CNN MODELS

AlexNet: Introduced in 2012, this was one of the first deep learning models to revolutionize computer vision. It uses convolutional layers, pooling, and ReLU activation to classify images, marking a significant improvement in image recognition tasks.

The key features of AlexNet are:

1. Deep Architecture: Introduced eight layers (five convolutional and three fully connected), which was groundbreaking at the time.
2. ReLU Activation: Used Rectified Linear Units (ReLU) to speed up training by introducing non-linearity.
3. Dropout Regularization: Added dropout layers to reduce overfitting during training.
4. GPU Acceleration: Utilized GPUs to handle large-scale computations efficiently.
5. Max Pooling: Implemented max pooling to down-sample feature maps while retaining important spatial features.

The architecture of AlexNet consists of the following key components:

1. Input Layer: Takes an input image of size 227x227x3 (RGB image).
2. Convolutional Layers:
    - Layer 1: A convolutional layer with 96 filters of size 11x11 and a stride of 4, followed by ReLU activation.
    - Layer 2: A convolutional layer with 256 filters of size 5x5, followed by ReLU activation and local response normalization.
    - Layer 3, 4, and 5: Three convolutional layers (384, 384, and 256 filters, respectively), each with 3x3 filters and ReLU activation.
3. Max Pooling Layers: Pooling layers follow some convolutional layers to reduce spatial dimensions and computational cost.
4. Fully Connected Layers:
    - Three fully connected layers are used. The first two have 4096 neurons each, and the last layer maps to 1000 output categories (for ImageNet).
5. Activation Function: ReLU is applied after each layer to introduce non-linearity.
6. Dropout: Dropout is used in the fully connected layers to prevent overfitting.
7. Softmax Classifier

VGGNet: Developed by the Visual Geometry Group (VGG), it builds on AlexNet by using very deep networks with smaller convolutional filters (3x3). This allows for better feature extraction, although it requires more computational resources.

The key features of VGGNet are:

1. Deep Architecture: VGGNet uses a very deep structure with up to 19 layers, allowing it to learn detailed features from images.
2. Small Convolutional Filters: It uses 3x3 convolutional filters consistently, which improves feature extraction while keeping computations efficient.
3. Uniform Design: The network maintains a simple and uniform architecture, making it easy to extend or modify.
4. Max Pooling: Similar to AlexNet, it uses max pooling to down-sample feature maps and retain critical spatial features.
5. Pretrained Models: VGGNet provides pretrained versions that can be fine-tuned for different tasks, saving time for researchers.

The architecture of VGGNet :

1. Input Layer: Takes an image input of size 224x224x3 (RGB image), resized as needed.
2. Convolutional Layers:
    - Uses small 3x3 convolutional filters with a stride of 1 and padding to preserve spatial resolution.
    - Stacked multiple convolutional layers (2 or 3) before pooling, allowing it to capture complex features.
3. Max Pooling Layers:
    - After every group of convolutional layers, a 2x2 max pooling layer is applied to reduce spatial dimensions while retaining key features.
4. Fully Connected Layers:
    - Two fully connected layers, each with 4096 neurons, followed by one fully connected layer with 1000 neurons for classification (ImageNet categories).
5. Activation Function:
    - ReLU (Rectified Linear Unit) is used after every convolutional and fully connected layer, introducing non-linearity.
6. Softmax Classifier:
    -

GoogleNet: Also known as Inception, this model introduced the Inception module, which uses a combination of different-sized convolutional filters in parallel to efficiently capture features at multiple scales.

The key features of GoogleNet (Inception Network) are:

1. Inception Modules: Uses a combination of different filter sizes (1x1, 3x3, 5x5) in parallel to capture features at multiple scales, making it highly efficient.
2. Reduced Parameters: By incorporating 1x1 convolutions and avoiding fully connected layers, GoogleNet dramatically reduces the number of parameters compared to models like VGGNet.
3. Deeper Architecture: With 22 layers, GoogleNet achieves great depth while maintaining computational efficiency.
4. Auxiliary Classifiers: Employs additional classifiers at intermediate layers to improve gradient flow and act as regularization.
5. High Performance: Achieved state-of-the-art results in the ImageNet competition of 2014.

The architecture of GoogleNet (Inception Network) :

1. Inception Modules:
    - The cornerstone of GoogleNet's architecture, these modules process data using multiple filter sizes (1x1, 3x3, and 5x5) in parallel, capturing features at different scales.
    - Includes 1x1 convolutions for dimensionality reduction, minimizing the number of parameters without losing spatial information.
2. Deep Structure:
    - GoogleNet comprises 22 layers, including convolutional, pooling, and fully connected layers, enabling deep feature extraction.
3. Auxiliary Classifiers:
    - Intermediate classifiers are added to prevent vanishing gradients in deeper networks and provide regularization during training.
4. Global Average Pooling:
    - Instead of fully connected layers at the end, GoogleNet employs global average pooling, which reduces the risk of overfitting and creates a compact representation for classification.
5. Reduced Parameters:
    - By using techniques like 1x1 convolutions and replacing fully connected layers