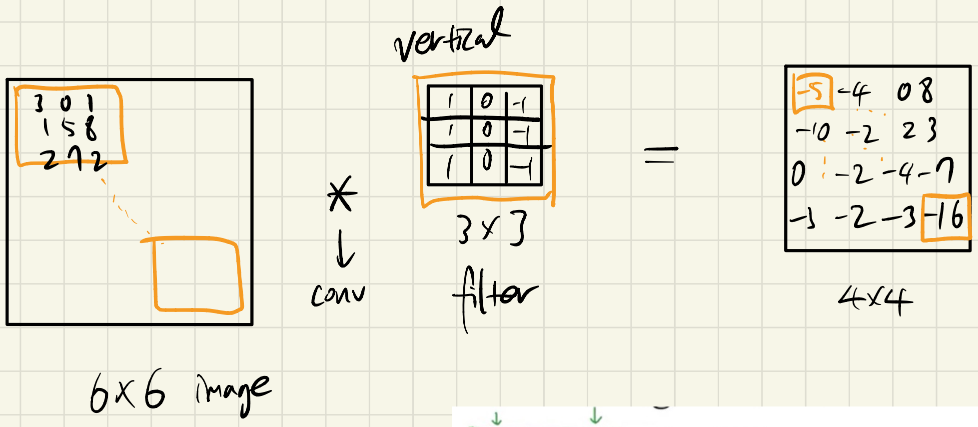
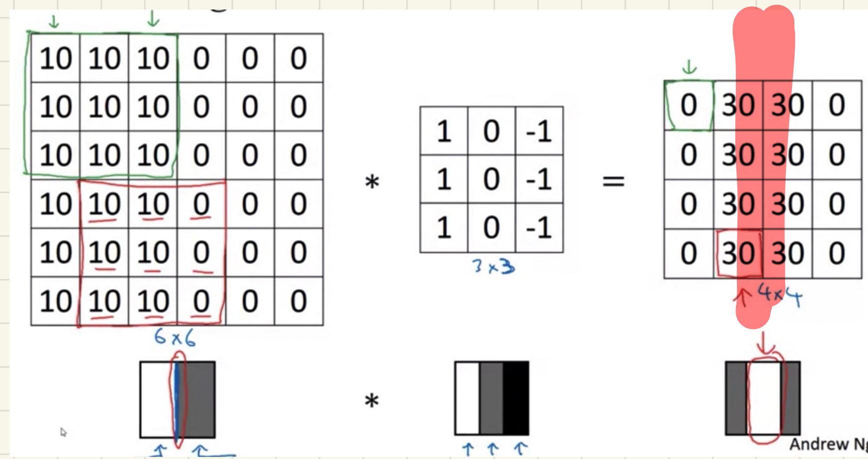
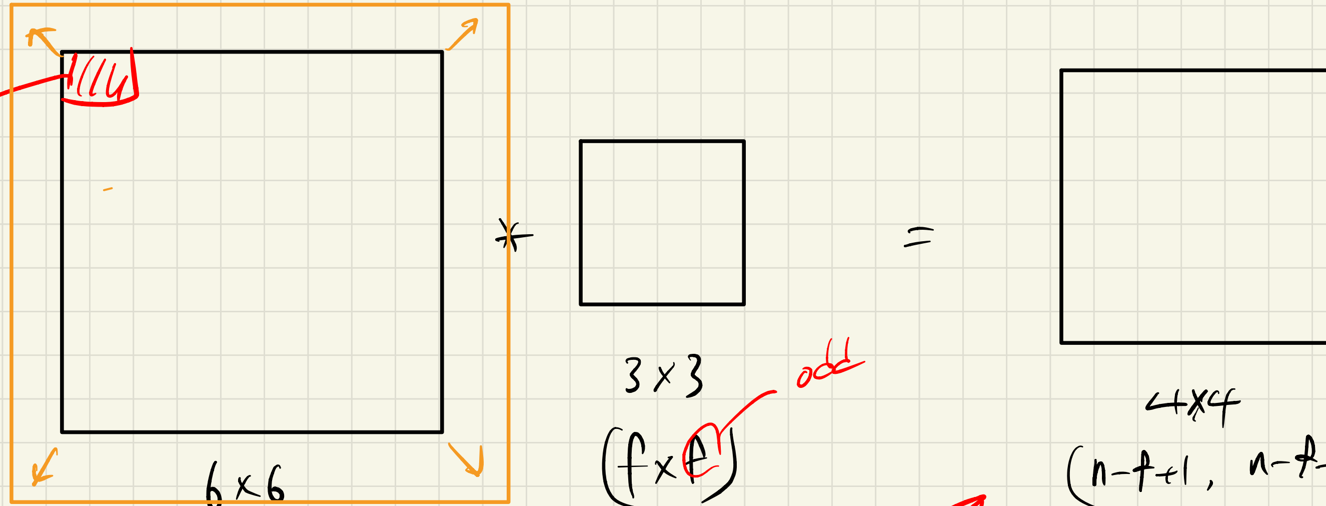
[edge detection]



Vertical line



[padding]



② 넓히기

6×6 $(n \times n)$ $*$ 3×3 $(f \times f)$ odd $=$ 4×4

valid conv

$(n-f+1,\ n-f+1)$

$(6-3+1,\ 6-3+1)$

① small

same conv 8×8 (padding) $p=1$

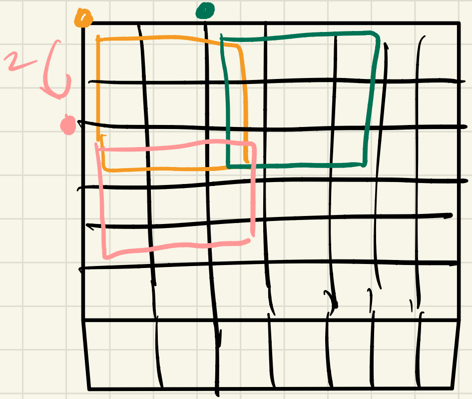
$\underset{\sim\sim}{6 \times 6}$ 유지!

$(n+2p-f+1,\ n+2p-f+1)$

$(6+2-3+1,\ 6+2-3+1)$

[Stride]

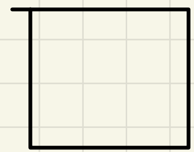
2 (stride)



7×7

$n \times n$

$*$

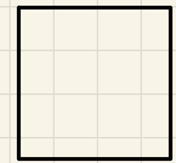


3×3

$f \times f$

Stride $= 2$

$=$



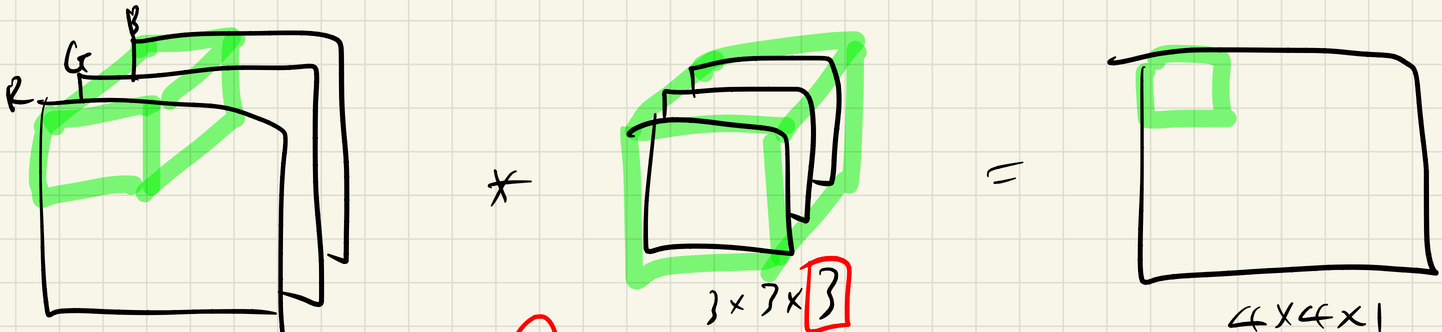
3×3

$$\left(\frac{n+2p-f}{s}+1 \quad \times \quad \frac{n+2p-f}{s}\right)$$

$$\left(\frac{7+2\cdot 0-3}{2}+1 \quad \times \quad \text{same}\right)$$

$$(3 \quad \times \quad 3)$$

[Conv in RGB images]



$6 \times 6 \times \boxed{3}$

$h \times w \times c$ (channel)

=

If you have 10 filters that are 3 x 3 x 3 in one layer of a neural network, how many parameters does that layer have?

$3 \times 3 \times 3$

$= 27$

$+ b \quad = 28$

$= 280$ params

$f^{[l]}$ = filter size of layer l

$p^{[l]}$ = padding

$s^{[l]}$ = stride

Input : $n_H^{[l-1]} \times n_W^{[l-1]} \times n_c^{[l-1]}$

Output : $n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$

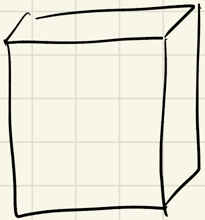
$$n_H^{[l]} = \frac{n_H^{[l-1]} + 2p^{[l]} - f^{[l]}}{s^{[l]}} + 1$$

Each filter : $f^{[l]} \times f^{[l]} \times n_c^{[l-1]}$

Activations : $a^{[l]}$
$\rightarrow n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$

Weights : $f^{[l]} \times f^{[l]} \times n_c^{[l-1]} \times n_c^{[l]}$

bias : $n_c^{[l]} - (1,1,1,n_c^{[l]})$



$39 \times 39 \times 3$

$n_H^{[0]} = n_W^{[0]} = 39$

$n_C^{[0]} = 3$

$f^{[1]} = 3$

$s^{[1]} = 1$

$p^{[1]} = 0$

10 filters

$$\frac{n + 2p - f}{s} + 1$$

$$= \frac{39 + 0 - 3}{1} + 1 = 37$$

$37 \times 37 \times 10$

$n_H^{[1]} = n_W^{[1]} = 37$

$n_C^{[1]} = 10$

$f^{[2]} = 5$

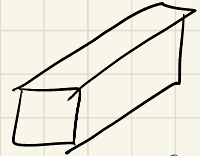
$s^{[2]} = 2$

$p^{[2]} = 0$

20 filters

$$\frac{n+2p-f}{s} + 1$$

$$= \frac{37 + 0 - 5}{2} + 1 = 17$$



$17 \times 17 \times 20$

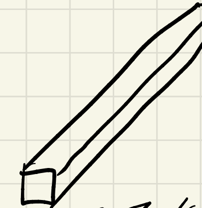
$n_H^{[2]} = n_W^{[2]} = 17$

$n_C^{[2]} = 20$

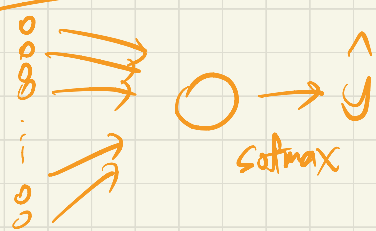
$f^{[3]} = 5$

$s^{[3]} = 2$

40 filters

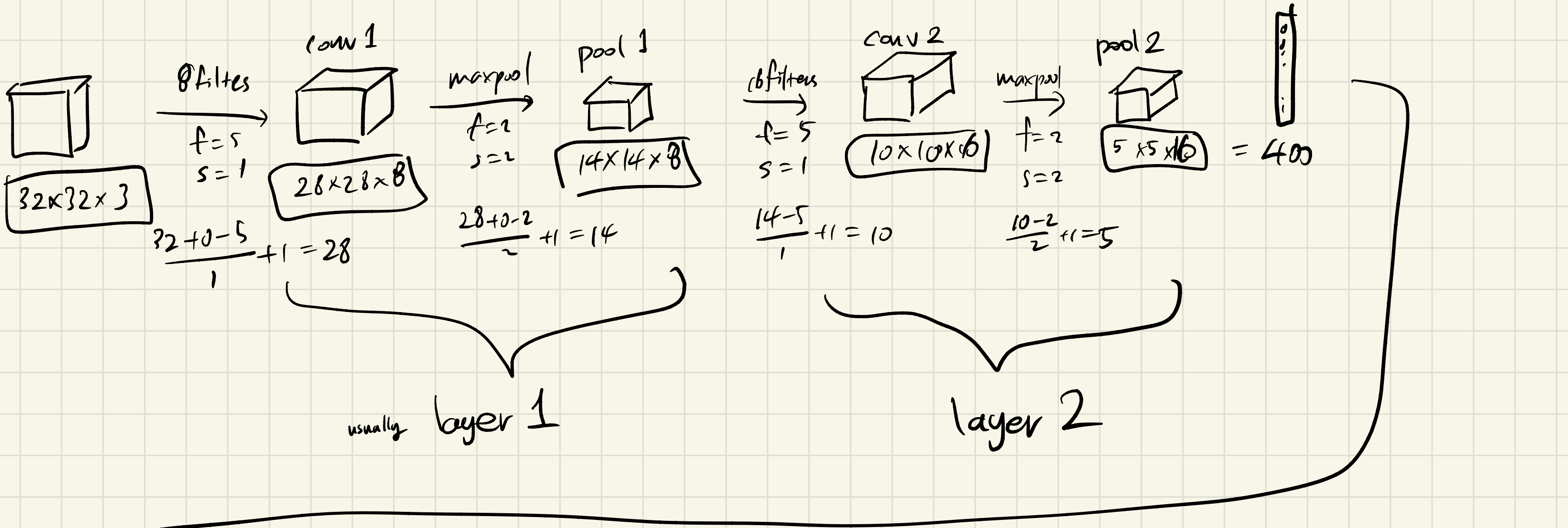


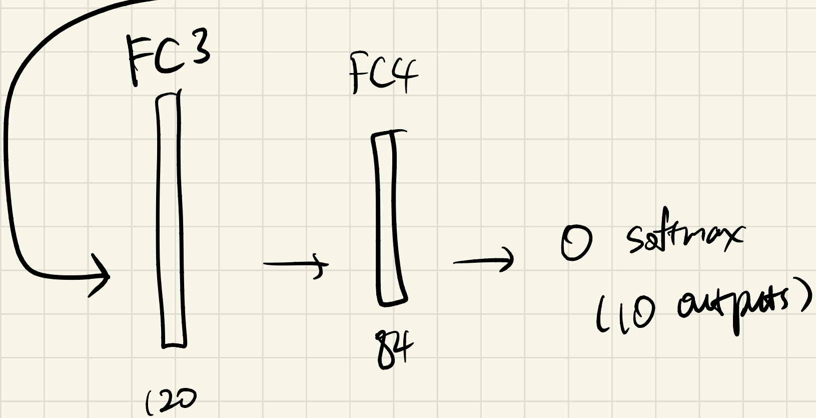
$7 \times 7 \times 40$



1960

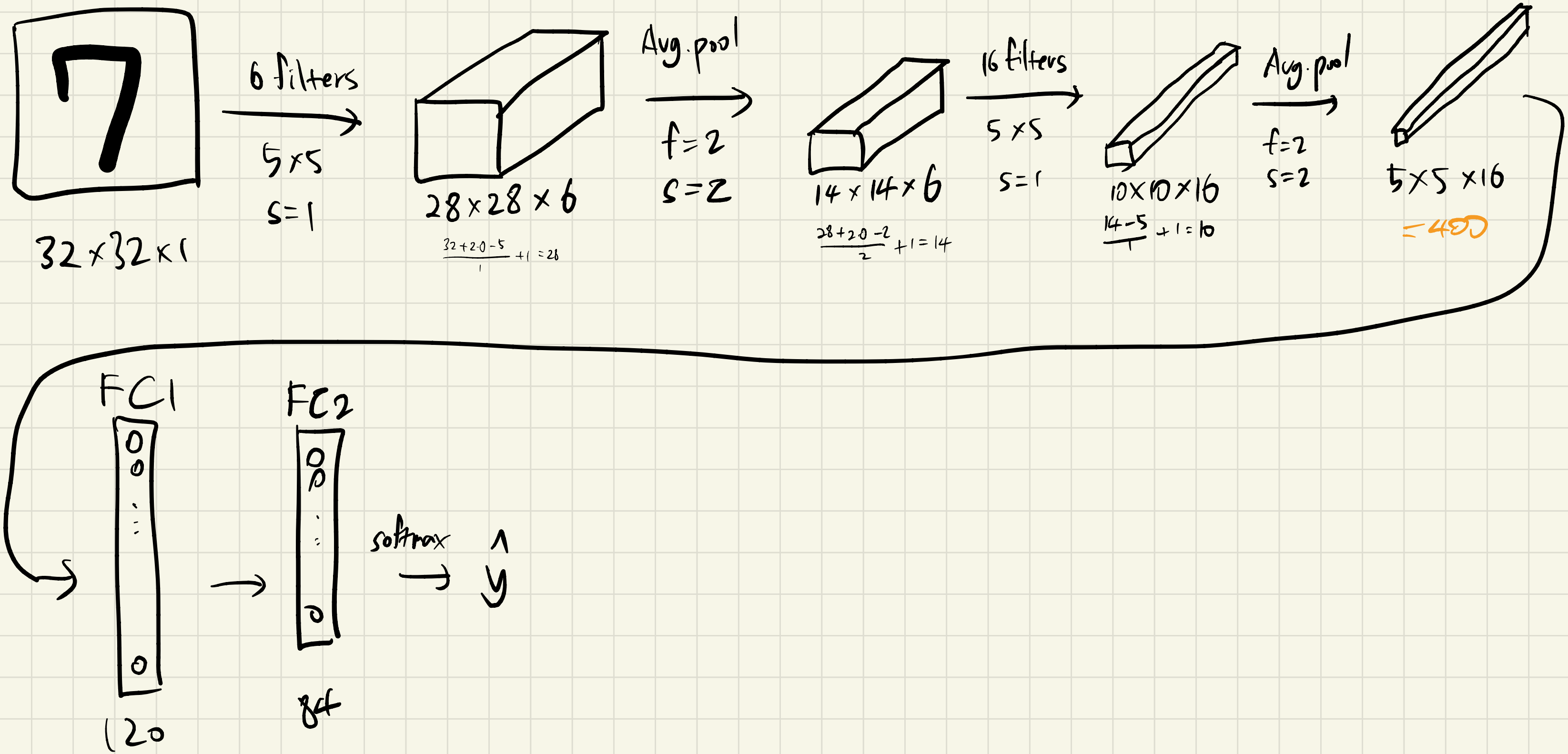
softmax





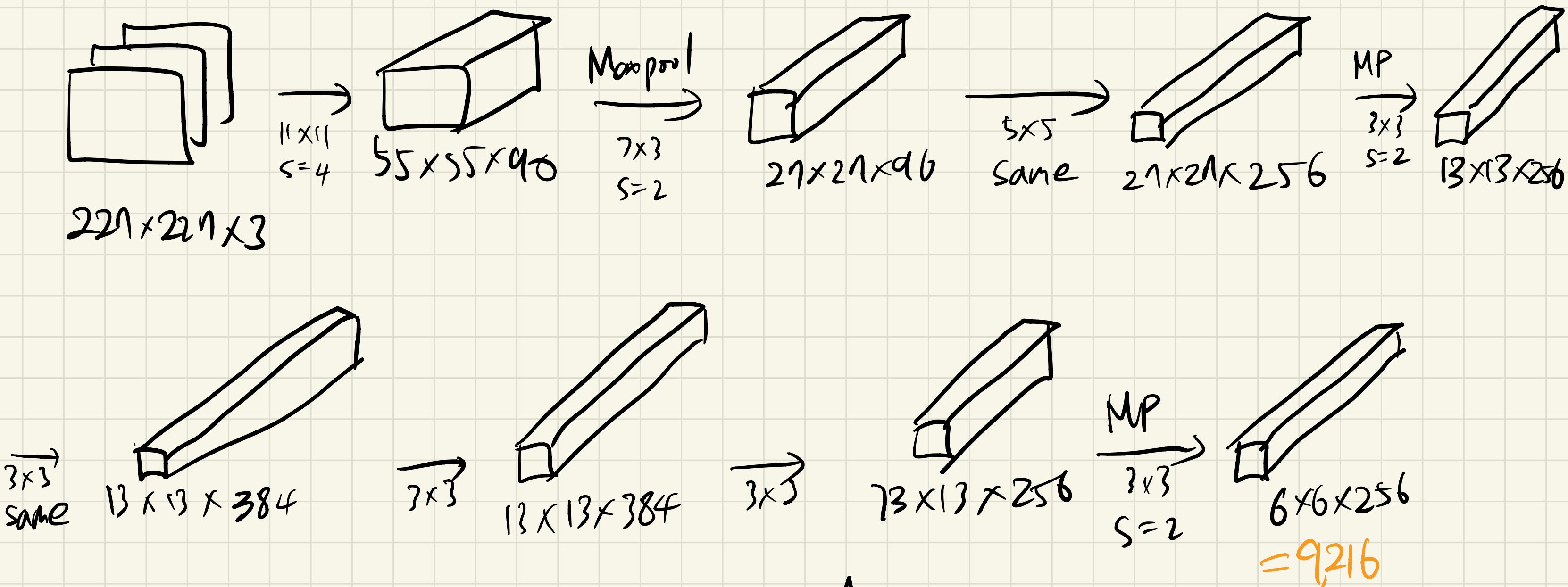
	Activation shape	Activation Size	# parameters
Input:	(32,32,3)	— 3,072 $a^{[0]}$	0
CONV1 (f=5, s=1)	(28,28,8)	6,272	(5*5*3+1)*8 ~~208~~ 608
POOL1	(14,14,8)	1,568	0
CONV2 (f=5, s=1)	(10,10,16)	1,600	(5*5*3+1)*16 ~~416~~ 3216
POOL2	(5,5,16)	400	0
FC3	(120,1)	120	(400×120)+120 ~~48,001~~ 48120
FC4	(84,1)	84	(120×84)+84 ~~10,081~~ 10164
Softmax	(10,1)	10	(84×10)+10 ~~841~~ 85

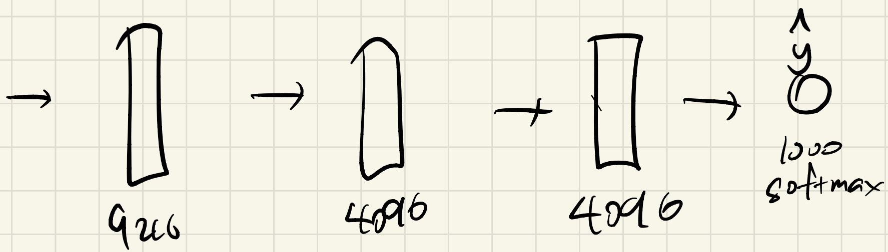
[LeNet 5]



7
32×32×1
6 filters
5×5
s=1
28×28×6
(32+2·0−5)/1 +1 = 28
Avg.pool
f=2
s=2
14×14×6
(28+2·0−2)/2 +1 = 14
16 filters
5×5
s=1
10×10×16
(14−5)/1 +1 = 10
Avg.pool
f=2
s=2
5×5×16
=400
FC1
120
FC2
84
softmax
ŷ

[AlexNet]

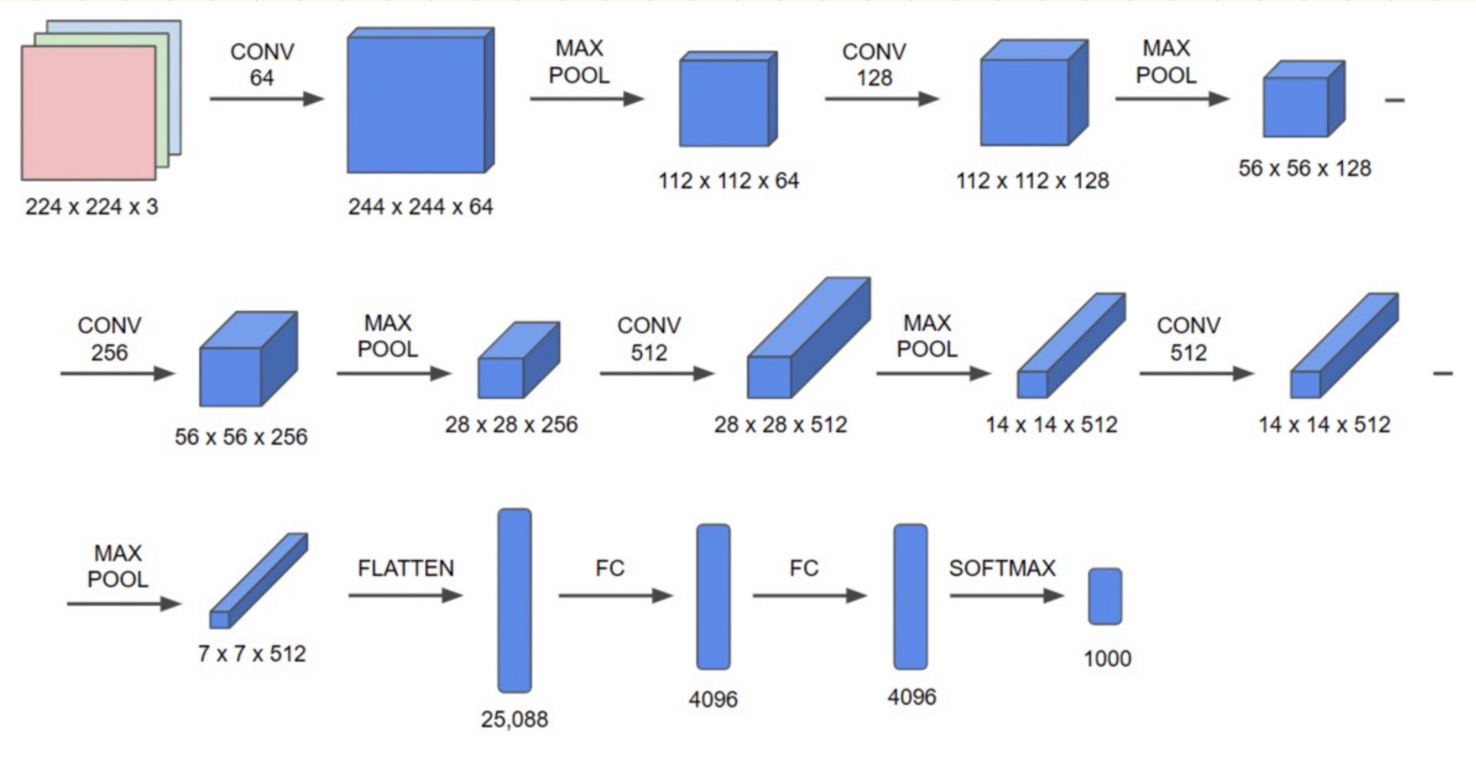




- ReLU
- Bigger
- Multi-GPUs

[VGG 16] < VGG19

CONV = 3×3, S=1, SAME / MP = 2×2, S=2



[ResNet]

skip connection

$a^{[l]} \rightarrow$ Linear $\rightarrow$ Activation $\rightarrow a^{[l+1]} \rightarrow$ Linear $\rightarrow$ activation $\rightarrow a^{[l+2]}$

$z^{[l+1]} = W^{[l+1]} a^{[l]} + b^{[l+1]}$

$a^{[l+1]} = g(z^{[l+1]})$

$z^{[l+2]} = W^{[l+2]} a^{[l]} + b^{[l+2]}$

$a^{[l+2]} = g(z^{[l+2]} + a^{[l]})$

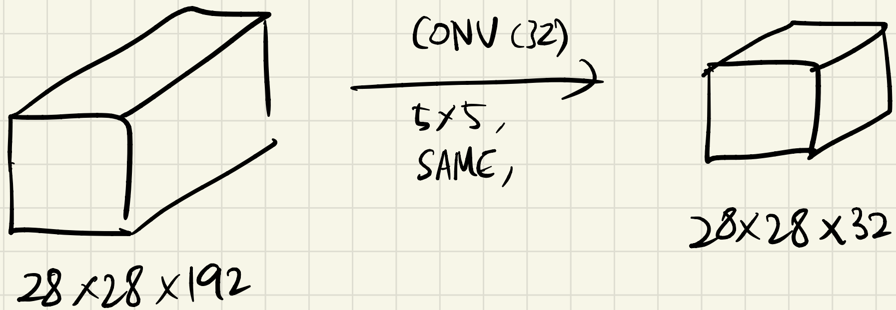
$g(\underbrace{W^{[l+2]}}_{0} \cdot a^{[l+1]} + \underbrace{b^{[l+2]}}_{0} + a^{[l]})$

If

$= g(a^{[l]}) = a^{[l]}$

No Vanishing!

[Using 1d Conv]

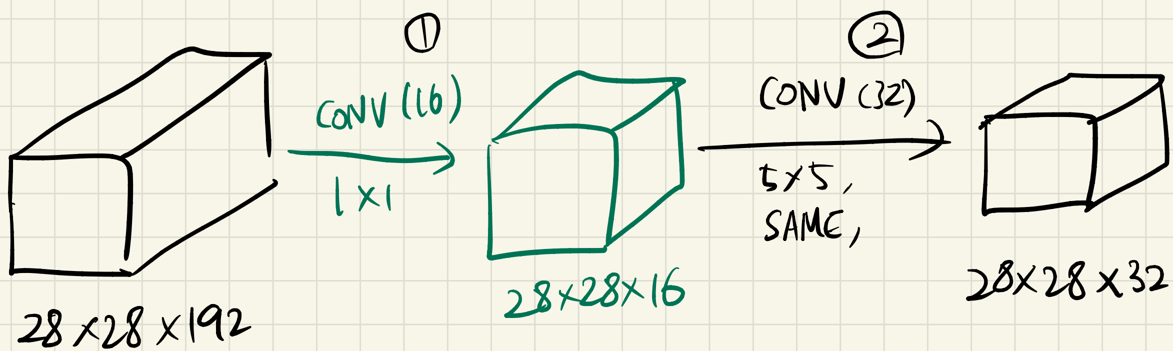


<computation>

$28 \times 28 \times 32$

$\times\ 5 \times 5 \times 192$

$\approx 120M$

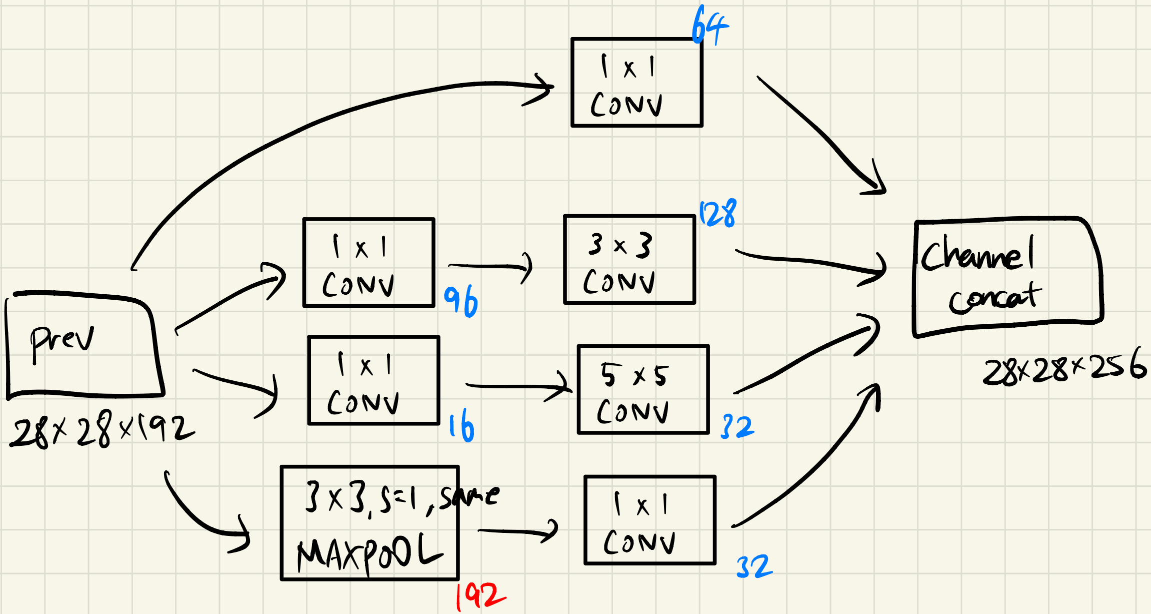


1/10

① $(28 \times 28 \times 16) \times (1 \times 1 \times 192) \simeq 2.4M$

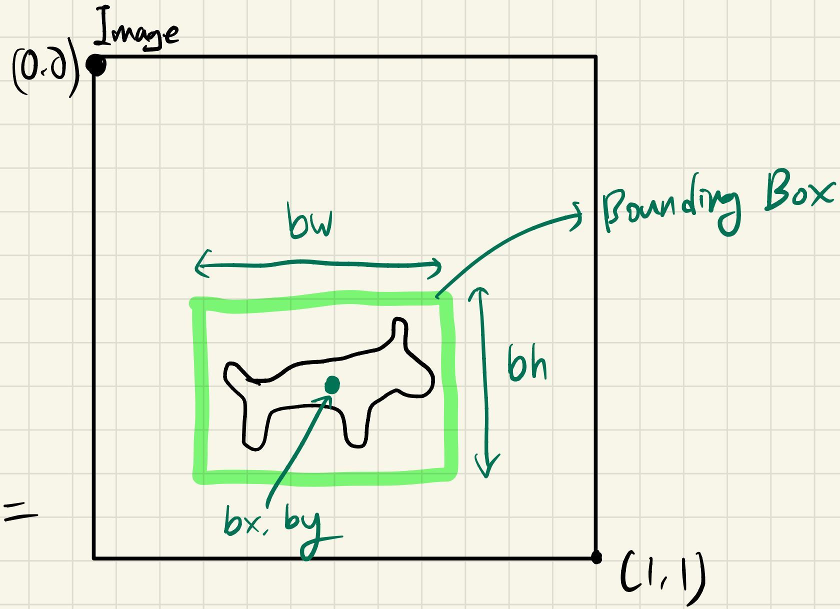
② $(28 \times 28 \times 32) \times (5 \times 5 \times 16) \simeq 10.0M$

[Inception Module]



[Object Detection]

$X =$

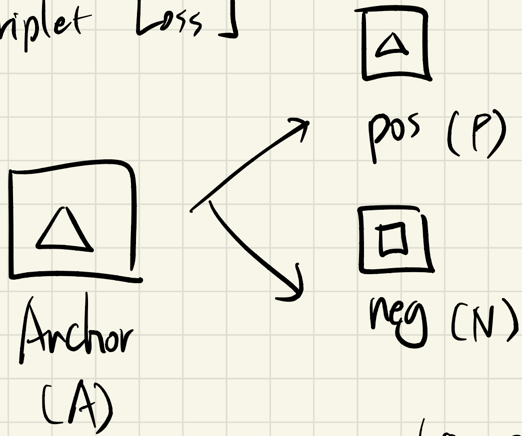


$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$$

$p_c \rightarrow$ object?

c_1, c_2, c_3, c_4: classification

[Triplet Loss]



$$\|f(A) - f(P)\|^2 \; : \; d(A,P)$$

$$\leq$$

$$\|f(A) - f(N)\|^2 \; : \; d(A,N)$$

$$d(A,P) - d(A,N) + \alpha \leq 0$$

if 0.5 0.51 ↓ must be 0.7↑ //

α → Margin 0.2

$$\mathcal{L}(A,P,N) = \max\left(\boxed{\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha}, \; 0\right)$$

≤ 0

[Neural Style transfer cost function]

$$J(G) = \alpha \underline{J_{content}(C, G)} + \beta \underline{J_{style}(S, G)}$$

$$\frac{1}{2} \| a^{[l](C)} - a^{[l](G)} \|^2$$

$$\| G^{[l](S)} - G^{[l](G)} \|_F^2$$

$$= \text{const} \cdot \sum_k \sum_{k'} \left(G_{kk'}^{[l](S)} - G_{kk'}^{[l](G)} \right)^2$$

Style Matrix

$$\boxed{G_{kk'}^{[l](S)} = \sum_i^{n_H^{[l]}} \sum_j^{n_W^{[l]}} a_{ijk}^{[l](S)} \cdot a_{ijk'}^{[l](S)} \qquad k = 1, \dots, n_c^{[l]}}$$

→ will be large if k & k' have correlation.