



[논문리뷰] MobileNets [2017]

Introduction

☀ 해당 논문은 Mobile 및 embedded Vision applications에 관한 MobileNets이라 불리는 효율적인 모델을 제시합니다.

MobileNet은 Depthwise separable convolution 기반으로 Model을 경량화(light weight)하였습니다.

메모리가 제한된 환경에서 MobileNet을 최적으로 맞추기 위해, 두 개의 Parameter를 소개합니다.

두 Parameter는 latency와 Accuracy의 균형을 조절합니다.



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

MobileNet Architecture

Depthwise Separable Convolution

Depthwise Separable Convolution은 Depthwise Convolution을 적용한 후, Pointwise Convolution을 결합한 것을 의미합니다.

이러한 방식은, drastically reducing Computation과 Model size를 감소시킵니다.

Depthwise convolution의 연산량은 다음과 같습니다.

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F$$

D_K 는 입력값 크기, M 은 입력의 채널수, D_F 는 피쳐맵 크기입니다.

Network Structure and Training

MobileNet에서 사용하는,

Depthwise separable Convolution은 Depthwise 이후, Pointwise를 적용한 것이다.

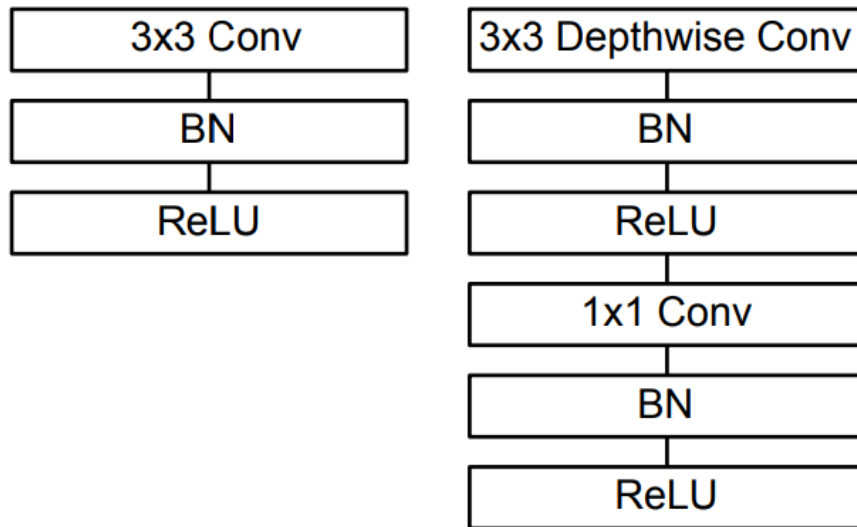


Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.

전체 연산량은 $D_k * D_k * M * D_F * D_F + M * N * D_F * D_F$

기존 **Conv**연산량에 비해 훨씬 적습니다.

M채널의 $D_k * D_k$ 크기의 입력값에 M개의 3x3 Conv 묶음을 N번 수행하여 $D_F * D_F$ 크기의 FeatureMap을 생성합니다.

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5×	Conv dw / s1	$3 \times 3 \times 512$ dw
	Conv / s1	$1 \times 1 \times 512 \times 512$
	Conv dw / s2	$3 \times 3 \times 512$ dw
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Table 2. Resource Per Layer Type

Type	Mult-Adds	Parameters
Conv 1×1	94.86%	74.59%
Conv DW 3×3	3.06%	1.06%
Conv 3×3	1.19%	0.02%
Fully Connected	0.18%	24.33%

Two Parameters

1. Width Multiplier : Thinner Models

- MobileNet Model이 smaller하고 fast하기 위해서는, 여러 **case** 및 **application**이 필요합니다.
- Width Multiplier라고 불리는 Simple Parameter를 도입합니다.

✨ MobileNet은 1을 적용합니다! (0 ~ 1)

Table 6. MobileNet Width Multiplier

Width Multiplier	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
0.75 MobileNet-224	68.4%	325	2.6
0.5 MobileNet-224	63.7%	149	1.3
0.25 MobileNet-224	50.6%	41	0.5

1. Resolution Multiplier : Reduced Representation

- 두 번째 하이퍼파라미터는 Resolution Multiplier p입니다.
- 모델의 연산량을 감소시키기 위해 사용합니다.
p는 입력 이미지에 적용하여 해상도를 낮춥니다.

Table 7. MobileNet Resolution

Resolution	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
1.0 MobileNet-192	69.1%	418	4.2
1.0 MobileNet-160	67.2%	290	4.2
1.0 MobileNet-128	64.4%	186	4.2

Conclusion

Depthwise Separable Convolutions을 기반으로 **MobileNets Model**을 제안합니다.

또한, 다른 Model과 Comparison할 때, **상당히 성능이 Improve**을 알 수 있습니다.

Table 8. MobileNet Comparison to Popular Models

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

Reference

MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

We present a class of efficient models called MobileNets for mobile and embedded vision applications. MobileNets are based on a streamlined architecture that uses depth-wise separable convolutions to build light weight deep neural networks. We introduce two simple global hyper-

📄 <https://arxiv.org/abs/1704.04861>



[논문 읽기] MobileNet(2017) 리뷰, Efficient Convolutional Neural Networks for Mobile Vision Applications

이번에 읽어볼 논문은 MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Application 입니다.

MobileNet은 Depthwise separable convolution을 활용하여 모델을 경량화했습니다. Xception은 Depthwise separable convolution을 활용하여 감소한 파라미터 수 만큼 층을 쌓아 성능을 높이는데 집중했는데요. MobileNet은 반대로 경량화에 집

📄 <https://deep-learning-study.tistory.com/532>

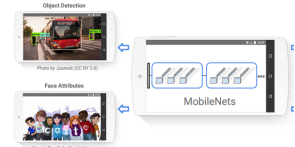


Figure 1. MobileNet models can be applied to various recognition tasks for effi

[Deep Learning] 딥러닝에서 사용되는 다양한 Convolution 기법들

또한, 영상 내의 객체에 대한 정확한 판단을 위해서는 Contextual Information 이 중요하다. 가령, 객체 주변의 배경은 어떠한 환경인지, 객체 주변의 다른 객체들은 어떤 종류인지 등. Object Detection 이나 Object Segmentation 에서는 충분한 Contextual Information을 확보하기 위해 상대적으로 넓은 Receptive Field

📄 <https://eehoeskrap.tistory.com/431>

