Report

Proeblem 1:

- 1.How different scales specialize for different object sizes

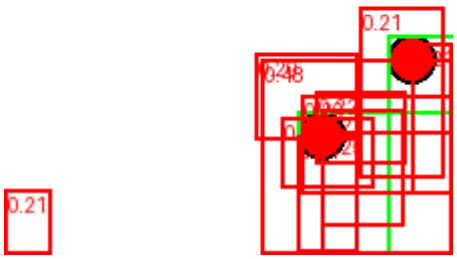
The detector uses three heads:56*56,28*28,14*14. For 56*56, it specialized on small objects. 28*28 specialized on medium objects. 14*14 specialized on large objects.

- 2.The effect of anchor scales on detection performance

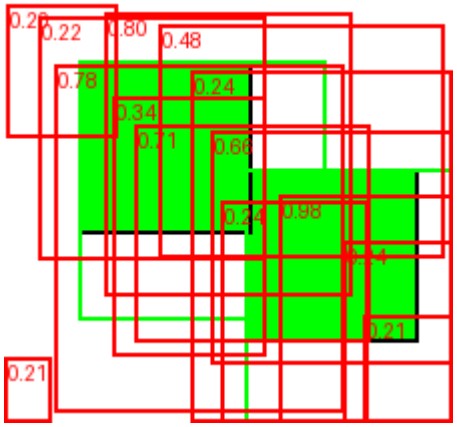
The result shows mAp@0.5 = 0.219 and AP per class: class 0 = 0.006, class1 = 0.104, class2 = 0.548. The strong AP for one class indicates its typical sizes are well-matched to our anchor sets and receptive fields. The very low AP for another class suggests mismatch at the small end.

- 3.Visualization of the learned features at each scale

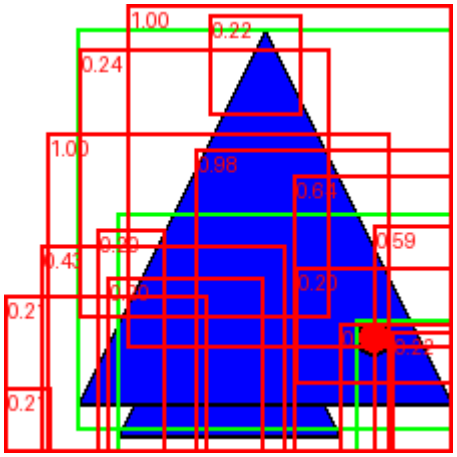
56*56:



28*28:



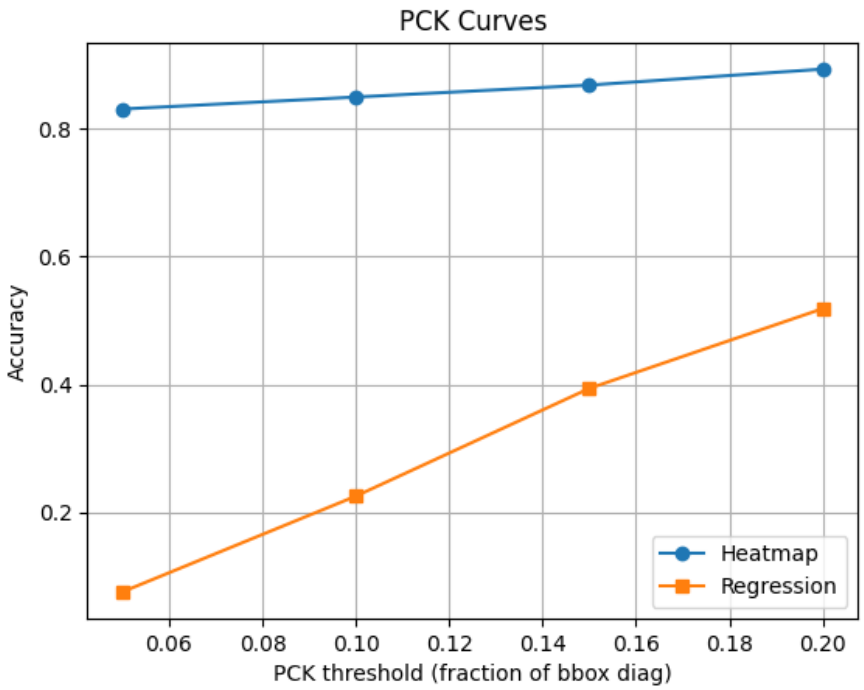
14*14:



So for 56*56, activations align with edges and corners, its helpful for detecting small objects. For 28*28, captures part-level patterns, boxes are steadier on medium objects. For 14*14, emphasizes global shape and context, which stabilizes large-object boxes.

Problem2

- 1. PCK curves at thresholds [0.05, 0.1, 0.15, 0.2]



From the image above we can see that the heatmap model's curve stays above the regression curve across the shown thresholds, higher PCK from 0.05 to 0.20.

- 2.Analysis of why heatmap approach works better (or worse)

Based on the result, heapmap approach works better. As we can see, higher PCK from 0.05 to 0.2, which indicates better localization over a wide tolerance range. Also, the network learns to spread probability near likely positions instead of commiting to a single coordinate early.

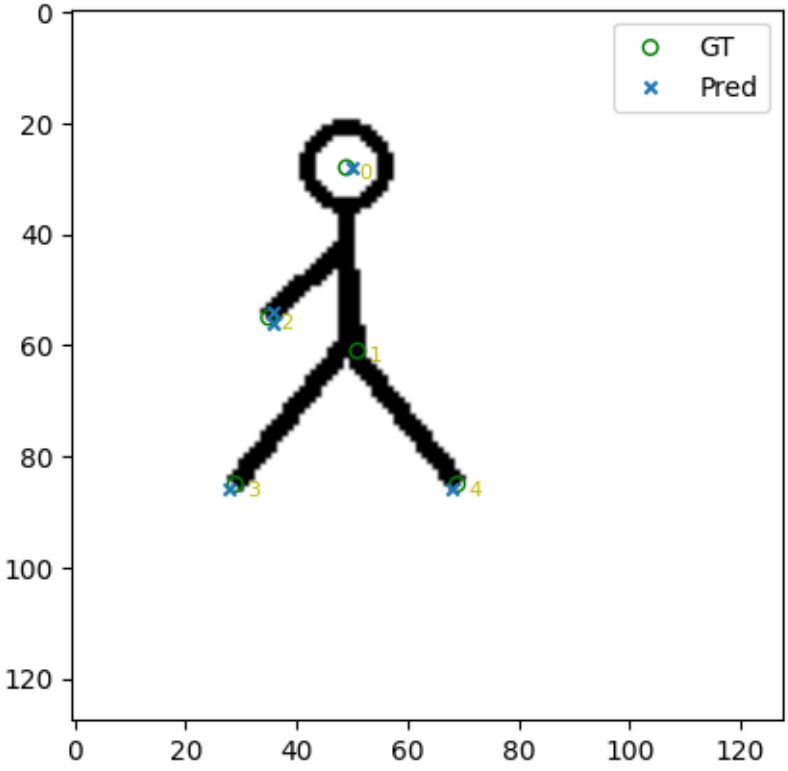
- 3.Ablation study results showing effect of sigma and resolution

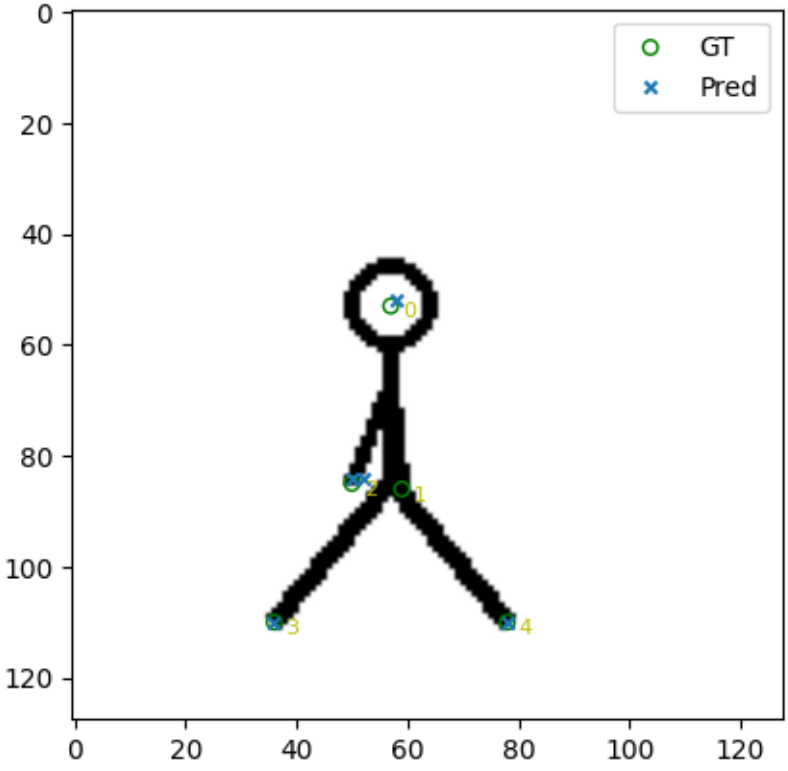
Sigma: If sigma is too small, it will lead to very unstable learning. And if it's too large, it will cause worse precision at 0.05 and 0.1.

Resolution: Higher resolution can give sharper heapmap peaks, so that the keypoint can do it more precisely. Lower resolution can quantize the predictions to a coarser grid.

- 4.Visualization of learned heatmaps and failure cases

Heatmaps:


GT
Pred
0
20
40
60
80
100
120
0
1
2
3
4


GT
Pred
0
20
40
60
80
100
120
0
1
2
3
4

Failure case: