are connected by depth concatenation. After the last module, the result encounters a 7x7 average pooling layer following which there is the Fully Connected layer. The result of the FC layer is then fed to a softmax layer to classify.
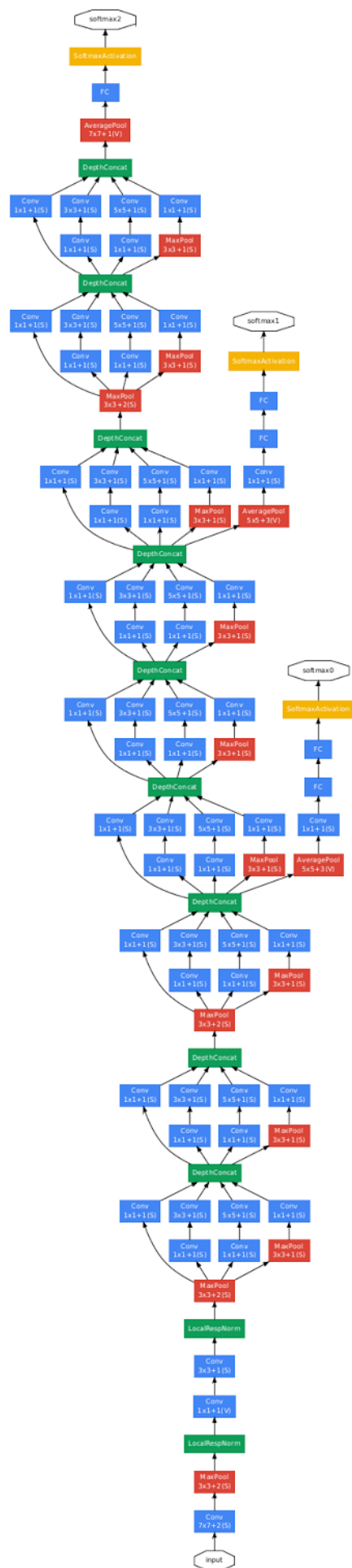
Inception Module:

- The "naive" inception module performs the following operations on the input:
  - Convolution with 1x1 kernel size
  - Convolution with 3x3 kernel size
  - Convolution with 5x5 kernel size
  - Max pooling
- The problem with this is that computation of these convolutions is very expensive. Thankfully, there's a solution to it.
- The solution is to first convolve the input with 1x1 kernels to reduce the number of channels and then convolve the effective output with 5x5 kernels. This method reduces the computation by ~10x.
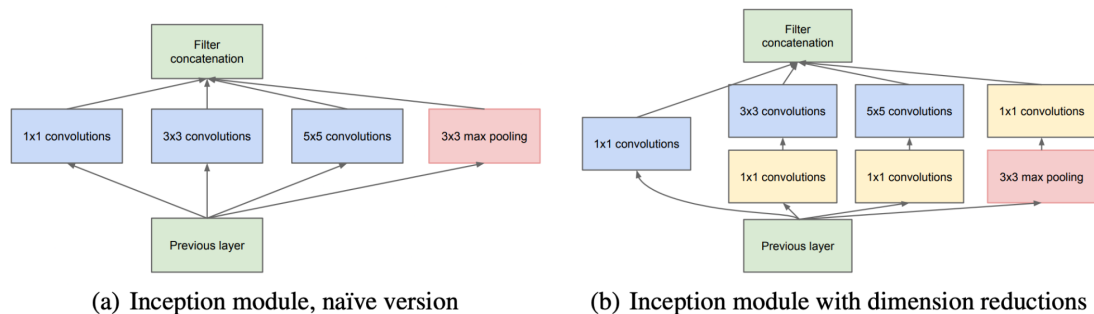

(a) Inception module, naïve version   (b) Inception module with dimension reductions

Figure 2: Inception module

Auxiliary Classifier:

- Another problem with such a deep network is vanishing gradients and poor flow of gradients during BackProp.
- To tackle this problem, the paper uses Auxiliary Classifiers. They are connected to the intermediate layers to help the gradient signals to propagate back. These are added during training and discarded during inference.
- Every auxiliary classifier's architecture is:
  - Average pooling layer of 5x5 kernel with a stride of 3.
  - Convolution layer of 128 1x1 kernel size.
  - FC layer of 1024 units and ReLU activation.
  - A dropout layer of 70%.
  - A softmax layer for classification.
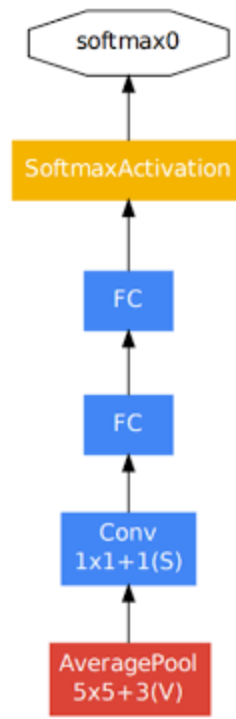
Conclusion

Below is the result of the ILSVRC classification challenge:

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

GoogLe Net is thus a novel architecture which provides for significant quality gain for a little extra computation expense when compared to shallower and less wide networks.