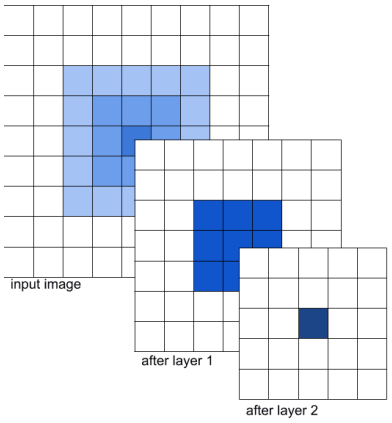
VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

대규모 이미지 인식 분야에서 합성곱 신경망의 깊이가 정확도에 미치는 영향에 대해 기술한 논문

탄생 배경

기존에 존재하였던 AlexNet은 ReLU, drop out, augmentation 등의 기법을 통하여 딥러닝의 기반다져왔고, VGGNet은 적절한 필터의 크기나 레이어의 수에 초점을 두어 딥러닝 분야를 발전시킴

특이사항으로는 3*3 크기의 커널 만을 사용하여 모델을 구성: receptive field



-특성 맵의 특정 영역이 input data에서 얼마만큼의 영역으로 도출됐는지를 나타내는 지표

5*5의 데이터에서 3*3크기의 커널을 2차례 사용. 마지막 한 개의 특성이 input image에서 수용하는 영역 receptive field

- 3*3크기의 커널을 2번 사용한 receptive field와 5*5 크기의 커널을 1번 사용한 receptive field가 동일하게 되지만, 파라미터 수의 차이 발생

Cf. 128채널의 데이터에

3*3커널을 2번 사용하여 특성 맵 구성 시

- (3*3*128*128+128+3*3*128*128), 약 29만 5천 개

5*5커널을 1번 사용하여 특성 맵 구성 시

- (5*5*128*128), 약 41만 개

➔ VGGNet(3*3)을 활용 시 더 적은 파라미터로 더 넓은 receptive field 구축 가능

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					

구성

레이어의 수에 의거, 각각 16(19)로 구성된 VGG16(19)로 나뉨

Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

자료에서는 D열이 VGG16, E열이 VGG19의 아키텍처

VGG16 – 13개의 합성곱 레이어와 3개의 완전 연결 레이어로 구성

VGG19 – 16개의 합성곱 레이어와 3개의 완전 연결 레이어로 구성

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**

해당 아키텍처에 사용된 최대 풀링은 stride와 kernel 크기를 같게 하는 non-overrapping 풀링을 활용 – stride 값이 커널 크기보다 작게하여 overrapping(중복)이 발생하는 AlexNet과는 대조적

결론

VGGNet은 상대적으로 적은 수의 파라미터로 더 넓은 receptive field를 갖는 효율적인 아키텍처를 제시

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**

- 이후 1*1의 커널을 가진 합성곱 레이어를 활용해 파라미터 수를 줄인 인셉션 네트워크(GooLeNet)이 등장

Table 7: **Comparison with the state of the art in ILSVRC classification**. Our method is denoted as "VGG". Only the results obtained without outside training data are reported.

Method	top-1 val. error (%)	top-5 val. error (%)	top-5 test error (%)
VGG (2 nets, multi-crop & dense eval.)	**23.7**	**6.8**	**6.8**
VGG (1 net, multi-crop & dense eval.)	24.4	7.1	7.0
VGG (ILSVRC submission, 7 nets, dense eval.)	24.7	7.5	7.3
GoogLeNet (Szegedy et al., 2014) (1 net)	-	7.9	
GoogLeNet (Szegedy et al., 2014) (7 nets)	-	**6.7**	
MSRA (He et al., 2014) (11 nets)	-	-	8.1
MSRA (He et al., 2014) (1 net)	27.9	9.1	9.1
Clarifai (Russakovsky et al., 2014) (multiple nets)	-	-	11.7
Clarifai (Russakovsky et al., 2014) (1 net)	-	-	12.5
Zeiler & Fergus (Zeiler & Fergus, 2013) (6 nets)	36.0	14.7	14.8
Zeiler & Fergus (Zeiler & Fergus, 2013) (1 net)	37.5	16.0	16.1
OverFeat (Sermanet et al., 2014) (7 nets)	34.0	13.2	13.6
OverFeat (Sermanet et al., 2014) (1 net)	35.7	14.2	-
Krizhevsky et al. (Krizhevsky et al., 2012) (5 nets)	38.1	16.4	16.4
Krizhevsky et al. (Krizhevsky et al., 2012) (1 net)	40.7	18.2	-