Very Deep Convolutional Network for Large-Scale Image Recognition

팀 명: 은혜로운 민트나라
발표자: 박혜민

1. Introduction

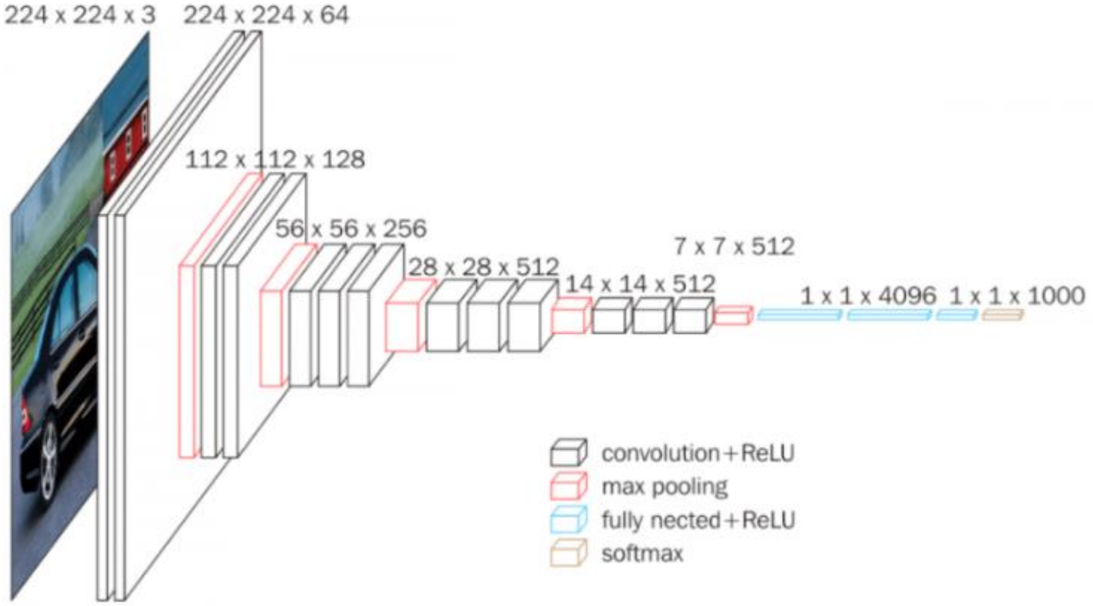
- 다른 parameter을 같게 하고, convolutional layer의 depth를 늘려가며 ConvNet architecture 를 모델링함
- (3 x 3)의 매우 작은 필터를 이용하여 깊은 모델을 만듦

- ILSVRC2014의 classification과 localisation 분야에서 좋은 성적을 거둠
- 다양한 dataset에 적용 가능

2. ConvNet Configurations

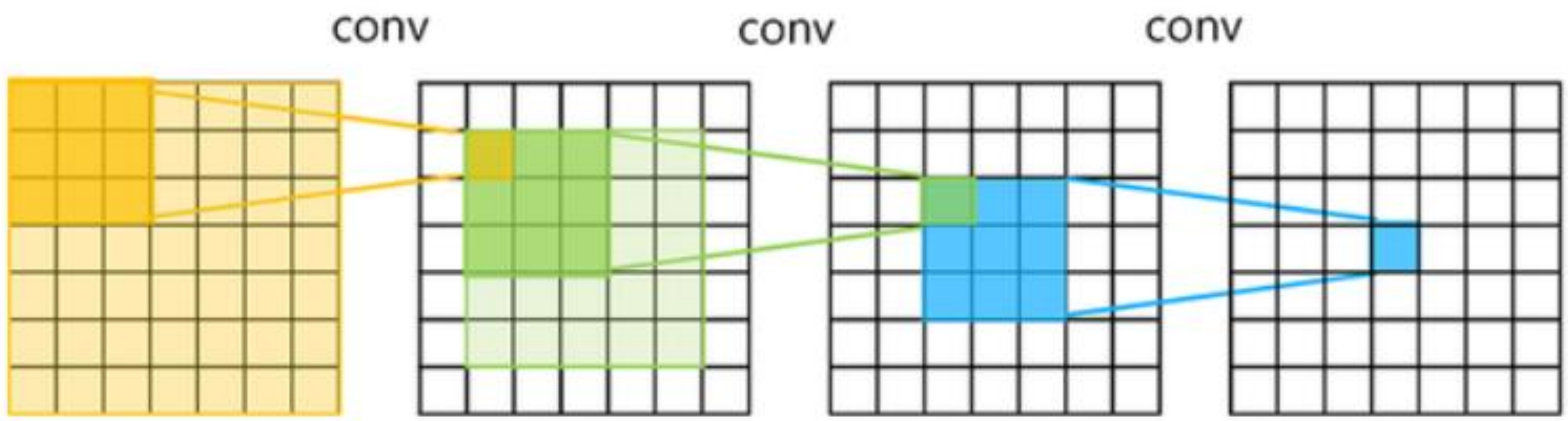
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	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					

- Depth 만 다른 netA~netE
- 224 x 224 RGB 이미지



- Normalisation을 하지 않음
- Parameter의 수 크지 않음.

2. ConvNet Configurations



- 3개의 (3 x 3) conv layer의 receptive field
  = 1개의 (7 x 7) conv layer의 receptive field
- 결정함수의 non-linearity 증가
- parameter 수 감소: $3(3^2C^2) = 27C^2 < 49C^2 = 7^2C^2$

3.1. Training

- back-propagation에 기반한 mini-batch gradient descent를 사용하여 다항 logistic regression을 최적화
- Batch size 256, Momentum 0.9
- Weight decay(L2), Drop out
- Learning rate 는 10^{-2}에서 시작하여, validation set의 정확도 향상이 둔화될 때마다 10^{-1}을 곱함.
- 74 epochs
- Random initialization을 통해 netA를 먼저 훈련시켜, 다른 net들에 대해 처음 4개의 conv layer와 모든 FC layer를 A의 훈련 결과로 초기화함.

3.1. Training

- S: 원본 비율을 유지하며 re-scale한 training 이미지의 짧은 변
- Re-scale한 이미지에서 224 x 224 크기의 image를 crop
- Single-scale training
  ① S = 256
  ② S = 384
- Multi-scale training (scale jittering)
  ③ S ∈ $[S_{min}, S_{max}]$ $(S_{min} = 256, S_{max} = 512)$
- Initialization : S=256 → S = 384 → S∈[256,512]

3.2. Testing

- Q: 원본 비율을 유지하며 re-scale한 test 이미지의 짧은 변(S≠Q)

Dense		Multi-crop
fully-convolutional net (3FC layer를 conv layer 로 변환)	Layer	변환 없음
whole(uncropped) image	Image Crop	50 crops per scale
Image의 이웃한 부분	padding	Zeros

- Dense 와 Multi-crop은 상호보완적
- 연산이 간단한 Dense방식으로 모델 검증 후, 성능 좋은 모델에 대해 Multi-crop 방식도 적용

4.0. Dataset

- ILSVRC-2012 dataset
  - 1000 classes
  - training 1.3M, validation 50K, testing 100K
- Measurements
  - top-1 error
  - top-5 error

4.1. Single Scale Evaluation

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**

- Single-scale training: Q=S
- Multi-scale training:
  Q=0.5($S_{min}+S_{max}$)

- LRN이 도움 안됨
- ConvNet의 depth가 깊을수록 에러 감소
- Multi-scale training이 에러 감소

4.2. Multi Scale Evaluation

- Single-scale training: Q={S-32, S, S+32}
- Multi-scale training: Q={S_{min} , 0.5($S_{min}+S_{max}$), S_{max}}
- Multi scale evaluation 이 더 좋은 성능을 보임

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**

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**

4.3. Multi Crop Evaluation

- Multi-crop 방식이 Dense방식보다 에러가 약간 작음
- 위의 두 방식을 함께 이용할 때, 에러가 가장 작음.

Table 5: **ConvNet evaluation techniques comparison.** In all experiments the training scale S was sampled from $[256; 512]$, and three test scales Q were considered: $\{256, 384, 512\}$.

ConvNet config. (Table 1)	Evaluation method	top-1 val. error (%)	top-5 val. error (%)
D	dense	24.8	7.5
	multi-crop	24.6	7.5
	multi-crop & dense	**24.4**	**7.2**
E	dense	24.8	7.5
	multi-crop	24.6	7.4
	multi-crop & dense	**24.4**	**7.1**

4.4. Convnet Fusion

- 기존의 single model 방식과 달리, 여러 Model들의 결과를 평균 내어, softmax로 분류함.

Table 6: **Multiple ConvNet fusion results.**

Combined ConvNet models	Error		
	top-1 val	top-5 val	top-5 test
ILSVRC submission			
(D/256/224,256,288), (D/384/352,384,416), (D/[256;512]/256,384,512) (C/256/224,256,288), (C/384/352,384,416) (E/256/224,256,288), (E/384/352,384,416)	24.7	7.5	7.3
post-submission			
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), dense eval.	24.0	7.1	7.0
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop	23.9	7.2	-
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop & dense eval.	**23.7**	**6.8**	**6.8**

4.5. Comparison with the State of the Art

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	-

5. Conclusions

- Large-scale image 에 대한 very deep convolutional network에 대해 연구함
- Depth 가 분류의 정확도 향상에 큰 역할을 함