Very Deep Convolutional Networks For Large-Scale Image Recognition

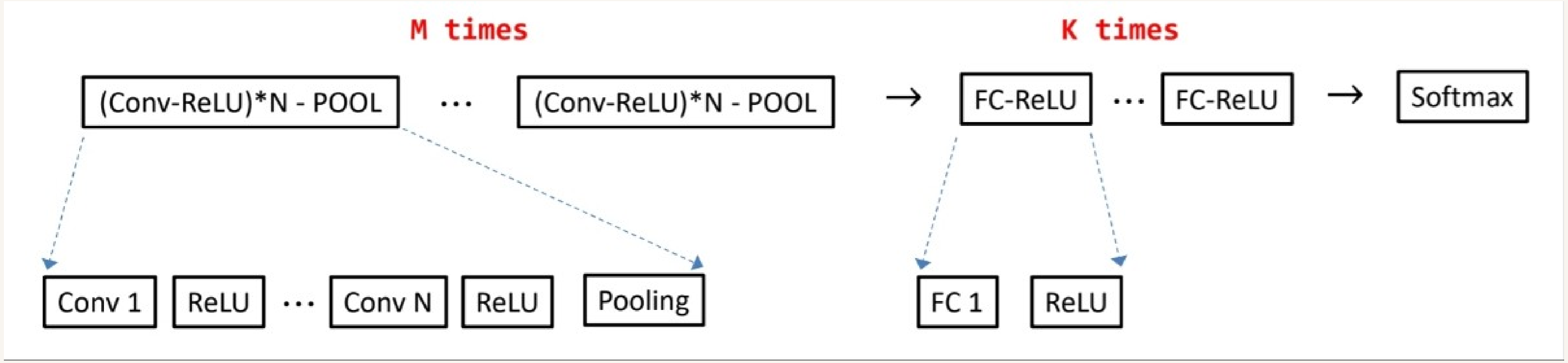
Karen Simonyan   Andrew Zisserman

박민경

BackGround

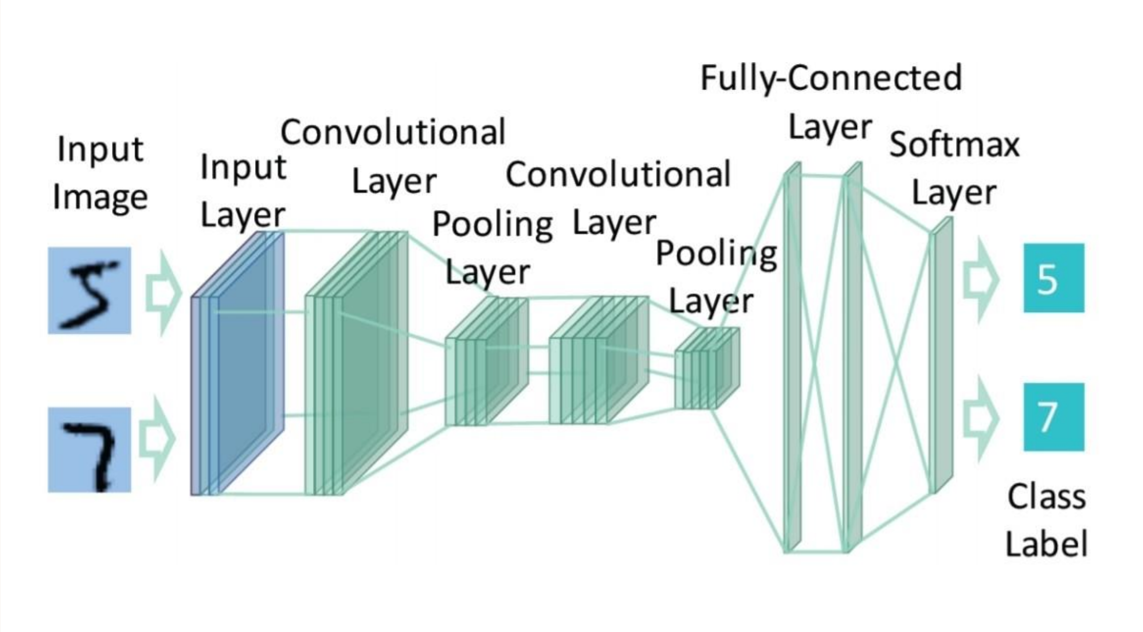
Architecture

[(Conv-ReLU)*N - POOL]*M - (FC-ReLU)*K – Softmax



BackGround

Architecture



Convolution Layer

- computes the output of neurons that are connected to local regions in the input.

Non-Linear Layer

- activates relevant responses

Pooling Layer

- performs a down sampling operation along the spatial dimensions

Fully-Connected Layer

- each neuron in this layer will be connected to all the numbers in the previous volume

VGG-16

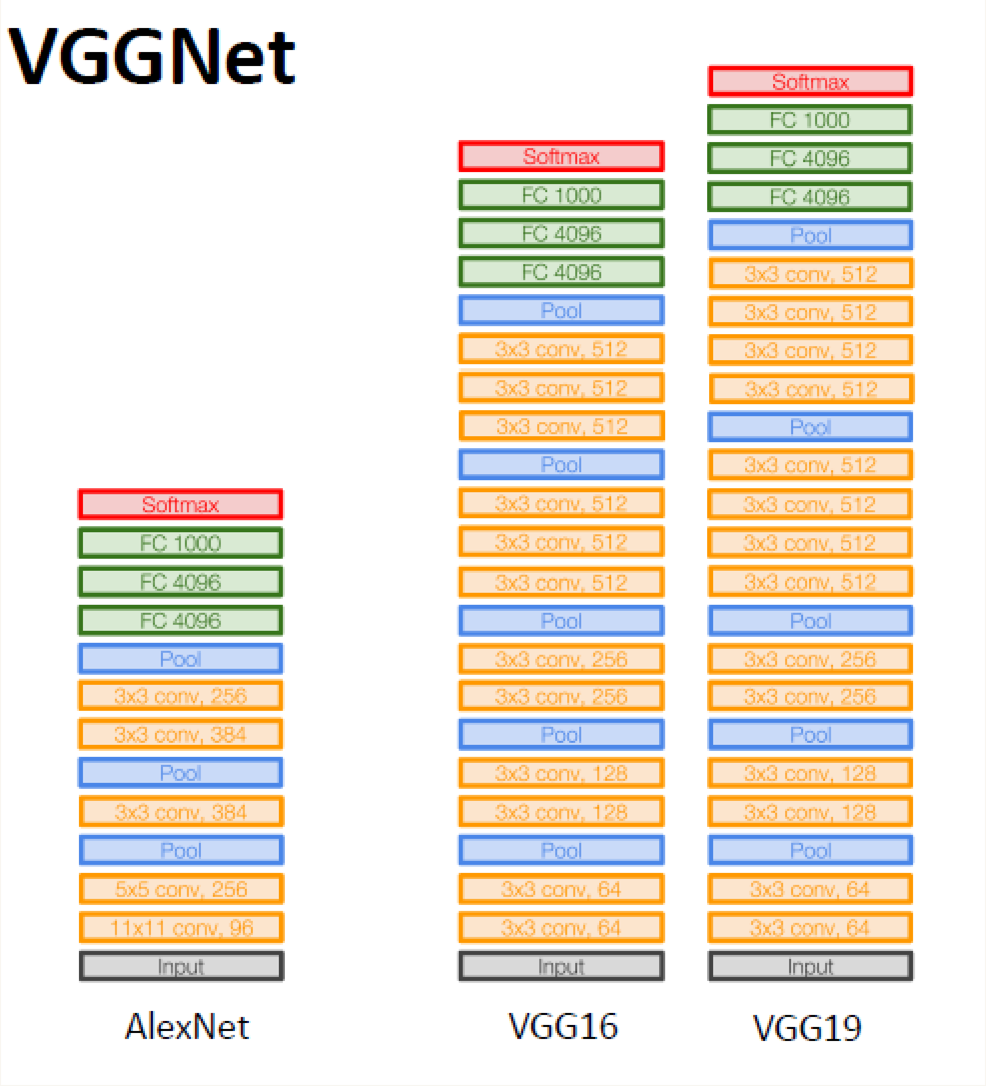
Architecture

Small Filters & Deeper Networks 사용
No Local Response Normalisation
Training process는 AlexNet과 유사함

모든 은닉층에서는 activation으로 ReLU사용

Small Filter: 3x3 conv (stride 1)
→ same receptive fields (7x7), fewer parameters

Deeper Networks: 16-19layer
→ more non-linearity

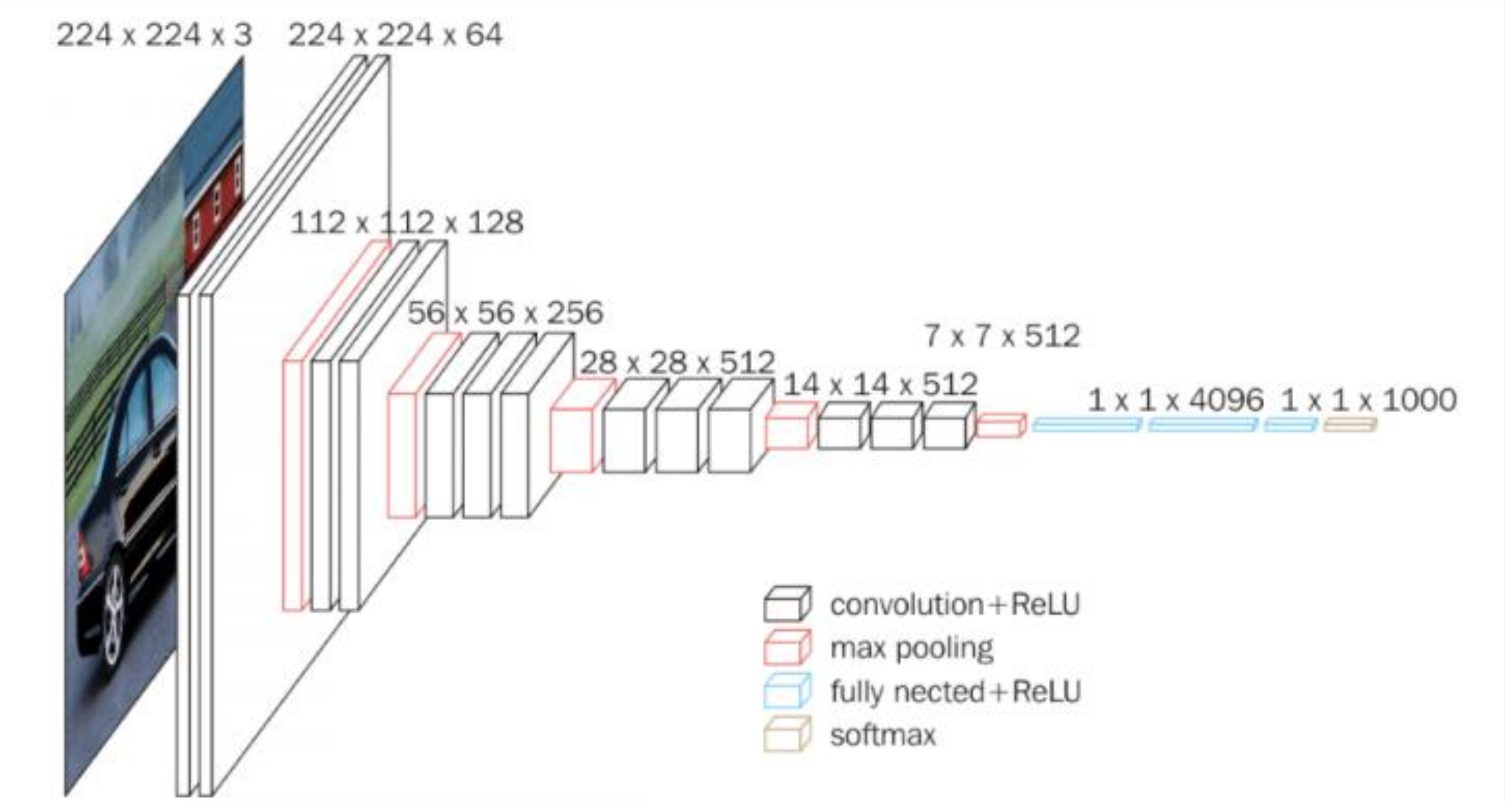


VGG-16

Architecture

Input: 224 x 224 RGB
preprocess: sub mean RGB val

labels: 1000



VGG-16

Convolution Layer: Filter

3x3 filter를 사용

Smallest Size to Capture the Notion of L/R,U/D, C

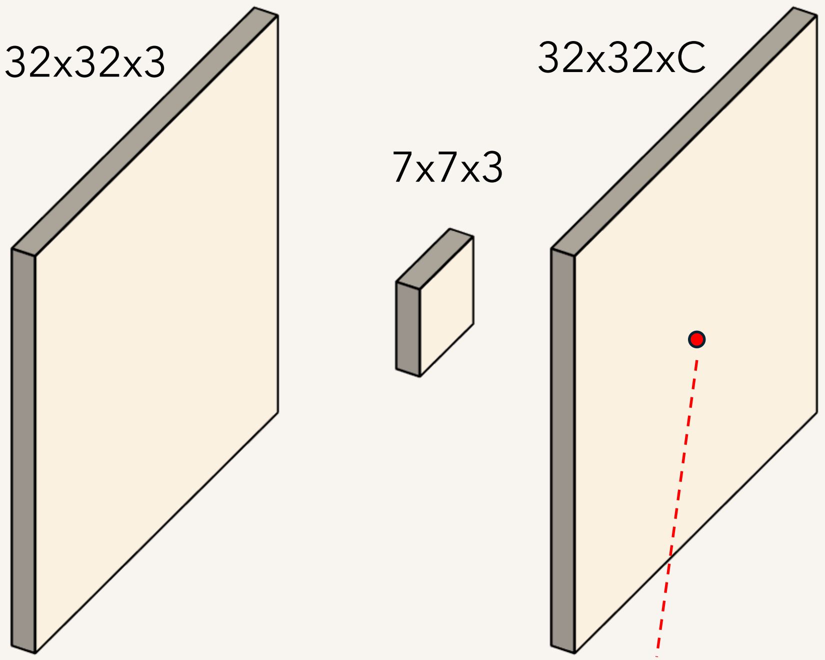
7x7과 동일한 receptive field를 가지지만,
더 적은 수의 parameter가 사용된다.

	Layer	Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	224 x 224 x 3	-	-	-
1	2 X Convolution	64	224 x 224 x 64	3x3	1	relu
	Max Pooling	64	112 x 112 x 64	3x3	2	relu
3	2 X Convolution	128	112 x 112 x 128	3x3	1	relu
	Max Pooling	128	56 x 56 x 128	3x3	2	relu
5	2 X Convolution	256	56 x 56 x 256	3x3	1	relu
	Max Pooling	256	28 x 28 x 256	3x3	2	relu
7	3 X Convolution	512	28 x 28 x 512	3x3	1	relu
	Max Pooling	512	14 x 14 x 512	3x3	2	relu
10	3 X Convolution	512	14 x 14 x 512	3x3	1	relu
	Max Pooling	512	7 x 7 x 512	3x3	2	relu
13	FC	-	25088	-	-	relu
14	FC	-	4096	-	-	relu
15	FC	-	4096	-	-	relu
Output	FC	-	1000	-	-	Softmax

VGG-16

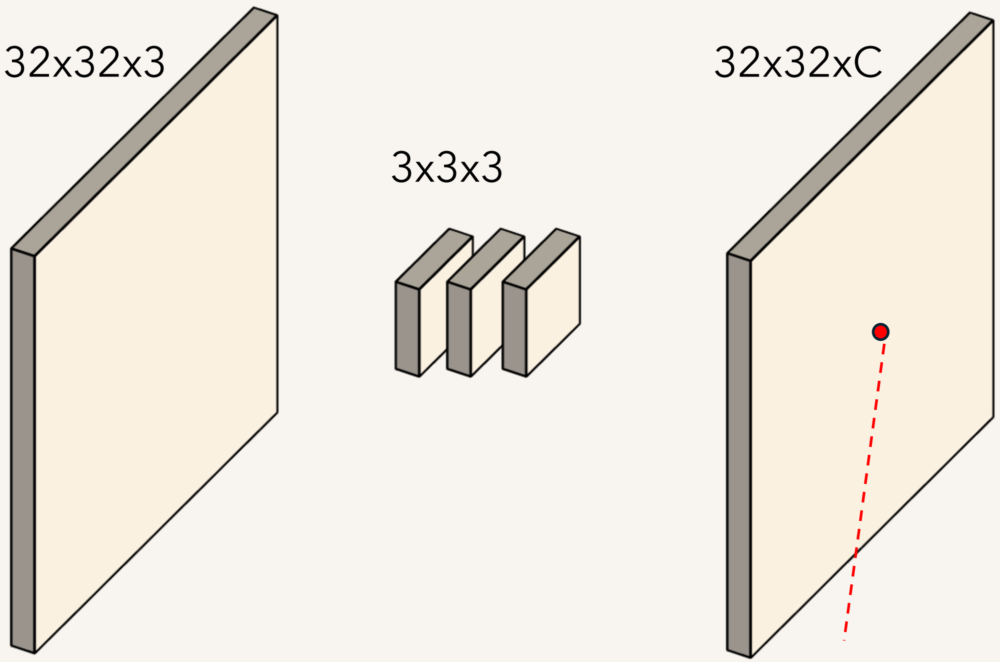
Convolution Layer: Receptive Fields

7x7 Filter 사용, C channels



receptive field: 7^2
parameter: 7x7xC = 49C

3x3 Filter(Stride 1) 사용, C channels



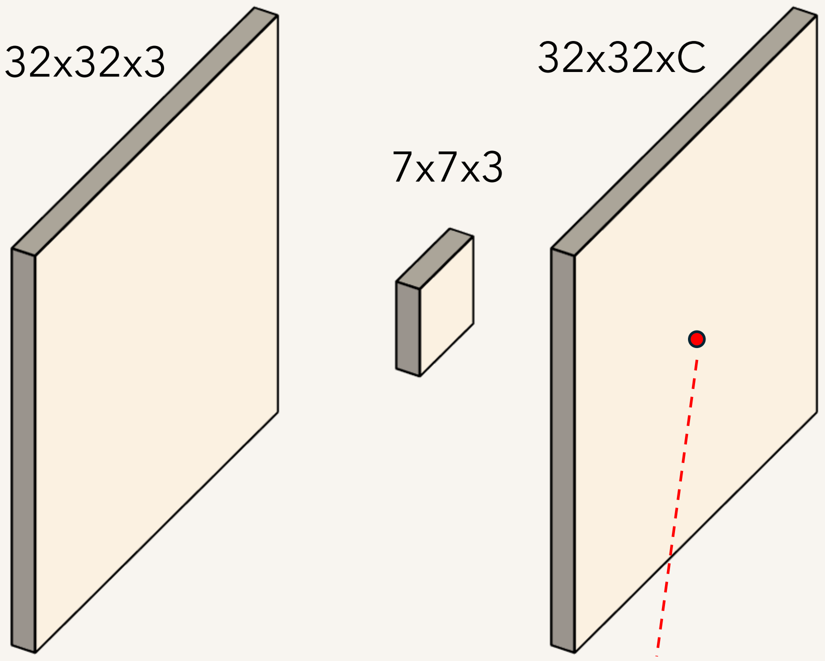
receptive field: $3+(3-1)+(3-1)^2 = 7^2$
parameter: 3x3X3xC = 27C

VGG-16

Convolution Layer: Receptive Fields

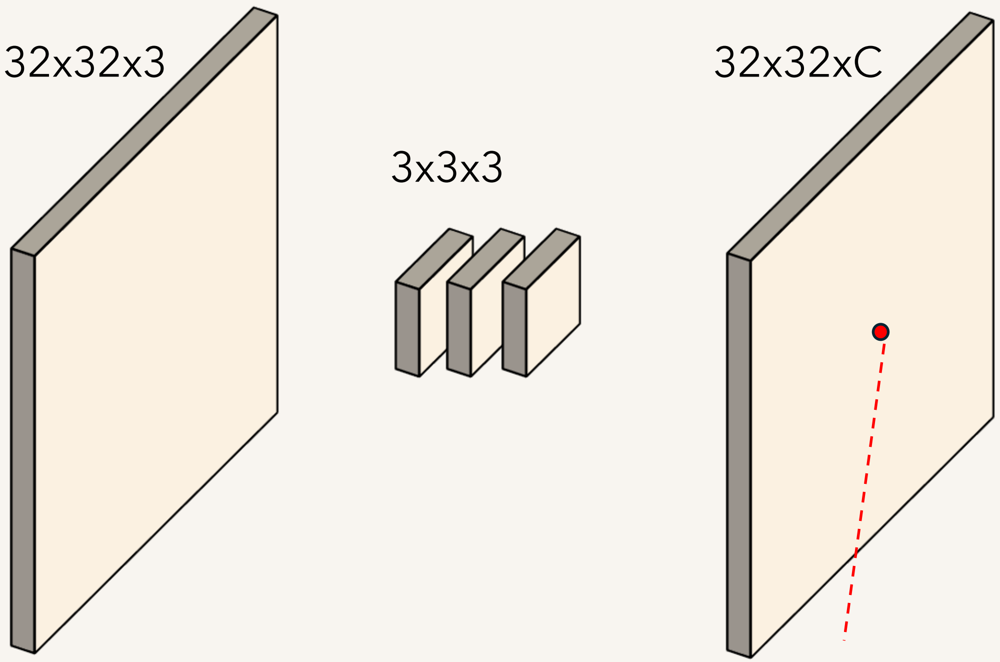
1. **Three Non-Linear layer – more discriminative**
2. **Decrease Parameters**

7x7 Filter 사용, C channels



receptive field: 7^2
parameter: 7x7xC = 49C

3x3 Filter(Stride 1) 사용, C channels

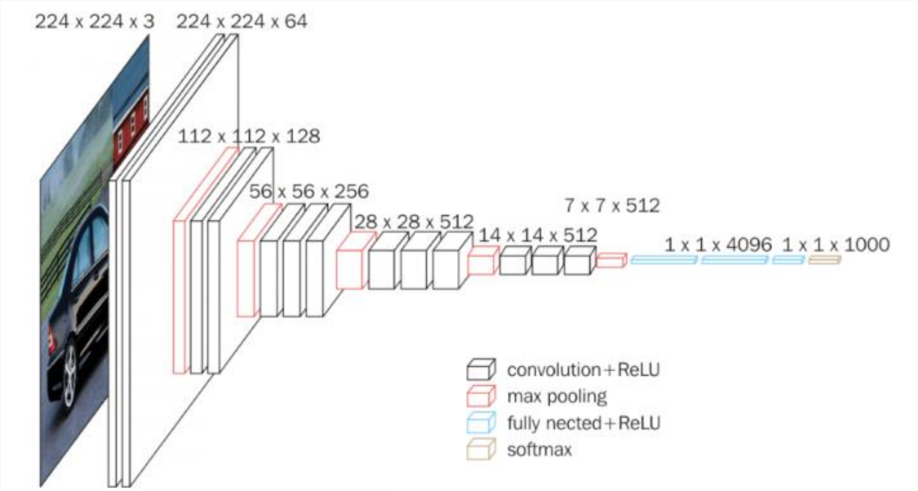


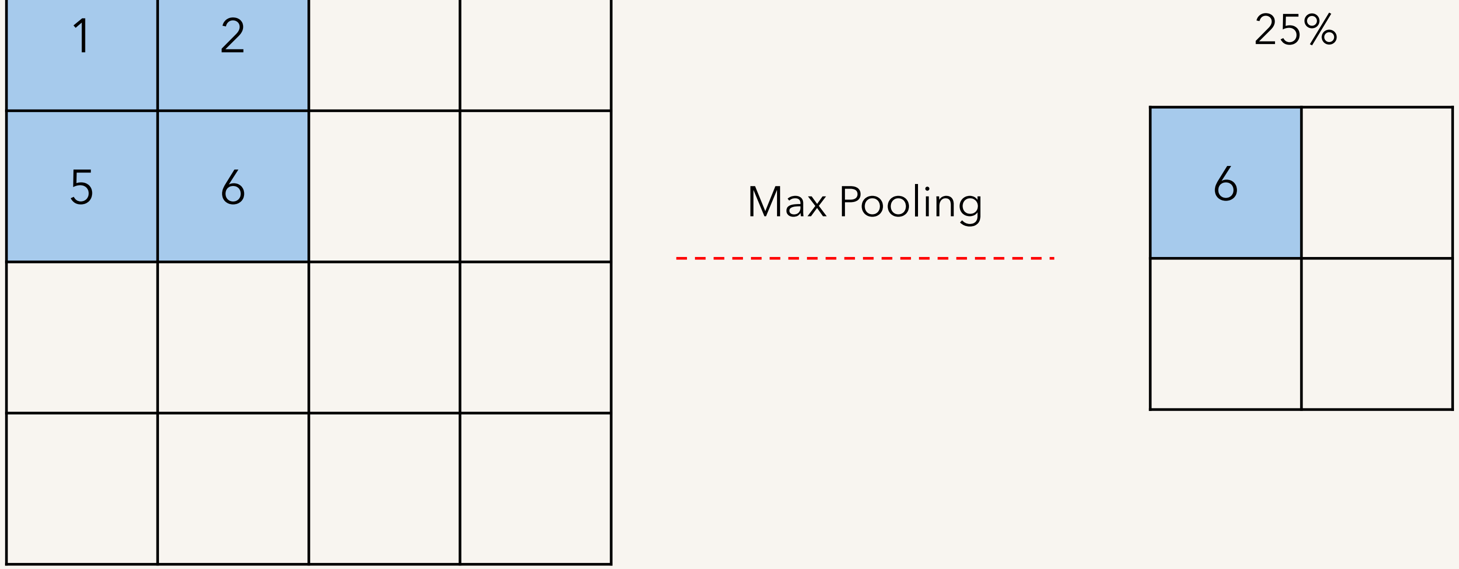
receptive field: $3+(3-1)+(3-1)^2 = 7^2$
parameter: 3x3X3xC = 27C

VGG-16

Pooling

Max Pooling, 2x2 window, stride 2 사용하여 Down Sampling
Representations Smaller > More Manageable

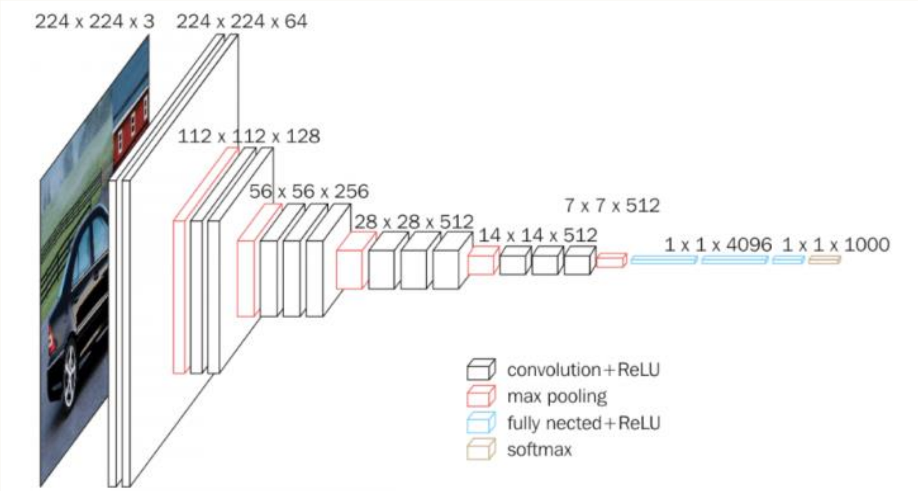


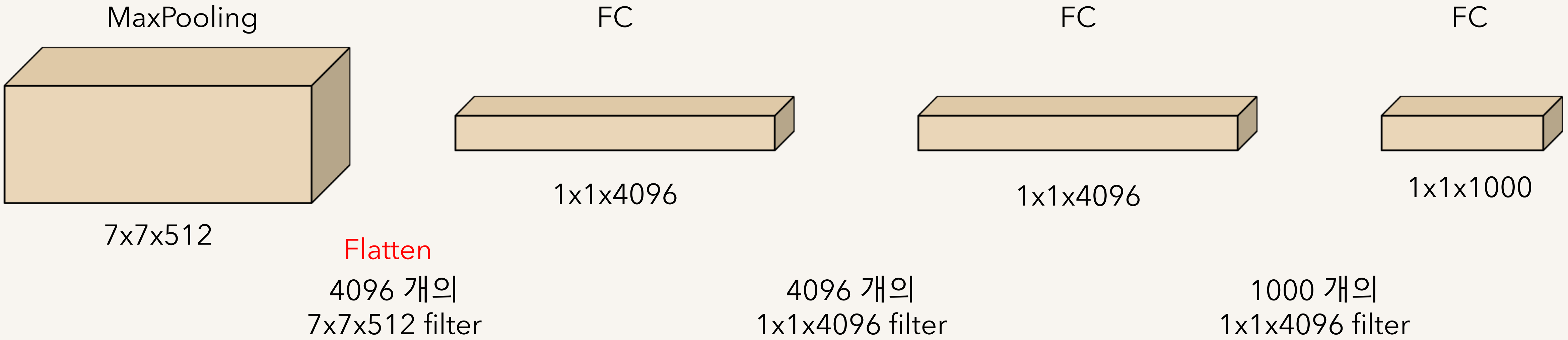


VGG-16

Fully Connected Layer

3 Fully Connected Layer 사용
(implemented as conv)





VGG-16

Test

FC Layer들을 Conv Layer로 바꿔서 테스트 한다.

- Training 시에는 이미지를 Crop하여 크기를 맞췄지만, Test 시에는 여러 크기의 이미지를 사용할 수 있게 한다.

- 동일 Filter 사용 시에, output layer의 크기가 다르게 나올 수 있음 > 이 경우에 sum pooling으로 1x1x1000form으로 만들게 된다.

VGG-16

Test & Result

Table 1: **ConvNet configurations** (shown in columns). The depth of the configurations increases from the left (A) to the right (E), as more layers are added (the added layers are shown in bold). The convolutional layer parameters are denoted as "conv⟨receptive field size⟩-⟨number of channels⟩". The ReLU activation function is not shown for brevity.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Table 3: **ConvNet performance at a single test scale.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
A	256	256	29.6	10.4
A-LRN	256	256	29.7	10.5
B	256	256	28.7	9.9
C	256	256	28.1	9.4
	384	384	28.1	9.3
	[256;512]	384	27.3	8.8
D	256	256	27.0	8.8
	384	384	26.8	8.7
	[256;512]	384	25.6	8.1
E	256	256	27.3	9.0
	384	384	26.9	8.7
	[256;512]	384	**25.5**	**8.0**

AlexNet의 LRN사용 시 성능 좋아지지 않음

Table 4: **ConvNet performance at multiple test scales.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
B	256	224,256,288	28.2	9.6
C	256	224,256,288	27.7	9.2
	384	352,384,416	27.8	9.2
	[256; 512]	256,384,512	26.3	8.2
D	256	224,256,288	26.6	8.6
	384	352,384,416	26.5	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**
E	256	224,256,288	26.9	8.7
	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**