No.

Date

H.W - 4

Sahil Sheikh - A205 18693

Deep learning - fall 22

Q.1

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad G = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix}$$

3+6=9

6+9=12

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix} \qquad f = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$f * R = \begin{bmatrix} x & x & x & x \\ x & 9 & 9 & x \\ x & 9 & 9 & x \\ x & x & x & x \end{bmatrix} = \begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix}$$

$\uparrow$ convolution

$$f * G = \begin{bmatrix} x & x & x & x \\ x & 18 & 18 & x \\ x & 18 & 18 & x \\ x & x & x & x \end{bmatrix} = \begin{bmatrix} 18 & 18 \\ 18 & 18 \end{bmatrix}$$

$$f * B = \begin{bmatrix} x & x & x & x \\ x & 18 & 18 & x \\ x & 27 & 27 & x \\ x & x & x & x \end{bmatrix} = \begin{bmatrix} 18 & 18 \\ 27 & 27 \end{bmatrix}$$

$\therefore$ final img $= (f * R) + (f * G) + (f * B) = \begin{bmatrix} 45 & 45 \\ 54 & 54 \end{bmatrix}$

2. with zero padding.

$$R = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

$$G = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

$$f * R = \begin{bmatrix} 4 & 6 & 6 & 4 \\ 6 & 9 & 9 & 6 \\ 6 & 9 & 9 & 6 \\ 4 & 6 & 6 & 4 \end{bmatrix} \qquad f * G = \begin{bmatrix} 8 & 12 & 12 & 8 \\ 12 & 18 & 18 & 12 \\ 12 & 18 & 18 & 12 \\ 8 & 12 & 12 & 8 \end{bmatrix}$$

$$B = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{pmatrix} \qquad f * B = \begin{bmatrix} 6 & 9 & 9 & 6 \\ 12 & 18 & 18 & 12 \\ 18 & 27 & 27 & 18 \\ 14 & 21 & 21 & 14 \end{bmatrix}$$

6+9
=18
12
27

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

Q.3 done later

No.

Date

zero padding → creates special activation at edges (dis-adv)
→ does not reduce img size (adv)

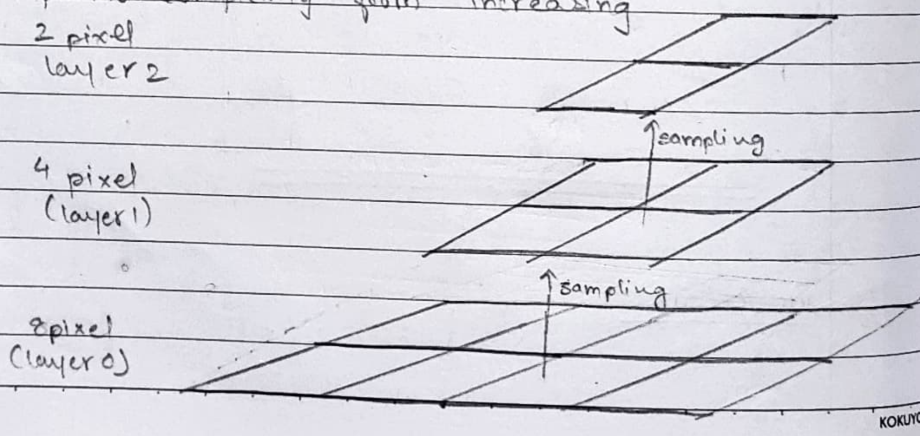
3.4 Template matching is process of moving the template over entire image and calculating the similarity b/t the template and covered window on image.
Template maching is implemented two - dimension convolution

filter dimension

$$I_A(i,j) = I(i,j) * A(i,j) = \sum_{h=-\frac{m}{2}}^{\frac{m}{2}} \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} A(h,k)\, I(i-h, j-k)$$

convolution

5. Multiscale analysis means processing of image at diff^n scales.
To achieve this with a fixed filter the image size needs to reduced as it gets sample. This would result in formation of an image pyramid, i.e image with different sizes (resolutions) at different convolution layer.
In this way a fixed size convolution filter cover larger details (area) in upper layers and smaller details in the lower layers.
This helps the complexity from increasing



7.
W = 12
K = 3
S = 0
P = 0

Output size

∴ size of
∴ size of

8. W = 128 × 1
K = 3
S = 2
P = 0

∴ Output size

∴ Output size

6.
As spatial information we compense decrease b So as we g the number the depth.
The main for reduced

KOKUYO

~~200~~

168

7.

$W = 128 \times 128 \times 3$

$K = 3$

$S = \cancel{0}\ 1$

$P = 0$

Output size $= \dfrac{W - K + 2P}{S} + 1 = \left(\dfrac{128 - 3 - 0}{1}\right) + 1 = 126$

$\therefore$ size of img after 1 filter $= 126 \times 126 \times 1$

$\therefore$ size of img after 16 filter $= \underline{\underline{126 \times 126 \times 16}}$

8. $W = 128 \times 128 \times 3$

$K = 3$

$S = 2$

$P = 0$

$\therefore$ Output size $= \dfrac{W - K + 2P}{S} + 1 = \left(\dfrac{128 - 3}{2}\right) + 1 = \cancel{64 \times 64 \times 1}\ \ 63.5 + 1 \quad 64 \times 64 \times 1$

$\therefore$ Output size after 16 filter $= \cancel{64 \times 64 \times 16}\ \ \cancel{63 \times 63 \times 16}$

$\underline{\underline{64 \times 64 \times 16}}$

6. As spatial resolution decrease, we end up with less information as we loss information. To counter this we compensate for spatial ~~resoulti~~ resolution decrease by using higher depth. This is done by:- So as we go deeper in the network, we increase the number of filters in each layer to increase the depth.

The main purpose of doing this to compensate for reduced coefficients ie keep the same number of coefficients.

No.

Date

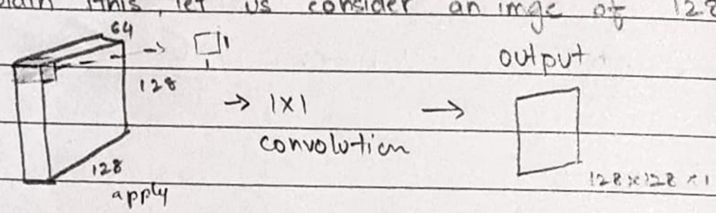
9.

1 x 1 convolution means the kernel size = 1 x 1

This means while calculating weighted average, the neighbouring pixels are not considered.

1 x 1 convolutions allows to reduce size of image, by reducing the number of channels i.e reducing the depth and allows the user to control the complexity of the implementation. ~~and also~~

To explain this, let us consider an imge of 128 x 128 x 64



now when we apply kernel of size 1x1, it will span the entire depth of image (i.e 64) here and will produce a single output after the weighted sum.

~~10.~~

3.

dilation rate = 2

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

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

$$f * R = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$f * G = \begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}$$

$$f * B = \begin{bmatrix} 8 & 8 \\ 12 & 12 \end{bmatrix}$$

$$\text{final} = \begin{bmatrix} 20 & 20 \\ 24 & 24 \end{bmatrix}$$

10.

- Convolution layer are implemented to extract features of the image. In neural networks these features are learned as the model is trained.
- The early convolution layers extract simple features such as an vertical or horrizontal edge.
- The deeper convolution layers extract complex features for example eyes, nose, ears, etc from an image of an animal.

No.

Date

11.

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

∴ After ~~pooling~~ Max-pooling,

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

12.

The purpose of pooling is to reduce the spatial size of the image. Pooling only reduces the spatial dimensions of the image, ~~ori~~ i.e width and height of the image but does not ~~change/ effect~~ reduce the depth of the image.

13.

The purpose of data ~~augme~~ augmentation is to increase the existing amount of data by creating copies of data by augmenting the data. This helps the model to generalize better.
Data Augmentation is helpful when the dataset is ~~s~~ small.

14.

Transfer learning reffers to utilizing a pre-trained network, trained on ImageNet dataset for the purpose of object classification.

No.

Date

The purpose to transfer learning ~~*~~ is to utilize pretrained weights from a pre'trained model. This is most useful when the task / problems are solved by untrained and pretrained network are similar.

15.

It is important to freeze coefficient of pretrained models as not freezing the coefficients would result in destruction of weights of the pre trained network.

16.

Steps to fine tune :-

1) Add custom network on top of trained layers
2) freeze trained layers
3) Train custom networks
4) Unfreeze top layers in the base network.
5) Jointly train the custom network and unfrozen layers.

17. The purpose of inception blocks are :-

1) have multiple receptive feilds instead of a single receptive feild
2) it helps to reduce the number of parameters of the model. using ~~dim~~ 1X1 convolutions.

KOKUYO

No.

Date

18. The advantage of residual block (residual connection) are:-

(1) helps with vanishing gradient.

(2) zero weights in block produces identity instead of

(2) destroying the signal

(3) Network can learn to zero block to eliminate un-needed layers.

(4) Information passes through units with zero weights.

19. Intermediate activations of convolution layers can by visualized by :-

i) creating a new model from existing one with new outputs

2) making use of "model" class instead of "sequential" class.

Visualization helps to understand the working of the convolution layers.

20.

- Using gradient descent will find the input which will maximize the response at the filter ie correlated with the filter.
- Using gradient descent can also find the input which will minimize the loss at the filter

This helps to interpret the working of filter

No.

Date

21.

1) feed image to the network
2) Compute the gradient of selected output note with respect to each channel of the target layer where activation is to be computed.
3) Compute the average of each channel
4) Add the activation of each channel weighted by their average gradient magnitude
5) Superimpose activation on input image.