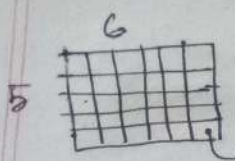
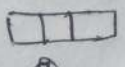


(single channel Image)

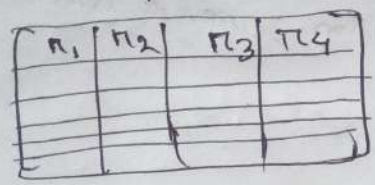
→ 0-255 (int)
0-1 (IR) [Normalized]



filter (1×3)

$I \otimes F = \begin{bmatrix} & & 4 \\ & & \\ & & 6 \end{bmatrix}$

$I \otimes F =$ $\quad r_1 = f_1 * i_1 + f_2 * i_2 + f_3 * i_3$; $r_2 = f_1 * i_2 + f_2 * i_3 + f_3 * i_4$
; $r_3 = f_1 * i_3 + f_2 * i_4 + f_5 * i_5$
; $r_4 = f_1 * i_4 + f_2 * i_5 + f_3 * i_6$

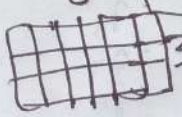
$I \otimes F =$ 

Stride → shifting step (1, 2)

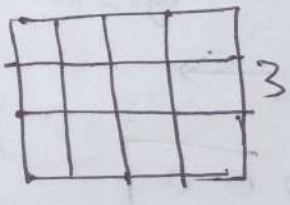
padding → increasing dimension (r, c)

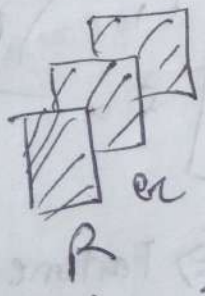
For filter  3×1

$$n/e = \frac{img.d - f.d}{stride} + 1$$

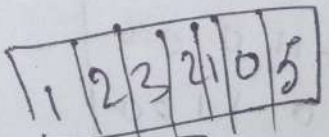
$I \otimes F =$ 

For filter  3×3

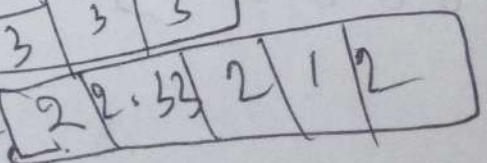
$I \otimes F =$ 

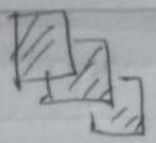


Tensor → (3, 5, 4)
↓ channel
↓ row
column

$I =$ 

$F =$ 

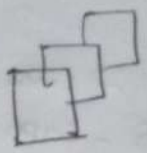
$R =$ 

(3,5,4) → Tenson 

(3*, x, y) → Filter

3* - fixed for considering Tenson

x, y → determined by owner

Tenson ⊗ Filter ⇒ Response 

n x filter = n x Response's

so, Response shape (# of filter, r, c)

↓ number of filter

(c, r1, c1) → Tenson

(c, r2, c2) → Filter

(ch, r3, c3) → Response

ch = Number of filter.

Conv - 1

48, 11, 11

Stride = 4

Output ⇒ 48, 55, 55

48, 55, 55

Maxpooling

Region(2) 3x3

Strid 2

$\frac{55-3}{2}+1 = 27$

ImageNet paper

AlexNet paper

Input layer
3, 227, 227

Conv-1

48, 11, 11 — 48, 11, 11

stride = 4

Output ⇒ 48, 55, 55 – GPU 1
48, 55, 55 – GPU 2

Max pooling
Region (z) 3×3
stride 2
$\frac{55-3}{2} + 1 = 27$

Output shape 48, 27, 27 – GPU 1
48, 27, 27 – GPU 2

Relu

Conv-2

128, 48, 5, 5 — 128, 48, 5, 5

stride = 1

Region (z) 5×5

$\frac{27-5}{1} + 1$

$= 23$ with padding

$\frac{(27+2P)-5}{1} + 1 = 27$

$P = 2$

Relu

$\frac{(27+2\times2-5)}{1} + 1 = 27$

with padding = 2

128, 27, 27 — 128, 27, 27

Conv-3

Max pooling 2=3, stride=2

output $\frac{27-3}{2}+1=13$

Output shape 128,13,13 → 128,13,13 Relu

Conv-33 (Merging both)

192,256,3,3 192,256,3,3 [384 filters]

padding=1, stride=1

Output = 192,13,13 192,13,13

Conv-44

padding -1 | 192,192,3,3 → 192,192,3,3

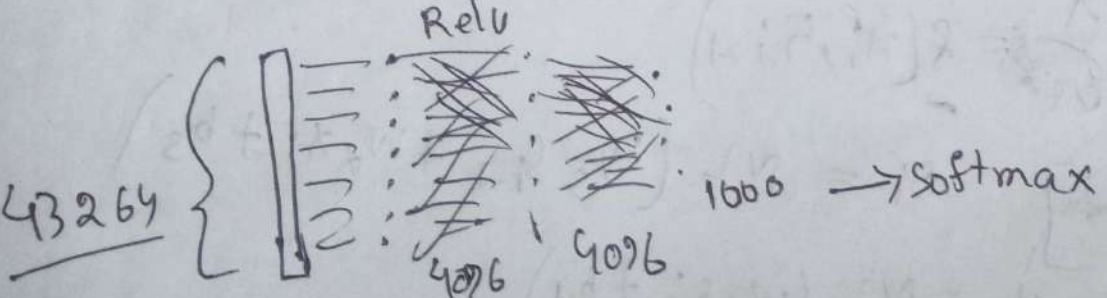
output 192,13,13 192,13,13 Relu

Conv-5

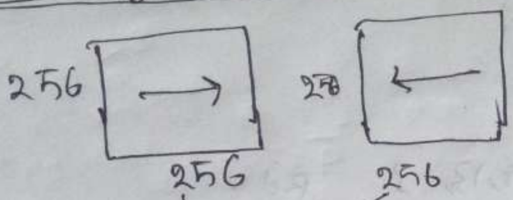
128,192,3,3 128,192,3,3

padding -1, stride -1 Relu

output: 128,13,13



Data Augmentation



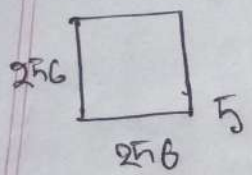
227 samples 227×227 (crossed out) ... $2 \times 27 \times 27$

⇓

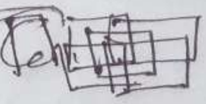
2×841 images

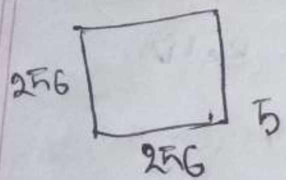
(Both) ⇓ ~~1682 im~~

1682 images



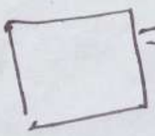
Test Time Augmentation (5 types)





$= 10$ 227×227

⇓ each image

$x_i \Rightarrow$  $\Rightarrow \vec{p}_i$

$$\frac{1}{10} \sum_{i=1}^{l} P_l$$

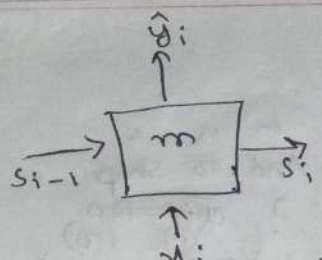
RNN (Recurrent Neural Network) [Background]

$$\hat{y}_i, s_i = \hat{f}(x_i, s_{i-1})$$

$$s_i = NL(W_s s_{i-1} + W_x x_i + b_s)$$

s_0 [Zero Vector]

$$y_i = NL(W_y s_i + b_y)$$



$s_i = NL_1(w_s \cdot s_{i-1} + w_x x_i + b_s)$

$\hat{y}_i = NL_2(w_y s_i + b_y)$

$s_0 \rightarrow$ is zero vector

$\hat{y}_{10}, \hat{y}_1, \hat{y}_2$

(x_1, x_2, x_3) s_0 is zero vector

(y_1, y_2, y_3)

$s_1 = NL_1(w_s s_0 + w_x x_1 + b_s)$

$\hat{y}_1 = NL_2(w_y s_1 + b_y)$

$s_2 = NL_1(w_s s_1 + w_x x_2 + b_s)$

$\hat{y}_2 = NL_1(w_y s_2 + b_y)$

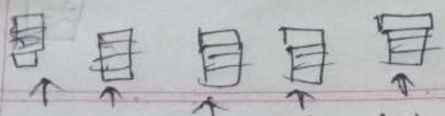
$s_3 = NL_3(w_s s_2 + w_x x_3 + b_y)$

$\hat{y}_3 = NL(w_y s_y + b_y)$

$$L = \frac{1}{3}\left(nll(y_1, \hat{y}_1) + nll(y_2, \hat{y}_2) + nll(y_3, \hat{y}_3)\right)$$

nll = negative log likely hood.

NER → Named Entity Recognition.



I a am going to Australia

Person – B / I
location – B / I
Institution – B / I
Other – B / I

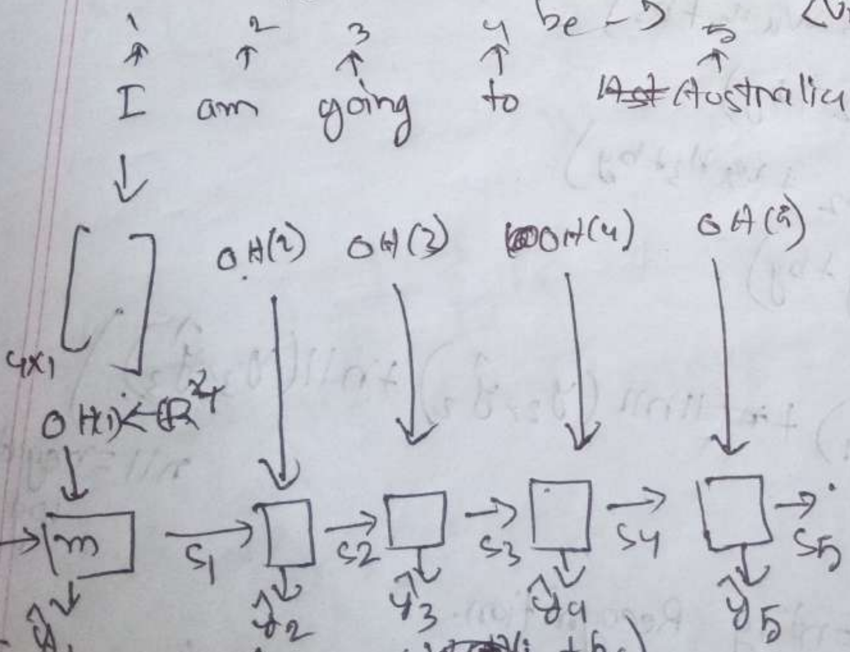
1	2	3	4	5	6
I	am	going	to	And to	Shop
P	O	O	O	?	O B
(B)	(B)	(B)	(B)	(B)	(B)

→ (Depends)

Meena Bazar
L L
(B) (I)

7	8	9	10	11	12
There	I	will	also	be	going
13	2,14		15	16	
to	Melbourne	and	Sydney		

vocabulary [I – 1, am – 2, going – 3, to – 4, Australia – 5, There – 6, will – 7, also – 8, be – 9, Melbourne – 10, and – 11, Sydney – 12, <Unk> – 13]



$s_i = NL_1(W_s s_{i-1} + W_u x_i + b_s)$

$\hat{y}_i = NL_2(W_y s_i + b_y)$

NL → Activation function