

# (간단) VGG, AlexNet

## AlexNet (2012)

- 8 learned(with weight) layers, 5 convolutinal and three FC layers, multinomal logistic regression with softmax
- **ReLU Activation Function**: 몇배는 빠른 학습, tanh 함수와 동등한 성능 → 큰 네트워크를 학습할 수 있게 함 && saturating (graient descent, exploding과 같은 현상)을 막아줌
- ReLU Activation Function을 통해 몇배는 빠른 학습, tanh 함수와 동등한 성능 → 큰 네트워크를 학습할 수 있게 함, saturating (graient descent, exploding과 같은 현상)을 막아줌
- GPU 병렬 학습: 특정 레이어에서 GPU 간의 통신, e.g. layer 4에서 layer 3의 커널 맵을 인풋으로 넣을 때 같은 GPU 에 있는 커널 맵을 사용할 수 있음
- LRN(Local Response Normalisation)

### 구조

첫번째 conv layer에서는 11x11x3의 96개의 필터 사용, stride=4

두번째conv layer에서는 5\*5\*48의 256개 필터 사용

3,4,5번째 conv layer는 어떠한 풀링이나 normalization layer가 들어가 있지 않음

단, 두번째 레이어에서 세번째 레이어로 연결될 때 3x3x256의 필터는 384개 사용하는 데 애네가 nomalized, pooling 되어 있음, z

네번째는 3x3x192의 384개의 커널, 5번째는 256개의 3x3x192

마지막 FC는 4092 뉴런을 가지고 있고 최종적으로 1000개의 logit 값 산출

### Discussion

하나의 conv layer를 뺐을 때 2%가량 top1 성능이 줄었음, *layer의 depth가 정말 중요하게 작용했다*

그렇다고 무작정 많이 쌓아서 큰 네트워크를 만들면 컴퓨터에서날 파워가 부족했는데, ReLU나 GPU 병렬 사용을 통해 이를 극복했다.

그래서 여기서 더 많이 레이어 쌓아서 더 큰 네트워크를 만들고 싶지만, 훈련 시간을 생각하면 무리가 있다. 비디오 시퀀스와 같은 곳에서도 이 deep conv net이 쓰일 수 있게 하는 것이 궁극적 목표

## VGG (2015.04)

224\*224\*3 이미지를 받아, 3x3이나 1x1 필터를 사용, stride는 1로 고정., maxpolling은 (2,2)의 stride 2로 적용, AlexNet 처럼 ReLU를 사용하만

LRN : 계산이나 리소스 사용 대비 성능 잘 안오르더라

커널의 사이즈로 5x5, 7x7을 적은 레이어로 해서 해보기도 했는데, 확실히 non-linear 함수를 여러번 쓰면서 하는게 좀 더 분별력 있게 만들어주더라, 그리고 파라미터의 개수도 적어져서 베토 더라 그래서 3x3에 이정도 레이어 쌓는게 제일 좋더라

가장 레이어를 깊게 많이 쌓은 네트워크의 성능이 가장 좋더라

rs are added (the added layers are shown in bold). The as “conv<receptive field size>-<number of channels>”. or brevity.

et Configuration			
	C	D	E
ght	16 weight	16 weight	19 weight
s	layers	layers	layers
× 224 RGB image)			
54	conv3-64	conv3-64	conv3-64
64	conv3-64	conv3-64	conv3-64
maxpool			
28	conv3-128	conv3-128	conv3-128
28	conv3-128	conv3-128	conv3-128
maxpool			
56	conv3-256	conv3-256	conv3-256
56	conv3-256	conv3-256	conv3-256
	<b>conv1-256</b>	<b>conv3-256</b>	conv3-256
			<b>conv3-256</b>
maxpool			
12	conv3-512	conv3-512	conv3-512
12	conv3-512	conv3-512	conv3-512
	<b>conv1-512</b>	<b>conv3-512</b>	conv3-512
			<b>conv3-512</b>
maxpool			
12	conv3-512	conv3-512	conv3-512
12	conv3-512	conv3-512	conv3-512
	<b>conv1-512</b>	<b>conv3-512</b>	conv3-512
			<b>conv3-512</b>
maxpool			
FC-4096			
FC-4096			
FC-1000			

E는 Conv layer<필터 크기 3x3>-<채널 개수 즉, 필터개수> 총 파라미터는 144m

총 19개의 학습 가능한 파라미터를 가진 (weight layers) 레이어들로 구성, FC레이어까지 포함, maxpooling은 learnable 파라미터를 갖고 있지 않음.

	[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	<b>24.8</b>	<b>7.5</b>
	256	224,256,288	26.9	8.7
E	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	<b>24.8</b>	<b>7.5</b>