

# Эффективные сверточные сети: MobileNet

Диц Даниил Денисович

3 апреля 2025 г.

# Что такое MobileNet?

- Архитектура нейросети, разработанная Google в 2017 году.
- Цель: **высокая точность при низком потреблении ресурсов.**
- Ориентирована на мобильные и встроенные устройства.

# Зачем нужен MobileNet?

- Классические сети (ResNet, VGG) **тяжёлые**.
- Нужна компактная альтернатива для телефонов, IoT и edge-устройств.
- Решение: использовать **depthwise separable convolutions**.

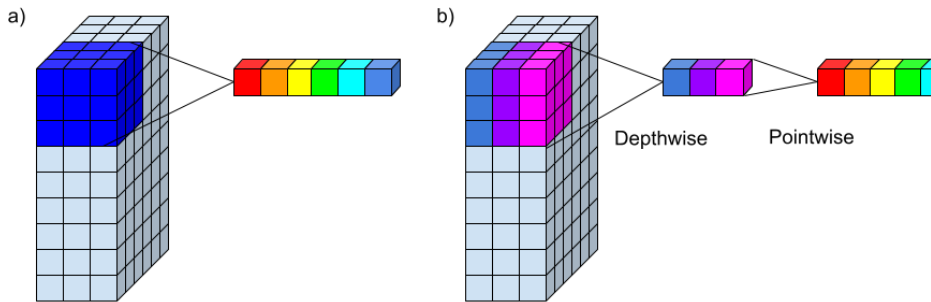
# Depthwise Separable Convolutions

Вместо обычной свёртки: две операции

- 1 **Depthwise Convolution** — отдельная свёртка на каждый канал
- 2 **Pointwise Convolution** —  $1 \times 1$  свёртка по всем каналам

Это позволяет резко **уменьшить количество параметров** и ускорить модель.

# Графическое сравнение



a) Обычная свёртка vs b) Depthwise + Pointwise

# Интуитивно: почему это работает?

## Обычная свёртка = толстый фильтр

- Один фильтр обрабатывает сразу все каналы.
- Много операций и параметров.

# Интуитивно: почему это работает?

## Обычная свёртка = толстый фильтр

- Один фильтр обрабатывает сразу все каналы.
- Много операций и параметров.

## Разделим на 2 шага: легче и быстрее

- **Step 1 — Depthwise:** по отдельности для каждого канала
- **Step 2 — Pointwise:** всё склеиваем через  $1 \times 1$  свёртку

# Что делает свёртка 1x1?

- Обработывает каждый пиксель индивидуально
- Учитывает все каналы RGB одновременно
- Может уменьшать или увеличивать количество каналов
- Используется как альтернатива полносвязным слоям



# Channel-wise Bottleneck (MobileNetV2)

- 1 Depthwise Conv ( $3 \times 3$ )
- 2 BatchNorm
- 3 ReLU

Почему это важно:

- Меньше параметров
- Быстрее обучение
- Хорошая обобщающая способность

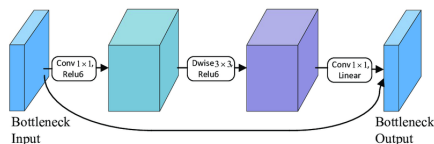


Рис.: Enter Caption

- ReLU помогает бороться с **затухающими градиентами**
- Обеспечивает **нелинейность**, без неё сеть = линейная модель
- Облегчает обучение глубоких сетей

## Плюсы

- Компактность и скорость
- Подходит для мобильных и встроенных систем
- Используется в object detection, segmentation и др.

## Ключевая идея

Depthwise Separable Convolution = меньше операций, тот же результат