

MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

Background

- 핸드폰이나 임베디드 시스템 같이 **저용량 메모리환경**에 딥러닝을 적용하기 위해서는 **모델 경량화**가 필요하다.

Motivation

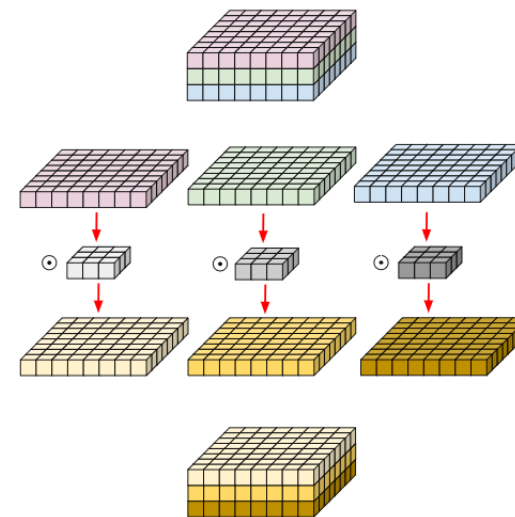
- Depthwise convolution은 각 입력 채널에 대하여 3x3 conv **하나의 필터**가 연산을 수행하여 **하나의 피처맵**을 생성합니다. 그리고 **연산된 피처맵을 합쳐준다**. 이 과정의 파라미터를 표현하면 다음과 같다.

- $D_K \cdot D_K \cdot M \cdot D_F \cdot D_F$
- 일반 convolution: $D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$

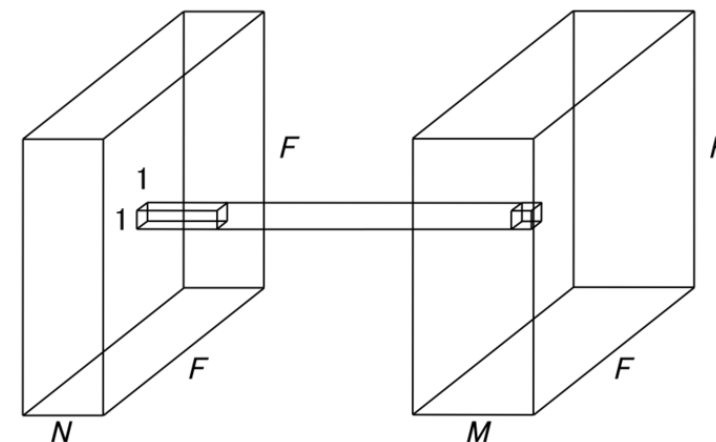
- Pointwise Convolution 은 공간방향의 convolution은 진행하지 않고, 채널 방향의 convolution을 진행하는 것이다. 즉, Channel Reduction 시에 주로 사용된다. **1x1xN 크기의 커널을 사용하여 입력의 특징을 M개의 채널로 압축** 시키는 효과를 가지고 있다. 그리고 이 과정의 파라미터를 표현하면 다음과 같다.

- $M \cdot N \cdot D_F \cdot D_F$

(D_K : 입력 값 크기, M : 입력 채널 수, N : 출력 채널 수, D_F : 피처맵 크기)



< Depthwise convolution >

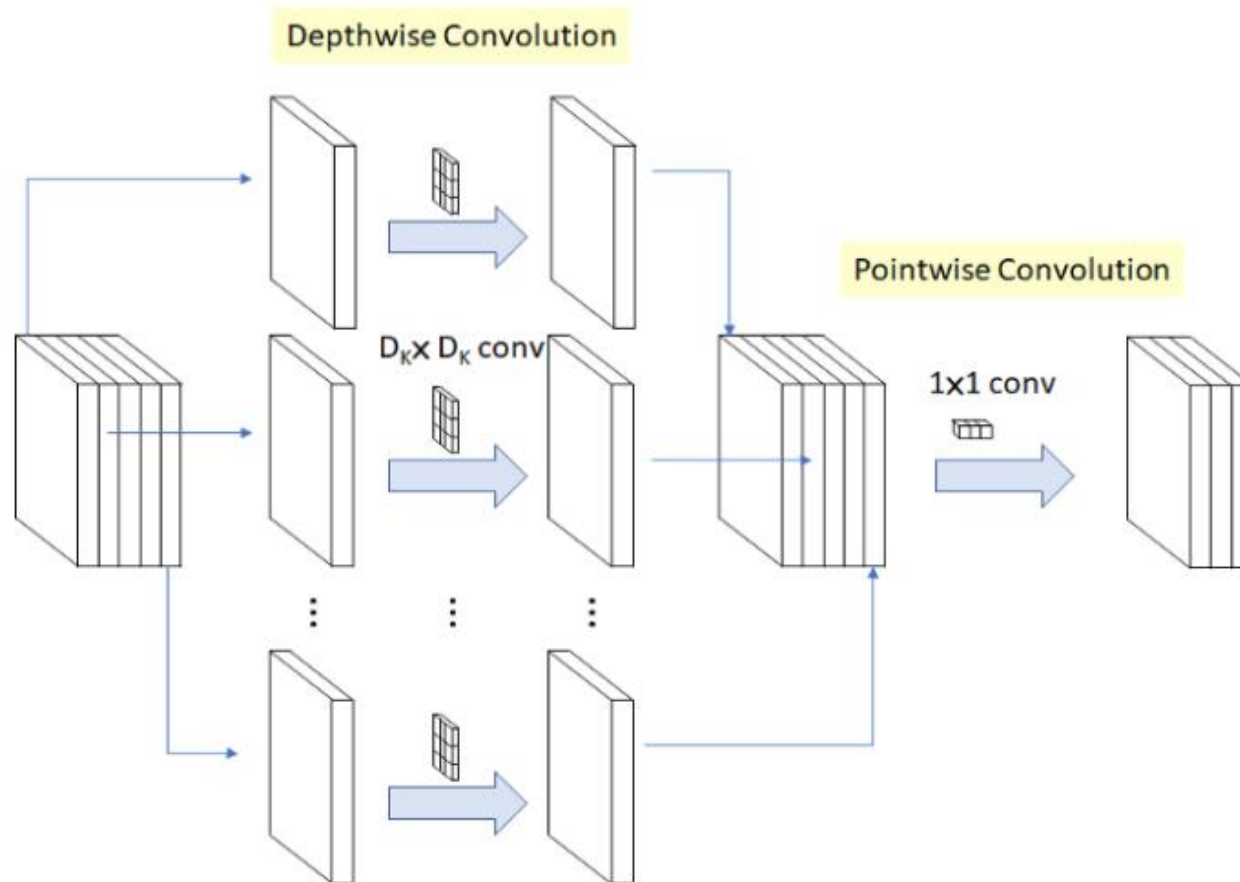


< Pointwise convolution >

Model

- Depthwise convolution과 Pointwise convolution을 차례로 배치한 Depthwise Separable Convolution을 제안하였고, 전체 구조는 이 블록을 쌓아서 만든다. 이 블록의 파라미터는 다음과 같다.

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



< Depthwise Separable Convolution 블록 구조 >

Hyper parameter

- 첫 번째 하이퍼파라미터 α 는 MobileNet의 두께를 결정한다. Convolution net에서 두께는 **각 레이어에서 필터 수**를 의미한다. 이 width Multiplier α 는 더 얇은 모델이 필요할 때 사용한다. 입력 채널 M 과 출력 채널 N 에 적용하여 αM , αN 이 됩니다. 그리고 이를 수식을 표현하면 다음과 같다.

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

- 두 번째 하이퍼파라미터는 Resolution Multiplier ρ 입니다. 모델의 연산량을 감소시키기 위해 사용한다. ρ 는 **입력 이미지에 적용하여 해상도를 낮춘다**. 범위는 0~1이고, 논문에서는 입력 이미지 크기가 224, 192, 169, 128 일때 비교한다. 기본 MobileNet은 $\rho=1$ 을 사용한다.

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

<Width Multiplier α 에 따른 MobileNet성능 변화>

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

< Resolution Multiplier ρ 에 따른 MobileNet성능 변화>

Result

- **좋은 성능**을 보이는 GoogleNet과 VGG을 MobileNet과 비교하였다. GoogleNet과 비교하였을 때는 성능이 좋을 뿐더러 파라미터 또한 더 적었다. 그리고 VGG와 비교하였을 때 성능이 조금 안 좋지만 파라미터는 약 1/32로 줄었다.
- **경량화를 목표로** 한 모델인 Squeezenet과 AlexNet을 MobileNet과 비교하였다. Squeezenet과 비교하였을 때 약간 더 많은 파라미터를 가졌지만 더 좋은 성능을 가졌고, AlexNet과 비교하였을 때 더 좋은 성능을 보이고, 파라미터는 약 1/45로 줄었다.

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

<성능이 좋은 모델과 성능 및 파라미터 비교>

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

<작은 모델과의 성능 및 파라미터 비교>