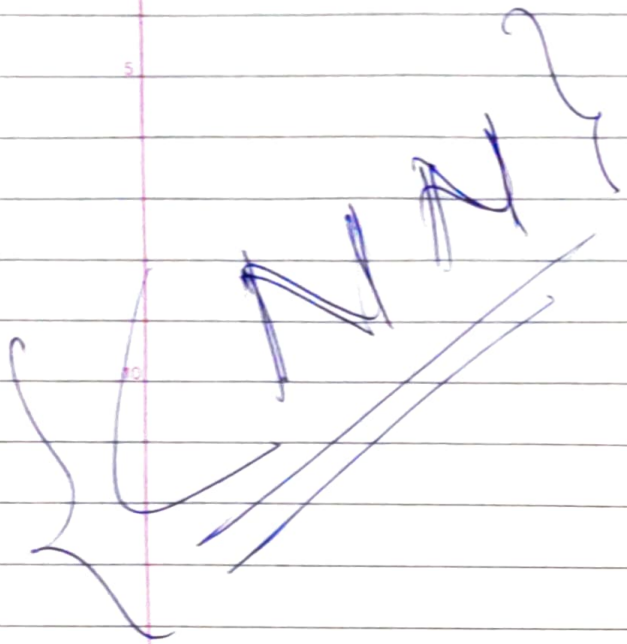
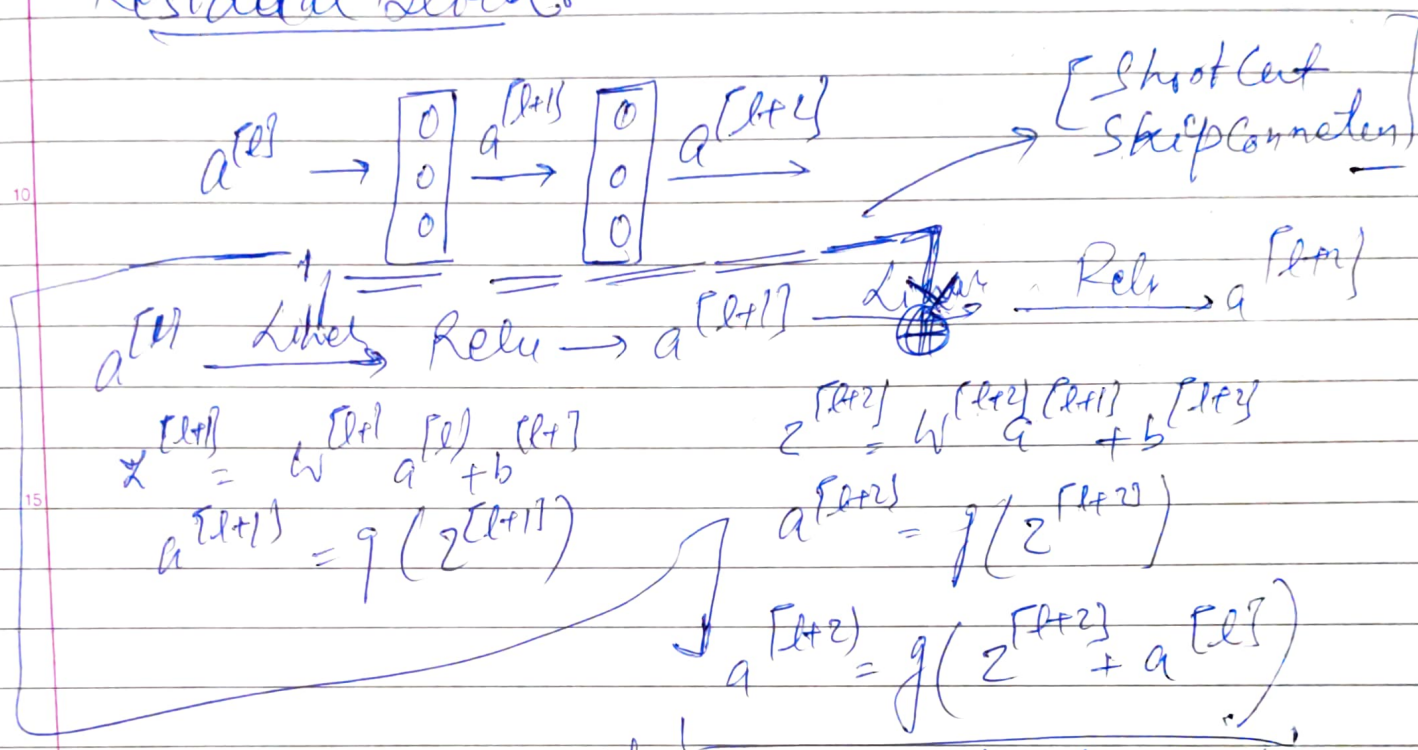


Residual Network & [ResNets]

⊗ LEARN Vanishing & Exploding Gradient.

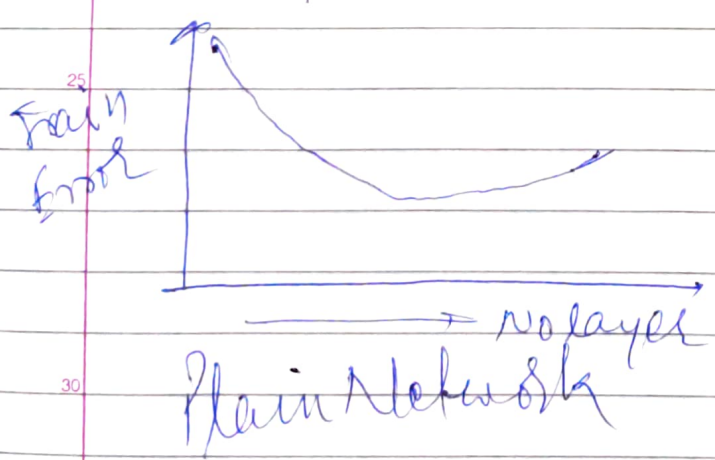
ResNet :- He et al. 2015 Deep residual Network.

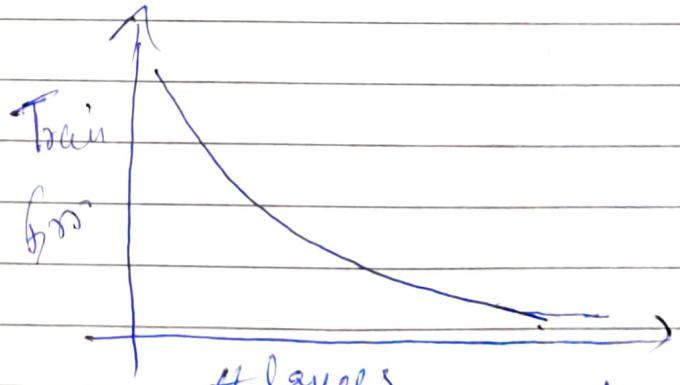
Residual Block :-



This is Residual Block.

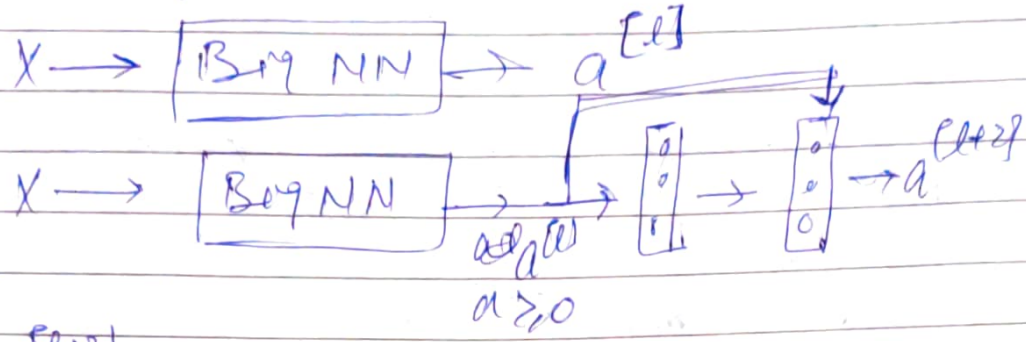
⊗ Using residual blocks allow to train much deeper network.





Help vanishing & exploding problem.

Why Resnet Work



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

$$= g[w^{[l+2]} a^{[l+1]} + b^{[l+2]} + a^{[l]}]$$

if there is L2 regularization this will shrink $w^{[l+2]}$ & may $b^{[l+2]}$ hence $w^{[l+2]}, b^{[l+2]} \approx 0$

$$\rightarrow g[0 + 0 + a^{[l]}]$$

$$\approx g(a^{[l]})$$ Residual block learn or turn into identity function

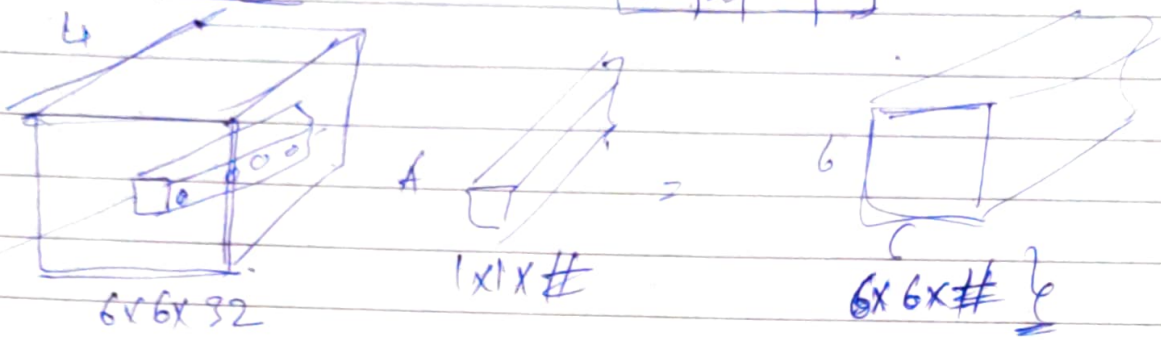
1x1 Convolution

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

* 2 =

2	4	6	8
10	12	14	16
18	20	22	24
26	28	30	32

Just a scale version.



→ We can think like 1 Neuron multiply each of the 32 Numbers & apply Relu.

* It is generally used to shrink the channels. ex.

(28x28x192) —Relu Conv (1x1)#32→ (28,28,32)