Lecture - 6

22/02/2026

Parameter Calculation in Neural Networks.

Input → [layer] → o/p feature map.

dimension calculation etc.

Pytorch → tensors.

Batch

Dimension : Batch x channels x Height x width

Data representation = B X C X H X W

Convolution layer's hyperparameters :-

#filters : K # size of filter = F #Stride = S.

zero padding = P.

H, W

Formula

width → i/p, filter size, padding

$$W_2 = \frac{W_1 - F + 2P}{S} + 1$$

W_2 → width - o/p

S → stride

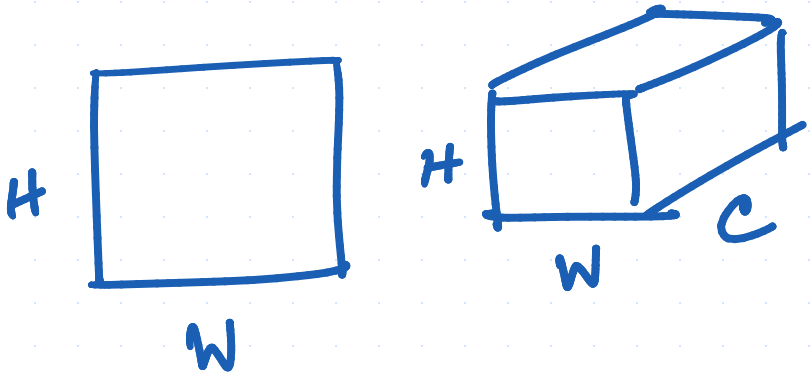
Height

$$H_2 = \frac{H_1 - F + 2P}{S} + 1$$

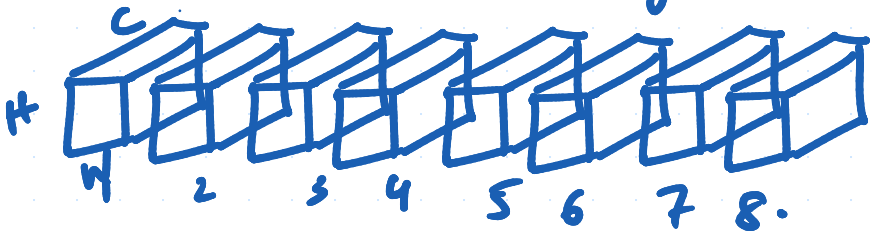
H_2 → new height

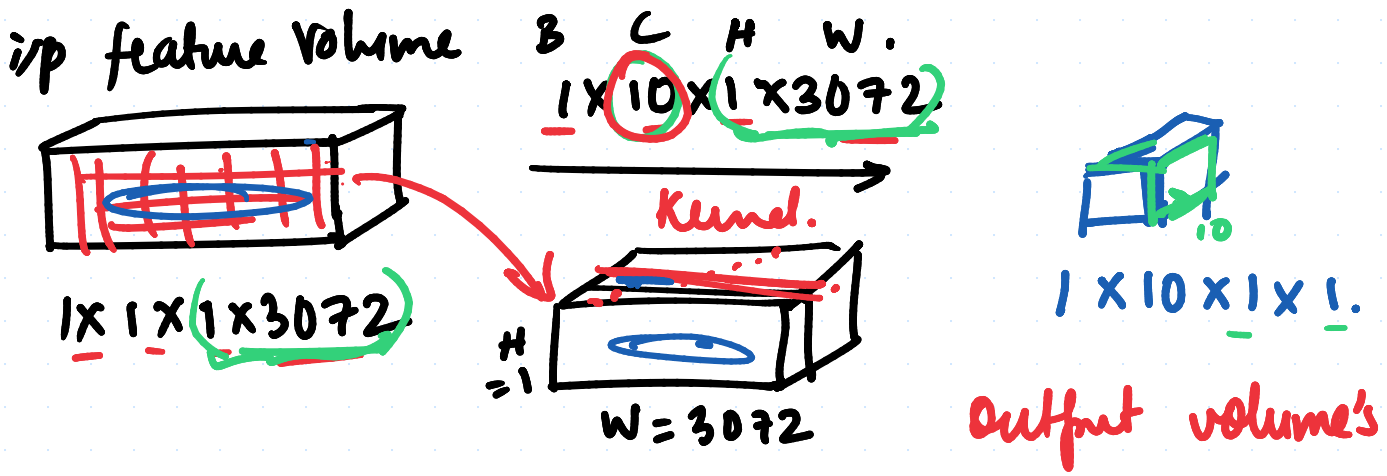
filter size, padding

stride



Batch size = 8.





Non-trivial example.

In general we have feature (i/p) size >> kernel.

Kernel >> Feature volume.

How to calculate the parameter of the given neural computation / layer.

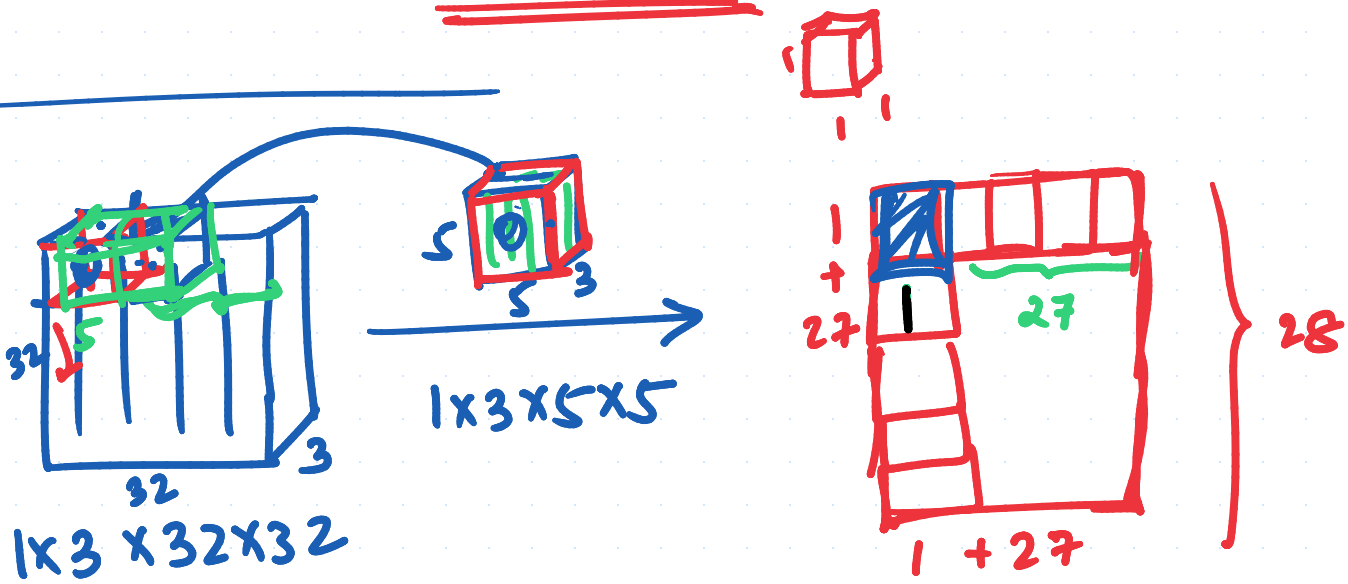
Kernel stores the parameter → which is intern the weights and biases of the layer.

$$1 \times 10 \times 1 \times 3072 = 30720.$$

For bias + 10.

from each channel.

Total params = 30730



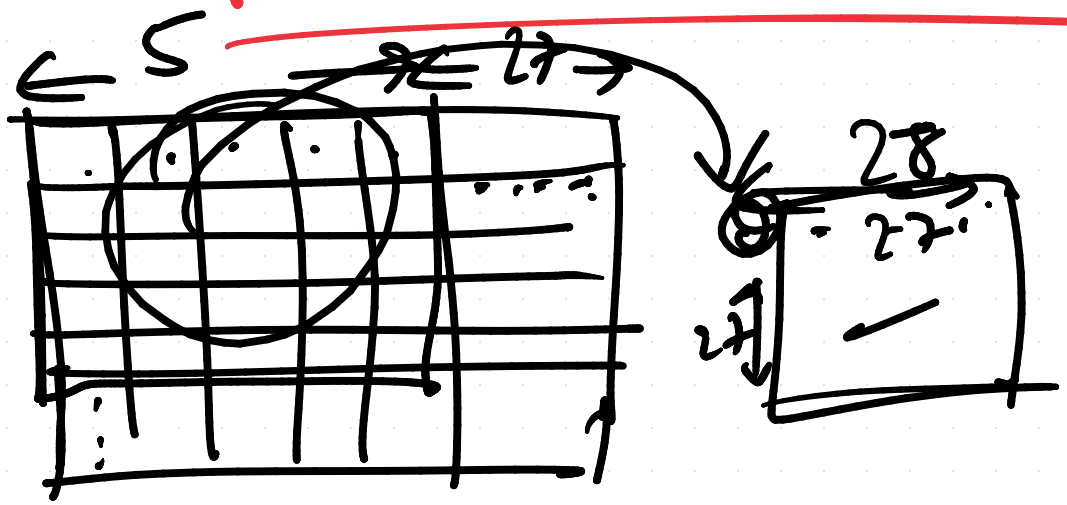
BxCxHxW

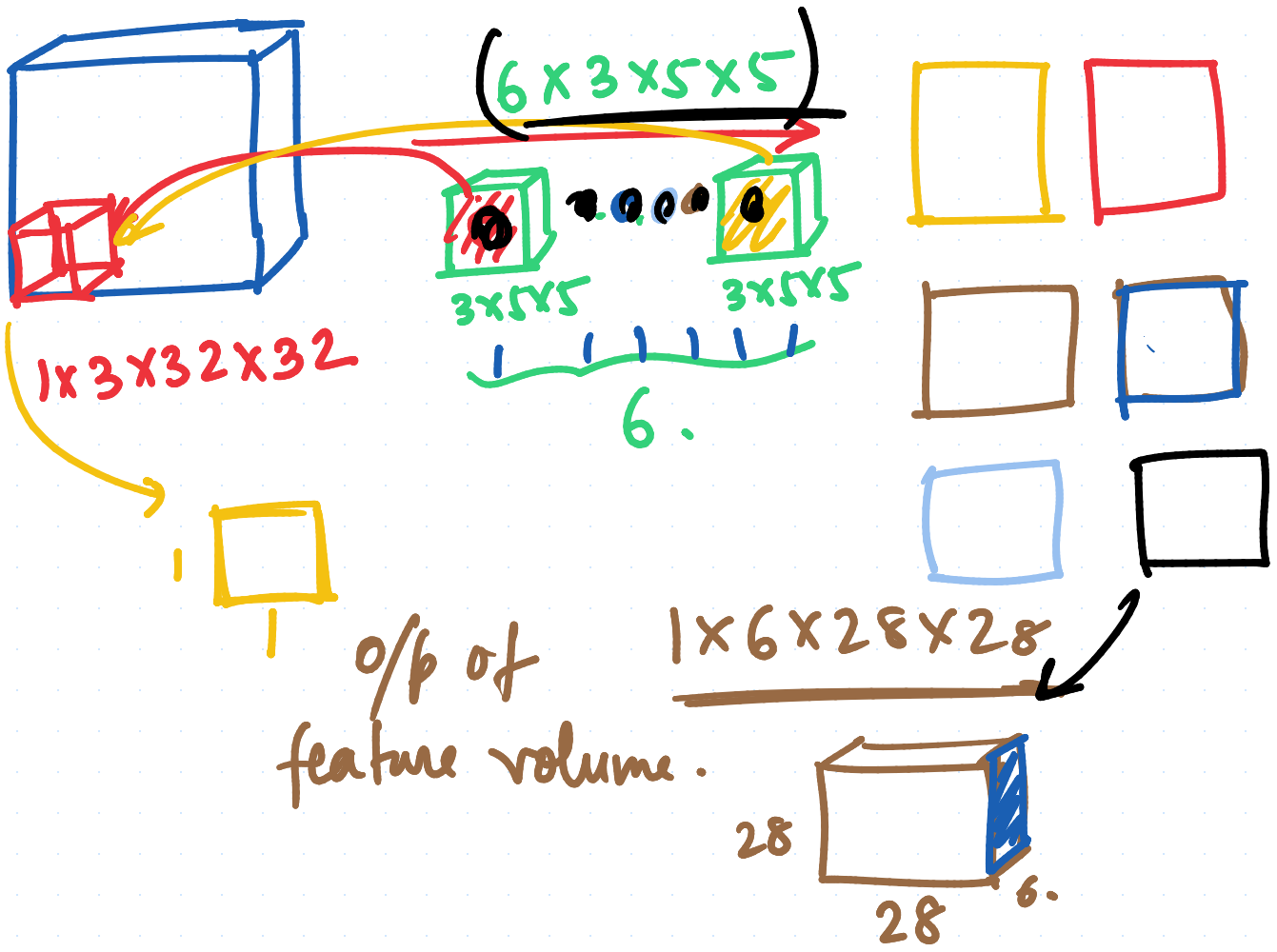
28

1x1x28x28.

Parameter in the kernel = ?

$5 \times 5 \times 3 + 1 = 75 + 1 = 76$



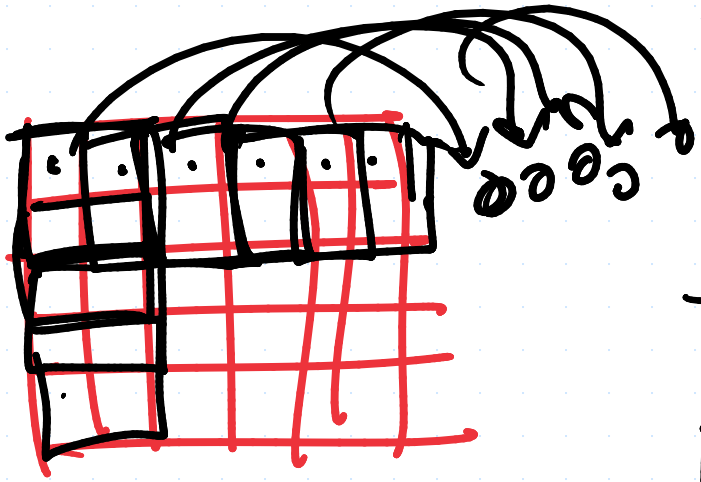


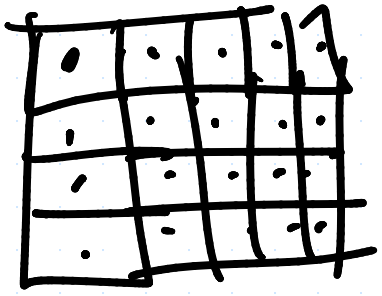
Parameters = ?

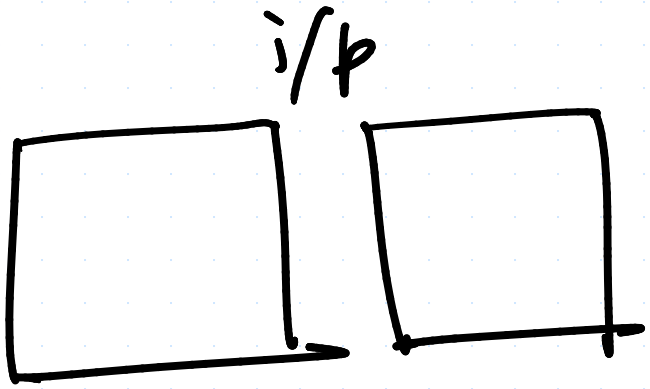
$3 \times 5 \times 5 \times 6 + 6 \Rightarrow 450 + 6$

25
x 18
200
25 x
450

$\Rightarrow 456$



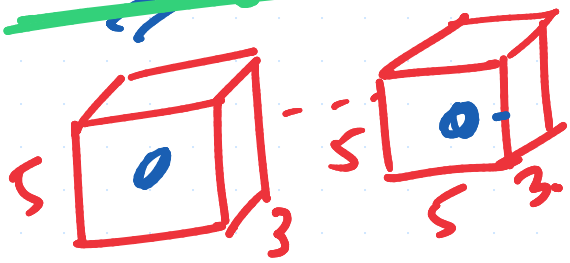




6 x 3 x 5 x 5

2 x 3 x 32 x 32

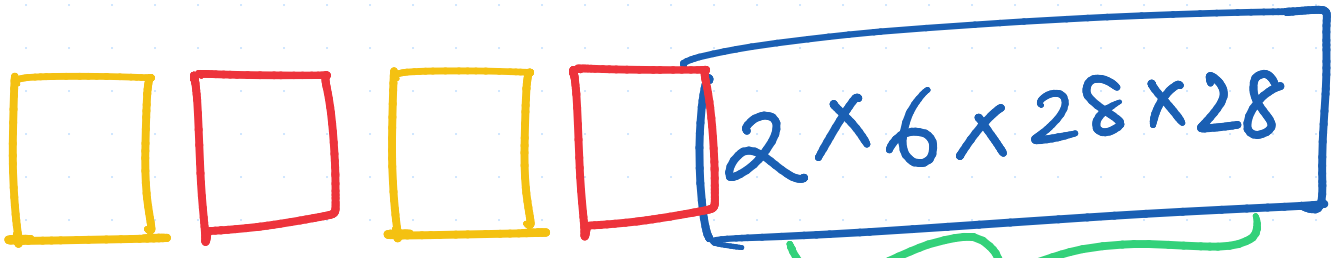
Batch size : 2



6.

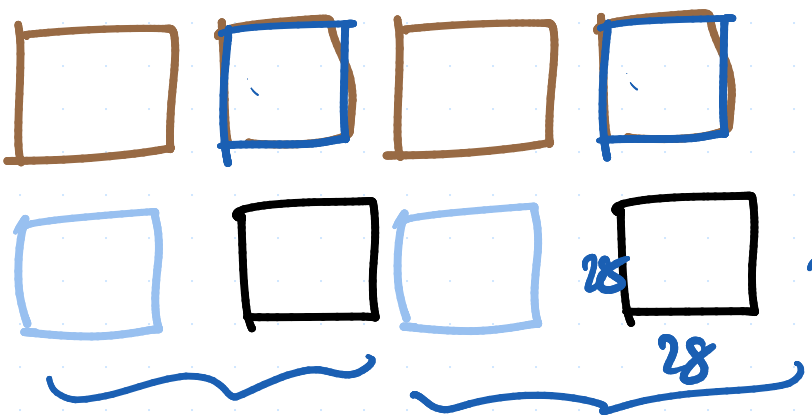
o/p feature volume

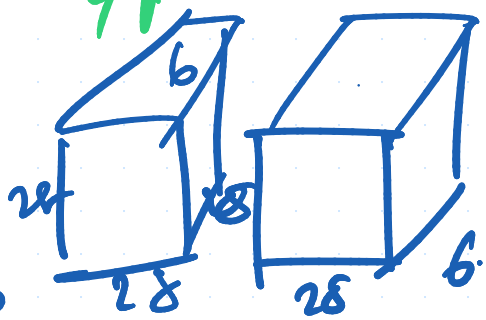
parameters = ?



2 x 6 x 28 x 28

o/p volume shape





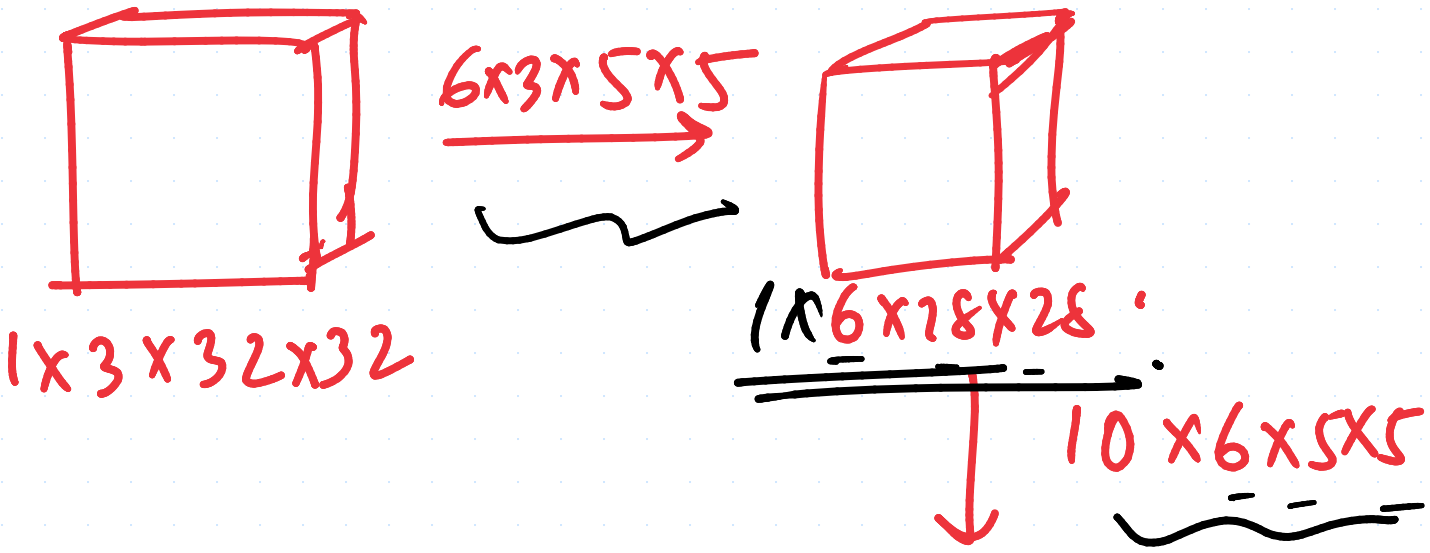
6 6

2-batch.

parameters. ⇒ 5×5×3×6+6.

456

If we increase the batch size, the parameters remains the same, but the o/p volume changes.

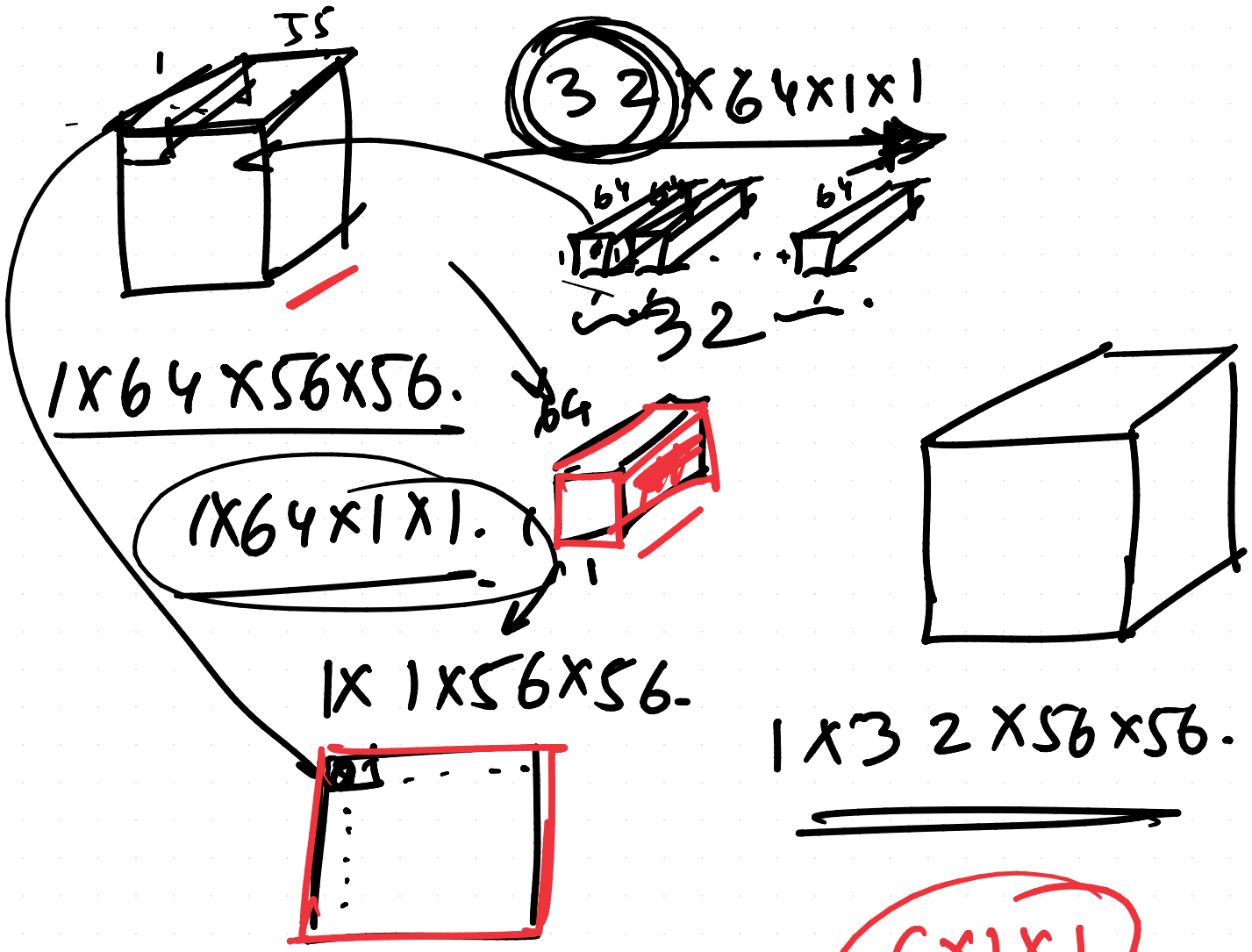


1x10x24x24.

Parameters

$(6 \times 3 \times 5 \times 5 + 6) + (10 \times 6 \times 5 \times 5 + 10)$

$\Rightarrow$ 1966.



C x 1 x 1

$32 \times 64 \times 1 \times 1 + 32.$

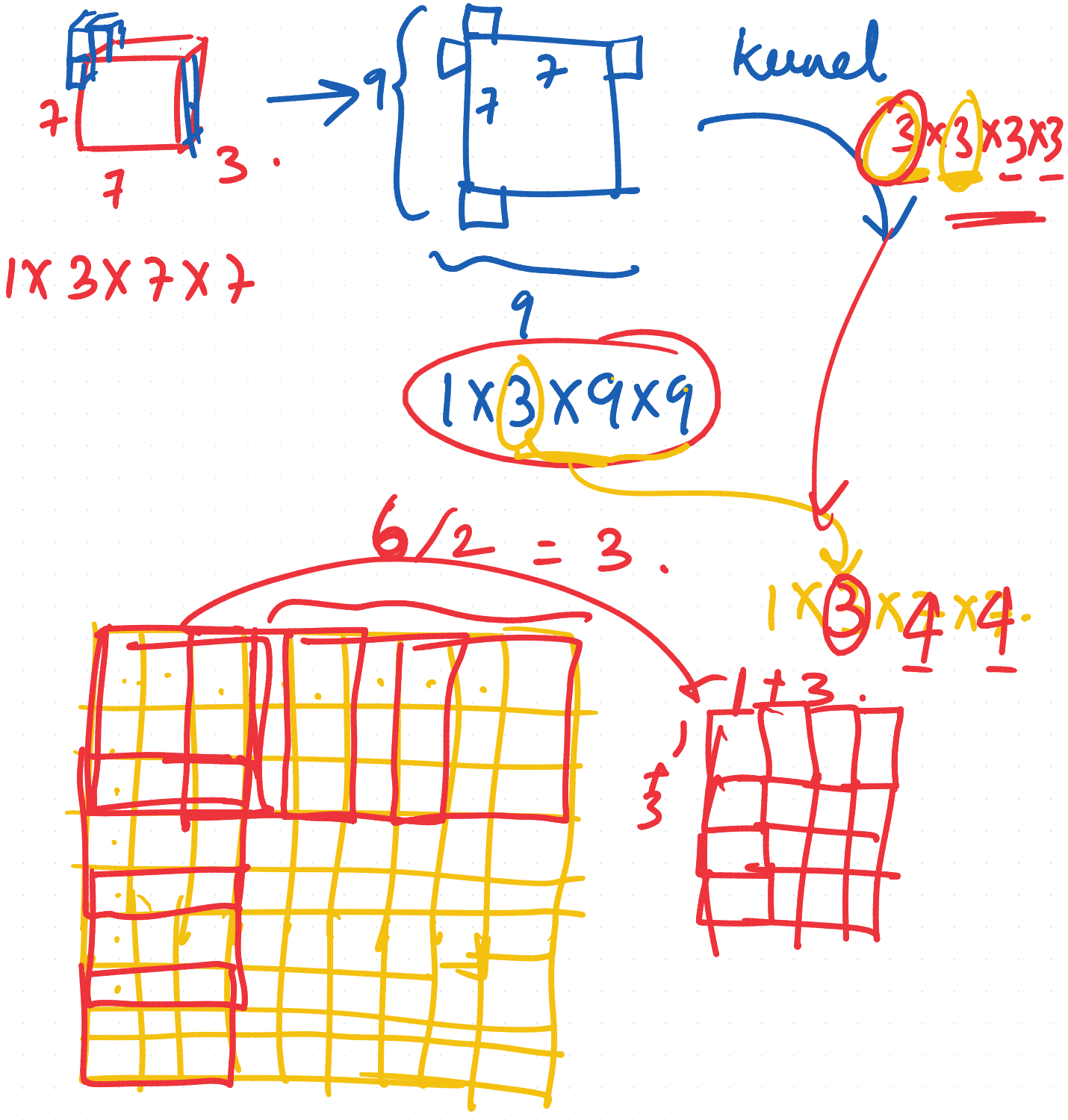
$\Rightarrow$ 2080.

C x 1 x 1 → filter's functionality is to transform an i/p volume

of dim (1 x C x H x W) to (1 x H x W)

thick → thin : feature volume

Calculating the parameters of a big neural network.

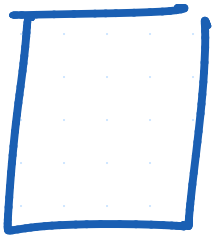


parameters $\Rightarrow 3\times3\times3\times3+3 \Rightarrow 81+3$

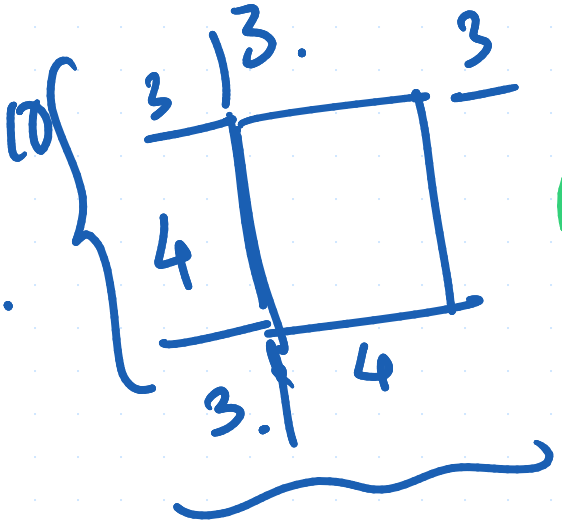
$\Rightarrow 84$

2
27
x 3
81.

1x3x4x4



pad = 3.



10.

3x3x3x3

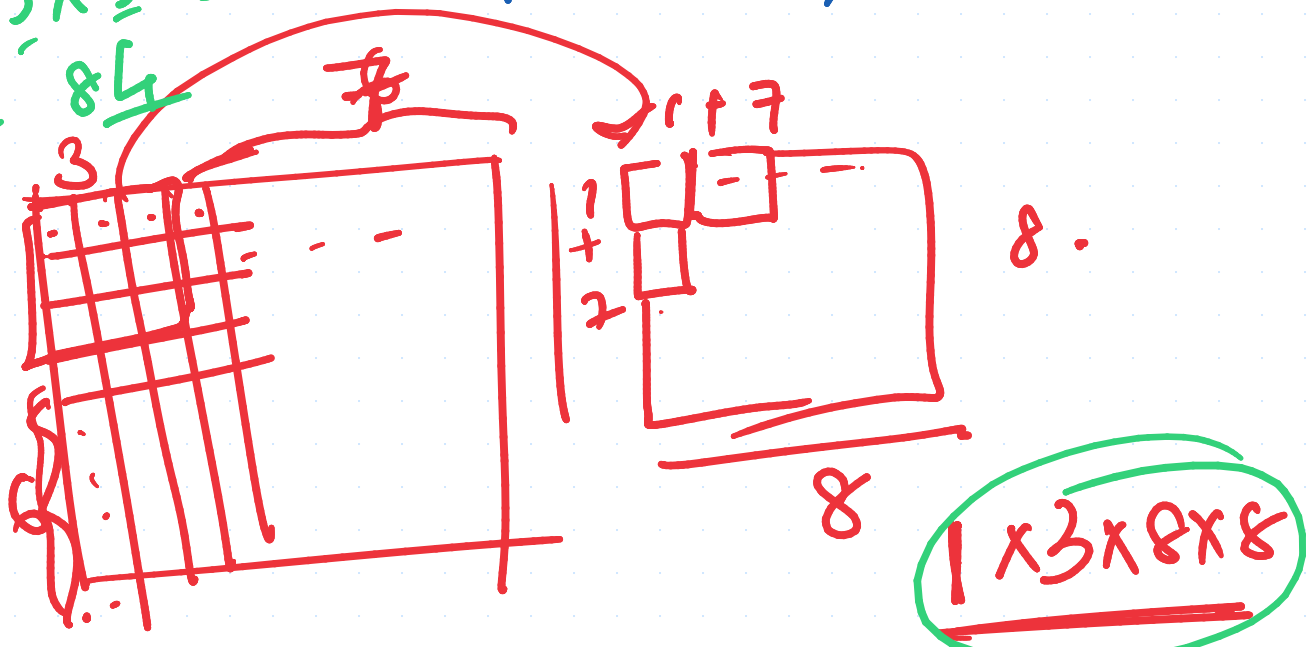
S=1,

params

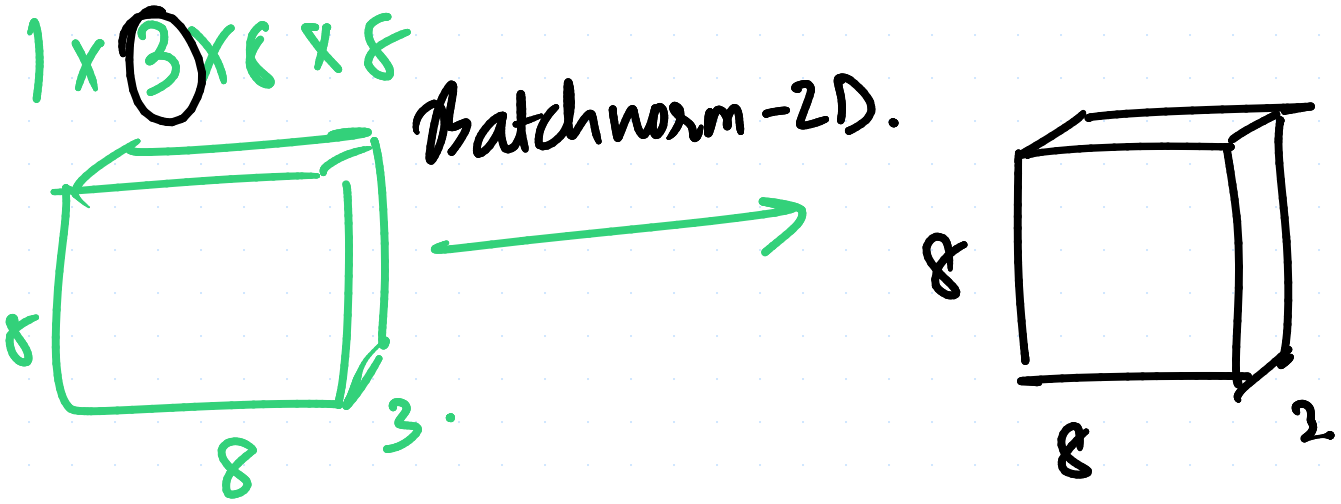
$= 3\times3\times3\times3+3$

$= 84$

1 x 3 x 10 x 10.



1x3x8x8

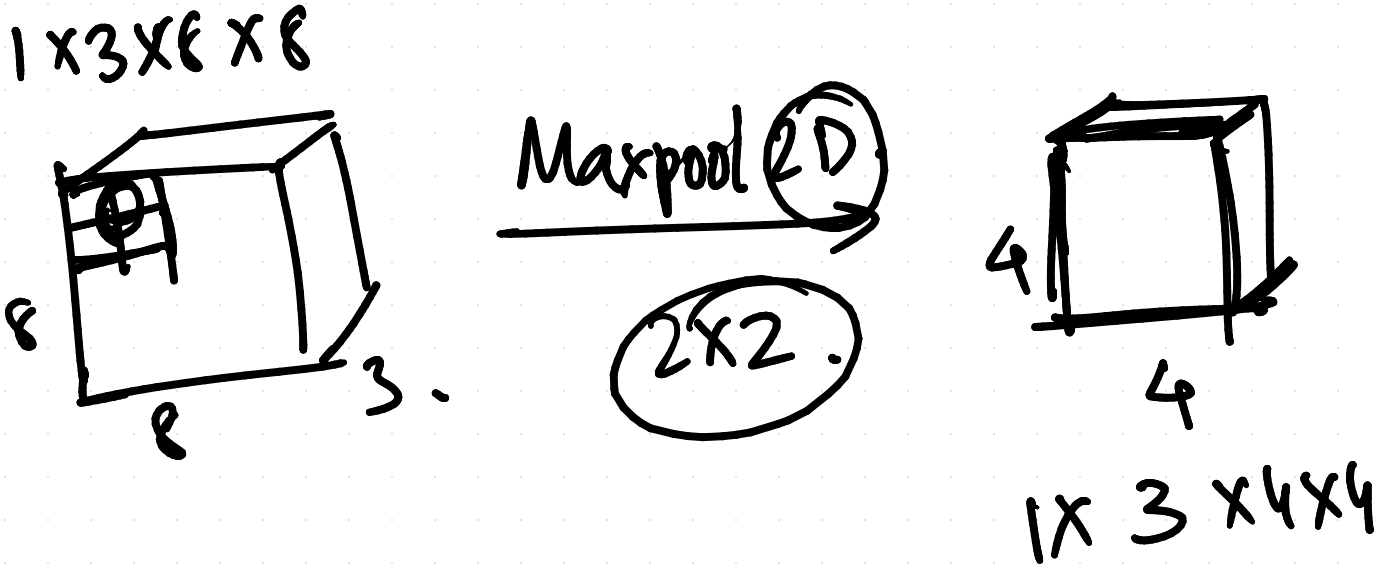


1 x 3 x 8 x 8

Batchnorm-2D doesn't change the o/p volume; it has only parameters.

(B), (D)

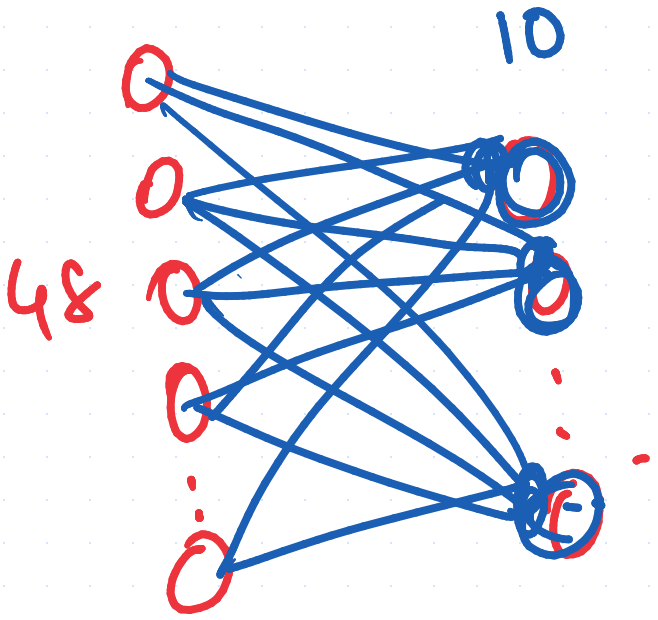
channels x 2 = 3 x 2 = 6.



Maxpool 2D doesn't have any extra parameters.

It just downsamples the feature volume.

No extra params.



$48 \times 10 + 10$

$\Rightarrow 480 + 10$

$\Rightarrow 490.$

0
0
⋮
0

1 × 10

BatchNorm 1D.

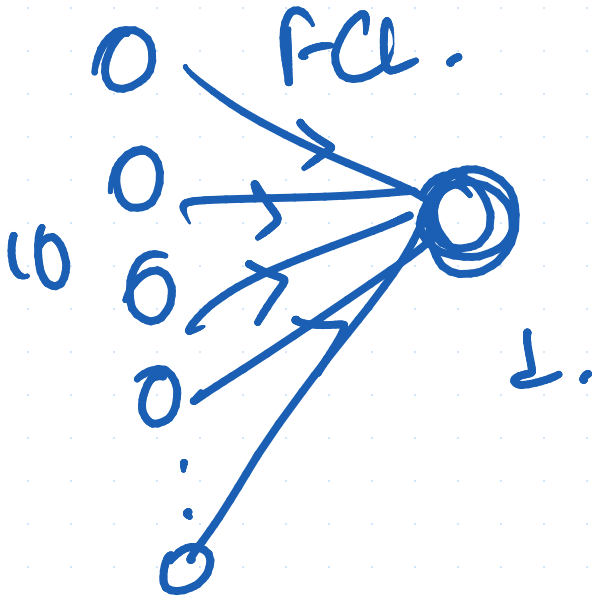
0
0
⋮
0

1 × 10

(γ, β)

$2 \times 10 \Rightarrow 20$

parameters.



$10 \times 1 + 1 = 11$

$84 + 84 + 6 + 0 + 0 + 490 + 20 + 11$

$= 695$