Q1.

There are two main components in a faster RCNN :

RPN & fast-RCNN network.

1. RPN:

1-1. classifier : The RPN uses 2 classifiers for each anchor box.

One of them determine whether the anchor contains an object, and the other determine if it doesn't contain object (background).

classifier loss : Cross-entropy loss for foreground / background classification.

1-2. Regressors : The RPN uses a pair of regressors for each anchor box.

These regressors adjust the position and size of the anchor box, aligning it more effectively with the ground truth object.

Regression loss : The RPN employs a smooth L2 loss to compute regression loss.

2. Fast R-CNN

2-1. classifier : The fast R-CNN network uses a classifier for each region proposal.

These classifiers predict the class probability for proposed region.

classifier loss : multi-class softmax loss.

2-2. Regressor : The fast R-CNN also employs for refining the bounding box coordinate of the proposed regions.

Regression loss : It uses smooth L2 loss.

# of anchor : $\left(\frac{800}{16}\right) \times \left(\frac{1000}{16}\right) \times 9 \doteq 50 \times 63 \times 9 = 28350$*.

Q2.

For original :

$$\begin{cases} W_e = 480/13 \fallingdotseq 36.92 \ , \ W_h = 480/13 \fallingdotseq 36.92 \\ b_x = \text{sigmoid}(t_x) \times W_c + C_x = \text{sigmoid}(1.6) \times 36.92 + 128 \fallingdotseq 158.72 \\ b_y = \text{sigmoid}(t_y) \times W_h + C_y = \text{sigmoid}(-2.8) \times 36.92 + 192 \fallingdotseq 194.14 \\ b_w = P_w \times e^{t_w} = 46 \times e^{-0.36} \fallingdotseq 32.09 \\ b_h = P_h \times e^{t_h} = 72 \times e^{0.27} = 94.32 \end{cases}$$

For transform to 416 x 416

$$\begin{cases} b_x = 158.72 \times (416/480) \fallingdotseq 137.56 \\ b_y = 194.14 \times (416/480) \fallingdotseq 168.25 \\ b_w = 32.09 \times (416/480) \fallingdotseq 27.81 \\ b_h = 94.32 \times (416/480) \fallingdotseq 81.74 \ \# \end{cases}$$

Q3.

1. Find the smallest value of the row (user minus)

	D_1	D_2	D_3
T1	45	0	75
T2	22	0	33
T3	0	8	16
T4	12	0	9

2. Find the smallest value of the col (user minus)

	D_1	D_2	D_3
T1	45	0	66
T2	22	0	24
T3	0	8	7
T4	12	0	0

Tracklet 3 → detection 1

Tracklet 4 → detection 3

∵ 399<437, ∴ Tracklet 2 → detection 2

Tracklet 1 can't find the corresponding detection since the asigment has excited the FOV.