Assignment 6_Solution

Q1. Consider the following statements regarding Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN):

1. There are sparse connections between inputs and outputs between two consecutive layers in a CNN.
2. Parameters are shared between output neurons in a CNN layer.
3. There are sparse connections between inputs and outputs between two consecutive layers of an ANN.
4. For any two layers with the same number of neurons an ANN will have fewer parameters than a CNN.

Which of the above statements are **TRUE**

A. 1 and 2
B. 1, 2 and 3
C. 1, 3 and 4
D. 2, 3 and 4

Ans: A

Q2. What will be the size of the output of a convolutional layer with :

Input size = [227 x 227 x 3],
Filter Size = [11 x 11 x 3],
Stride = 4

A. [54 x 54]
B. [55 x 55]
C. [216 x 216]
D. [68 x 68]

ANS: B

Solution:

Given,

$[W_i \times H_i \times D_i] = [227 \times 227 \times 3]$

$Filter\ size = [f_w \times f_h \times f_d] = [11 \times 11 \times 3]$

Stride = 4

Padding = 0

Suppose $output\ size = [W_o \times H_o \times D_o]$ then,

$W_0 = \frac{W_i - f_w + 2P}{S} + 1 = \frac{227-11+0}{4} + 1 = 55$

$H_o = \frac{H_i - f_h + 2P}{S} + 1 = \frac{227-11+0}{4} + 1 = 55$

Q3. Pooling layers are used to accomplish which of the following?

A. To progressively reduce the spatial size of the representation.
B. To reduce the amount of parameters and computation in network.
C. To select maximum value over pooling region always.
D. None of the above.

ANS: A & B

Answer questions 4-6 for the CNN architecture given below

The whole network is composed of CONV layers that perform 3x3 convolutions with stride 1 and padding is 'same'. POOL layers perform 2x2 max pooling with stride 2 (and no padding). Number of filters in the Conv layers and number of neurons in fully connected layers are shown in brackets.

Answer the following questions.

Q4. The output size after pool1, pool2 are

A. [111x111x128], [56x56x64]
B. [112x112x128], [56x56x64]
C. [114x114x128], [58x58x64]
D. [111x111x128], [58x58x64]

Ans: B

Solution:

SInce Padding is same, width and height of output of convolutional layer will be equal to the input of the convolutional layer. Max pooling has stride 2, this will reduce width and height of input by factor of 2. Therefore,

Input size of pool1 layer $= 224 \times 224 \times 128$

Output size of pool1 layer $= 112 \times 112 \times 128$

Similarly

Input size of pool2 layer $= 112 \times 112 \times 64$

Output size of pool2 layers = $56 \times 56 \times 64$

Q5. Number of parameters till pool1 are

A. 89186
B. 73570
C. 75648
D. 64898

Ans: C

Solution:

Calculation of number of parameters in given Network.

- # parameter in input layer = 0
- # parameters in conv1 layer
    A. Filter size $= 3 \times 3$
    B. # channels in the input = 3
    C. # filter fov conv1 layer = 64
    D. # bias parameter for each filter = 1
    E. Therefore,
    # parameters in conv1 layer $= 3 \times 3 \times 3 \times 64 + 64 \times 1 = 1792$
- # parameters in conv2 layer
    A. Filter size $= 3 \times 3$
    B. # channels in the input = 64
    C. # filter fov conv1 layer = 128
    D. # bias parameter for each filter = 1
    E. Therefore,
    # parameters in conv1 layer $= 3 \times 3 \times 64 \times 128 + 128 \times 1 = 73856$
- Total # parameter till pool1 layers $= 1792 + 73856 = 75648$

Q6. Total number of parameters in the given network except output layer are

E. 104097392
F. 206081344
G. 326789108
H. 207816190

Ans: B

Solution:

Calculation of number of parameters in given Network.

- # parameter in input layer = 0
- # parameters in conv1 layer
    F. Filter size $= 3 \times 3$
    G. # channels in the input = 3
    H. # filter fov conv1 layer = 64
    I. # bias parameter for each filter = 1
    J. Therefore,
    # parameters in conv1 layer $= 3 \times 3 \times 3 \times 64 + 64 \times 1 = 1792$
- # parameters in conv2 layer
    F. Filter size $= 3 \times 3$
    G. # channels in the input = 64
    H. # filter fov conv1 layer = 128
    I. # bias parameter for each filter = 1
    J. Therefore,
    # parameters in conv1 layer $= 3 \times 3 \times 64 \times 128 + 128 \times 1 = 73856$

- Total # parameter till pool1 layers $= 1792 + 73856 = 75648$
- # parameters in conv3 layer
    A. Filter size $= 3 \times 3$
    B. # channels in the input = 128
    C. # filter fov conv1 layer = 128
    D. # bias parameter for each filter = 1
    E. Therefore,
    # parameters in conv1 layer $= 3 \times 3 \times 128 \times 128 + 128 \times 1 = 147584$
- # parameters in conv4 layer
    A. Filter size $= 3 \times 3$
    B. # channels in the input = 128
    C. # filter fov conv1 layer = 64
    D. # bias parameter for each filter = 1
    E. Therefore,
    # parameters in conv1 layer $= 3 \times 3 \times 128 \times 64 + 64 \times 1 = 73792$
- #parameters in Fully Connected layer 1
    A. Input to the FC1 $= 56 \times 56 \times 64 = 200704$
    B. Output neurons of FC1 = 1024
    C. Total # parameters in FC1 layer
    $= 200704 \times 1024 + 1024 = 205521920$
- #parameters in Fully Connected layer 2
    A. Input to the FC2 $= 1024$
    B. Output neurons of FC2 = 256
    C. Total # parameters in FC1 layer $= 1024 \times 256 + 256 = 262400$

- #parameters in output layer
    A. Input to the output layer $= 256$
    B. Output neurons of FC2 = 10
    C. Total # parameters in FC1 layer $256 \times 10 + 10 = 2570$

The given below list all possible trainable parameters in the given network except output layer.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
conv1 (Conv2D)	(None, 224, 224, 64)	1792
conv2 (Conv2D)	(None, 224, 224, 128)	73856
pool1 (MaxPooling2D)	(None, 112, 112, 128)	0

conv3 (Conv2D)	(None, 112, 112, 128)	147584
conv4 (Conv2D)	(None, 112, 112, 64)	73792
pool2 (MaxPooling2D)	(None, 56, 56, 64)	0
Flatten (Flatten)	(None, 200704)	0
Dense1 (Dense)	(None, 1024)	205521920
Dense2 (Dense)	(None, 256)	262400

==========

Total params: 206,081,344
Trainable params: 206,081,344
Non-trainable params: 0

Q7. Which of the following is true for most CNN architectures?

A. Size of input (height and width) decreases, while depth increases
B. Multiple convolutional layers followed by pooling layers.
C. Fully connected layers in the first few layers
D. Fully connected layers in the last few layers
E. Multiple pool layers followed by a convolutional layer

Ans: A,B,D

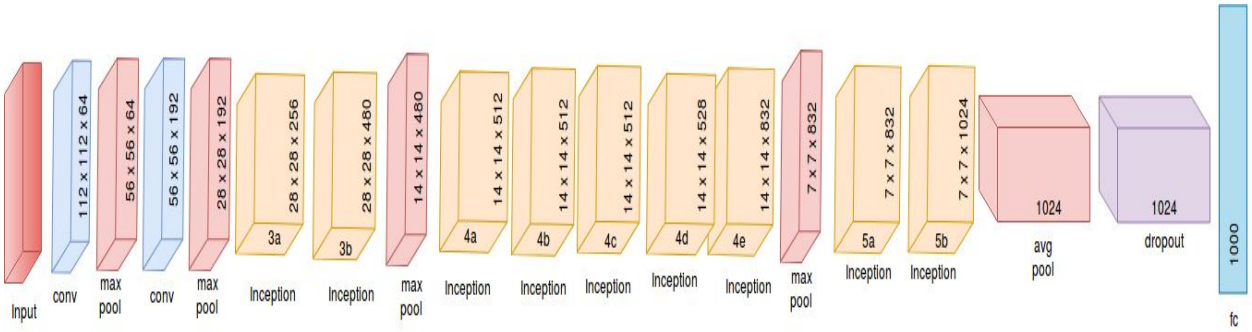
Consider the architecture shown below and answer Questions 8-10.

Type	path size/stride	output size	depth	#1X1	#3X3 reduce	#3X3	#5X5 reduce	#5X5	pool proj
convolution	7X7/2	112X112X64	1						
max pool	3X3/2	56X56X64	0						
convolution	3X3/1	56X56X192	2		64	192			
max pool	3X3/2	28X28X192	0						
(3a)		28X28X256	2	64	96	128	16	32	32

Q8. The network shown is popularly known as:

A. AlexNet
B. VGG
C. GoogLeNet
D. ResNet

Answer: C
Solution:
The given network is GoogleNet.



Q9. What are the number of parameters and number of operations, for layer (3a) in above question?

A. #parameters=163 K (approx.), #operations=128 M (approx.)
B. #parameters=159 K (approx.), #operations=128 M (approx.)
C. #parameters=128 M (approx.), #operations=159 K (approx.)
D. #parameters=128 K (approx.), #operations=159 M (approx.)

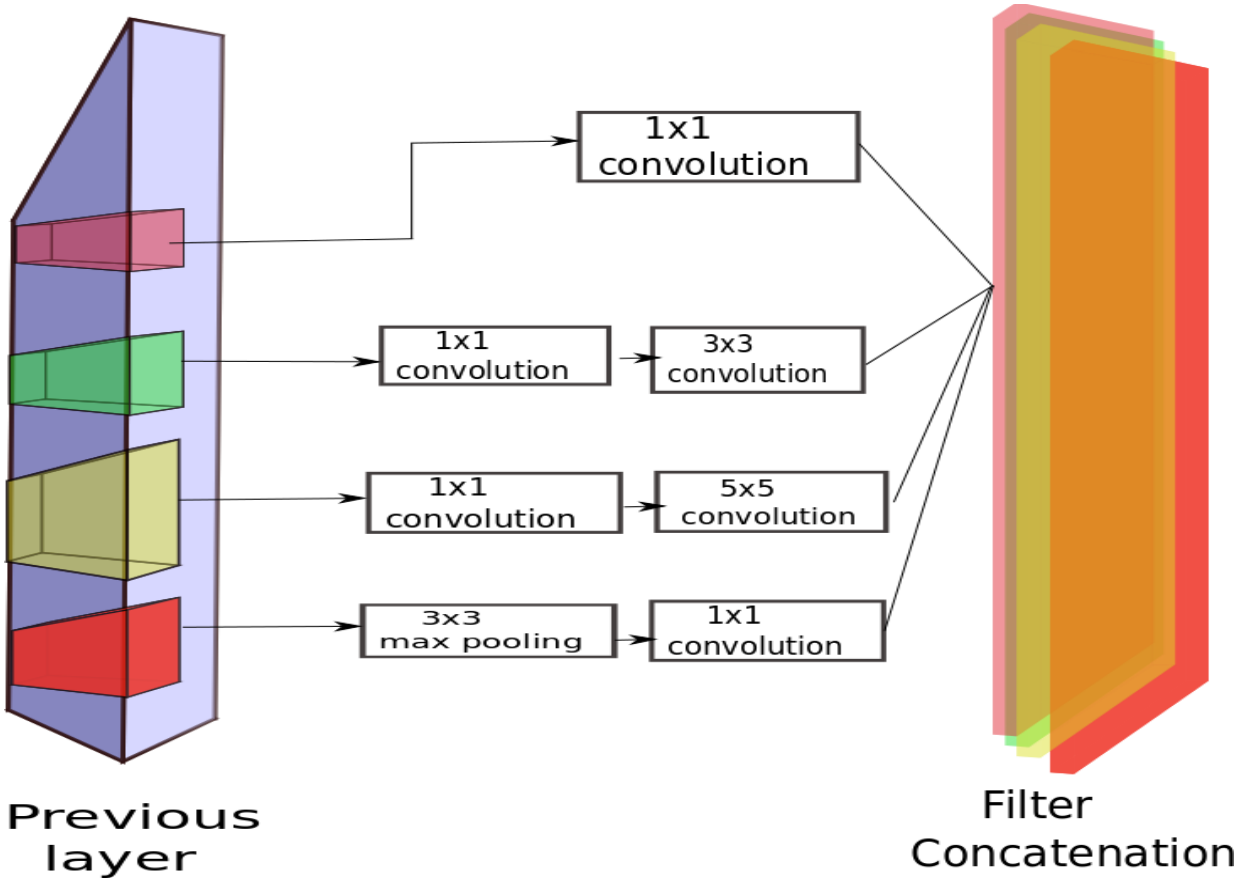
Answer: A

Q10. The importance of “reduce” in the table is that it

A. Reduces no. of feature maps in the previous layer
B. Reduces no. of operations
C. Reduces no. of parameters
D. All of the above

Answer: D.

Solution 9

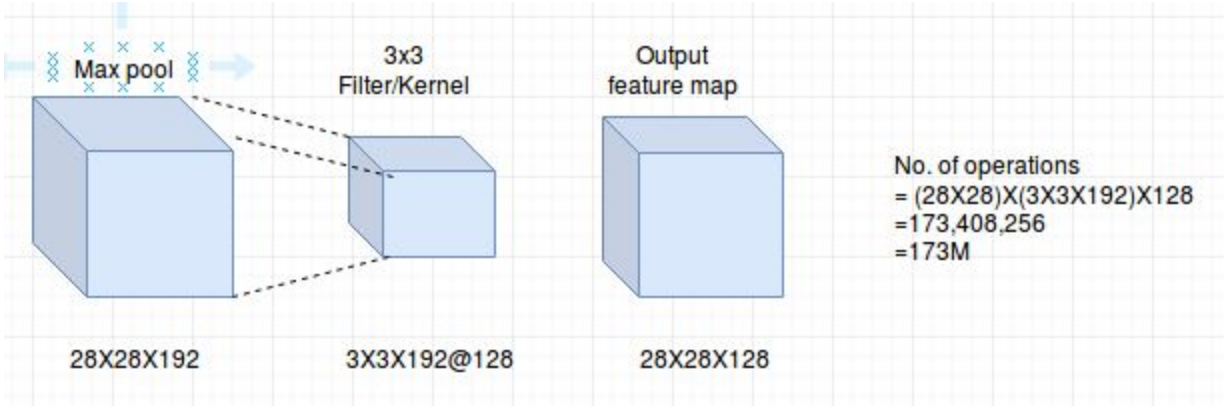


Considering the inception module (layer 3(a)) as shown in the above figure, the number of parameters and no. of operations are shown in the table below:

Kernels/Filters in 3(a)	#1X1	#3X3 reduce	#3X3	#5X5 reduce	#5X5	pool proj	Total
No.of filters (n)	64	96	128	16	32	32	
Input size (heightXwidthXdepth)	28X28X192	28X28X192	28X28X96	28X28X192	28X28X16	28X28X192	
size of each filter (s)	1X1X192	1X1X192	3X3X96	1X1X192	5X5X16	1X1X192	
No. of parameters (sxn)	(1X1X192)X64	(1X1X192)X96	(3X3X96)X128	(1X1X192)X16	(5X5X16)X32	(1X1X192)X32	
No. of parameters (p=s x n)	12288	18432	110592	3072	12800	6144	**163328**
No. of operations (height x depth x s x n)	28X28X12288	28X28X18432	28X28X110592	28X28X3072	28X28X12800	28X28X6144	
No. of operations	9633792	14450688	86704128	2408448	10035200	4816896	**128049152**

Solution 10: Importance of reduce

When reduce is not used (i.e., 1X1 convolutions) are not used:



After reduce: