

- **#paper** ~ [Computer Vision](#)
 - <https://arxiv.org/abs/1704.04861>
- ☐ Section 2 contains a great **#list** of prior works.
 - Sequel papers:
 - [MobileNetV2](#)
 - [Searching for MobileNetV3](#)
 - Mentioned papers:
 - [Going Deeper with Convolutions](#)
 - [Flattened Convolutional Neural Networks](#)
 - Mentioned topics:
 - ☐ Side heads?
 - [Label Smoothing](#)
 - [Knowledge Distillation](#)

• Summary

- The model uses [Depthwise Separable Convolutions](#) to reduce the number of parameters and multiply-accumulate operations.
- There are also 2 model-specific hyperparameters:
 - **Width** multiplier α where $\alpha \in (0, 1]$.
 - The number of both input and output channels scales by α .
 - **Resolution** multiplier ρ where $\rho \in (0, 1]$.
 - The input image and the internal representation of every layer are reduced by ρ .

• Usage

- [Object Detection](#)
 - Under both [Faster-RCNN](#) and [Single Shot MultiBox Detector, SSD](#).
- [Fine-grained Image Classification](#)
- [Photo Geolocation Estimation](#)
- [Facial Attribute Classification](#)
 - Using [Triplet Loss](#).

- [Face Recognition \(Embeddings\)](#)
- **Implementation details**
 - [General Matrix Multiply, GEMM](#)
- **Training Process**
 - MobileNets were trained using [RMSprop Optimization](#) with [Asynchronous Stochastic Gradient Descent](#).
 - Very little or no [weight decay](#) on the depthwise kernels.
 - Because there are too few parameters in them.

• Ideas

- What if we used 3×3 depthwise convolutions instead of 1×1 ?
 - Or probably $k \times k$ with a stride of k for small k ?
-

