Imperial College London

60006 Computer Vision (Term 2)

Tutorial 5: Object Detection, Motion

1. We have developed an object detection system and plan to evaluate its performance.

   1.1 In the following figure, three regions of interest (ROIs) are defined, namely A, B and C. The numbers next to the ROI denote its size in pixels. Calculate the aspect ratio (width to height ratio) of the ROIs.

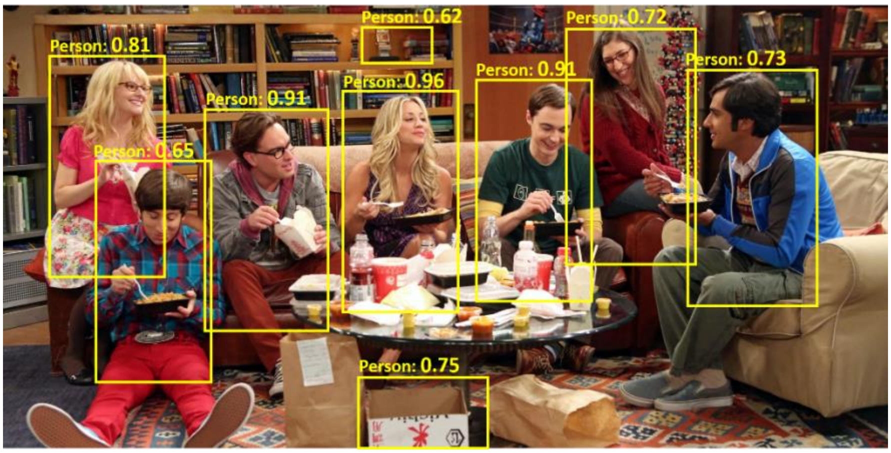
Solution:

Aspect ratios are 1:1 for A, 2:1 for B and 1:2 for C.

   1.2 Precision and recall are two commonly used metrics for evaluation object detection performance. They are defined as,

$$precision = \frac{TP}{TP + FP}$$
$$recall = \frac{TP}{TP + FN}$$

where TP denotes the number of true positives (items correctly detected as belonging to the positive class), FP denotes the number of false positives (items wrongly detected) and FN denotes the number of false negatives (items missed in detection but belonging to the positive class). In the following figure, a detection method produces a number of bounding boxes for the class of person, each associated with a probability score.



Count the number of ground truth persons in the figure. Suppose we use a threshold of 0.9 to perform detection, only keeping bounding boxes with a probability higher than or equal to the threshold. Calculate TP, FP, FN, precision and recall.

Solution:

Number of ground truth persons = 7.

If threshold 0.9 is used, TP = 3, FP = 0, FN = 4.

$$precision = \frac{TP}{TP + FP} = \frac{3}{3 + 0} = 1$$
$$recall = \frac{TP}{TP + FN} = \frac{3}{3 + 4} \approx 0.429$$

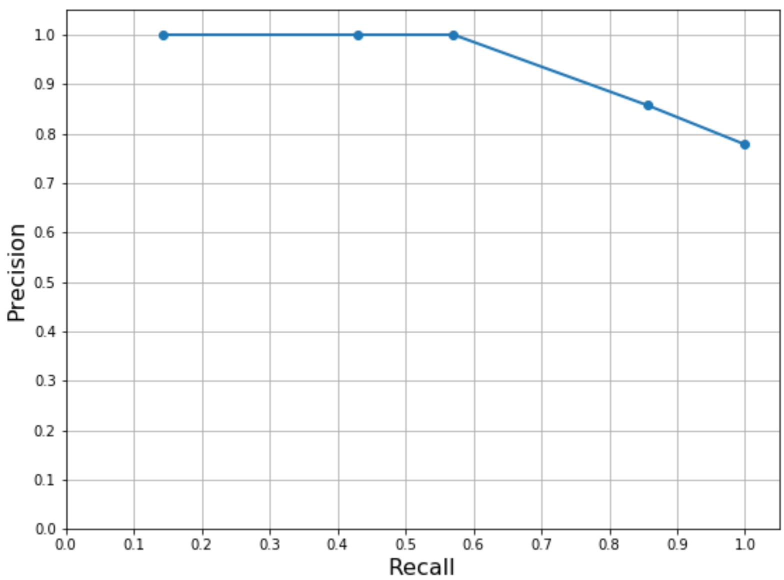
1.3 Calculate TP, FP, FN, precision and recall using thresholds of 0.6, 0.7, 0.8, 0.9 and 0.95. Fill in the following table.

Solution:

Threshold	TP	FP	FN	Precision	Recall
0.6	7	2	0	0.778	1
0.7	6	1	1	0.857	0.857
0.8	4	0	3	1	0.571
0.9	3	0	4	1	0.429
0.95	1	0	6	1	0.143

1.4 Plot the precision-recall curve using the previous table.

Solution:



1.5 As we can see, at different recall values, the precision values are different. We can define an average precision (AP), which samples the precision-recall curve at recall values of 0.2, 0.4, 0.6, 0.8 and 1 and evaluates the average precision values at these points. Estimate the average precision using the precision-recall curve.

Solution:

$$AP = \frac{1}{5}(1 + 1 + 0.98 + 0.88 + 0.78) \approx 0.93$$

2. Given an image sequence, the images at time t and time t+1 are respectively,

0	0	0	0	0	0
0	1	2	3	4	5
0	2	2	3	4	5
0	3	3	3	4	5
0	4	4	4	4	5
0	5	5	5	5	5

time t

0	0	0	0	0	0
0	0	0	0	0	0
0	0	1	2	3	4
0	0	2	2	3	4
0	0	3	3	3	4
0	0	4	4	4	4

time $t + 1$

2.1 Calculate the spatial and temporal image gradients I_x, I_y, I_t at the shaded pixel using the finite differences.

$$I_x = \frac{I(x+1, y, t) - I(x-1, y, t)}{2}$$
$$I_y = \frac{I(x, y+1, t) - I(x, y-1, t)}{2}$$
$$I_t = I(x, y, t+1) - I(x, y, t)$$

Solution:

The spatial and temporal gradients at the shaded pixel are,

$$I_x = 1/2, I_y = 1/2, I_t = -1$$

2.2 Write down the optic flow constraint equation at the shaded pixel.

Solution:

Plug into $I_x u + I_y v + I_t = 0$, we have,

$$u + v = 2$$

2.3 Can you solve the equation? If not, assume that the flow is constant with in 3x3 neighbourhood, add an additional equation at the striped pixel and solve the flow.

Solution:

The gradients at the striped pixel are,

$$I_x = 1, I_y = 0, I_t = -1$$

The optic flow constraint equation is

$$u = 1$$

Combining the two equations, we can solve the local flow,

$$u = 1$$
$$v = 1$$

3. Video denoising is a process to remove noise and other imaging artefacts such as scratches from videos. Unlike single image denoising, where the only information available is in the current picture, video denoising can borrow information from adjacent time frames. In order to do this without introducing blur, video denoising requires accurate pixel-wise motion estimates. Describe your idea of a possible video denoising algorithm. The input is a noisy video of a moving object and the expected output is a clean video.

Solution:

A potential video denoising algorithm can consist of the following steps:

1. Estimate the optic flow between adjacent time frames using the Lucas-Kanade method.

2. For each image to be denoised, warp its adjacent frames to this image according to the optic flow field.

3. Perform denoising by averaging the current frame with the warped adjacent frames.