

Overview of VGG16, ResNet50, Xception and MobileNet Neural Networks

Introduction

Image classification is a crucial task in the field of computer vision, and the performance of neural architectures plays a significant role in achieving high accuracy and efficiency. In this essay, we introduced four popular **Convolutional Neural Network(CNN)** architectures including **VGG16**, **ResNet50**, **Xception**, and **MobileNets**, and explored their key features and performance on the ImageNet benchmark dataset.

VGG16

VGG16, developed by the Visual Geometry Group at the University of Oxford, is a deep convolutional neural network known for its simplicity and uniformity. Consisting of 16 layers, 13 convolutional layers with 3x3 filters, and 3 fully connected layers, it demonstrated remarkable performance in the 2014 ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), achieving 92.7% classification accuracy. It is able to capture fine-grained details due to small filter sizes. Although computationally intensive, VGG16 set the stage for deeper networks and influenced subsequent models.

The architecture of VGG16 (Simonyan and Zisserman, 2014)

ResNet50

ResNet50 is a powerful deep convolutional neural network architecture introduced by Microsoft Research in 2015. It consists of 50 layers and features an innovative use of residual blocks, which allows the network to skip connections and mitigate the vanishing gradient problem during training. This design facilitates the training of very deep networks, enabling better performance in various computer vision tasks. It achieves 92.2% top-5 classification accuracy on the ImageNet benchmark dataset.

The architecture of ResNet50 (He et al., 2015)

Xception

Xception, introduced by Google in 2017, is a deep convolutional neural network known for its novel feature learning. It uses depth-separable convolutions, which decompose standard convolutions into depth and pointwise convolutions. In particular, a spatial convolution is performed independently on each channel of an input, followed by a pointwise convolution (1x1 convolution) that projects the channels output by the depth convolution to a new channel space. This unique design reduces the number of parameters while maintaining significant performance. It consisted of 36 convolutional layers organized into 14 modules. All modules except the first and last have linear residual connections around them. It achieves 95.5% top-5 classification accuracy on the ImageNet benchmark dataset.

The architecture of Xception (Chollet, 2017)

MobileNets

MobileNets are a family of efficient convolutional neural networks designed for mobile and edge devices. Introduced by Google in 2017, MobileNets prioritize computational efficiency without compromising performance. They use depth-separable convolutions to reduce the number of parameters and computational cost, while maintaining competitive accuracy. MobileNets are available in several versions, including MobileNetV1, MobileNetV2, and MobileNetV3, each offering improvements in speed and efficiency. MobileNetV1 achieves a classification accuracy of 89.9% top-5 on the ImageNet benchmark dataset.

The architecture of MobileNetV1(Howard et al., 2017)

Performance Comparison on ImageNet

The ImageNet dataset is a benchmark for assessing image recognition models. the following table provides a quick comparison of VGG16, ResNet50, Xception, and MobileNet on ImageNet, including their sizes, top-5 accuracies, and notable features.

Table created by author. The information in the table is based on the original papers.

Conclusion

In this essay, we provide an overview on the distinctive features and individual performances of four prominent neural architectures VGG16, Xception, MobileNet, and ResNet50 in image classification on the ImageNet data set. Each model exhibits unique characteristics, encompassing accuracy, efficiency, and model complexity. The choice of architecture hinges on the specific requirements of the image recognition task, considering factors such as available computational resources and the desired balance between accuracy and efficiency. While VGG16, ResNet50 and Xception demonstrate exceptional performance on ImageNet, MobileNet is tailored for scenarios with limited resources.