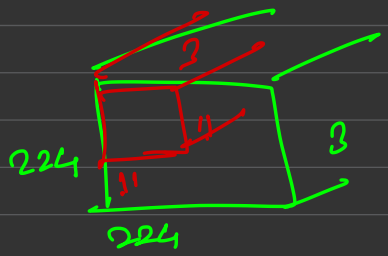
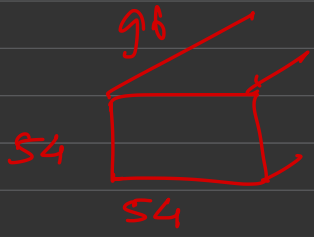
Alexnet Architecture.



$\frac{364+1}{}$

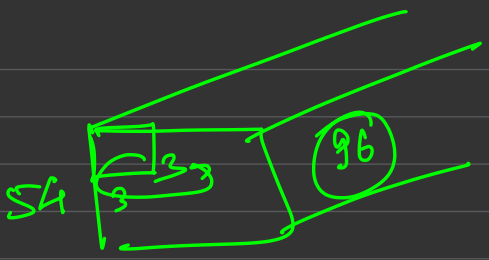
364×96



$\left(\frac{224-11}{4}+1\right)$

$\frac{213}{4}$ $53.25+1$

$54.25 \approx 54 \times 54$



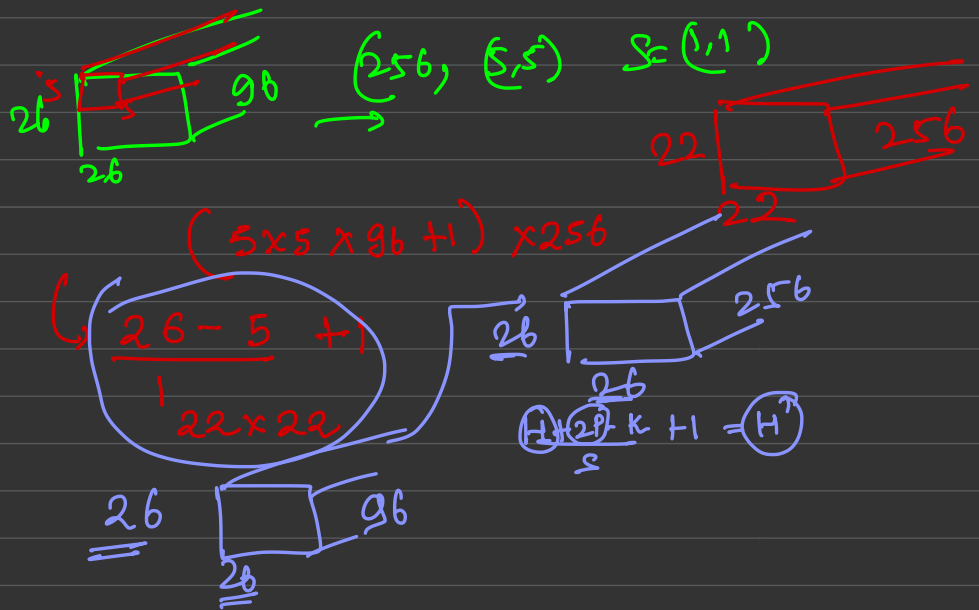
map(2D) → valid $\left(\frac{H-f}{s}+1\right)$

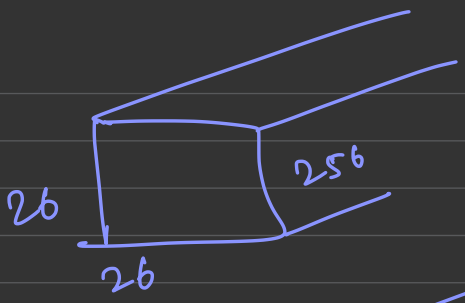
$\frac{54-3}{2}+1$

same $H = \frac{H+2P-f}{s}+1$

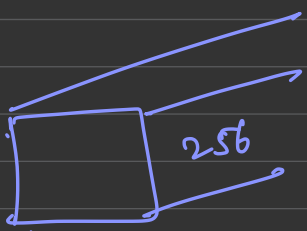
$\frac{51}{2}+1$

$= 26 \times 26 \times 96$

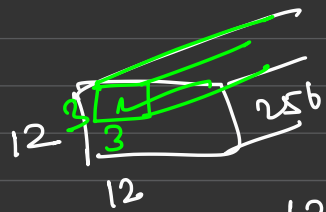




→ MP ((3,3) 5=2,2))



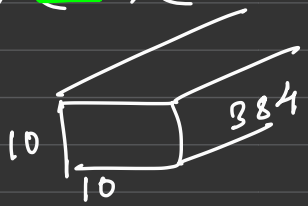
$\frac{26-3}{2}+1$

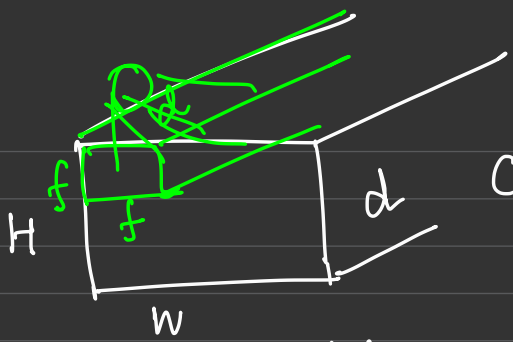


384, (3,3), (5=1,1)

$\frac{12-3}{1}+1$

(3×3×256 +1) × 384





Conv2D $\left(f \times f,\ s=(s,s),\ n\right)$ no. of filters

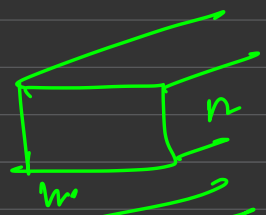
Total learnable parameters

$= (f \times f \times d + 1) \times n$

padding = 'valid' H' (output) $= \frac{H-f}{s} + 1$

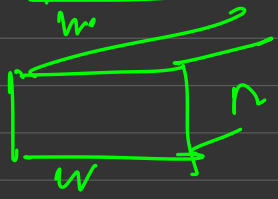
W' (output) $\frac{W-f}{s} + 1$

output shape $H' \times W' \times n$



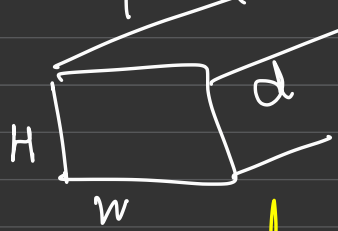
padding = 'same' $= (H \times W) \times n$

output shape



Max Pool (2D)

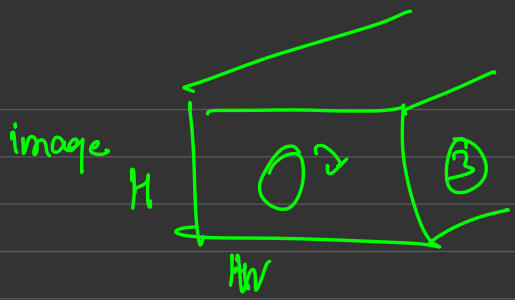
$(f \times f,\ s \times s)$



$H' = \frac{H-f}{s} + 1$

$W' = \left(\frac{W-f}{s} + 1\right)$

learnable parameters = 0



data / image augmentation