GoogleNet (Inception-V1)

GoogleNet also known as Inception-V1, was a groundbraking CNN architecture introduced in 2014 by researchers at Google. It marked a significant shift in CNN design by focusing on **computational efficiency** and **multi-scale processing**, allowing for deeper and wider networks without an explosion in computational cost and parameters. GoogleNet won the ImageNet challenge in 2014, as it achieved a top-5 error rate of 6.67% which was a significant improvement over previous winners like ZFNet (2013 winner, 11.7% error rate) and AlexNet (2012 winner, 15.3% error rate).

The core idea behind GoogleNet was that most activations in a deep network are either unnecessary (zero value) or redundant due to correlations. Therefore, a more efficient architecture would have **sparse connections** between activations. This led to the development of Inception module.

GoogleNet achieved better performance with drastically fewer parameters: VGGNet had 138 million parameters, while GoogleNet had only 6.8 million parameters. This efficiency was a major improvement, enabling deeper and wider connections/network without the associated computational burden.

Architecture details: The Inception Module

The most distinctive feature of the GoogleNet is the Inception module. Instead of choosing a single filter size at each layer, the inception module performs multiple convolution operations with different filter sizes (1X1, 3X3, 5X5) and max pooling operation (3X3) in parallel on the same input feature map. The output of these parallel operations are then concatanated along the depth dimension. This allows the network to capture features at various scales and resolutions simultaneously.

A crucial component in the Inception module is the use of **1X1 convolutions** (bottleneck layers). These are applied before thee larger 3X3 and 5X5 convolutions. The purpose of 1X1 convolution is to:

→ **Reduce dimensionality:** They reduce the number of channels (the feature maps) of the input before feeding them into larger, more computationally expensive convolutions. This significantly reduces the computation cost.
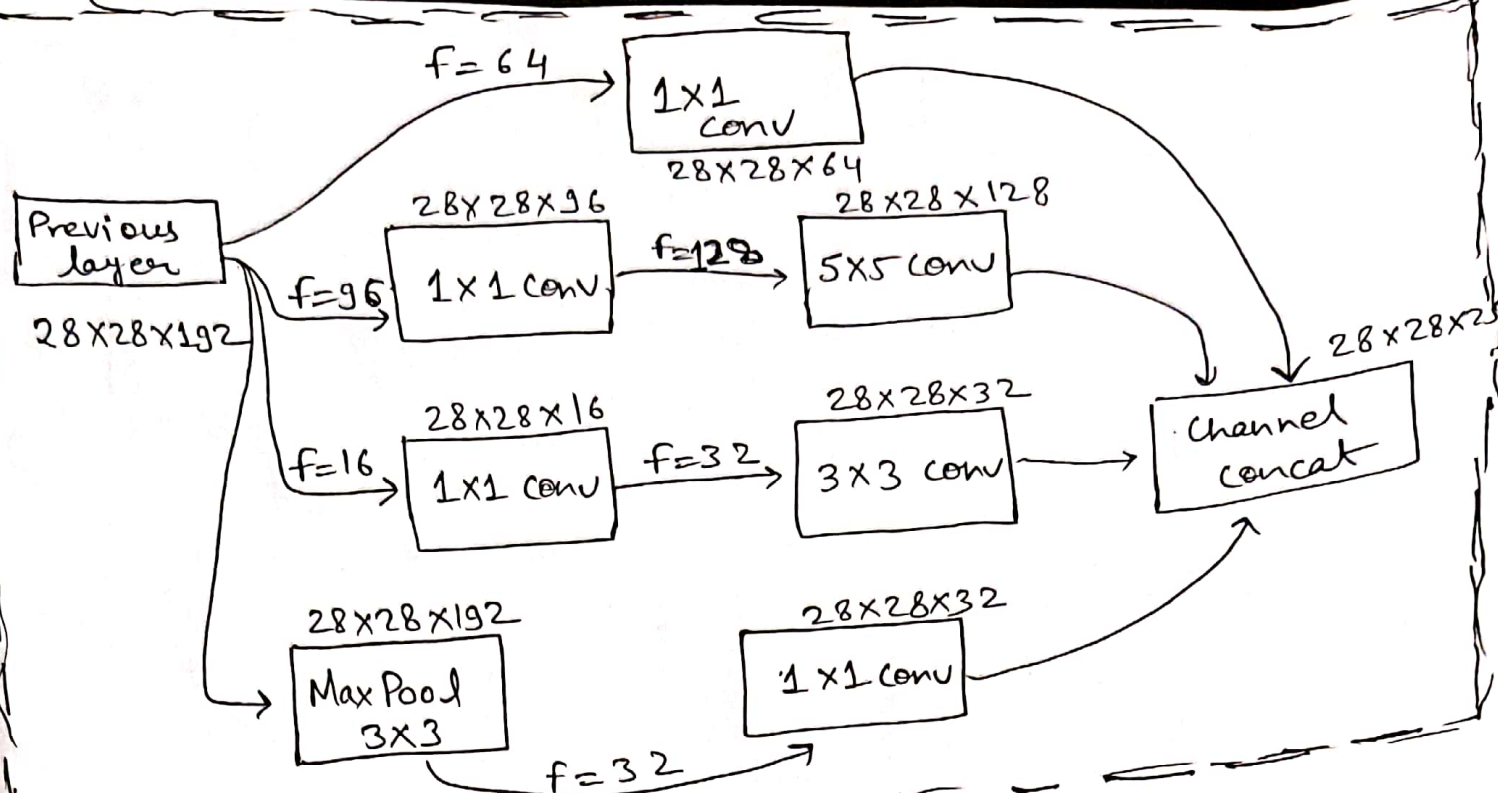
→ **Add Non-linearity:** They incorporate a ReLU activation function, adding more non-linearity to the network without affecting the receptive field.

Other key architectural features of Google Net include:

(a) Global Average Pooling

(b) Auxiliary classifiers at intermediate layers.

✱ The overall GoogleNet architecture is **22 layers deep** (excluding pooling layers) and consists of **9 Inception modules** stacked linearly.

f = 64

1x1 Conv
28X28X64

28X28X96

Previous layer
28X28X192

f=96 1x1 Conv → f=128

28 X28 X128
5X5 Conv

28 X 28 X2[5]

28X28X16

f=16 1x1 Conv → f=32

28X28X32
3X3 Conv

Channel concat

28X28X192

Max Pool 3X3

28X28X32
1X1 Conv

f=32

Example of an inception module

*Pros (Advantages):—

(a) High Accuracy ⇒ GoogleNet achieved state of the art accuracy in ILSVRC 2014 challange.

(b) Computationally Efficient

(c) Fewer Parameters

(d) Multi-scale feature extraction ⇒ The Inception module effectively captures features at different scales, making the network robust to variation in object sizes.

(e) Regularization ⇒ Auxiliary classifiers provided a regularization effect, helping to prevent overfitting and ensure that intermediate layers learned useful features.

* Cons (Disadvantages):—

(a) High Complexity ⇒ The inception module with its parallel branches and 1x1 convolutions, is more complex to design and understand than sequential architectures like AlexNet and VGG Net.

(b) It is still a deep network.