

VGGNet for Image Classification

Built using Transfer Learning (VGG16)

Ankit Shrestha

What is VGGNet?

- A Convolutional Neural Network (CNN) developed by Visual Geometry Group (VGG), Oxford.
- Known for its simplicity and depth.
- Introduced in the 2014 ImageNet Challenge.
- Uses very small (3×3) convolution filters stacked deeply.

VGG Architecture

- Uses many layers (16 or 19) of convolution
- Input size: $224 \times 224 \times 3$ (RGB).
- Pretrained on ImageNet (1.2M images, 1000 classes).
- Sequence: Conv \rightarrow ReLU \rightarrow Pool (repeated) \rightarrow Fully Connected \rightarrow Softmax.

Sequence

1. Input: You give the model an image (say, a cat).
2. Conv (Convolutional layer): finds simple patterns like lines or corners.
3. ReLU (Rectified Linear Unit): keeps only positive activations, which helps the model learn complex patterns.
4. Pool (Pooling layer): shrinks the image size so the model focuses on the main parts.
5. Steps 2 – 4 repeat several times. Each time, the model learns more complex things from edges → to shapes → to full objects.
6. After that, the output is flattened and passed to Fully Connected (Dense) layers to combine features and make final predictions.
7. Softmax (or Sigmoid for binary tasks): Converts final scores into probabilities for each class.

Why VGGNet Matters

- Simple design: Easy to extend and modify
- High accuracy: Benchmark model on ImageNet
- Transfer learning: Often used as a base for custom image classifiers

Why Use Transfer Learning

- Training deep networks from scratch requires huge data and compute.
- Pretrained models like VGG16 already learned general image features (edges, shapes, textures).
- We reuse these layers and only train a small classifier head for our custom task (cats vs dogs).
- Reduces training time and improves accuracy.

Advantages

- Uniform structure (only 3×3 filters)
- Good at extracting visual features
- Pretrained models widely available

Limitations

- Large number of parameters (~138 million)
- High memory and computation cost
- Training time is long

Applications

- Image classification
- Object detection
- Feature extraction for other models

Thank You !