② CONVOLUTIONAL NEURAL NETWORKS

(a).

GIVEN IMAGE :

$$R : \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad G : \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix} \qquad B : \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix}$$

CONVOLUTION FILTER :

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

CONVOLUTION ON DIFFERENT CHANNELS :

$$R : \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad R : \begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix}$$

$$G : \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix} \qquad G : \begin{bmatrix} 18 & 18 \\ 18 & 18 \end{bmatrix}$$

$$B : \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix} \qquad B : \begin{bmatrix} 18 & 18 \\ 27 & 27 \end{bmatrix}$$

OUTPUT : (R+G+B)

$$\Rightarrow \begin{bmatrix} 9+18+18 & 9+18+18 \\ 9+18+27 & 9+18+27 \end{bmatrix} = \begin{bmatrix} 45 & 45 \\ 54 & 54 \end{bmatrix}$$

(b.)

WITH ZERO PADDING

$$R : \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow R : \begin{bmatrix} 4 & 6 & 6 & 4 \\ 6 & 9 & 9 & 6 \\ 6 & 9 & 9 & 6 \\ 4 & 6 & 6 & 4 \end{bmatrix}$$

INPUT $\qquad$ $(\text{OUTPUT})_R$

INPUT

$$G: \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$(OUTPUT)_G$

$$G: \begin{bmatrix} 8 & 12 & 12 & 8 \\ 12 & 18 & 18 & 12 \\ 12 & 18 & 18 & 12 \\ 8 & 12 & 12 & 8 \end{bmatrix}$$

INPUT

$$B: \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 3 & 3 & 3 & 3 & 0 \\ 0 & 4 & 4 & 4 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$(OUTPUT)_B$

$$B: \begin{bmatrix} 6 & 9 & 9 & 6 \\ 12 & 18 & 18 & 12 \\ 18 & 27 & 27 & 18 \\ 14 & 21 & 21 & 14 \end{bmatrix}$$

FINAL_OUTPUT:(R+G+B):

$$\begin{bmatrix} 18 & 27 & 27 & 18 \\ 30 & 45 & 45 & 30 \\ 36 & 54 & 54 & 36 \\ 26 & 39 & 39 & 26 \end{bmatrix}$$

(C).

INPUT

$$R: \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

FILTER

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

RED_OUTPUT

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

INPUT

$$G: \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

FILTER

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

GREEN_OUTPUT

$$\begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}$$

$$B = \overset{\text{INPUT}}{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 3 & 3 & 3 & 3 & 0 \\ 0 & 4 & 4 & 4 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}} \quad \overset{\text{FILTER}}{\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}}$$

BLUE_OUTPUT

$$\begin{bmatrix} 12 & 12 \\ 8 & 8 \end{bmatrix}$$

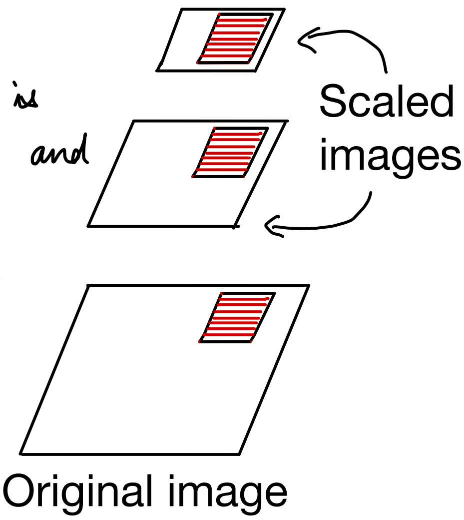
OUTPUT : R + G + B

$$\begin{bmatrix} 4+8+12 & 4+8+12 \\ 4+8+8 & 4+8+8 \end{bmatrix} = \begin{bmatrix} 24 & 24 \\ 20 & 20 \end{bmatrix}$$

(d). Template matching is performed through convolution by Selecting the filter as the template itself. If at all the feature we are trying to capture matches with any part of the image then the values at those regions have higher value compared to the areas where the template doesn't match. This is how convolution is interpreated as template matching.

(e). Multiple scale analysis can be acheived through the same window by pooling.

As seen in the figure alongside it is seen that by retaining the original filter and scaling the images the same template (filter) can be detected over different image scale.



This is how multiple scale analysis is performed through the same window.

(f). Performing multiple convolutions results in spacial resolution decrease as the edges of the images gets clipped. This is compensated by increasing the number of filters resulting in increasing the channels.
More number of channels results in better feature extraction because more filters captures more information.

(g.) Given tensor 128 x 128 x 32 (W = 128) filter size 3 x 3 x 32 (K = 3)
Number of filters = 16. Padding : 0 (P) Stride : 1 (S)

we have the formula ⇒

$$\text{output size} = \frac{w - k + 2p}{s} + 1 = \frac{128 - 3 + 0}{1} + 1 = 126$$

Output tensor = 126 X 126 X 16

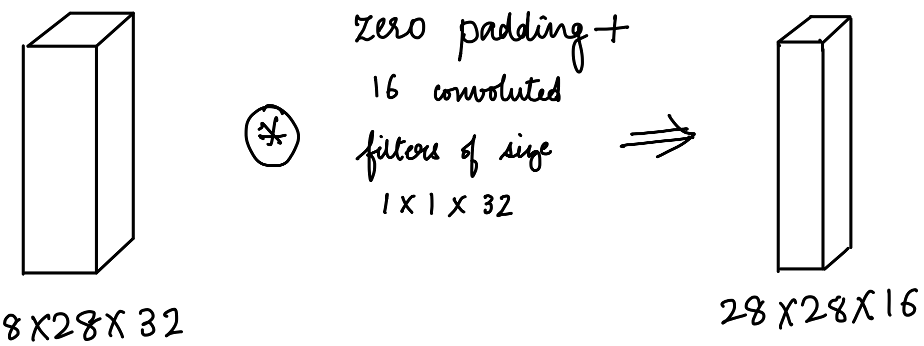
(h.) Same data as previous question with stride : 2

$$\text{output size} = \frac{w - k + 2p}{s} = \frac{128 - 3 + 0}{2} + 1 = 64$$

Output tensor = 64 X 64 X 16

(i) To perform channel reduction we perform convolutions using 1 x 1 filters (desired number to achieve o/p size) over the original image. The process is illustrated in the diagram below.

Eg:



(j) Early and deeper convolutional layers are interpreted to complexities of the patterns that are being recognised. More convolutional layers means more features has to be extracted.
If we take the example of recognizing a human face compared to recognizing a simple shape like a circle, the number of convolutional layers needed to recognize a human face would be deeper compared to recognizing a circle.

(l) The main purpose of pooling is to scale the images after every convolutional layer to reduce the amount of features that are extracted and to prevent overfitting.

(k) Given image:

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix}$$

Convolution filter: $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

After pooling we get the following results

$$R = \begin{bmatrix} \boxed{\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}} & \boxed{\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}} \\ \boxed{\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}} & \boxed{\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}} \end{bmatrix} \quad G = \begin{bmatrix} \boxed{\begin{matrix} 2 & 2 \\ 2 & 2 \end{matrix}} & \boxed{\begin{matrix} 2 & 2 \\ 2 & 2 \end{matrix}} \\ \boxed{\begin{matrix} 2 & 2 \\ 2 & 2 \end{matrix}} & \boxed{\begin{matrix} 2 & 2 \\ 2 & 2 \end{matrix}} \end{bmatrix} \quad B = \begin{bmatrix} \boxed{\begin{matrix} 1 & 1 \\ 2 & 2 \end{matrix}} & \boxed{\begin{matrix} 1 & 1 \\ 2 & 2 \end{matrix}} \\ \boxed{\begin{matrix} 3 & 3 \\ 4 & 4 \end{matrix}} & \boxed{\begin{matrix} 3 & 3 \\ 4 & 4 \end{matrix}} \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix}$$