

**Deep Learning**

**GoogLeNet CNN architecture**

**Prepared by: Yared Terefe (BSc)**

**Submitted to: Fantahun Bogale (PhD)**

**Submitted date:09/08/2024**

# GoogLeNet (Inception-v1) Architecture

**GoogLeNet**, also known as **Inception-v1**, was a groundbreaking convolutional neural network (CNN) architecture introduced by Google in 2014. It made significant strides in the field of image classification and object detection.

## Historical Context

At the time, there was a trend towards deeper neural networks, as researchers believed that increased depth would lead to better performance. However, simply stacking more layers could lead to vanishing gradients, a problem where the gradients, used to update the network's weights, become very small, hindering the learning process.

On top of that the deeper the model, the more parameters that had to be learned, which slowed down the training process and needed much more computing resources.

## GoogLeNet's Innovation: The Inception Module

GoogLeNet addressed this challenge by introducing the **Inception module**. This module was designed to efficiently extract features at multiple scales simultaneously. It consists of a parallel combination of convolutional filters with different sizes (1x1, 3x3, and 5x5) and pooling layers. The outputs of these parallel layers are then concatenated to form a single output tensor.

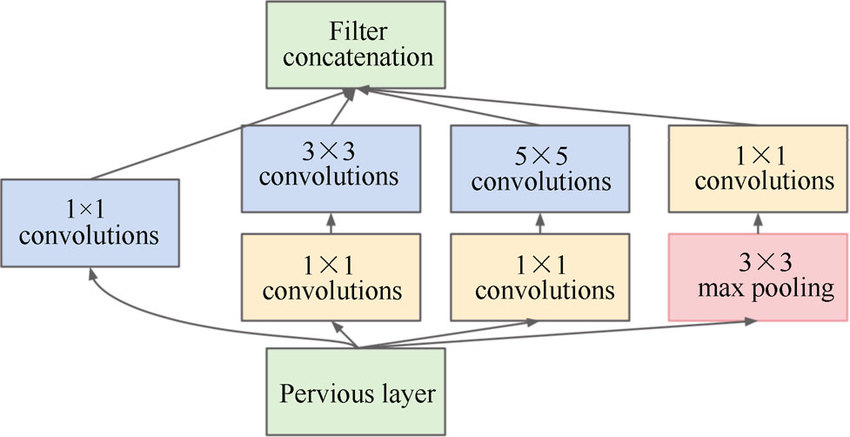


Figure 1. Inception module – taken from [Image-link](https://www.researchgate.net/figure/nception-module-of-GoogLeNet-This-figure-is-from-the-original-paper-10_fig3_312515254)

It accepts data from previous layer and uses mainly 3 kernel sizes and a max pooling function so that It’s able to extract features based on the different sizes of the kernels. The network then can learn effective features by adjusting the parameters of these kernels accordingly.

Before applying the 3x3 and 5x5 kernels, it first applies a reducing convolution with the 1x1 kernel this is mainly for the purpose of size reduction on our original image, the output from this kernel then can be effectively used to apply the 3x3 and 5x5 kernels, which are later concatenated into one.

## Key advantages of the Inception module

* Efficient feature extraction – Because of the different kernel sizes it can extract multiple area features in one go.
* Computational efficiency – This is mainly achieved by first using the 1x1 kernel to reduce the size of the image and the feature map from the previous layers
* Reduction in Overfitting: The inception module and global average pooling help in reducing overfitting by introducing sparsity in the connections and reducing the total number of parameters.

## The GoogLeNet architecture

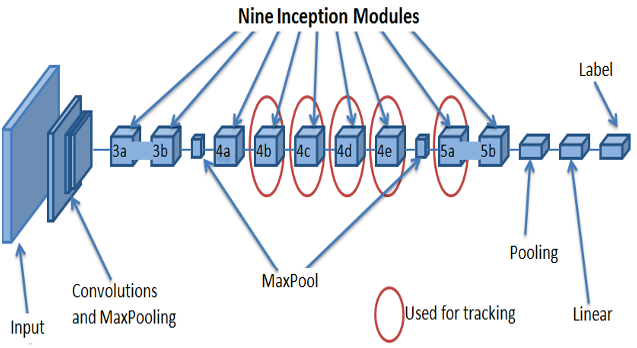


Figure 2 – Main GoogLeNet Architecture – image taken from [image-link](https://research.iitj.ac.in/publication/deep-googlenet-features-for-visual-object-tracking)

As seen in the image it incorporates 9 inception modules, named from 3a to 5b. And max pooling and convolutional layers are applied along the way. In the code file u can see the inception modules being prepared with different parameters for the kernel sizes, and max pooling size, that’s taken directly from their architecture design which they found with trial and errors adjusting the values as they saw fit.

For larger datasets with color images, this network is highly efficient and captures complex relationships, which would have been harder for the regular CNN’s to learn cause of their static single sized kernels.