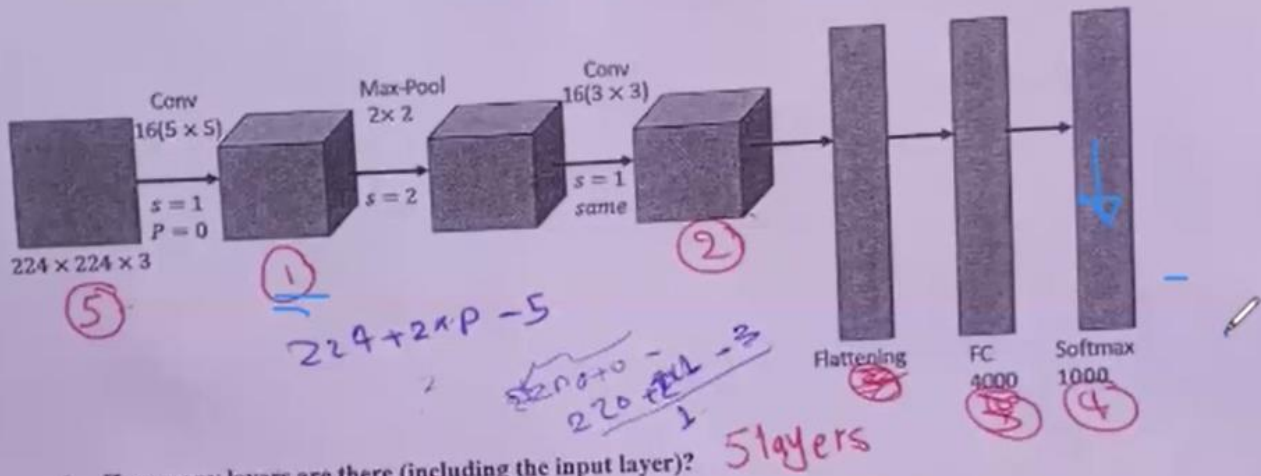


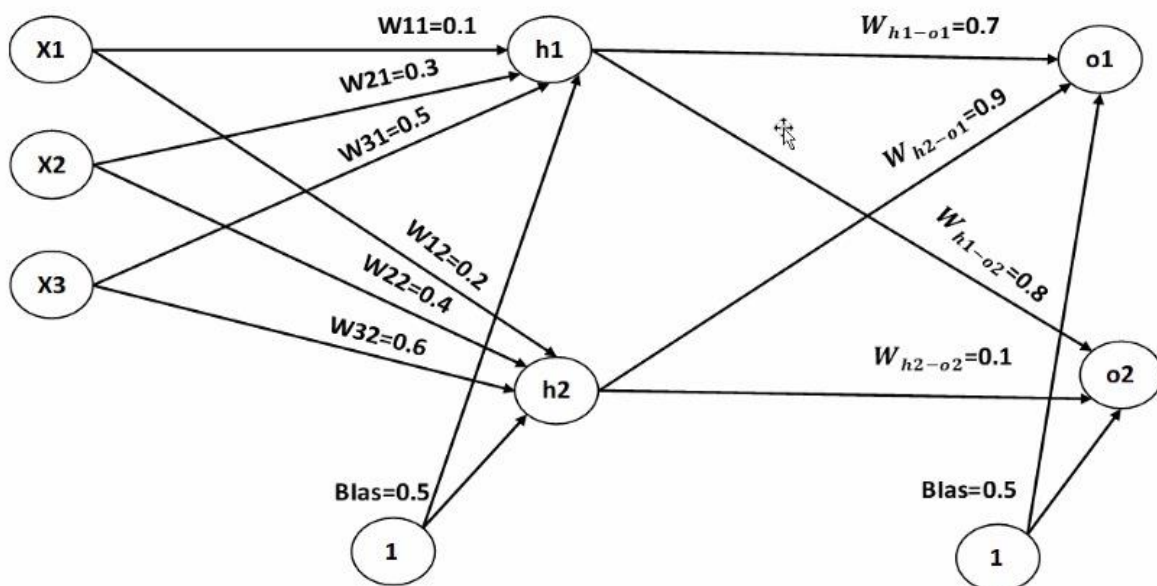
Multimedia mining questions:



1. How many layers are there (including the input layer)?
2. How many classes are there? 1000
3. Fill-in the following table:

Layer	Size of Output	Number of Parameters
Input	224x224x3	0
Conv1	220x220x16	$(5 \times 5 \times 3 + 1) \times 16 = 1216$
Pool1	110x110x16	0
Conv2	110x110x16	$(3 \times 3 \times 16 + 1) \times 16 = 2320$
Flattening	110x110x16 = 193600	0
FC	4000	$(193600 \times 4000) + 4000 = 774400000$
Softmax	1000x1	$(4000 \times 1000) + 1000 = 4001000$

Quiz1 Model A



- 1- What is the number of classes in the above graph? 2
- 2- Use Activation Function Relu \rightarrow to Find h_1 and h_2
- 3- Given $X = [1.5, 1]$ what is the recommended class? And Why?

h

Image Matrix					Filter Matrix		
1	0	0	1	1	1	0	0
1	0	1	1	1	1	0	0
0	1	1	0	1	0	1	1
0	1	0	1	0			
1	1	1	0	1			

Figure 3

- Write the output image after applying Right filter?
- What is the size of the output image after applying the same filter with stride 4? 1×1
- Write the image after applying maxpooling with kernel size 2×2 . 3×3 all 1
- If you use three CNN layers with following filters [35,30,100] what the size of feature map resulted from each layer where you use kernel size 5×5 .

questions:

- why we need to do "data preprocessing" and how ? 2 ways Normalization
Augmentation
- how can you solve overfitting problem ? 3 ways = early stop, dropout Regularization
- what the difference between segmentation and detection ?
- what is the difference between ?
 - object classification
 - object localization
 - object detection
- if you have 4 classes what is the shape of the output ? and what is the loss function ?

4 os
Softmax

C-
C-
C-
C-

cross entropy

$$\begin{array}{r}
 P_1 R_1 \\
 \hline
 50
 \end{array}
 \quad
 \begin{array}{r}
 P_1 R_2 \\
 \hline
 0
 \end{array}
 \quad
 \begin{array}{r}
 P_1 R_3 \\
 \hline
 30
 \end{array}$$

Given the following confusion matrix, choose the correct answer:

	P_1	P_2	P_3
Predicted	P_1	P_2	P_3
	50	20	10
	90	0	200
	80	290	130

R_1

R_2

R_3

140

120

240

1. The number of classes included in this confusion matrix is 3.

$$\begin{array}{r}
 80 \\
 \hline
 500
 \end{array}$$

- a. 2
2. The worst classification performance is at class
a. 1
3. The best discrimination between classes is between classes (actual and predicted)
a. 1 and 2
4. The accuracy of the classification is
a. 35 %
5. This dataset is ...
a. Balanced
6. Accuracy can be a sufficient measure for the classification performance for this dataset.
a. True
7. True positive of class 1 is equal to ...
a. 330
8. False positive of class 1 is equal to ...
a. 330
9. False negative of class 1 is equal to ...
a. 330
10. True negative of class 1 is equal to ...
a. 330

c. 9

d. None

c. 3

d. None

c. 2 and 1

d. None

c. 66 %

d. None

c. Sufficient

d. None

c. 50

d. None

c. 50

d. None

c. 50

d. None

c. 50

d. None

c. 50

d. None

c. 50

d. None

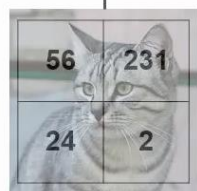
FP → row so 90 + 0

FN → col so 20 + 10

cancel row, col and count remaining

Example with an image with 4 pixels, and 3 classes (cat/dog/ship)

Stretch pixels into column



Input image

0.2	-0.5	0.1	2.0
1.5	1.3	2.1	0.0
0	0.25	0.2	-0.3

W

56
231
24
2

+

1.1
3.2
-1.2

b

=

Cat score

Dog score

Ship score

	very similar dataset	very different dataset
very little data	?	?
quite a lot of data	?	?

Why we need Early Stopping? – To prevent overfitting by tracking model performance on validation set, it stops when the validation error starts increasing

Why we need Drop Out? – To prevent overfitting of the tiny unnecessary tiny details by dropping some neurons of the learning trained data (up to 50% of it)

Activation Functions

Sigmoid:

Symmetric sigmoids are faster:

- Tanh
- Sigmoid

Straightforward sigmoids not good

Rectified Linear Unit (ReLU):

- Much faster computations
- No {Saturation, Vanishing, Exploding}
- Consist of: {Comparison, addition, multiplication}

Why yolo is more efficient than sliding window technique ?

