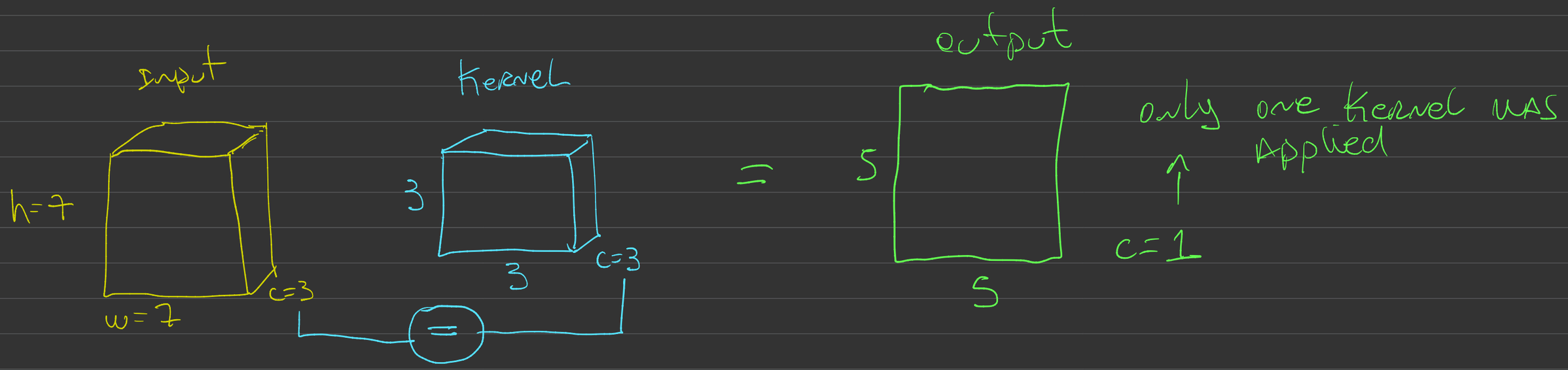
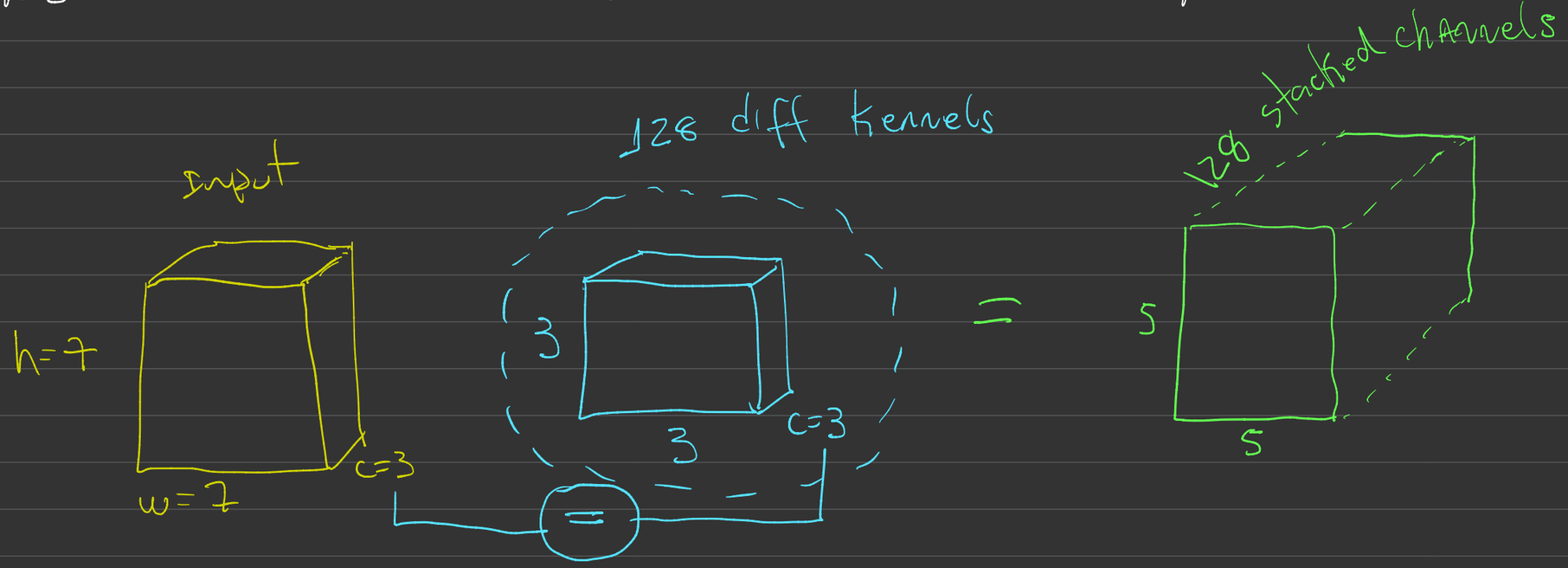
Standard 2D Convolutions (1)



To generate an output of 128 channels we must apply 128 kernels of 3x3x3, so we will have 128 output channels.



Standard 2D Convolutions (2)

Number of operations

$C_{in} = 3$
$H_{in} = 7$
$W_{in} = 7$

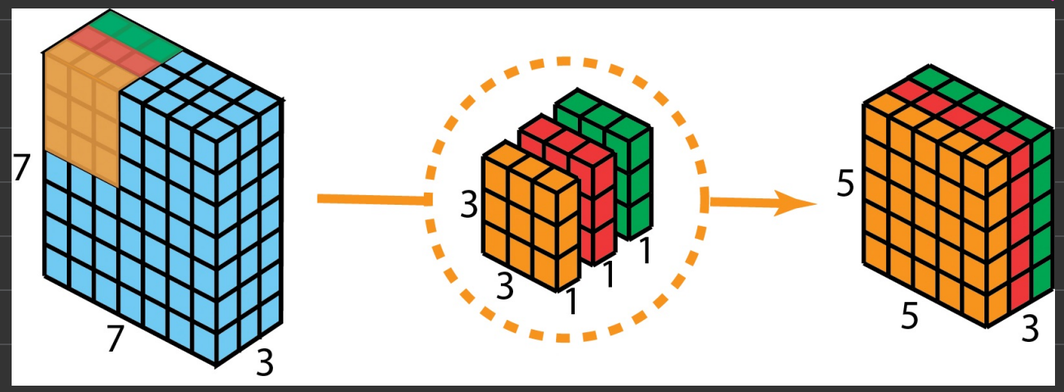
$C_{out} = 128$
$H_{out} = 5$
$W_{out} = 5$

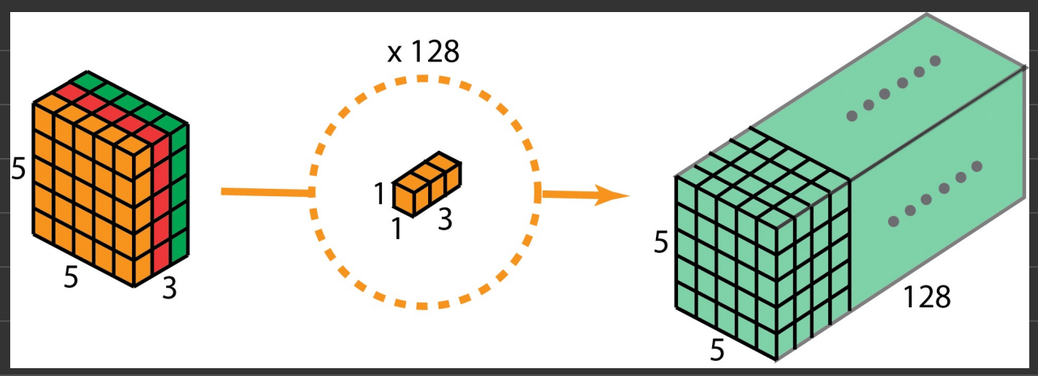
Filter
$C_{in} = 3$
$H = 3$
$W = 3$
$* C_{out}$

There are 128 (C_{out}) $3 \times 3 \times 3$ (Filter) kernels that moves 5×5 (C_{out}) times

$$= 128 * 3 \times 3 \times 3 * 5 \times 5 = 86{,}400 \text{ multiplications}$$

Depthwise Separable Convolutions (1)





1. Split the Input Image in channels I_1 I_2 I_3
2. Split the Filter in channels k_1 k_2 k_3
3. Apply each kernel in the respective In channel

$I_1 \otimes k_1 = Out_1$ [5,5,1]
$I_2 \otimes k_2 = Out_2$ [5,5,1]
$I_3 \otimes k_3 = Out_3$ [5,5,1]

4. Stack the output. This produces an output [5,5,3]
5. Apply Cout filters of 1x1xCin → $C_{IN} = 3$
   Each output will be of 5x5x1
6. Stack the 128 outputs will produce (5,5,128)

- Number of Operations

$C_{in} = 1$, $H_{in} = 7$, $W_{in} = 7$] x3

$C_{out} = 128$, $H_{out} = 5$, $W_{out} = 5$

Filter_1: $C_{in} = 1$, $H = 3$, $W = 3$] x3

Filter_2: $C_{in} = 3$, $H = 1$, $W = 1$] x 128

1) Filter_1: $3 \times 3 \times 3 \times 1 \times 5 \times 5 = 675$ mul
2) Filter_2: $128 \times 1 \times 1 \times 3 \times 5 \times 5 = 9600$ mu

] = **10.275 mult**

So, for an image with arbitrary size, how much time can we save if we apply depthwise separable convolution. Let's generalize the above examples a little bit. Now, for an input image of size H x W x D, we want to do 2D convolution (stride=1, padding=0) with Nc kernels of size h x h x D, where h is even. This transform the input layer (H x W x D) into the output layer (H-h+1 x W-h+1 x Nc). The overall multiplications needed is

Nc x h x h x D x (H-h+1) x (W-h+1)

On the other hand, for the same transformation, the multiplication needed for depthwise separable convolution is

D x h x h x 1 x (H-h+1) x (W-h+1) + Nc x 1 x 1 x D x (H-h+1) x (W-h+1) = (h x h + Nc) x D x (H-h+1) x (W-h+1)