CNN MODELS

ALEXNET

What is Alexnet?

- AlexNet is a pioneering deep learning architecture introduced in 2012 by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton.
- It marked a major breakthrough in computer vision, particularly by winning the ImageNet Large Scale Visual Recognition Challenge with impressive results.

Key Features:

1. **Layer Architecture**: AlexNet comprises 8 layers—5 convolutional layers (to extract features like edges and textures) and 3 fully connected layers (for classification).
2. **ReLU Activation**: Uses Rectified Linear Units (ReLU), which speed up training by introducing non-linearity.
3. **Dropout**: Introduces dropout in fully connected layers to reduce overfitting.
4. **Max-Pooling**: Combines convolutional layers with max-pooling to reduce the spatial dimensions and retain important features.
5. **Parallel Processing**: Utilizes two GPUs for training, splitting the network to handle large datasets.

Architecture:

- **Input Layer**:
  - Accepts images resized to 227x227x3 (RGB images with three color channels). Image preprocessing involves normalization to scale pixel values.
- **First Convolutional Layer**:
  - Applies 96 filters of size 11x11 with a stride of 4.
  - Extracts low-level features like edges and textures.
  - **Max-pooling** follows with a 3x3 filter and stride of 2, reducing spatial dimensions.
- **Second Convolutional Layer**:
  - Consists of 256 filters of size 5x5.
  - Enhances feature detection for more complex patterns.
  - Max-pooling follows, similar to the first layer.

- **Third, Fourth, and Fifth Convolutional Layers**:
  - Third layer: 384 filters of size 3x3.
  - Fourth layer: 384 filters of size 3x3.
  - Fifth layer: 256 filters of size 3x3 followed by max-pooling.
  - These layers capture increasingly abstract features, such as shapes and objects.
- **Fully Connected Layers**:
  - Three dense layers: the first two have 4096 neurons each; the last layer outputs class probabilities using Softmax for classification.
- **ReLU Activation**:
  - Every layer employs Rectified Linear Units (ReLU), speeding up training by preventing the vanishing gradient problem.
- **Dropout Regularization**:
  - Dropout is used in fully connected layers to deactivate random neurons, reducing overfitting.

VGG NET:

What is VGG Net ?

- VGGNet, developed by the Visual Geometry Group at the University of Oxford, is a deep convolutional neural network (CNN) architecture that gained recognition for its excellent image classification performance.
- Introduced in the 2014 paper "Very Deep Convolutional Networks for Large-Scale Image Recognition," it became a key model in computer vision.

Key Features:

1. **Depth**:
   - VGGNet is much deeper than earlier architectures, such as AlexNet.
   - Versions include VGG16 (16 layers) and VGG19 (19 layers), consisting of convolutional, max-pooling, and fully connected layers.
2. **Small Filters**:
   - Uses 3x3 convolutional filters stacked together, which improves feature extraction and enables deeper models.
3. **Max-Pooling**:
   - Reduces spatial dimensions while preserving features, aiding in computational efficiency.

4. **Fully Connected Layers**:
    - The last layers are dense (fully connected), leading to classification of images into categories.
5. **Consistency**:
    - Maintains uniform architecture and simplicity, making it easy to implement.

Architecture :

- **Input Layer**
    - ➢ Accepts fixed-size images (e.g., 224x224x3 for RGB images).
    - ➢ Preprocessing includes resizing the image and subtracting the mean RGB value from each pixel.
- **2. Convolutional Layers**
    - ➢ Stacks multiple **3x3 convolutional filters** with a stride of 1.
    - ➢ These small filters extract complex features while keeping the model's depth manageable.
    - ➢ The depth increases as layers progress, starting with fewer filters (e.g., 64) and increasing to 128, 256, and 512 in deeper layers.
- **3. Max-Pooling Layers**
    - ➢ Introduced after every few convolutional layers.
    - ➢ Uses a 2x2 filter with a stride of 2 to reduce spatial dimensions, lowering computational complexity while retaining essential features.
- **4. Fully Connected Layers**
    - ➢ Three dense layers are included at the end of the architecture:
    - ➢ The first two layers have 4096 neurons.
    - ➢ The final layer outputs class probabilities using the **Softmax activation function**.
- **5. Depth and Uniformity**
    - ➢ VGGNet comes in two main versions: VGG16 (16 layers) and VGG19 (19 layers), consisting of convolutional, pooling, and fully connected layers.
    - ➢ It emphasizes depth for feature extraction and maintains a consistent use of 3x3 filters.
- **6. Strengths**
    - ➢ Achieves high accuracy due to its depth and small filters.
    - ➢ The simplicity of its architecture made it a benchmark for future CNNs.

GOOGLE NET :

What is GoogLeNet ?

- GoogLeNet is a deep convolutional neural network (CNN) introduced by Google researchers in 2014 as part of the Inception architecture.
- It gained prominence for its innovative design and performance in tasks like image recognition, especially during the ImageNet Large Scale Visual Recognition Challenge, where it achieved top results.

Key Features :

- **Inception Modules**:
  - ➢ The architecture introduces "Inception modules," which apply multiple convolutional operations (1x1, 3x3, 5x5 filters) and pooling in parallel. This captures features at different scales in the same layer.
  - ➢ These operations are combined efficiently to avoid computational overload.
- **Efficiency**:
  - ➢ Utilizes **1x1 convolutions** for dimensionality reduction before applying larger filters. This reduces computational costs while retaining critical feature information.
- **Depth**:
  - ➢ The network is deeper than earlier architectures, consisting of **22 layers** (including convolutional and fully connected layers).
- **Global Average Pooling**:
  - ➢ Replaces traditional fully connected layers with global average pooling at the end, reducing the number of parameters and preventing overfitting.
- **Performance**:
  - ➢ It is highly efficient in extracting features, achieving superior results on large-scale datasets like ImageNet.

Architecture :

1. Input Layer

- Accepts images resized to 224x224x3 (height, width, and RGB color channels).
- Preprocessing includes normalization for consistent training.

2. Inception Modules

- The core innovation of GoogLeNet lies in its **Inception modules**:
  - Each module processes the input with **1x1**, **3x3**, and **5x5 convolutional filters**, as well as **max-pooling**, in parallel.
  - These outputs are concatenated, enabling the network to extract features at multiple scales.
  - **1x1 convolutions** are applied beforehand to reduce the input's dimensionality, minimizing computational costs.

3. Depth

- GoogLeNet has **22 layers**, including multiple Inception modules, making it significantly deeper than earlier models like AlexNet.

4. Global Average Pooling (GAP)

- Instead of traditional fully connected layers, GoogLeNet employs GAP after convolutional layers.
  - GAP averages feature maps, reducing parameters and preventing overfitting.
  - This leads to lightweight yet highly efficient computations.

5. Auxiliary Classifiers

- To improve training and combat gradient vanishing problems:
  - Intermediate softmax classifiers are added to deeper layers during training.
  - They act as regularizers and provide additional gradient signals for earlier layers.

6. Efficiency

- With fewer parameters (12 times fewer than AlexNet) and greater computational efficiency, GoogLeNet achieved exceptional performance on large datasets like ImageNet.

CNN MODELS AFTER 2017

DENSE NET :

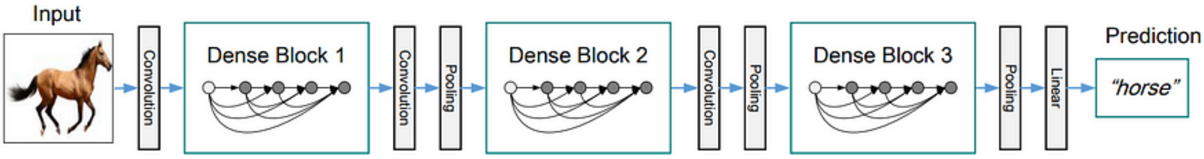
What is DenseNet ?

- DenseNet was introduced by a team of researchers from Cornell University and Tsinghua University.
- The key contributors were Gao Huang, Zhuang Liu,

  Laurens van der Maaten, and Kilian Q. Weinberger.
- They presented DenseNet in their paper titled **"Densely Connected Convolutional Networks"**, which was published in 2017 and gained significant attention in the field of deep learning.

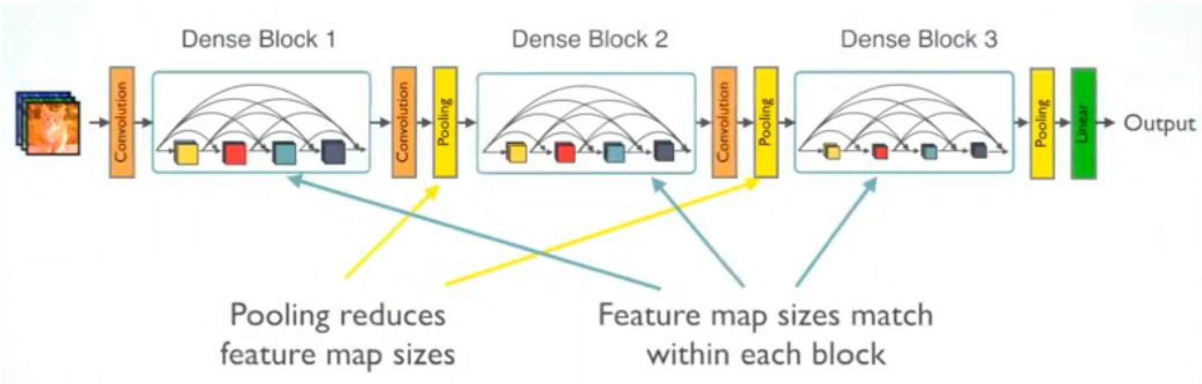
DenseNet Architecture

- **Dense Blocks**:
  - The network consists of several dense blocks, which are groups of layers connected in a specific way. Within a dense block, each layer is connected to every other layer, meaning the input to each layer comes not only from the previous layer but also from all earlier layers in the same block.



- **Feature Concatenation**:
  - Unlike traditional neural networks that sum the outputs, DenseNet concatenates the feature maps from all preceding layers. This ensures maximum feature propagation and reuse.

- **Growth Rate**:
  - The growth rate determines how much information each layer contributes to the network. If the growth rate is 32, each layer outputs 32 feature maps, which are then added to the existing ones.
- **Transition Layers**:
  - Between dense blocks, there are transition layers that reduce the number of feature maps using a combination of batch normalization, 1x1 convolution, and pooling. These layers help control the model's complexity and improve efficiency.
- **Bottleneck Layers**:
  - DenseNet often uses bottleneck layers (1x1 convolutions) to reduce the number of input feature maps before applying the 3x3 convolutions. This helps in making the model more parameter-efficient.
- **Final Classification Layer**:
  - After passing through all the dense blocks and transition layers, a global average pooling layer and a fully connected layer are used for the final classification.



MOBILENET :

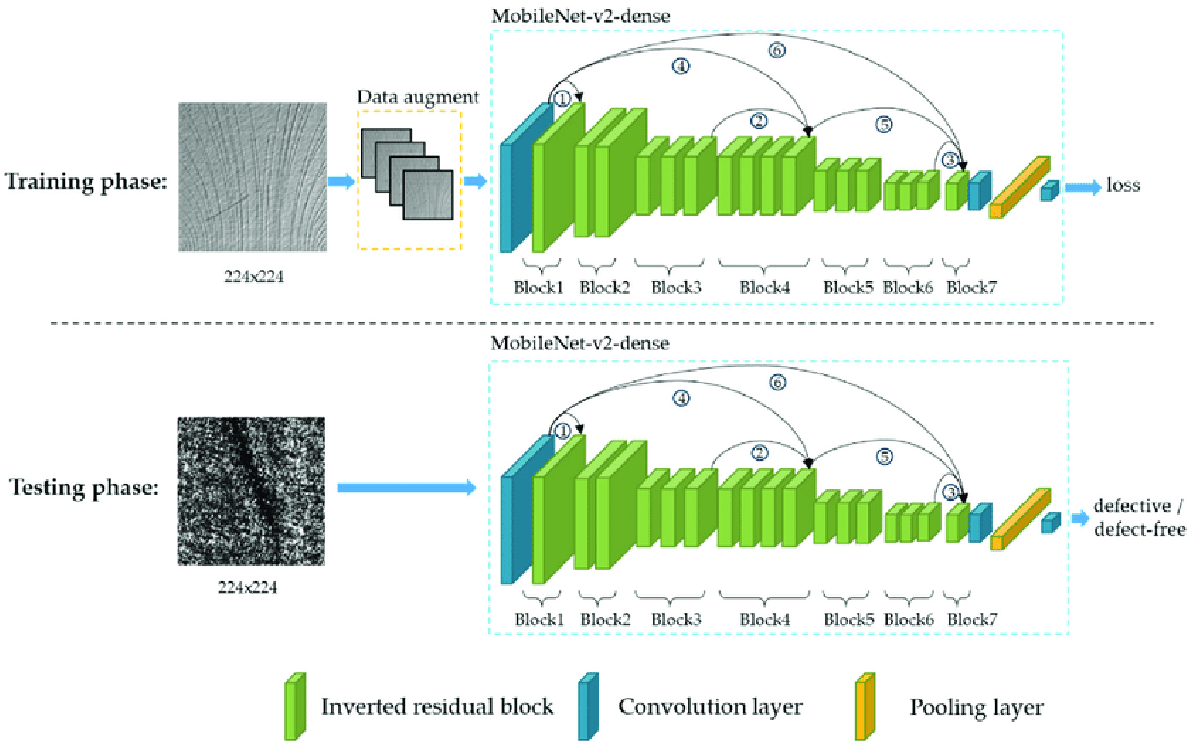
What is MobileNet ?

- MobileNet is a family of lightweight deep learning models designed specifically for mobile and edge devices, where computational resources like memory and processing power are limited.
- Introduced by Google, MobileNet achieves high efficiency without significantly compromising accuracy, making it ideal for tasks such as image classification, object detection, and more.

MobileNet Architecture

1. **Depthwise Separable Convolution**:
    - A key innovation in MobileNet is the use of depthwise separable convolutions, which break a standard convolution into two operations:
        - **Depthwise Convolution**: Applies a single filter to each input channel independently.
        - **Pointwise Convolution**: Uses 1x1 convolutions to combine the outputs from the depthwise convolution.
    - This approach drastically reduces computational complexity and the number of parameters.
2. **Convolutional Building Blocks**:
    - The network is built using several convolutional layers based on depthwise separable convolutions.
    - Each block includes batch normalization and ReLU activation for efficient learning.
3. **Width Multiplier (α)**:
    - A parameter that scales the number of filters in each layer. Smaller width multipliers lead to fewer parameters and lower computational cost, trading off some accuracy for efficiency.

4. **Resolution Multiplier (ρ)**:
    - This parameter adjusts the input image resolution. Reducing the resolution reduces the number of computations further, enabling deployment on resource-constrained devices.
5. **Global Average Pooling**:
    - After passing through all the convolutional layers, global average pooling is applied to reduce the spatial dimensions to a single feature vector.
6. **Fully Connected Layer and Softmax**:
    - The final fully connected layer maps the feature vector to class probabilities using a softmax function for classification tasks.



Variants of MobileNet:

The basic structure has evolved over time:

- **MobileNetV1**: Focused on depthwise separable convolutions.
- **MobileNetV2**: Introduced inverted residual blocks and linear bottlenecks to enhance efficiency and performance.

- **MobileNetV3**: Incorporated Neural Architecture Search (NAS) to optimize the architecture further and added hard-swish activation for better accuracy.