EE641 Homework 1

Aditi Surendra Tarate

Problem 1:

1. **How different scales specialize for different object sizes**

From the IoU vs. size curve, we see:

Scale 0 (blue, finest feature map): specializes in small objects (~20–40 px). IoU is high for tiny GT but drops sharply as size increases.

Scale 1 (orange, mid-level feature map): best for medium objects (~50–100 px). Provides good overlap for mid-sized ground truths.

Scale 2 (green, coarsest feature map): dominates for large objects (>100 px). IoU is high for big objects, while performance on small ones is poor.

This shows specialization across scales: each feature map level focuses on a size range.

2. **The effect of anchor scales on detection performance**

The fraction of GTs matched per scale highlights coverage:

Scale 2 covers nearly 50% of GTs, Scale 1 ~45%, Scale 0 ~35%. Anchors at coarser levels are crucial for larger objects.

The histograms (per scale) show anchor side lengths (dashed lines) relative to GT size distribution.

At Scale 0, anchors overlap well with the dense cluster of small GTs. At Scales 1 & 2, anchors are broader, covering the medium and large GT range.

If anchor sizes were poorly chosen (too small or too large), some GTs would fall outside anchor coverage leading to missed detections. So anchor scale choice directly affects how balanced the detector is across small, medium, large objects.

3. **Visualization of the learned features at each scale**

The detections per scale plot shows which scale the detector actually used in practice:

Here, most detections came from Scale 2, it means the model heavily relied on large-object anchors.

Very few detections from Scale 0 or 1 suggests the network struggled on small/medium shapes, despite anchor availability. This aligns with AP@0.5 results (near zero): the detector is failing small/medium cases, only picking up some large ones.

Problem 2:

1. PCK Curves:

PCK curves (at thresholds 0.05, 0.1, 0.15, 0.2) show that:

Regression consistently outperforms heatmap regression at all thresholds in this dataset. For example, at threshold = 0.2, regression reaches ~0.52 while heatmap stays below ~0.3.

2. Analysis of why heatmap approach underperforms:

Normally, heatmap regression works better because it allows spatial uncertainty and smoother gradients. Here, performance is worse likely because:

The dataset is low resolution (224×224) and keypoints are sparse/simple.

Small heatmaps (e.g., 32×32) lose precision when downsampled.

With few training examples, the network struggles to learn sharp Gaussian peaks.

3. Ablation Study

a. Effect of heatmap resolution:

Increasing resolution (32 to 64 to 128) reduces validation loss (best at 128). Higher resolution preserves precision for keypoint localization.

b. Effect of sigma:

Too small σ makes targets too sharp; gradients vanish. Too large σ makes targets too blurry; localization imprecise.

c. Skip connections:

With vs without skip connections improve gradient flow and spatial detail, reducing validation loss.

4. Visualization of heatmaps and failure cases

a. Learned heatmaps:

Some heatmaps peak at correct locations, but others are diffuse or off-center.

b. Failure cases:

Heatmap predicts correct limb but inaccurate localization.

Regression snaps to correct keypoint while heatmap spreads probability mass. Both fail when keypoint is occluded or very close to another.