Convolutional Neural Network(CNN)

AlexNet :

- AlexNet is a deep convolutional neural network (CNN) architecture that revolutionized computer vision by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 with a significant margin over previous models.
- It was developed by **Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton**.

Architecture of AlexNet :

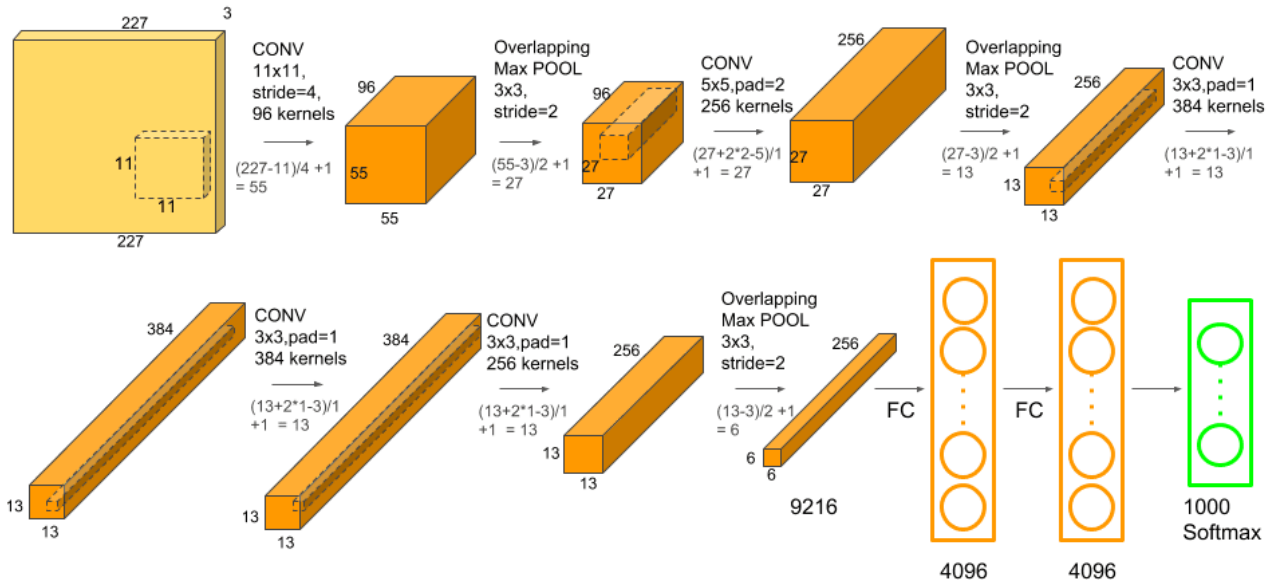
AlexNet consists of Eight layers, including :

1) **Five convolutional layers (Conv Layers)** – Used for feature extraction.
2) **Three fully connected (FC) layers** – Used for classification.

Important Parameters

- **Input :** 227×227×3 image (RGB)
- **Conv Layers :** Feature extraction.
- **Max Pooling :** Reduces spatial dimensions while preserving important features.
- **Dropout Layers :** Prevent overfitting.
- **Fully Connected Layers :** Convert extracted features into class probabilities.
- **Output :** Provides probability distribution over 1000 classes.

Architecture Diagram :



Innovations of AlexNet :

- **ReLU Activation Function:** Faster training compared to traditional sigmoid/tanh.
- **Overlapping Max Pooling**: Better feature extraction.
- **Dropout Regularization**: Prevents overfitting.
- **GPU Acceleration**: Trained using **two NVIDIA GTX 580 GPUs**.

VGGNet :

- VGGNet is a deep convolutional neural network (CNN) architecture developed by the Visual Geometry Group (VGG) at the University of Oxford.
- It was introduced in 2014 and became famous for its simplicity and effectiveness in image classification.

Architecture of VGGNet :

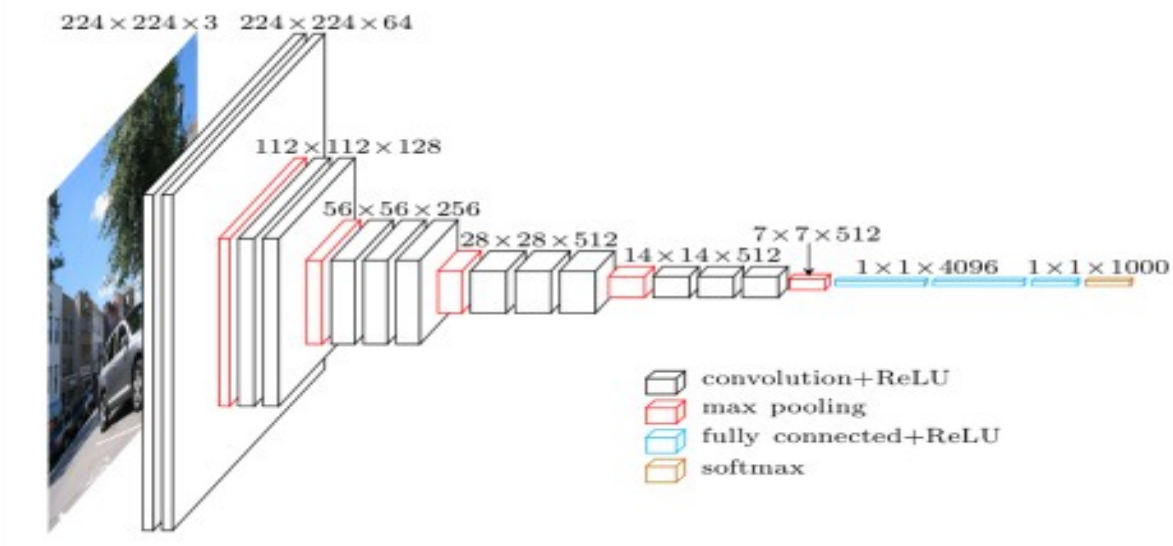
VGGNet comes in different versions, such as **VGG-16** and **VGG-19**, where the numbers refer to the total number of layers.

- **Convolutional Layers (Conv Layers) –** Uses multiple **3×3 convolutional filters** stacked back-to-back.
- **Max Pooling Layers** – Reduces spatial dimensions using **2×2 pooling**.
- **Fully Connected (FC) Layers** – Used for classification.
- **Softmax Output** – Predicts probabilities for **1000 classes** (ImageNet dataset).

Important Parameters :

- **Input**: 224×224×3 image (RGB).
- **Conv Layers**: Feature extraction using small **3×3 filters**.
- **Max Pooling**: Reduces dimensions, keeping important features.
- **Fully Connected Layers**: Three FC layers before classification.
- **Softmax Activation**: Outputs class probabilities.

Architecture Diagram :



Innovations of VGGNet :

- **Deep architecture**: Increased depth (16 or 19 layers) improved performance.
- **Small 3×3 filters**: Better feature extraction while keeping parameters manageable.
- **Uniform structure**: Uses only **3×3 conv layers** and **2×2 max pooling**, making it easier to understand.
- **Increased computation**: More layers lead to higher accuracy but require more resources.

GoogLeNet (Inception v1) :

- GoogLeNet is a deep convolutional neural network (CNN) architecture developed by Google's research team.
- It won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014, outperforming previous architectures like AlexNet and VGGNet in accuracy and efficiency.

Architecture of GoogLeNet :

GoogLeNet consists of **22 layers** (excluding pooling layers) and introduces the **Inception module**, which improves computational efficiency.

- **Inception Modules** – Uses multiple filter sizes (1×1, 3×3, 5×5) in parallel for better feature extraction.
- **1×1 Convolutions** – Used for dimensionality reduction and computational efficiency.
- **Global Average Pooling (GAP)** – Replaces fully connected layers to reduce parameters.
- **Softmax Output** – Predicts class probabilities for **1000 classes** (ImageNet dataset).

Important parameters :

- **Input**: 224×224×3 image (RGB).
- **Inception Modules**: Extracts multi-scale features in parallel.
- **1×1 Conv Layers**: Reduces the number of parameters.
- **Global Average Pooling**: Instead of fully connected layers, it averages feature maps.
- **Auxiliary Classifiers**: Extra classifiers during training to improve gradient flow.

Innovations of GoogLeNet :

- **Inception Module**: Processes multiple filter sizes at the same time.
- **Fewer Parameters**: Only **5 million parameters**, compared to **138 million in VGG-16**.
- **Better Accuracy with Less Computation**: Deeper but more efficient than AlexNet and VGGNet.
- **Auxiliary Classifiers**: Helps during training by preventing vanishing gradients.

Architecture Diagram :