11.25.24

MobileNet *CNN For Mobile Vision*

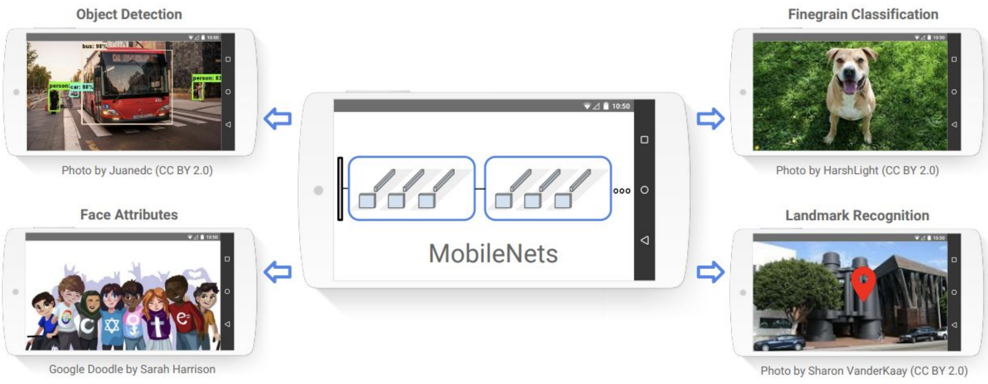

Figure 1. MobileNet models can be applied to various recognition tasks for efficient on device intelligence.

1. **Introduction**
2. **Methods of MobileNets**
3. **Experiments & Applications**
4. **Q&A**

Introduction

- The Problem
  - Traditional deep learning models are computationally expensive.
  - Require high-power GPUs and significant memory

- Why Does it Matter
  - The demand for efficient models is increasing in sectors like autonomous driving, augmented reality, and Internet of Things (IoT).

- The Solution
  - MobileNets promote computational efficiency, model accuracy, and run on resource-constrained devices.

1. Introduction
2. Methods of MobileNets
3. Experiments & Applications
4. Q&A

What are MobileNets?

- What are MobileNets?
    - Efficient models designed for mobile & embedded vision
    - Use depthwise separable convulsions for lightweight, efficient neural networks
- Key goals:
    - Reduce computational costs and model size
    - Optimize latency without compromising accuracy
    - Minimizing memory usage to fit in the memory constraints of mobile devices
    - Minimizing power consumption to extend battery life
    - Adaptability across platforms, such as phones, IoT devices, and embedded systems

MobileNet Architecture

- Depthwise separable convolutions
    - Breaks standard convolution into two operations
        - Depthwise convolution
        - Point convolution
    - Benefit: 8 to 9 times less computation than standard convolution
- Two key hyperparameters
    - Width multiplier (alpha)
        - Controls model size by thinning the network
    - Resolution multiplier (rho):
        - Reduces input image resolution for faster processing

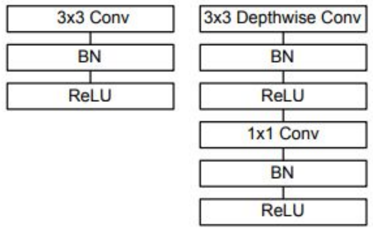

Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.

1. Introduction
2. Methods of MobileNets
3. **Experiments & Applications**
4. **Q&A**

Experimental Results

- Classification performance
  - MobileNet achieves comparable accuracy with fewer computations than traditional models like VGG16 & GoogleNet
  - MobileNet vs VGG16
    - 32x smaller, 27x less computation with minimal accuracy reduction
- Model shrinking with settings
  - Width & resolution multipliers allow scalable model size for specific application needs

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138

Example: Fine-Grained Recognition

- Dataset
  - Stanford Dogs dataset, containing images of 120 breeds
  - High intra-class similarity and subtle differences between breeds
- Performance comparison
  - Inception V3 Model
    - Accuracy: ~84%
    - 23.2 million parameters, 5 billion mult-adds
  - MobileNet
    - Accuracy: ~83.3%
    - 3.3 million parameters, 569 million mult-adds
- Implications
  - MobileNet achieves similar performance with 7x lighter framework and 8.8x less computation cost

Applications & Use Cases

- **Fine-Grained Recognition**
    - MobileNet achieved near state of the art recognition on the Stanford Dogs dataset
- **Large Scale Geolocation**
    - MobileNet powered a compact version of PlaNet (geolocalization model) with minimal performance loss
- **Face Attributes**
    - MobileNet compressed a large face attribute classifier with similar performance and lower computational cost
- **Object Detection**
    - MobileNet achieved comparable accuracy to other object detection models with fewer resources
- **Face Embeddings**
    - Distilled MobileNet excelled face recognition, rivaling FaceNet's performance

Citations

- https://production-media.paperswithcode.com/datasets/Stanford_Dogs-0000000577-91cb15b5_1ABtNf7.jpg
- https://arxiv.org/abs/1704.04861