CNN MODELS

ALEXNET

What is Alex Net?

Alex Net is a deep convolutional neural network (CNN) architecture that was designed by **Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton**. It was the winning model of the **ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012**, significantly outperforming traditional machine learning methods. This architecture helped popularize deep learning for image classification and computer vision tasks.

Key Features of Alex Net:

1. **Deep Architecture**: Consists of **8 layers** (5 convolutional + 3 fully connected).
2. **ReLU Activation**: Uses **Rectified Linear Unit (ReLU)** instead of sigmoid/tanh, speeding up training.
3. **Dropout Regularization**: Helps prevent overfitting by randomly deactivating neurons during training.
4. **Data Augmentation**: Uses techniques like flipping, cropping, and color shifting to improve generalization.
5. **Overlapping Max Pooling**: Reduces dimensions while retaining important spatial features.
6. **Parallel Processing**: Originally trained on **two GPUs**, splitting computations for efficiency.

Alex Net Architecture

Alex Net consists of **8 layers**:

1. **Input Layer**: Takes a **227×227×3** RGB image as input.
2. **Convolutional Layer 1**:
    - 96 filters of size **11×11×3**, stride **4**

- Followed by **ReLU activation** and **max pooling (3×3, stride 2)**

3. **Convolutional Layer 2**:
   - 256 filters of size **5×5**, stride **1**
   - Followed by ReLU and max pooling
4. **Convolutional Layers 3, 4, 5**:
   - Smaller filter sizes (**3×3**) with more channels
   - Layer 3: **384 filters**
   - Layer 4: **384 filters**
   - Layer 5: **256 filters** followed by max pooling
5. **Fully Connected Layers**:
   - **4096 neurons** in **FC6** and **FC7**, each followed by dropout
   - **SoftMax Layer (FC8)** for **1000-class classification (ImageNet dataset)**

Impact of Alex Net

- Revolutionized deep learning in **computer vision**.
- Demonstrated the power of **CNNs for image classification**.
- Inspired modern architectures like **VGG, ResNet, and EfficientNet**.

VGG NET

What is VGG Net?

- VGG Net is a deep convolutional neural network (CNN) architecture developed by the Visual Geometry Group (VGG) at the University of Oxford.

- It was introduced in the research paper *"Very Deep Convolutional Networks for Large-Scale Image Recognition"* by Simonyan and Zisserman in 2014.

- VGG Net gained prominence after achieving high performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014.

Key Features of VGG Net

1. **Deep Architecture** – VGG Net is deeper than its predecessors, with 16 or 19 layers (VGG-16 and VGG-19).
2. **Small 3×3 Convolutional Kernels** – Unlike previous networks that used larger filters (e.g., 5×5 or 7×7), VGG Net uses only 3×3 filters stacked on top of each other. This reduces parameters while increasing depth and non-linearity.
3. **Consistent Structure** – Uses multiple convolution layers, followed by max-pooling layers, with fully connected layers at the end.
4. **Uniform Padding and Stride** – Maintains spatial resolution better by using stride 1 and padding of 1 in convolutional layers.
5. **Pretrained Models Available** – Widely used in transfer learning for tasks like object detection, face recognition, and medical image analysis.
6. **High Computational Cost** – Requires significant memory and processing power due to its depth and fully connected layers.

VGG Net Architecture

VGG Net follows a uniform design, where convolutional layers are stacked in increasing depth. The architecture consists of the following components:

1. **Input Layer**:
    - Accepts images of size **224×224×3** (RGB).
2. **Convolutional Layers**:
    - Uses small **3×3 filters** with stride **1** and padding **1**.
    - Each convolutional block consists of **two or more convolution layers** followed by a **ReLU (Rectified Linear Unit) activation function**.
3. **Max-Pooling Layers**:
    - 2×2 pooling with stride 2 to reduce spatial dimensions.
4. **Fully Connected (FC) Layers**:
    - Three dense layers at the end: **two 4096-unit layers** and a **final 1000-unit SoftMax layer** (for ImageNet classification).
5. **SoftMax Classifier**:
    - Outputs probabilities for **1000 image classes** in the ImageNet dataset.

GOOGLE NET

What is Google Net?

• Google Net (or Inception v1) is a deep convolutional neural network (CNN) architecture developed by Google for image classification and recognition tasks.

• It was introduced in **2014** as part of the **ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 2014**, where it won the classification task with a top-5 error rate of only **6.67%**.

• The key innovation of Google Net was the **Inception module**, which allowed the network to capture multi-scale features efficiently while keeping computational costs low.

Key Features of Google Net

1. **Inception Module:**

    • Uses multiple convolution filter sizes (1x1, 3x3, 5x5) in parallel.
    • Captures features at different scales.
    • Includes 1x1 convolutions for dimensionality reduction and computational efficiency.

2. **Deeper but Computationally Efficient:**

    • 22 layers deep, but optimized to be computationally efficient.
    • Uses fewer parameters (~5 million) compared to VGG-16 (~138 million).

3. **Auxiliary Classifiers:**

    • Two additional SoftMax classifiers are placed in intermediate layers.
    • Helps in training by improving gradient flow and regularization.

4. **Global Average Pooling (GAP):**

    • Instead of fully connected layers, it uses average pooling before the final classification layer.
    • Reduces the number of parameters and prevents overfitting.

Architecture of Google Net

Google Net consists of several key components:

1. Input Layer

- Takes an input image of **224x224x3**.

2. Convolutional & Pooling Layers

- **First few layers:** Traditional convolutional and max pooling layers extract low-level features.

3. Inception Modules (Main Feature)

Each **Inception module** contains:

- **1x1 convolution** (for dimensionality reduction and feature extraction).
- **3x3 convolution** (for medium-scale feature detection).
- **5x5 convolution** (for large-scale feature detection).
- **3x3 max pooling** (to capture spatial information).
- Outputs from all these layers are concatenated.

4. Auxiliary Classifiers

- Two additional softmax classifiers (placed at intermediate layers) help with training deep networks.

5. Fully Connected Layer & Output

- Instead of traditional fully connected layers, **Global Average Pooling (GAP)** is used.
- Followed by a SoftMax layer for classification.

Advantages of Google Net

- **More efficient than previous architectures (e.g., VGG-16)**
- **Reduces overfitting with fewer parameters**

- **Captures multi-scale features using the Inception module**
- **Uses auxiliary classifiers to improve gradient flow**