ASSIGNMENT # 4 CS-577

Theoritical Questions A20519664

1.

RGB Image = $\begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix}$ filter = $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

convolution = $\begin{bmatrix} 15 \\ 18 \end{bmatrix}$

2. RGB IMAGE = $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 2 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ filter = $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

convolution = 6 9

6 9 7

15

$\begin{bmatrix} 6 & 9 & 7 \\ 9 & 15 & 12 \\ 9 & 18 & 15 \\ 6 & 13 & 11 \end{bmatrix}$

3) using dilated convolution

dilation rate = 2

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

$$= \begin{bmatrix} 6 & 7 \\ 9 & 15 \\ 6 & 17 \end{bmatrix}$$

(4) Template matching Interpretation

- convolution filter looks for similarity
- filter can be the pattern of the template and if we apply the filter to image, we get shapes/template highlighted at the result image.

(5) To resize filter, it becomes expensive and inefficient, instead we use fixed size convolution filter and make the image smaller and smaller, so the filter covers more area. The size is halved on every step.

6) Convolve the image with small filters, share the weight of filters between locations, this effectively increases the data

we want to reduce the spatial resolution because we want to have multiscale pyramid analysis, at the same time we don't want to loose the information, that is why we compensate for it to get higher depth. and reduced coefficients

7)

Input tensor: $128 \times 128 \times 32$

convolution filter = 16 filters

size = $3 \times 3 \times 32$

with zero padding: output tensor = $128 \times 128 \times 16$

without zero padding: output tensor = $126 \times 126 \times 16$

8) when using stride 2 = ~~64x64x16~~ $\frac{128-3}{2} + 1$

$\lfloor \frac{125}{2} \rfloor + 1 = 63 \times 63 \times 16$

9) 1x1 convolution is used for reducing dimensions so, we dont look at neighbours while calculating weighted average, we just look at that one pixel.

) The early layers can be interpreted for detecting edges and feature extraction and the deeper layers can be interpreted as classifying objects based on that inputs

11)

$$I = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix}$$

Result after max pooling

with Stride 1 $= \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$

with stride 2 $= \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$

12) It is used to reduce dimensions of the feature maps. Thus, it reduces the number of parameters to learn and amount of computation performed in network, It summarizes the features present in region.

13) It is used to increase the amount of data by adding slightly modified copies of already existing data, It is most useful when we have less amount of data and dont want to overfit.

14) use of pretrained model as a starting point of a model for new task, we use it to save time and resources by using model made for similar purpose

15) Freezing weights of a layer prevents its weights from being updated; it helps to preserve weights of a pretrained model during future training. The bad updates wont destroy fine trained model.

16) After training data, we unfreeze the top layers of pretrained model and retrain on freezed model to fit the model better.

17) we use inception block to get ~~dimension increase~~ multiple receptive fields to learn from their input. we control the number of parameters though it. To reduce number of channels.

Residual blocks allow information to flow from initial to last layers by skipping some layers. Its benefit is that, it allows to train extreme deep neural networks.

19) The purpose of visualizing intermediate activation is to understand the successive convent layers transformed from the input, they can give visual concepts of feature maps that are output of a layer. We can do so by plotting feature map of the layers output we can see what network is doing.

20) we find the input that will maximize the response of the filter, which means it is correlated with the filter. and then visualize input that maximizes the loss. The purpose is to visualize ~~the~~ how well we have achieved our goal.

) To visualize activation heatmaps, we use a weighted sum where more important channels get higher weights, higher weight means channels with higher gradients.

Algo:-

1) feed image to network
2) compute gradients of selected output node
3) compute average gradients of each channel.
4) Add activation of each channel
5) Super impose weights by their average gradient magnitude