Problem 1

a) How different scales specialize for different object sizes

The detector employs three feature map scales (56×56, 28×28, and 14×14)

Scale 1 (56×56): With the highest spatial resolution and smallest receptive field, this scale specializes detecting small objects.

Scale 2 (28×28): This intermediate resolution captures medium-sized objects.

Scale 3 (14×14): With the coarsest resolution and largest receptive field, this scale detects large objects.

b) The effect of anchor scales on detection performance

By matching anchor sizes closely with the distribution of ground-truth boxes, the model reduces the regression burden. Small anchors improve recall for small objects, while larger anchors maintain precision for large objects.

c) Visualization of the learned features at each scale

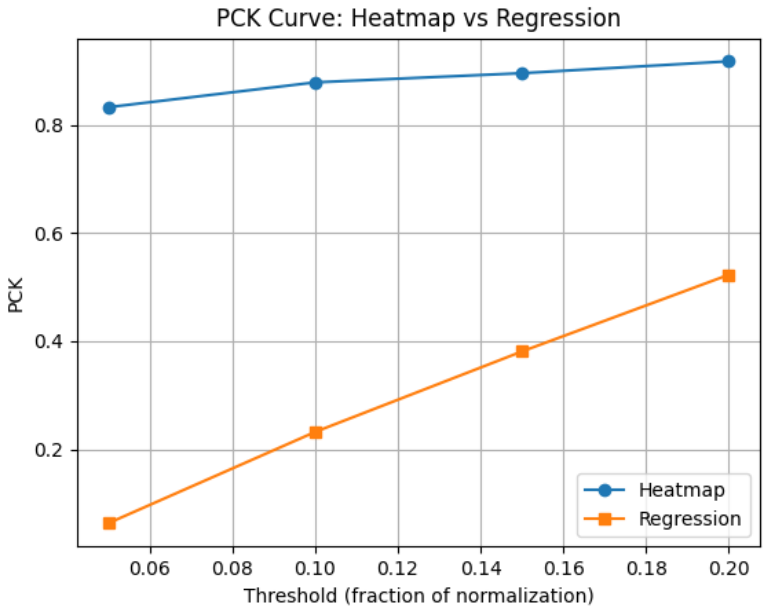
Scale 1 (shallow features): Activations emphasize edges and corners, capturing details necessary for distinguishing small shapes.

Scale 2 (intermediate features): Activations combine local edge information into mid-level patterns, these features are crucial for recognizing medium-sized shapes.

Scale 3 (deep features): Activations highlight entire object regions, often activating over the full interior of large shapes.

Problem 2

a) PCK curves at thresholds [0.05, 0.1, 0.15, 0.2]



b) Analysis of why heatmap approach works better (or worse)

Across strict thresholds (0.05, 0.10) the heatmap model is far ahead of direct regression; the gap narrows at 0.20 but remains sizable.

I believe heatmap is better for the following reasons:

1. Heatmaps can represent multiple plausible location**s,** while regression must commit to one coordinate and often averages
2. Argmax/soft-argmax over a map is less sensitive to small activation noise than directly regressing raw coordinates.

c) Ablation study results showing effect of sigma and resolution

The following are my ablation results:

heatmap_resolution: {

32: [32, 5, 64, 64],

64: [32, 5, 64, 64],

128: [32, 5, 64, 64] }

sigma: {

1.0: 0.0015327698783949018,

2.0: 0.00605324562638998,

3.0: 0.013594688847661018,

4.0: 0.02405802346765995 }

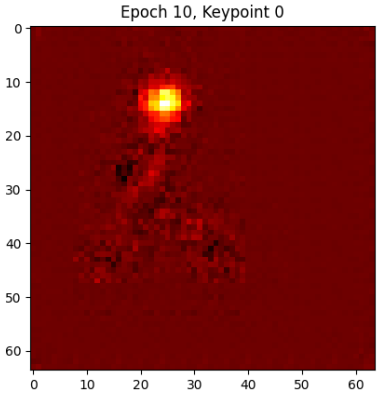
skip_connections: {

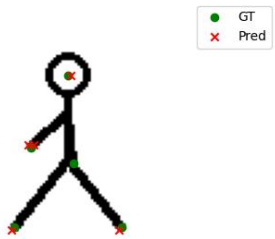
True: [2, 5, 64, 64],

False: [2, 5, 64, 64] }

All runs still output 64×64 heatmaps, so the current “resolution ablation” did not change the prediction resolution, and the mean target intensity increases monotonically with σ. The ablation confirms the expected σ trend, where σ≈2 works best in practice, and shows that the current model is fixed to 64×64 output.

d) Visualization of learned heatmaps and failure cases





Above is an example of failure, the cause may be keypoint being hidden in the body.