Alex Net,Google Net,Vgg

AlexNet:

AlexNet, developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, is a landmark model that won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2012. It introduced several innovative ideas that shaped the future of CNNs.

AlexNet Architecture:

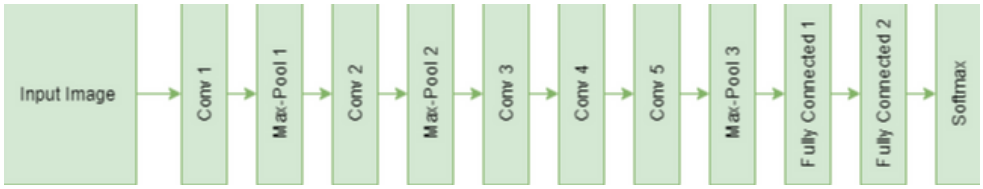
AlexNet consists of 8 layers, including 5 convolutional layers and 3 fully connected layers. It uses traditional stacked convolutional layers with max-pooling in between. Its deep network structure allows for the extraction of complex features from images.

- The architecture employs overlapping pooling layers to reduce spatial dimensions while retaining the spatial relationships among neighbouring features.
- Activation function: AlexNet uses the ReLU activation function and dropout regularization, which enhance the model's ability to capture non-linear relationships within the data.

The key features of AlexNet are as follows:-

- AlexNet was created to be more computationally efficient than earlier CNN topologies. It introduced parallel computing by utilising two GPUs during training.
- AlexNet is a relatively shallow network compared to GoogleNet. It has eight layers, which makes it simpler to train and less prone to overfitting on smaller datasets.
- In 2012, AlexNet produced ground-breaking results in the

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC). It outperformed prior CNN architectures greatly and set the path for the rebirth of deep learning in computer vision.
- Several architectural improvements were introduced by AlexNet, including the use of rectified linear units (ReLU) as activation functions, overlapping pooling, and dropout regularisation. These strategies aided in the improvement of performance and generalisation



Let's consider an image classification task of various dog breeds. AlexNet's convolutional layers learn features such as edges, textures, and shapes to distinguish between different dog breeds. The fully connected layers then analyze these learned features and make predictions.

GoogleNet:

GoogleNet is also known as (Inception v1), it was developed by a team at Google led by Christian Szegedy. It won the ILSVRC in 2014 and introduced several innovative concepts that aimed to address the challenges faced by deep neural networks.

Inception Modules: GoogleNet utilizes **inception** modules which use a deep, multi-branch architecture. It is composed of multiple parallel convolutional layers with different filter sizes. This allows the model to capture features at various scales and resolutions simultaneously.

- **Dimensionality Reduction:** To reduce computational complexity and improve efficiency, GoogleNet employs 1×1 convolutional layers for dimensionality reduction before apply

convolutions. This helps to preserve important spatial information while reducing the number of parameters. Auxiliary Classifiers: GoogleNet uses auxiliary classifiers at intermediate layers during training to combat the vanishing gradient problem and provide additional regularization.

The key features of GoogleNet are as follows:

- GoogleNet tried to overcome deep CNNs' computational inefficiencies. It uses the Inception module which reduces the number of parameters in the network and boosts computing efficiency. It outperformed AlexNet in terms of accuracy while using fewer parameters compared to AlexNet.
- GoogleNet is a considerably deeper network with 22 levels. Its depth enables it to collect more intricate characteristics and patterns from images, allowing it to perform better on larger and more complicated datasets.
- In 2014, GoogleNet won the ILSVRC, beating AlexNet. It demonstrated the efficiency of its Inception module by achieving improved accuracy while utilising fewer parameters.

In the context of image recognition, GoogleNet excels at capturing both fine-grained details and high-level features. For instance, when identifying objects within an image, GoogleNet's inception modules can simultaneously detect small-scale details. Details like facial features and larger-scale patterns like object shapes and textures.

VGG-Net:

The Visual Geometry Group (VGG) models, particularly VGG-16 and VGG-19, have significantly influenced the field of computer vision since their inception. These models, introduced by the Visual Geometry Group from the University of Oxford, stood out in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) for their deep convolutional neural networks (CNNs) with a uniform architecture. VGG-19, the deeper variant of the VGG models, has garnered considerable attention due to its simplicity and effectiveness.

Detailed Layer-by-Layer Architecture of VGG-Net 19:

1. Convolutional Layers: 3x3 filters with a stride of 1 and padding of 1 to preserve spatial resolution.
2. Activation Function: ReLU (Rectified Linear Unit) applied after each convolutional layer to introduce non-linearity.
3. Pooling Layers: Max pooling with a 2x2 filter and a stride of 2 to reduce the spatial dimensions.
4. Fully Connected Layers: Three fully connected layers at the end of the network for classification.
5. Softmax Layer: Final layer for outputting class probabilities.