**MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**

*Исследователи*: Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam.

Если пять лет назад нейронная сеть считалась «тяжеловесным» алгоритмом, требующим железа, специально предназначенного для высоконагруженных вычислений, то сегодня уже никого не удивить глубокими сетями, работающими прямо на мобильном телефоне.

В наши дни сети распознают ваше лицо, чтобы разблокировать телефон, стилизуют фотографии под известных художников и определяют, есть ли в кадре хот-дог.

Сегодня мы поговорим о MobileNet, передовой архитектуре сверточной сети, позволяющей делать всё это и намного больше.

**Структура MobileNet**

Слева нарисован блок обычной сверточной сети, а справа — базовый блок MobileNet.

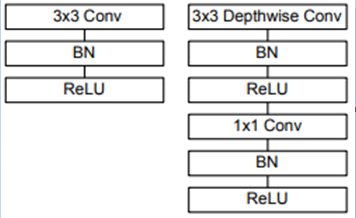


Рис.1 – структура свёрточной сети

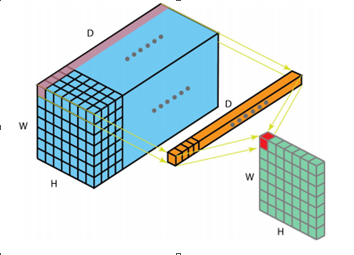


Рис.2 – пример 1х1 свёртки

Сверточная часть интересующей нас сети состоит из одного обычного свёрточного слоя с 3х3 свёрткой в начале и тринадцати блоков, изображенных справа на рисунке, с постепенно увеличивающимся числом фильтров и понижающейся пространственной размерностью тензора.

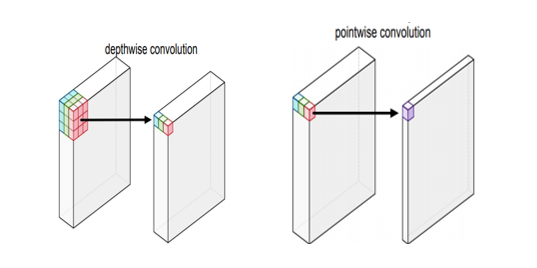
Особенностью данной архитектуры является отсутствие max pooling-слоёв. Вместо них для снижения пространственной размерности используется свёртка с параметром **stride**, равным 2.

**Stride** - целое число или кортеж/перечень одного целого, задающего длину шага свертки.

Говоря простыми словами, **stride**  — это количество сегментов, по которым одновременно проходится фильтр. Когда мы говорим об обработке фильтром входного изображения, то полагаем, что шаг фильтра равен 1 сегменту в заданном направлении. Мы можем сами регулировать количество сегментов (хотя обычно используется 1). В зависимости от условий конкретного случая можно выбрать более подходящее значение шага. Более широкие шаги обычно помогают уменьшать объём вычислений, обобщать результаты изучения признаков и т. д.

**Depthwise separable convolutions**

Обычная свертка представляет из себя фильтр Dk ∗ Dk ∗ Cin, где Dk — это размер ядра свёртки, а Cin — количество каналов на входе. Общая вычислительная сложность сверточного слоя составляет Dk ∗ Dk ∗ Cin ∗ Df ∗ Df ∗ Cout, где Df — это высота и ширина слоя (мы считаем, что пространственные размеры входного и выходного тензоров совпадают), а Cout — число каналов на выходе.  
  
Идея depthwise separable convolution состоит в том, чтобы разложить подобный слой на depthwise-свертку, которая представляет из себя поканальный фильтр, и 1x1-свёртку (также называемую **pointwise convolution**). Суммарное количество операций для применения такого слоя равно (Dk ∗ Dk+Cout) ∗ Cin ∗ Df ∗ Df.  
  
По большому счёту, это и всё, что нужно знать, чтобы успешно построить MobileNet.



Двумя **гиперпараметрами** архитектуры MobileNet являются α (множитель ширины) и ρ (множитель глубины или множитель разрешения).  
  
Множитель ширины отвечает за количество каналов в каждом слое. Например, α=1 даёт нам архитектуру, описанную в статье, а α=0.25 — архитектуру с уменьшенным в четыре раза числом каналов на выходе каждого блока.  
  
Множитель разрешения отвечает за пространственные размеры входных тензоров. Например, ρ=0.5 означает, что высота и ширина feature map, подаваемой на вход каждому слою будет уменьшена вдвое.  
  
Оба параметра позволяют варьировать размеры сети: уменьшая α и ρ, мы снижаем точность распознавания, но в то же время увеличиваем скорость работы и уменьшаем потребляемую память.