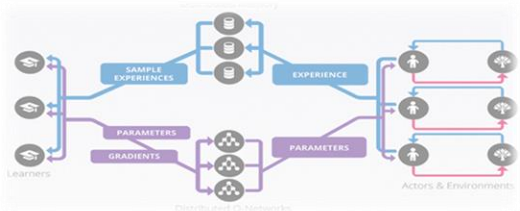



ShuffleNet 구현 및 실험 결과 보고

AutoML 연구실
석사과정 문아성

20.10.06

Introduction

ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile

2017.07 arXiv에 등록되어, 2018 CVPR Conference

"ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile" 로 발표

MobileNet "Depthwise Convolution", AlexNet "Group Convolution" 개념을 도입

- Pointwise Group Convolution
- Channel Shuffle

Arm-based Mobile device 에서 실험을 통해 정확성을 유지하면서 MobileNet과 유사하지만,

MobileNet보다 연산량을 줄임으로써 더욱 작은 네트워크 아키텍쳐를 제안

others Model: Inception

Inception-v4 network module

5x5 Conv. 대신 3x3 Conv. 를 2번 사용하는 것 처럼, nxn conv. 를 1xn conv., nx1 conv. 로 대신 사용하여 (Height, Width)를 줄이면서 동일한 receptive field를 가지는 Module을 제안

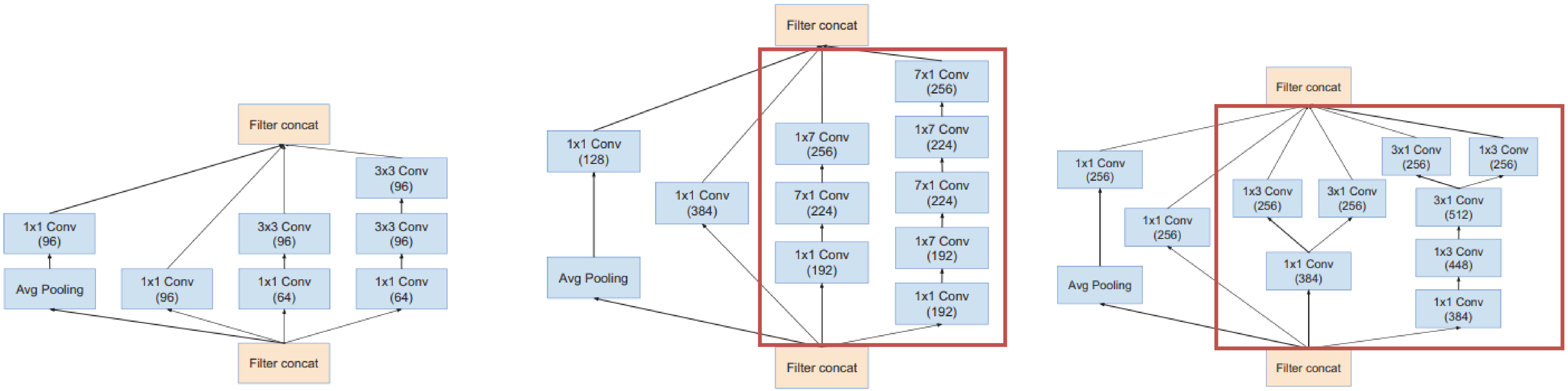


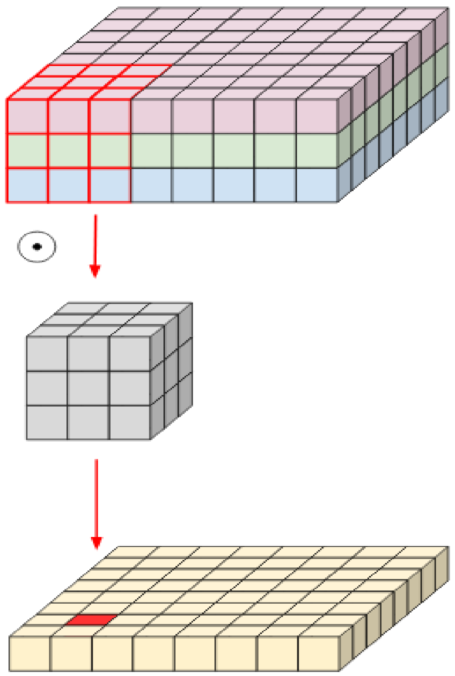
Figure 3: The schema for interior grid modules of the pure Inception-v4 network. The 35×35, 17×17 and 8×8 grid modules are depicted from left to right. These are the Inception-A, Inception-B, and Inception-C blocks of Figure 2 respectfully.

others Model: MobileNet (1)

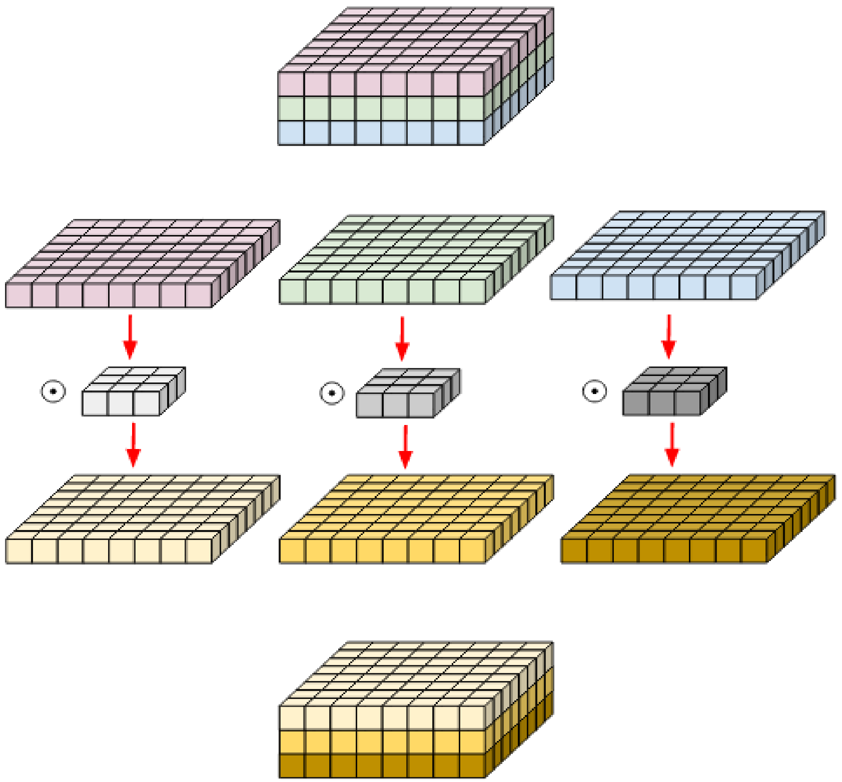
Depthwise Convolution of MobileNet

앞에서는 spatial한 방향에 변화를 주었다면, MobileNet에서는 Channel에 변화를 시도

채널별로 각각 Convolution 연산



Standard Convolution



Depthwise Convolution

others Model: MobileNet (2)

Depthwise Separable Convolution of MobileNet

3x3 Conv.을 통해 spatial한 방향과 1x1 Conv. 을 통해

Channel 방향(Pointwise Conv.)으로 분리하여 보는 방법을 제안

Depthwise Convolution + Pointwise Convolution

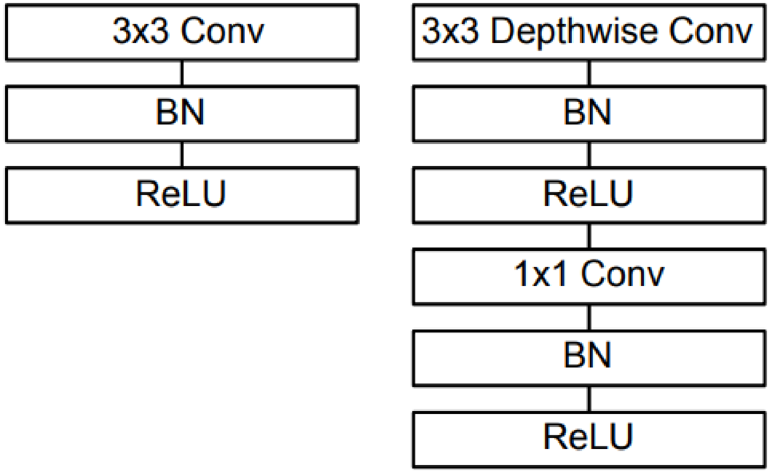
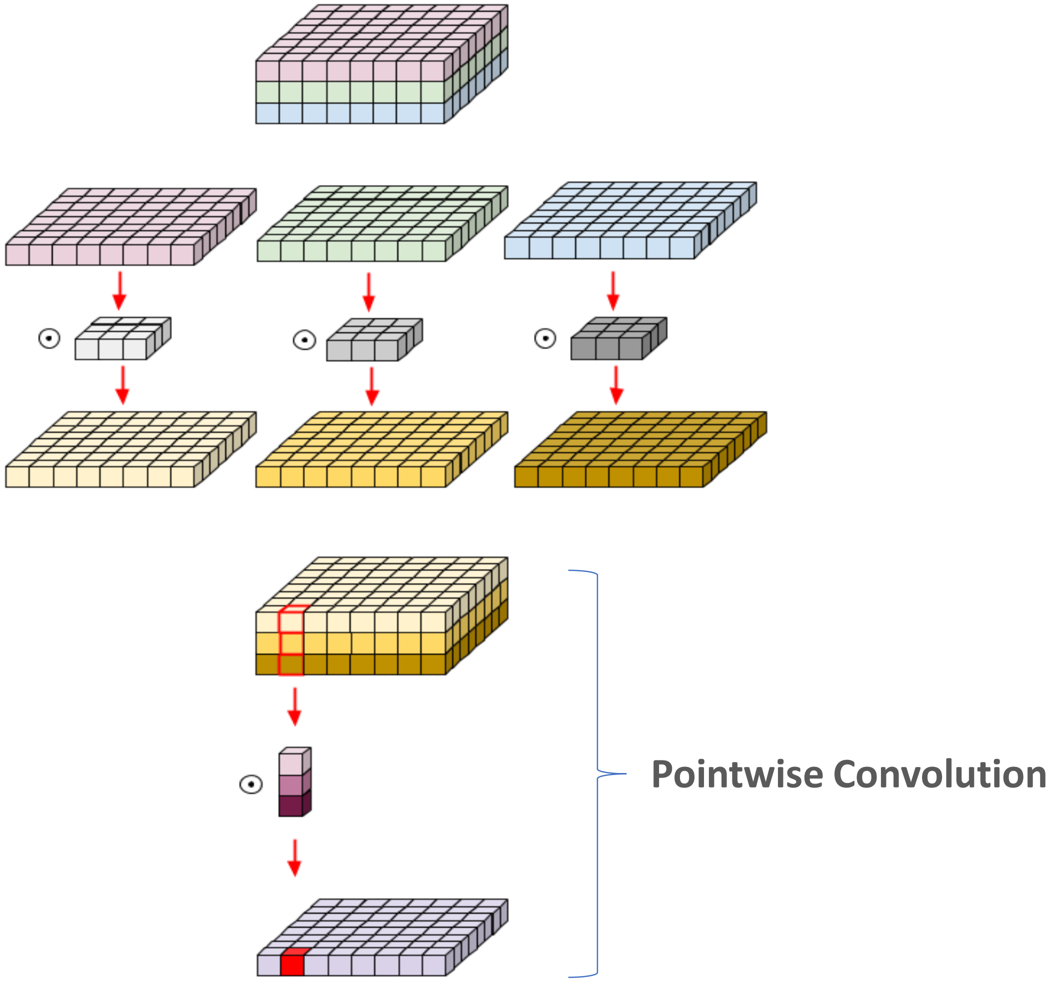


Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.



Depthwise Separable Convolution

others Model: MobileNet (3)

Can we reduce more?

대부분 1x1 Conv. Layer 에서 연산이 이루어지는 것을 알 수 있음

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%

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	$1 \times 1 \times 1000$

ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile (1)

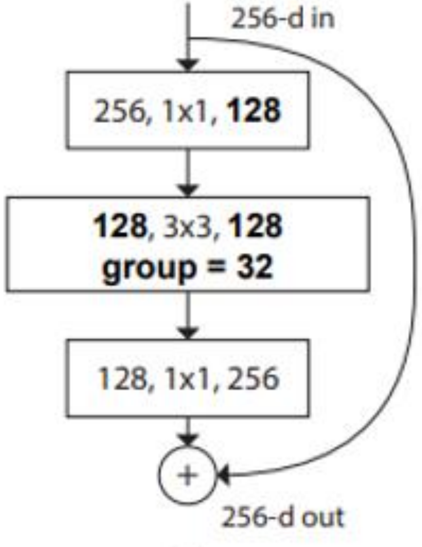
Grouped Convolution

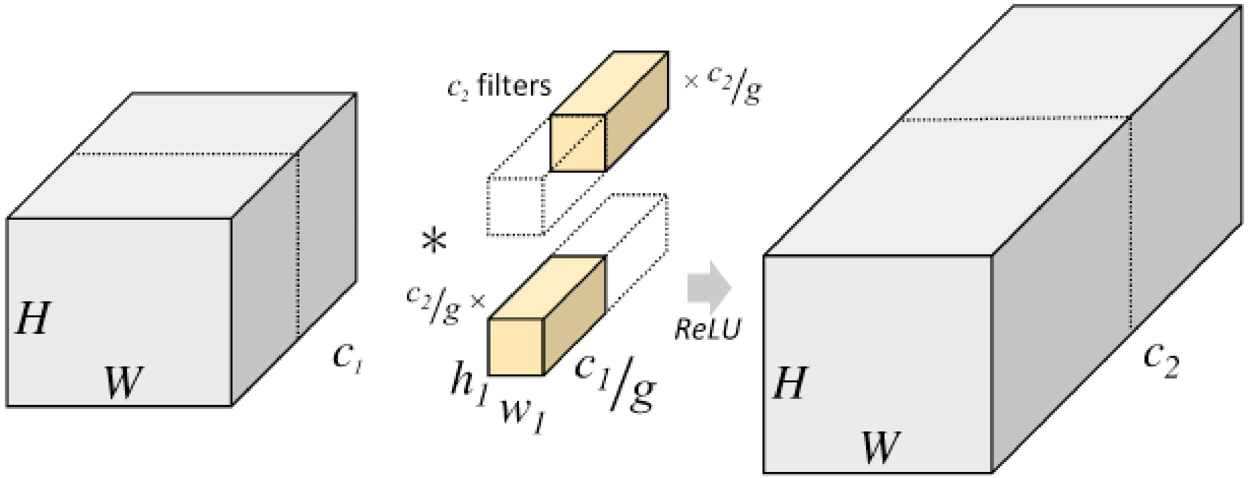
Input 의 channel들을 여러 개의 group으로 나누어 독립적으로 Convolution을 수행

Ex) Input Channel : 128 group : 32 Output Channel : 128 일 때,

128개 채널을 32개의 그룹으로 나누어 Convolution을 수행하며, 각 그룹은 4개의 채널을 갖는다.

그 다음, 그룹에서 나온 출력을 이어붙여 (concatenate) 다시 128 채널을 가진다.





g : 나눌 그룹의 수

ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile (2)

1 x 1 Grouped Convolution with Channel Shuffling

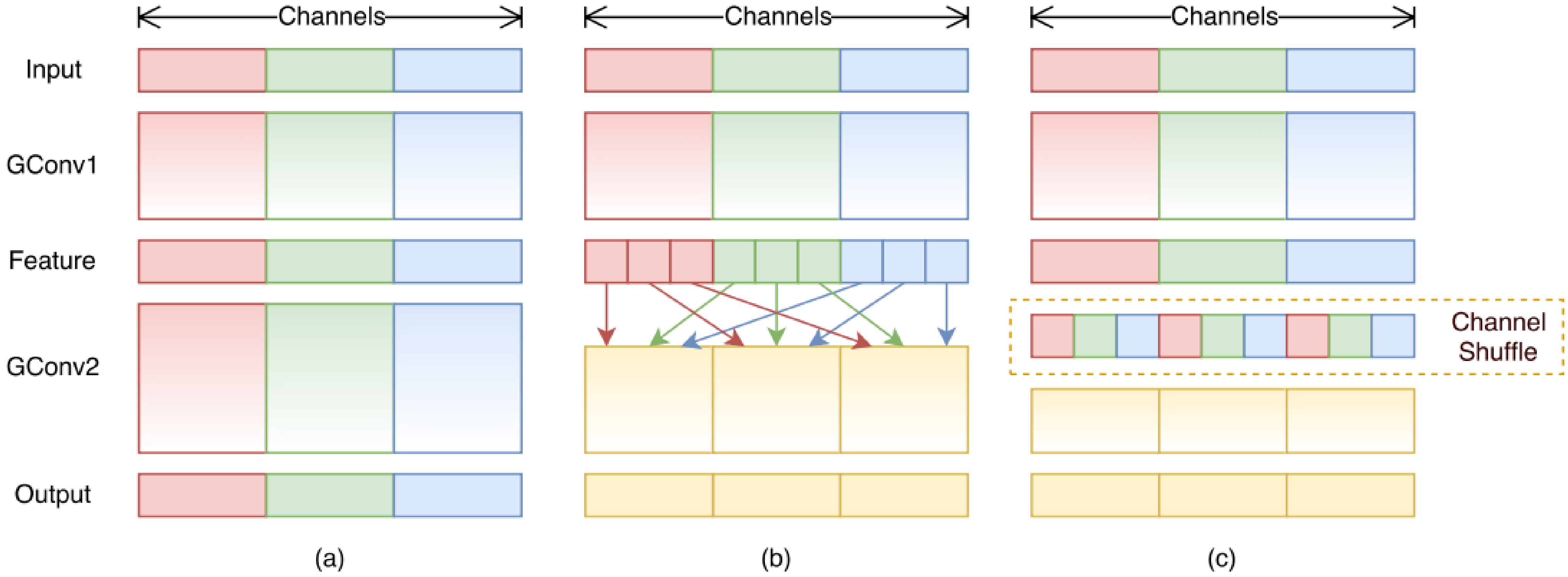


Figure 1. Channel shuffle with two stacked group convolutions. GConv stands for group convolution. a) two stacked convolution layers with the same number of groups. Each output channel only relates to the input channels within the group. No cross talk; b) input and output channels are fully related when GConv2 takes data from different groups after GConv1; c) an equivalent implementation to b) using channel shuffle.

ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile (3)

ShuffleNet Unit

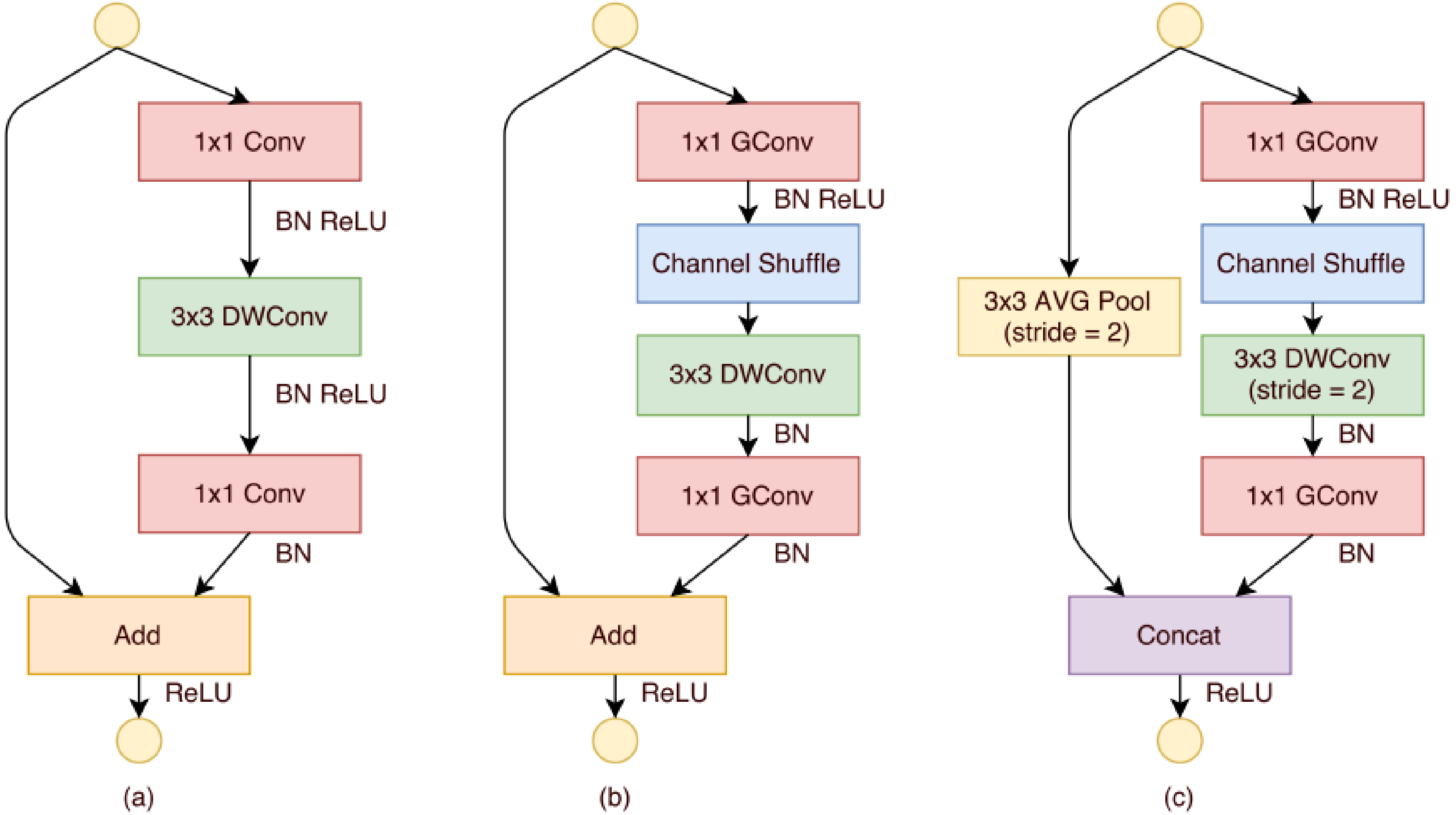


Figure 2. ShuffleNet Units. a) bottleneck unit [9] with depthwise convolution (DWConv) [3, 12]; b) ShuffleNet unit with pointwise group convolution (GConv) and channel shuffle; c) ShuffleNet unit with stride = 2.

ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile (4)

Experimental Results

동일한 Complexity 에서 다양한 모델을 적용하여 실험

Complexity (MFLOPs)	VGG-like	ResNet	Xception-like	ResNeXt	ShuffleNet (ours)
140	50.7	37.3	33.6	33.3	**32.4** ($1\times$, $g = 8$)
38	-	48.8	45.1	46.0	**41.6** ($0.5\times$, $g = 4$)
13	-	63.7	57.1	65.2	**52.7** ($0.25\times$, $g = 8$)

Table 4. Classification error vs. various structures (%, *smaller number represents better performance*). We do not report VGG-like structure on smaller networks because the accuracy is significantly worse.

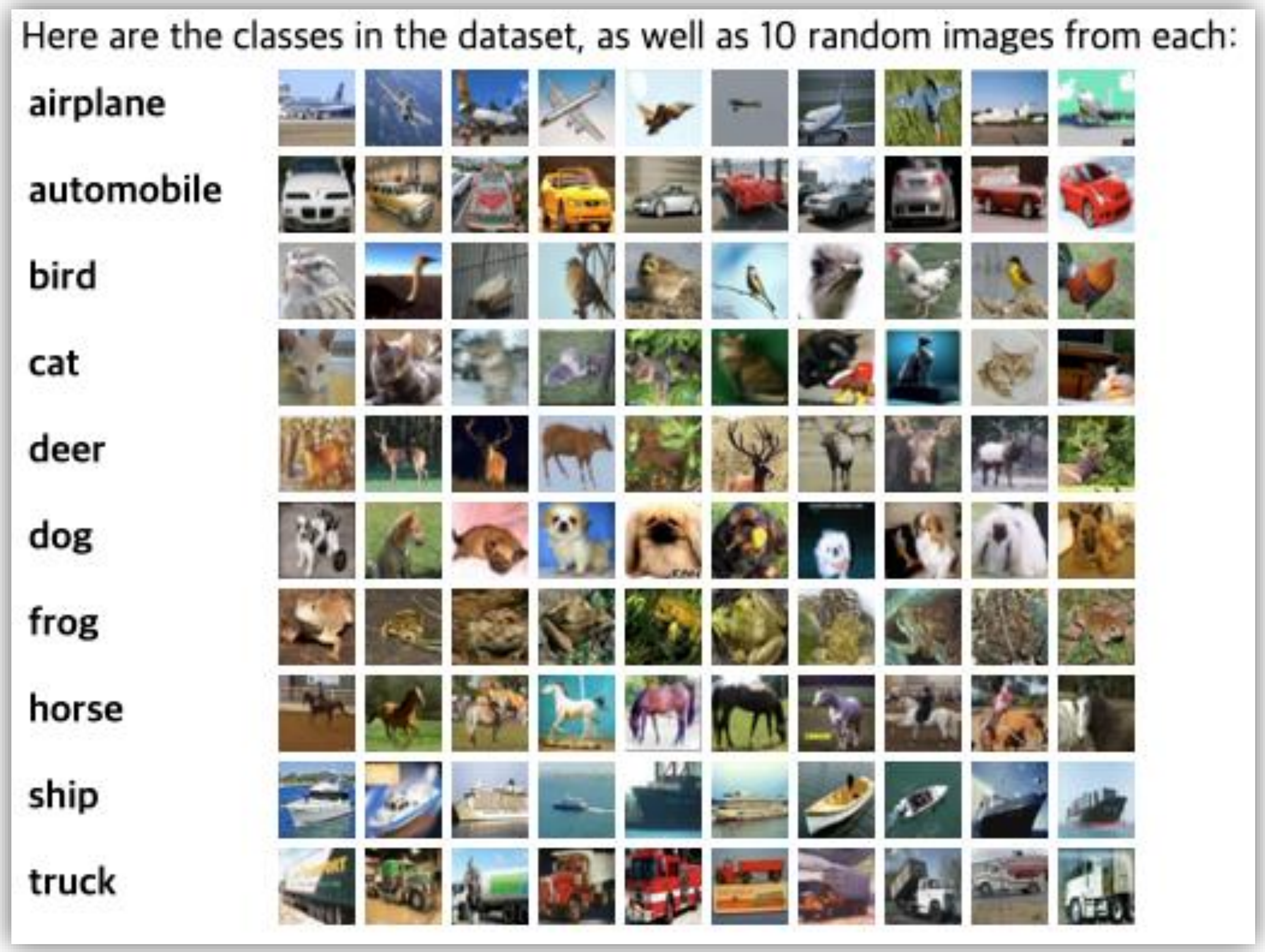
DATASET CIFAR-10

32x32사이즈의 RGB 이미지와

10가지의 레이블로 구성 $(32, 32, 3)$, (10)

INPUT DATA : 50000

TEST DATA : 10000



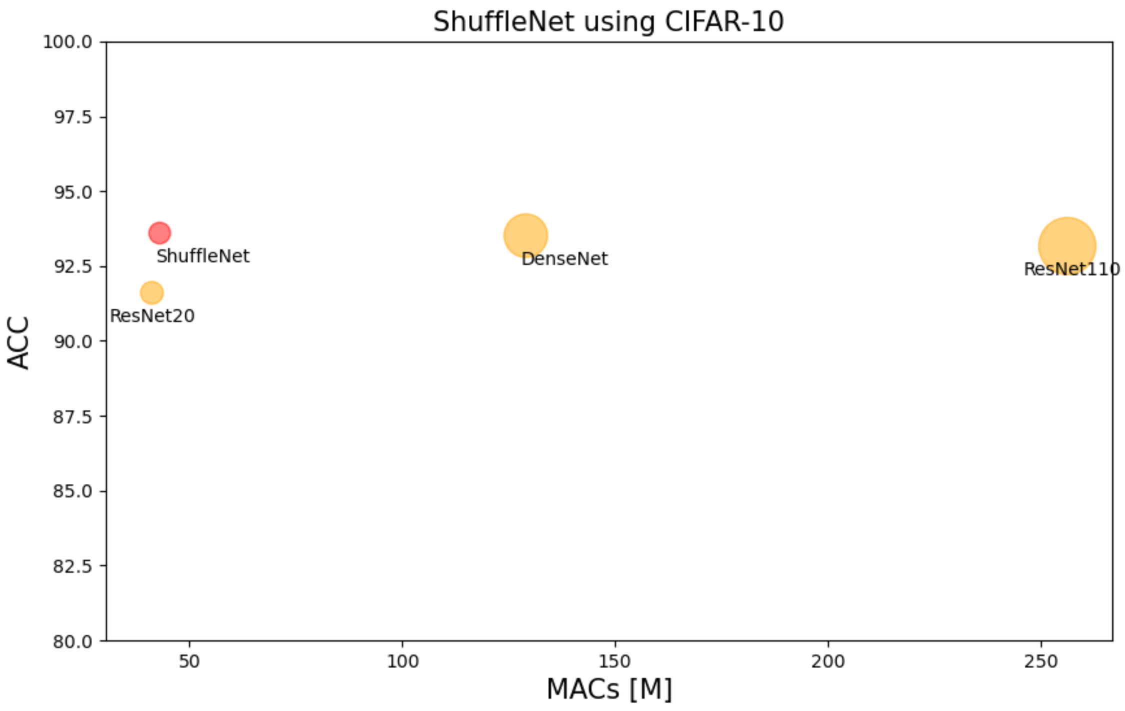
실험 결과

정밀도(Precision): $\frac{TP}{TP+FP}$

재현율(Recall): $\frac{TP}{TP+FN}$

정확도(Accuracy): $\frac{TP+TN}{TP+FP+FN+TN}$

Name	# params	precision	recall	accuracy
ResNet-20	0.27M	91.69%	91.62%	**91.63%**
ShuffleNet (0.5, g=3)	**0.26M**	**92.73%**	**92.72%**	**93.62%**
ResNet-110	1.7M	93.19%	93.17%	**93.17%**
DenseNet	1.78M	93.76%	93.77%	**94.22%**



Next Model

Xception