CVPR 2017
MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

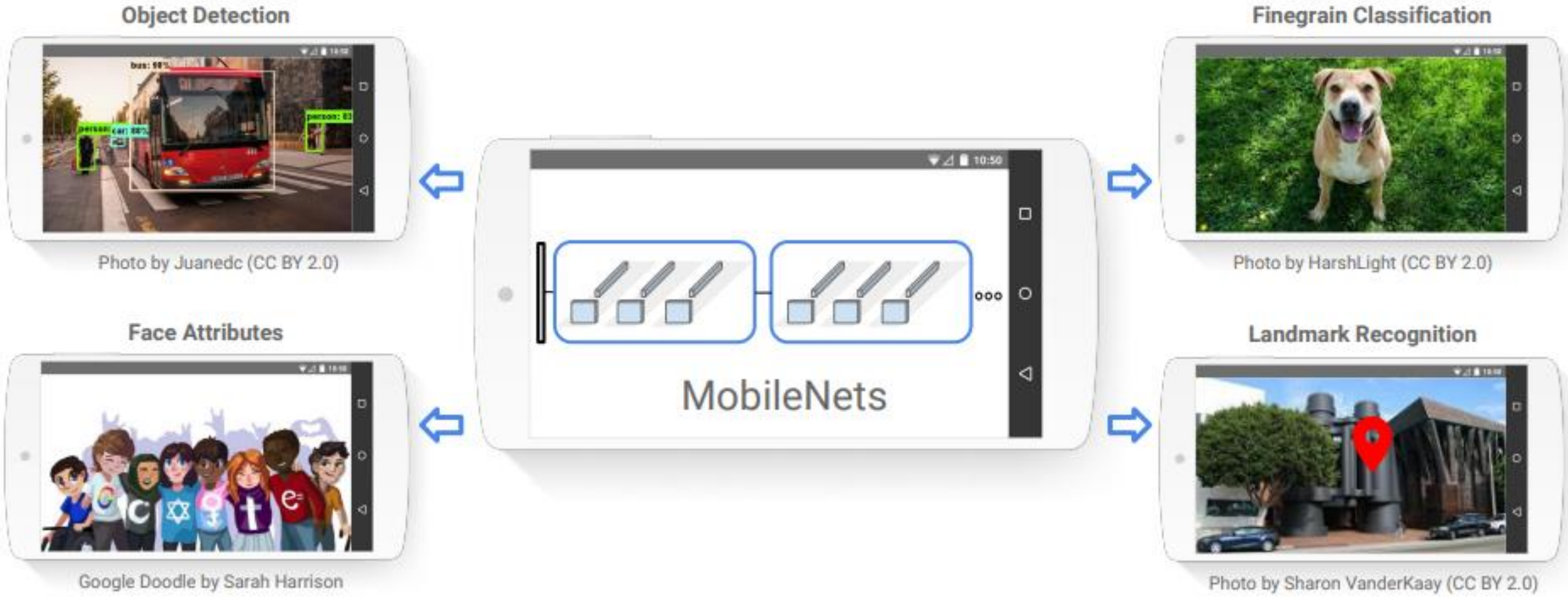
2022.07.29

논문 리뷰

배성훈

MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (CVPR 2017)

- **Research Background:**
  - 휴대폰이나 임베디드 시스템 같은 저용량 메모리 환경에 딥러닝 적용을 위한 **모델 경량화** 필수
  - **Depthwise Separable Convolution**을 사용한 **MobileNet** 제안
  - Application 환경에 따른 적절한 설계를 위한 2개의 hyperparameter(width multiplier, resolution multiplier)를 제안해 **latency, accuracy 균형 조절**



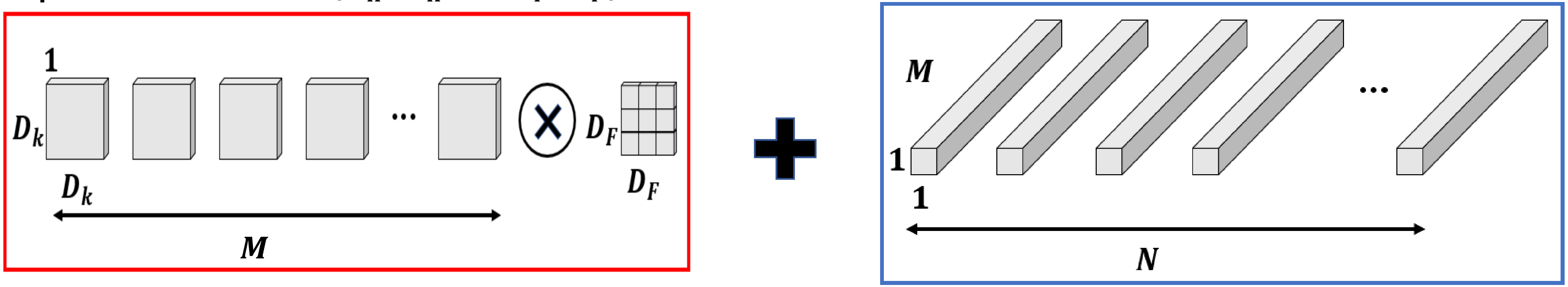
MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (CVPR 2017)

- **Method:**
  모델의 첫 번째 layer를 제외하고 모두 **<u>Depthwise Separable Convolution</u>**으로 변경해 적용.

Depthwise convolution + **Pointwise convolution** => 연산량↓ + 모델 크기↓

Depthwise Convolution $(D_K \cdot D_K \cdot M \cdot D_F \cdot D_F)$

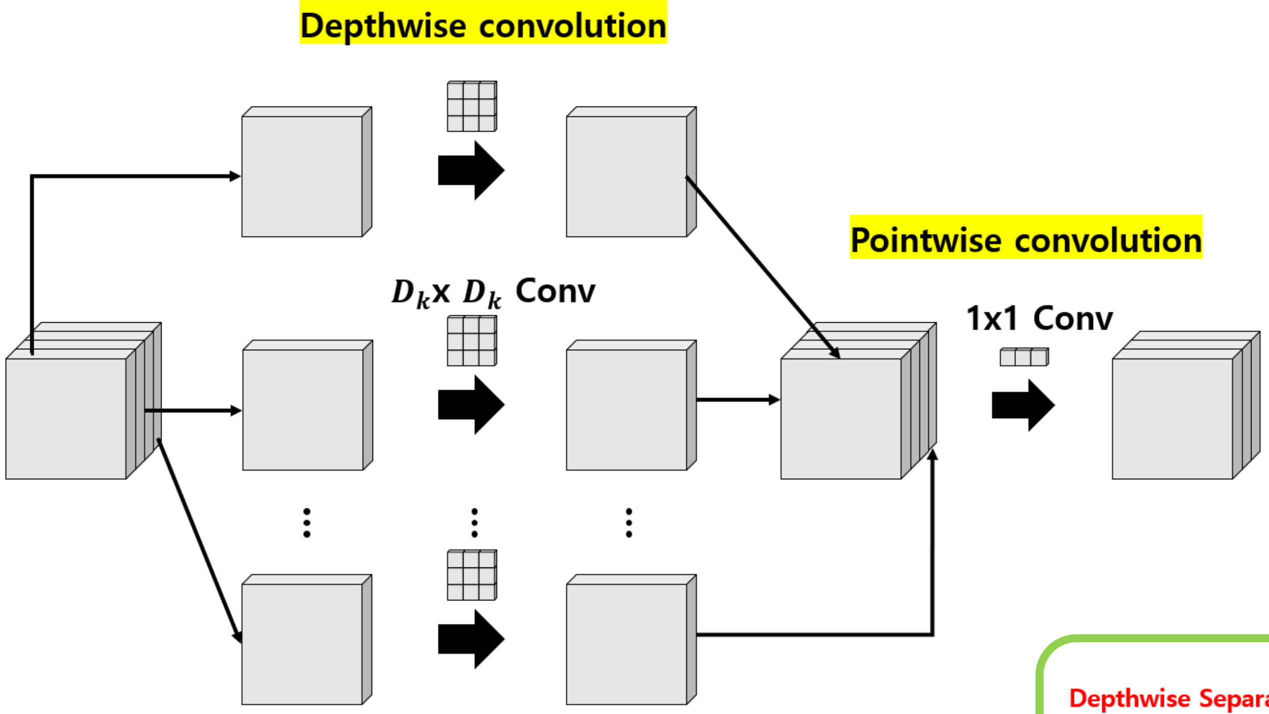
Pointwise Convolution $(M \cdot N \cdot D_F \cdot D_F)$



MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (CVPR 2017)

- **Method:**
  - 3x3 depthwise separable conv 기준 **<u>약 8~9배 적은 연산량</u>**

Depthwise Separable Convolution



Computational cost

Depthwise Separable Convolution

General Convolution

$$\frac{\boxed{M \cdot N \cdot D_F \cdot D_F + D_K \cdot D_K \cdot M \cdot D_F \cdot D_F}}{\boxed{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F}} = \frac{1}{N} + \frac{1}{D_K^2}$$

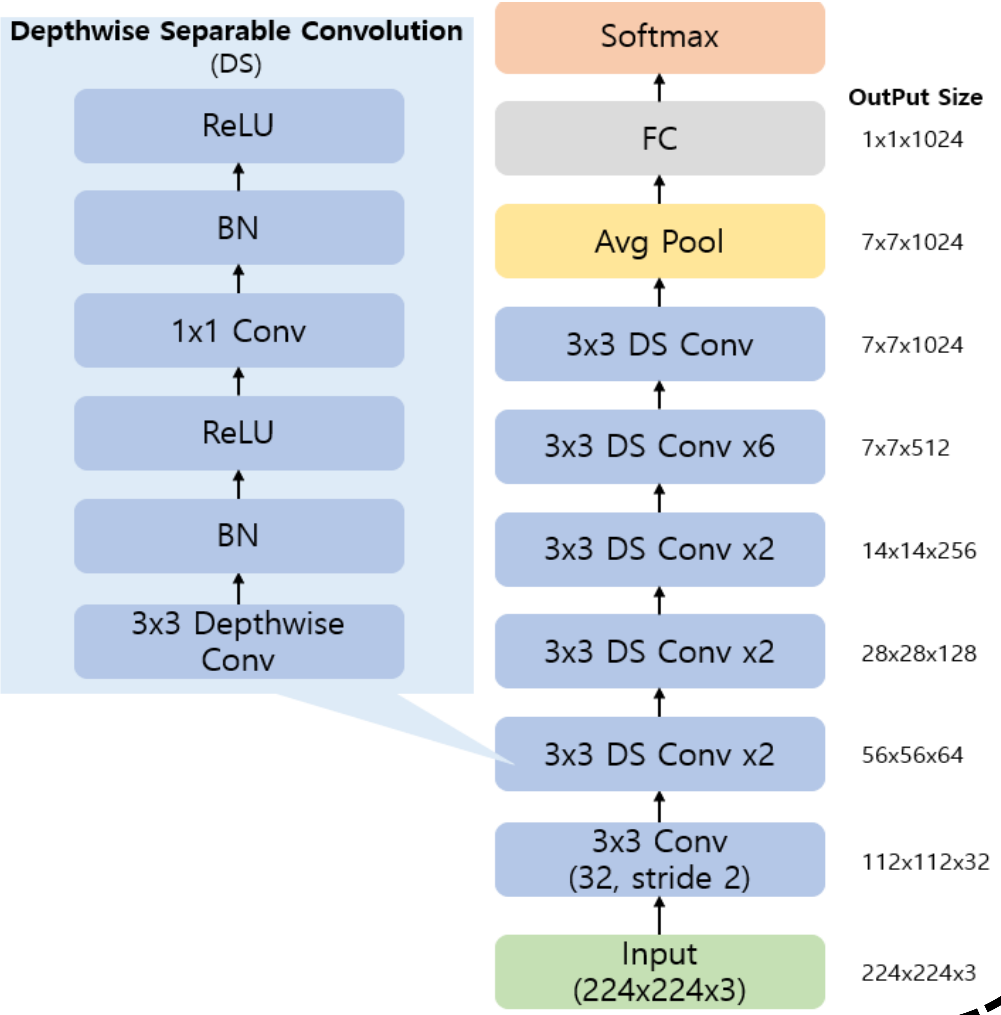
MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (CVPR 2017)

- **Method:**

MobileNet Architecture

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	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
$5\times$ Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
$5\times$ Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
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	$1 \times 1 \times 1000$



MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (CVPR 2017)

- **Method:**
  - 2개의 hyperparameter (width multiplier, resolution multiplier) **=>** **latency, accuracy** 균형 조절
  - **Width Multiplier** (α) **: Thinner Models** 네트워크를 균일하게 **얇게** 만듦

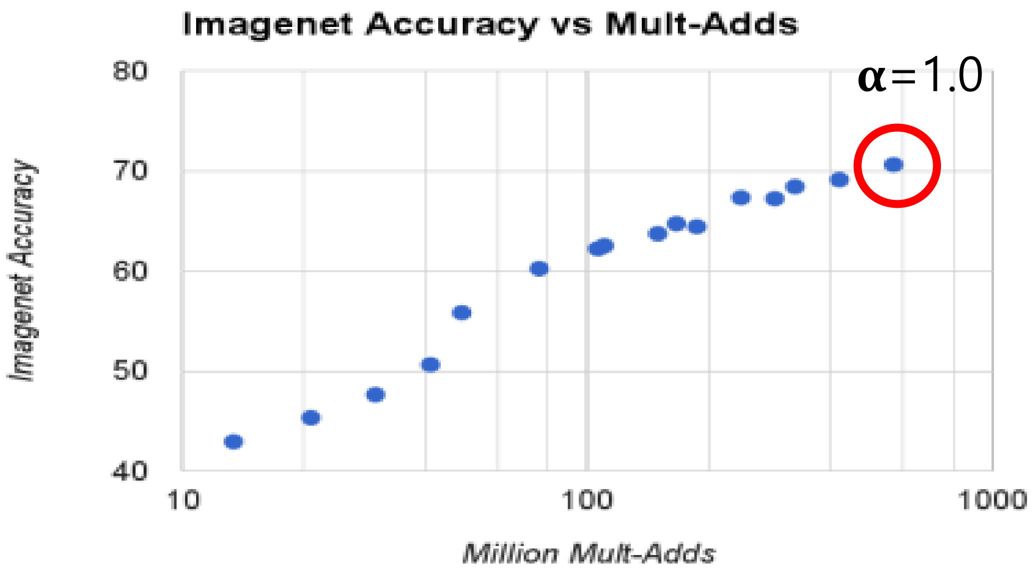
Computational cost using Width Multiplier (α)

$$\alpha M \cdot \alpha N \cdot D_F \cdot D_F + D_K \cdot D_K \cdot \alpha M \cdot D_F \cdot D_F$$

$\boldsymbol{\alpha} \in (\mathbf{0}, \mathbf{1})$, $\boldsymbol{\alpha}$ = **(**1, 0.75, 0.5, 0.25**)**

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



MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (CVPR 2017)

- **Method:**
  - 2개의 hyperparameter (width multiplier, resolution multiplier) **=> latency, accuracy 균형 조절**
  - **Resolution Multiplier (ρ) : Reduced Representation** 신경망의 **계산비용 감소**

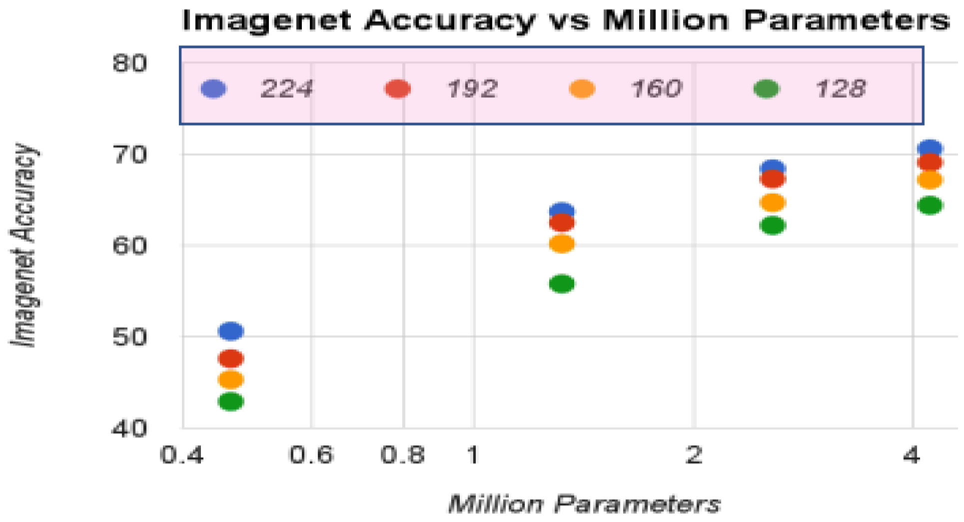
Computational cost using Resolution Multiplier (ρ)

$$\alpha M \cdot \alpha N \cdot \rho D_F \cdot \rho D_F + D_K \cdot D_K \cdot \alpha M \cdot \rho D_F \cdot \rho D_F$$

$\rho \in (0, 1)$, ρ = (224, 192, 160, 128)

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



MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (CVPR 2017)

- **Experiment:**
  - MobileNet(Depthwise Separable Convolution)이 Fully convolutional MobileNet보다 **더 적은 parameter를 나타내면서 합리적인 정확도**를 보임.
  - 기존의 very small network (squeezeNet, AlexNet) 보다 더 좋은 성능을 보임
  - 이 외에도 **다양한 task에서 성공적인 모델 경량화**

Table 4. Depthwise Separable vs Full Convolution MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
Conv MobileNet	71.7%	4866	29.3
MobileNet	70.6%	569	4.2

Better Performance

Table 5. Narrow vs Shallow MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
0.75 MobileNet	68.4%	325	2.6
Shallow MobileNet	65.3%	307	2.9

Better Performance

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

Table 9. Smaller MobileNet Comparison to Popular Models

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
0.50 MobileNet-160	60.2%	76	1.32
Squeezenet	57.5%	1700	1.25
AlexNet	57.2%	720	60

Table 10. MobileNet for Stanford Dogs

Model	Top-1 Accuracy	Million Mult-Adds	Million Parameters
Inception V3 [18]	84%	5000	23.2
1.0 MobileNet-224	83.3%	569	3.3
0.75 MobileNet-224	81.9%	325	1.9
1.0 MobileNet-192	81.9%	418	3.3
0.75 MobileNet-192	80.5%	239	1.9

한줄평:

MobileNet은 기존의 filter size로 parameter 수를 줄이려는 시각에서 벗어나 Depthwise Separable Convolution을 활용해 다양한 task에서 연산량을 줄이고 합리적인 성능을 달성했다.
하지만, 속도가 빨라진 대신 정확도가 낮아진 결과도 있어 추가적인 연구가 필요하다고 생각한다.