Case Studies

Classic networks

LeNet -> AlexNet -> VGGNet

deeplearning.ai

LeNet - 5

32×32×1   conv 5×5 $s = 1$ → 28×28×6

요즘에 max가 대세지만 당시엔 avg pooling을 많이 썼다고 함.

avg pool   f = 2   s = 2 → 14×14×6

당시엔 padding 거의 사용안했기 때문에 valid padding이라서 conv layer 적용할 때마다 크기가 shrink된다.

conv 5×5 $s = 1$ → 10×10×16

avg pool   f = 2   s = 2 → 5×5×16   400

FC → 120   FC → 84

논문에서는 softmax 안쓰고 딴걸 썼음.

$\hat{y}$

Softmax 10
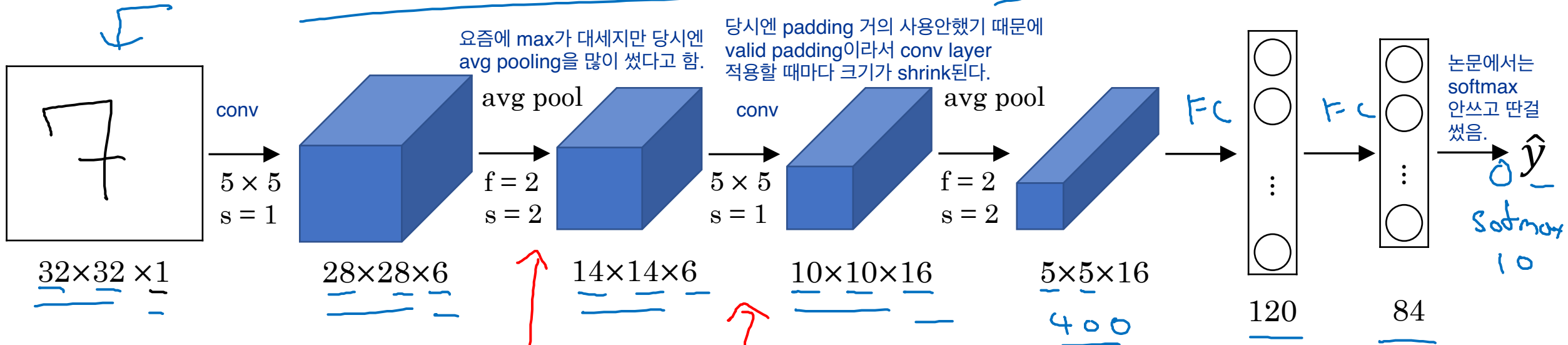
60K parameters.

파라미터 개수. (요즘엔 10~100 million도 흔함)

non-linearity after pooling

$n_H \times n_a \times n_c$ $f \times f \times n_c$

보통 이런 구성으로 모든 필터가 채널의 모든 것을 보도록 연산이 되지만 당시엔 computation이 비쌌기 때문에 지금은 거의 사용하지 않는 아주 복잡한 꼼수를 썼다.

$n_H, n_w \downarrow$ $n_c \uparrow$

네트워크 깊이 들어갈 수록 경향

Conv pool Conv pool fc fc output

NN에서 위와 같은 패턴. (요즘도 많이 쓰이는...)

Advanced: Sigmoid/tanh ReLU
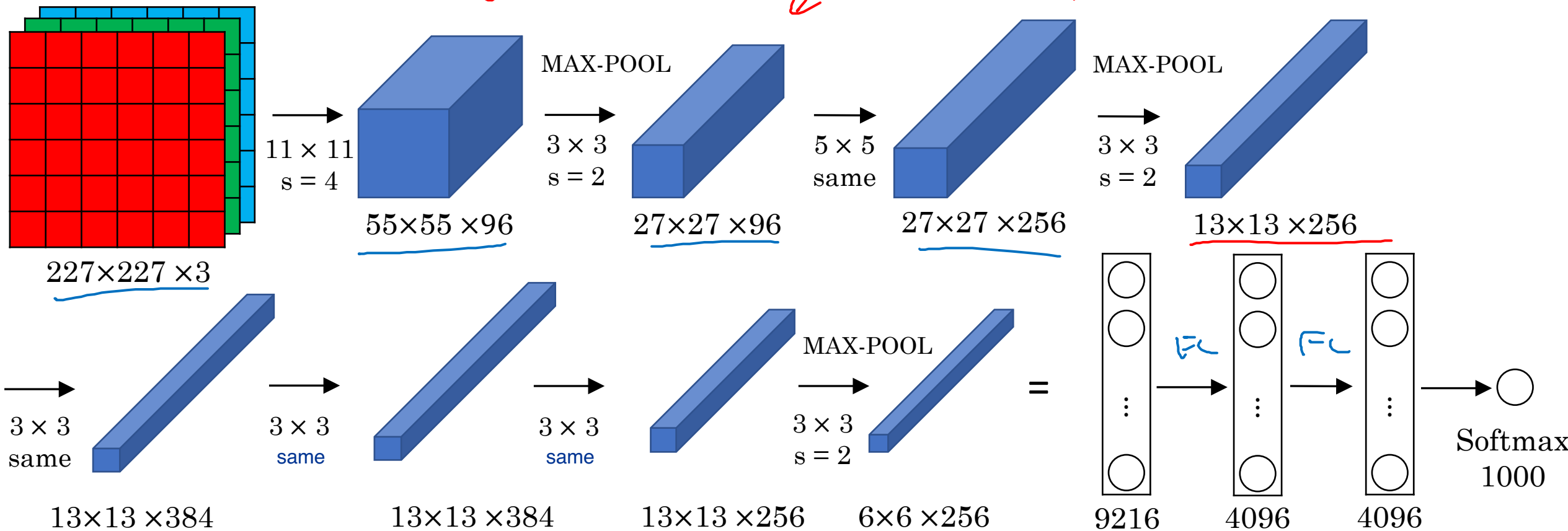
당시엔 sigmoid랑 tanh를 쓰고 ReLU는 안썼다.

혹시 이 논문 읽을거면 섹션 2, 3에 집중해서 봐라

Ⅱ Ⅲ

[LeCun et al., 1998. Gradient-based learning applied to document recognition]

Andrew Ng

AlexNet

227×227 ×3

11 × 11
s = 4

55×55 ×96

3 × 3
s = 2

27×27 ×96

MAX-POOL

5 × 5
same

27×27 ×256

3 × 3
s = 2

MAX-POOL

13×13 ×256

3 × 3
same

13×13 ×384

3 × 3
same

13×13 ×384

3 × 3
same

13×13 ×256

3 × 3
s = 2

MAX-POOL

6×6 ×256

9216

=

FC

FC

9216   4096   4096

Softmax
1000

LeNet과
두드러진
차이점들

- Similary to LeNet, but much bigger.
  Similarity to LeNet, but much bigger
- ReLU
- Multiple GPUs. 2 gpu로 학습하려고 복잡한 방법 썼음
- Local Response Normalization (LRN)
  LBN 레이어를 사용했는데 이거 요즘엔 많이 안쓴다.

~ 60 M parameter
(cf. LeNet 60 k)

13

13   256

이런 입력 볼륨에서 한 포지션에서 전체
채널에 걸친 값을 normalization한다.

특정 뉴런이 큰 값을 가지지 않게 하려는
의도였는데 후에 연구자들이 이게 별로 효과
없다는 것을 밝혀내서 요즘엔 안쓰임

[Krizhevsky et al., 2012. ImageNet classification with deep convolutional neural networks]

Andrew Ng

VGG - 16

weights를 가진
레이어가 16개라서

$VGG-19$

CONV = 3×3 filter, s = 1, same   MAX-POOL = 2×2 , s = 2

$224×224×3$ $224×224×64$ $224×224×64$

224×224 ×3

→ [CONV 64] ×2
64 filter
→ 224×224×64 → POOL → 112×112 ×64 → [CONV 128] ×2 → 112×112 ×128 → POOL → 56×56 ×128

→ [CONV 256] ×3 → 56×56 ×256 → POOL → 28×28 ×256 → [CONV 512] ×3 → 28×28 ×512 → POOL → 14×14×512

→ [CONV 512] ×3 → 14×14 ×512 → POOL → 7×7×512 → FC 4096 → FC 4096 → Softmax 1000
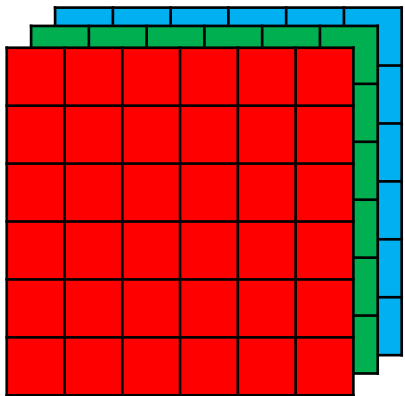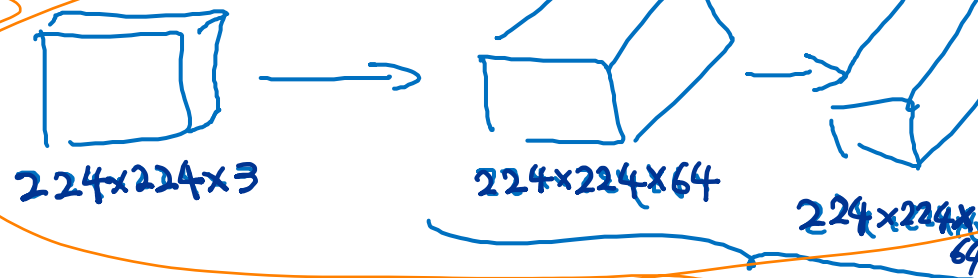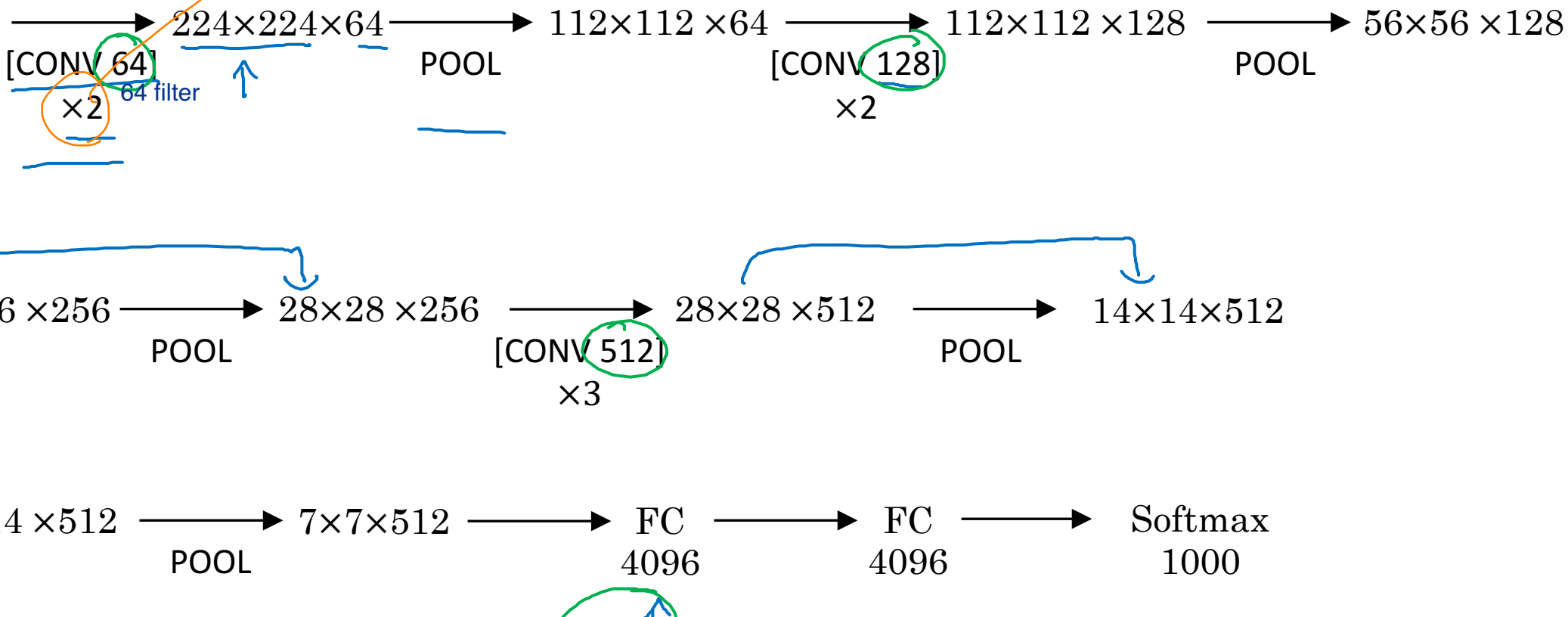
$n_H, n_W ↓$ $n_C ↑$

$\sim 138M$

위에 얘들을 보면 심플하고 규칙성이 있는게
꽤 uniformity를 가지고 있고 이런점이
연구자들에게 매력적으로 작용했다.

138M 파라미터를 가진 (최신
네트워크와 비교해도)꽤 큰
네트워크인 점이 단점

[Simonyan & Zisserman 2015. Very deep convolutional networks for large-scale image recognition]

Andrew Ng