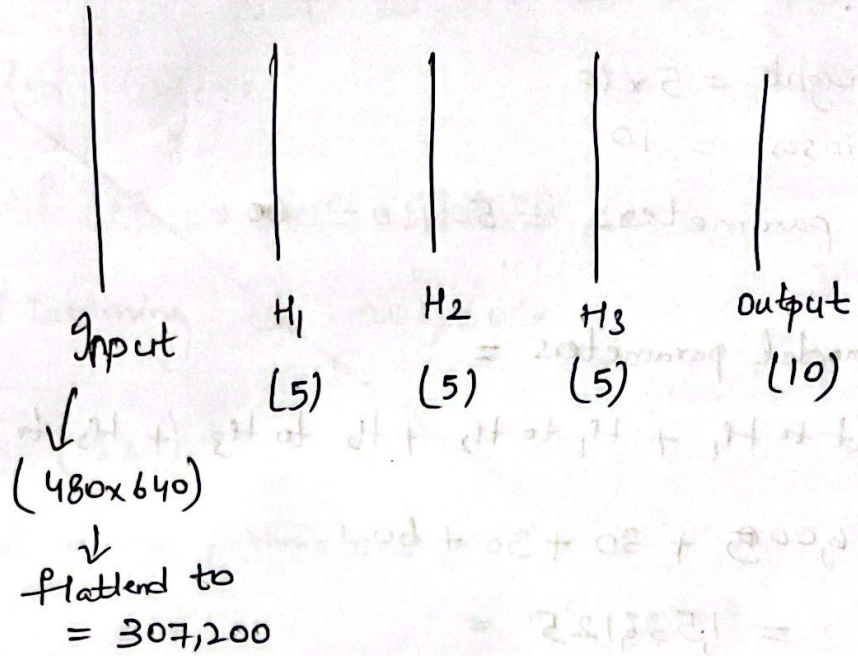
Q1)

a)



Since this is fully connected, no. of parameters between two layers of size (l_1) & (l_2) would be $l_1 \times l_2 + l_2$(bias)

Input to H_1 :

no. of weights = 480x640x5

no. of biases = 5

Total no. of parameters = 1,536,000 + 5

= 1,536,005

H_1 to H_2 :

no. of weights = 5x5

no. of biases = 5

Total no. of parameters = 5x5 + 5 = 30

H_2 to H_3 :

no. of weights = 5x5

no. of biases = 5

Total no. of parameters = 5x5 + 5

= 30

H_3 to output :-

no. of weights $= 5 \times 10$

no. of biases $= 10$

total parameters $= 50 + 10 = 60$

Total no. of model parameters =

Input to H_1 + H_1 to H_2 + H_2 to H_3 + H_3 to output

$1,536,005 + 30 + 30 + 60$

$= 1,536,125$

Input to H_1 :-

b) Input size : $1 \times 480 \times 640$

H_1 : 4 (3×3) filters

No. of model parameters $= 4(3 \times 3) + 4$

$= 40$

output shape of $H_1 = 4 \times \left(\frac{480-3}{1} + 1\right) \times \left(\frac{640-3}{1} + 1\right)$

$= 4 \times 478 \times 638$

H_1 to H_2 :

Since output shape of H_1 is $4 \times 478 \times 638$ H_2 should be 4 $(4 \times 3 \times 3)$ filters

No. of model parameters $= 4(4 \times 3 \times 3) + 4$

$= 148$

Output shape of $H_2 = 4 \times \left(\frac{478-3}{1} + 1\right) \times \left(\frac{638-3}{1} + 1\right)$

$= 4 \times 476 \times 636$

Flatten H_2 so as to connect to H_3 which is fully connected layer.

Shapee of H_2 = 4 x 476 x 636

after flattening = 1,210,944

H_3 : (5 x 1) filters

No. of model parameters = 1,210,944 x 5 + 5

= 6,054,725

H_3 to output :-

H_3 : ~~6,054,724~~ 5

output : 10

No. of model parameters = 5 x 10 + 10

= 60

Total no. of model parameters

= Input to H_1 + H_1 to H_2 + H_2 to H_3 + H_3 to outpr

= 40 + 148 + 6,054,725 + 60

= 6,054,973

22)

a) Input

4	5	3	2
6	3	0	1
8	4	1	2
9	6	2	3

filter

0	0	0
0.5	1.5	0
0	0.5	0

1x1 stride and no padding

size of output is gonna be $\frac{4-3}{1}+1, \frac{4-3}{1}+1$

$= (2, 2)$

Let A be the output

$A[0,0] = 4(0)+5(0)+3(0)$
$+$
$6(0.5)+3(1.5)+0(0) = 9.5$
$+$
$8(0)+4(0.5)+1(0)$

$A[0,1] = 5(0)+3(0)+2(0)$
$+$
$3(0.5)+0(1.5)+1(0) = 2$
$+$
$4(0)+1(0.5)+2(0)$

$A[1,0] = 6(0)+3(0)+0(0)$
$+$
$8(0.5)+4(1.5)+1(0) = 13$
$+$
$9(0)+6(0.5)+2(0)$

$A[1,1] = 3(0)+0(0)+1(0)$
$+$
$4(0.5)+1(1.5)+2(0) = 4.5$
$+$
$6(0)+2(0.5)+3(0)$

Output =

9.5	2
13	4.5

b) Zero padding layer & 3×3 stride

then input becomes

Input

0	0	0	0	0	0
0	4	5	3	2	0
0	6	3	0	1	0
0	8	4	1	2	0
0	9	6	2	3	0
0	0	0	0	0	0

filter

0	0	0
0.5	1.5	0
0	0.5	0

size of output is gonna be $\frac{4-3+2(1)}{3}+1$, $\frac{4-3+2(1)}{3}+1$

$= (2,2)$

Let A be the output

$A[0,0] = 0(0)+0(0)+0(0)$
$+$
$0(0.5)+4(1.5)+5(0)$
$+$
$0(0)+6(0.5)+3(0) = 9$

$A[0,1] = 0(0)+0(0)+0(0)$
$+$
$3(0.5)+2(1.5)+0(0)$
$+$
$0(0)+1(0.5)+0 = 5$

$A[1,0] = 0(0)+8(0)+4(0)$
$+$
$0(0.5)+9(1.5)+6(0)$
$+$
$0(0)+0(0.5)+0(0) = 13.5$

$$A[1,1] = 1(0)+2(0)+0(0)$$
$$+$$
$$2(0.5)+3(1.5)+0(0) = 5.5$$
$$+$$
$$0(0)+0(0.5)+0(0)$$

Output

9	5
13.5	5.5

Q3)

Image classification: It is a task wherein we want to assign a label or tag to an entire image from a given set of classes.

There are 3 types of Image classification:

- Binary classification : It is a special type of classification task where we would have only 2 different classes; the output or predicted class can be either of the two classes.
  Ex: Predict if the tumor is benign or malignant.
- Multiclass classification : Unlike binary classification, it categorizes items into multiple classes.
  Ex: Predict the breed of a dog from a predefined set of breeds.
- Multilabel classification : Unlike multiclass classification where a single label is assigned to an item, in multilabel classification multiple labels can be assigned to a single item.
  Ex: Predict all predefined animals present in a single image.

Object detection: Object detection is a task where one would like to identify/classify the objects present in the class and also localize them using a bounding box. It can be broken down into three steps of

1) To identify if an object is present in the item,
2) Localize the object
3) Predict which category the object belongs to

Ex: If an image consists of two dogs and an object detection algorithm would output a dog is present at top-left corner and another dog is present at the center of the image.

Semantic Segmentation: The goal of semantic image segmentation is to label each pixel of an image with a corresponding class from given predefined classes. We generally consider all other pixels apart from objects of predefined classes as background pixels. Generally uses transposed convolutional layers to retain input image size.

Ex: Semantic segmentation of cars for autonomous vehicles

Similarities:

All the above tasks use complex convolution neural networks to make predictions and all of them involve extracting features from the image. However, object detection and semantic segmentation understand and interpret the image more accurately than image classification because its trying to understand the difference between background and foreground as well

Difference:

Image classification	Object detection	Semantic segmentation
Outputs class of an image	Outputs the class of items of interest present in the image and corresponding bounding boxes location	Outputs class for each and every pixel present in the image.
Output is a vector of probabilities of length equal to number of predefined classes	Outputs multiple vectors of probabilities of length equal to number of predefined classes + 4 elements of (bounding boxes)	Output is a 3d matrix of probabilities of no.of classes * input width * input height