ASSIGNMENT → 4

NAME :- PUSHPENDRA SINGH CHAUHAN
STUDENT ID :- A20472647
COURSE NO. :- CS 577
SEMESTER :- SPRING 2021

Q. 1) Let I be a 4×4 RGB image where the R channel is all 1-s and G channel is all 2-s. The B channel has a value of 1 in its first row, a value of 2 in its second row, a value of 3 in its third row, and a value of 4 in its 4th row. Compute the convolution of this image with a 3×3 filter having all ones without zero padding.

Soln:- R-channel

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

G-channel

2	2	2	2
2	2	2	2
2	2	2	2
2	2	2	2

B-channel

1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4

multiply with each channel

3×3×3 filters:

1	1	1
1	1	1
1	1	1

1	1	1
1	1	1
1	1	1

1	1	1
1	1	1
1	1	1

Image after convolving with 3×3×3 filter [Without zero padding]:

45	45
54	54

2×2

Ans.

Q.2> Repeat the previous question with zero padding.

Ans:- Image after convolving with 3×3 filter: (With zero padding)

18	27	27	18
30	45	45	30
36	54	54	36
26	39	39	26

Ans.

Q.3> Repeat the previous question when using dilated (atrous) convolution with a dilation rate of 2.

Ans:- 2-dilated convolution filter

	1		1		1	
	1		1		1	
	1		1		1	

Image after convolving with 3×3 dilated filter (With zero padding):-

4×4

20	20	20	20
24	24	24	24
20	20	20	20
24	24	24	24

Ans.

Q.4> Explain the template matching interpretation of convolution.

Ans:- Template matching is the process of moving the template over the entire image and calculating the similarity between the template and the covered window on the image. Template matching is implemented through two-dimensional convolution filter.

In convolution, the value of an output pixel is computed by multiplying elements of two matrices and summing the result (i.e. dot product of two matrices). One of these matrices represents the image itself, while the other matrix is the template, which is known as a convolution kernel.

$$g(x,y) = f(x,y) \cdot t(x,y)$$

$g(x,y)$ → Image after convolution

$f(x,y)$ → original image

$t(x,y)$ → Convolution filter.

Q.5> Explain how multiple scale analysis can be achieved with a fixed window size (using a pyramid).

Ans:- By using pooling and stride (greater than 1), we can achieve multiple scale analysis with a fixed window size. (using a pyramid).

Q.6) Explain how to compensate for spatial resolution decrease using depth (no. of channels) and the purpose for doing so.

Ans :- As spatial resolution (dimensions) decreases, depth increases to compensate for reduced coefficients (keep the same number of coefficients).

The purpose behind doing so is that we don't want to loose information due to small no. of coefficients so, we increase depth resulting increase in number of coefficients which can easily store information coming from previous layer.

Q.7) Given a 128 X 128 X 32 tensor and 16 convolution filters of size 3x3x32, what will be the size of the resulting tensor when convolving without zero padding.

Ans :- 126 X 126 X 16 will be the size of the resulting tensor.

Q.8) Repeat the previous question when using a stride of 2.

Ans :- General formula for resulting convolved image :

$$\left[\left(\frac{W-m}{S}\right)+1\right] \times \left[\left(\frac{h-n}{S}\right)+1\right] \times \text{no. of filters}$$

Here,
W = width of original image
h = height of original image
s = value for stride
m = width of convolution filter
n = height of convolution filter

So, given:
$W = 128$, $h = 128$, $s = 2$
$m = 3$, $n = 3$

$$\left[\left(\frac{128-3}{2}\right)+1\right] \times \left[\left(\frac{128-3}{2}\right)+1\right] \times 16$$

$$\Rightarrow \left[\frac{125}{2}+1\right] \times \left[\frac{125}{2}+1\right] \times 16$$

$\Rightarrow \approx \underline{64 \times 64 \times 16}$ will be the size of the resulting tensor.

Q.9) Explain how the number of channels can be reduced using a 1×1 convolution.

Ans:- Let suppose we have an Image with dimension $m \times n \times 3$.
Now,
if we convolve this image with a $1 \times 1 \times 3$ convolution filter with stride = 1.
then,
$$\left[\left(\frac{m-1}{1}\right)+1\right] \times \left[\left(\frac{n-1}{1}\right)+1\right] \times 1$$

⇒ $m \times n \times 1$ is the dimensions of the resulting convolved image. Here, number of channels reduced from 3 to 1 without changing dimensions of the image (i.e. width and height).

Q.10) Explain the interpretation of convolution layers and the difference between early and deeper convolution layers.

Ans:- Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data.

→ Extract image patches (windows).

→ Vectorize image windows and filter and perform dot product (plus bias).

→ Filter extends full depth of image.

→ Multiple convolution filters per location (e.g. oriented edges).

→ Use stride to move filter → activation map may be smaller.

Early convolution layers detects or extracts low-level features such as lines from the raw pixel values.

While deeper convolution layers may extract features that are combinations of lower-level features, such as features that comprise multiple lines to express shapes. Very deep layers are extracting faces, animals, houses etc.

Q.11) Let I be an image as in question 1. Write the result obtained using max pooling with a 2×2 filter with a stride of 2.

Ans:- R-channel

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

G-channel

2	2	2	2
2	2	2	2
2	2	2	2
2	2	2	2

B-channel

1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4

Image after max-pooling with a 2×2 filter with a stride of 2.

R-channel

1	1
1	1

G-channel

2	2
2	2

B-channel

2	2
4	4

Dimension of image before max-pooling = 4×4×3

Dimension of image after max-pooling } = 2x2x3

Note :- Max-pooling does not affect depth or number of channels of the image.

Q.12) Explain the purpose of pooling.

Ans :- The purpose of pooling is to progressively reduce the spatial size of the representation or downsampling spatial dimensions without affecting depth in order to reduce the amount of parameters and computation in the network.

Pooling layer operates on each feature map independently.

Q.13) Explain the purpose of data augmentation and when it is most useful.

Ans :- The purpose of data augmentation is to augment / increase data by applying various transformations ~~on origin~~ like rotations, horizontal flip, zoom, sheer etc. on the original data for better generalization.

It is most useful when model is overfitting due to small dataset.

It helps us in better generalization and improving model performance when we have less data.

Q.14). Explain the purpose of transfer learning and when it is most useful.

Ans:- The purpose of transfer learning is to use a pretrained convnet trained on a large data set (e.g. ImageNet object classification) compared with small available data (cats/dogs).

It is most useful when we have small dataset and want to improve model performance.

Q.15). Explain the need for freezing the coefficients of the pre-trained network.

Ans:- We want not to destroy weights/coefficients of the pre-trained network by gradients from untrained fully connected (FC) layers on top due to this reason it is needed to freeze the coefficients of the pre-trained network.

Q.16). Explain how the coefficients of a pre-trained network can be fine-tuned.

Ans:- x Fine tuning:- After training the fully connected (FC) layers, unfreeze some top

layers in the base network and retrain to allow the model to fit the data.

* Steps :-

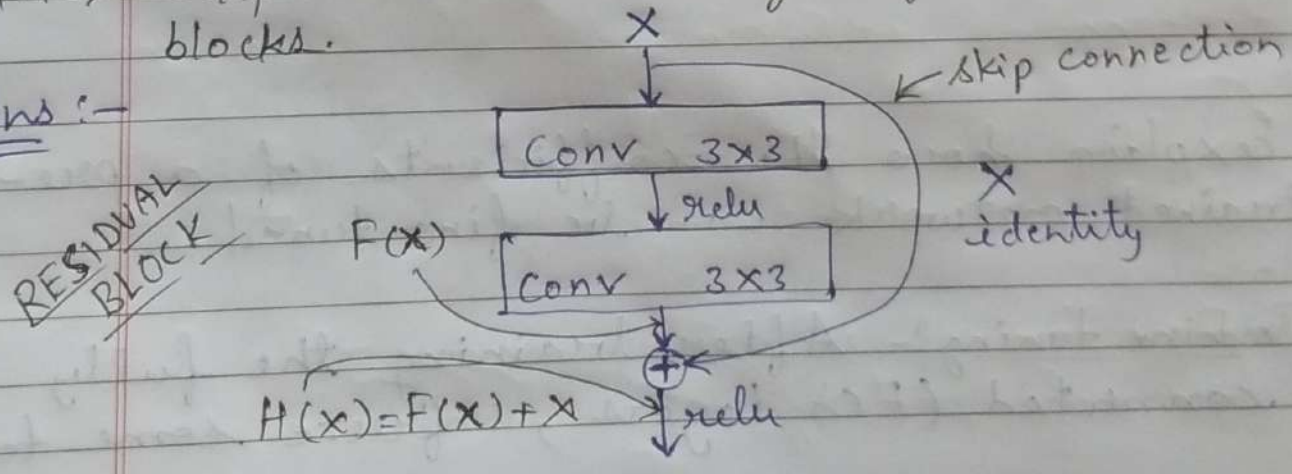
(i). Add custom network on top of trained layers.
(ii). Freeze trained layers.
(iii). Train custom network.
(iv). Unfreeze top layers in the base network.
(v). Jointly train the custom network and unfreeze layers.

Q.17). Explain the purpose of inception blocks.

Ans :- The purpose of inception blocks is to solve the problem of computational expense as well as overfitting by reducing the number of parameters using 1x1 convolutions.

Q.18). Explain the advantage of residual blocks.

Ans :-



Advantage of residual blocks:-

(a). Easier to learn $F(x)$ residual compared with $H(x)$ (learn deviation from identity instead of function).

(b). Skip connections help with vanishing gradients.

(c). Zero weights in the block produce identity instead of destroying the signal.

(d). The network can learn to zero blocks to eliminate un-needed layers.

(e)

(f). If coefficients in a regular network decay to zero they shut down information whereas here the information passes through units with zero weights.

(f). Gradients are passed directly through skip connections ⟹ quicker training.

Q:19). Explain how intermediate activations of convolution layers can be visualized given an input. What is the purpose for doing so?

Ans:-