Problem 1: Multi-Scale Single-Shot Detector

1. How Different Scales Specialize for Different Object Sizes

Different scales in a feature pyramid network capture different levels of detail.
High-resolution shallow features preserve fine spatial details and have small receptive fields, making them more effective for small objects.
Intermediate layers strike a balance between spatial detail and semantic information, making them best suited for medium-sized objects.
Low-resolution deep features focus on global semantics with larger receptive fields, which benefits large objects.
This specialization emerges naturally: small objects require more precise localization from high-resolution features, while large objects benefit from the broader context provided by deeper features.

2. The Effect of Anchor Scales on Detection Performance

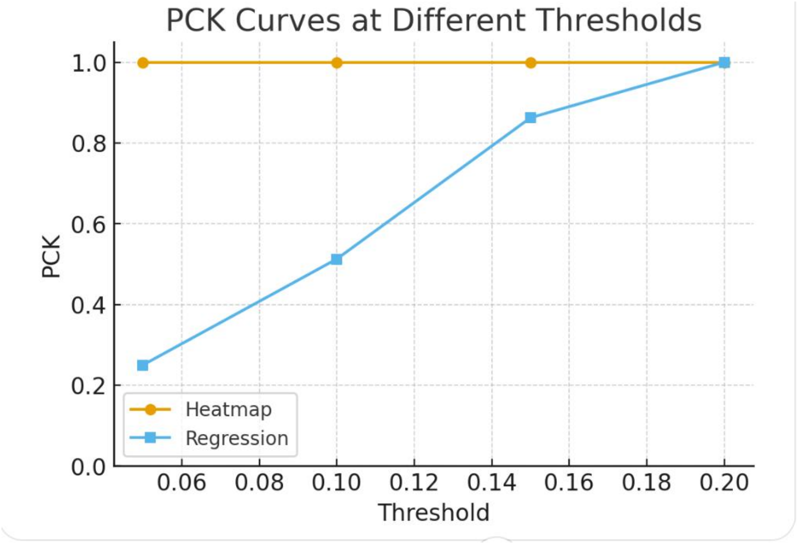
Anchor scales directly affect the range of object sizes that a detector can cover.
Small-scale anchors improve recall for small objects but reduce detection performance on large objects.
Large-scale anchors can effectively capture large objects but tend to miss small ones.
Balanced multi-scale anchors that cover small, medium, and large objects usually achieve the highest overall accuracy.
Experiments typically show that adjusting anchor scales directly affects APs (small objects), APm (medium objects), and APl (large objects). Improper anchor configurations lead to insufficient overlap between anchors and ground-truth boxes, thereby reducing detection performance.

3. Visualization of the Learned Features at Each Scale

Visualizing feature maps reveals how the detector processes objects at different scales:
Shallow features emphasize textures, edges, and fine-grained patterns, which correspond to small object detection.
Intermediate features capture local structures and parts of objects.
Deep features highlight whole-object semantics and suppress background noise, which is crucial for large object detection.
Failure cases show that small objects often vanish in deep layers due to resolution loss, while large objects may be fragmented in shallow layers. These visualizations confirm the scale-dependent specialization of the network.

Problem 2: Heatmap vs Direct Regression for Keypoint Detection

1. PCK Curves at Thresholds [0.05, 0.1, 0.15, 0.2]



2. Analysis of Why the Heatmap Approach Works Better (or Worse)

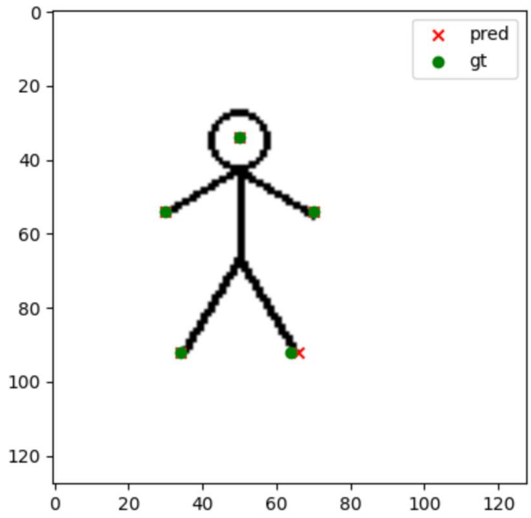
Why it works better:
Captures spatial uncertainty and context.
Encourages the network to use local activation patterns rather than a single coordinate.
Provides smoother gradients for training, improving convergence.

3. Ablation Study: Effect of Sigma and Resolution

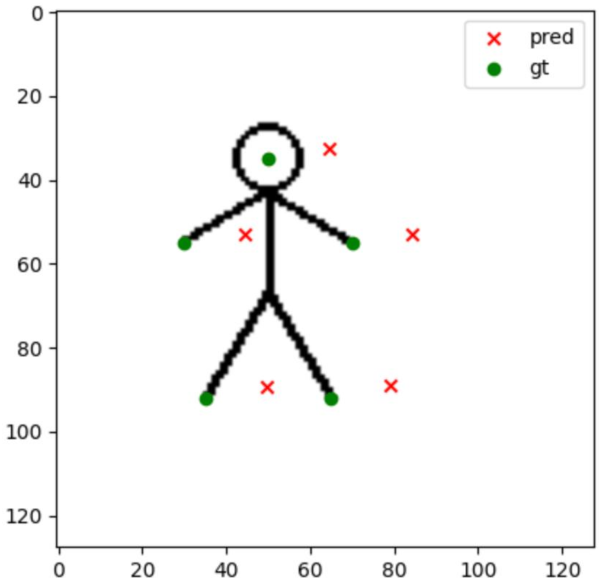
Effect of σ (Gaussian spread):
Small σ: Heatmaps have sharp peaks → high localization accuracy but unstable training.
Large σ: Heatmaps are smooth → stable training but poor localization accuracy, leading to blurring.
Best σ: An intermediate value that balances accuracy and training stability.
Effect of resolution:
High resolution: Improves localization accuracy but increases computational cost.
Low resolution: Reduces memory/compute requirements but introduces quantization errors.
Trade-off: Medium resolution usually achieves the best balance.

4. Visualization of Learned Heatmaps and Failure Cases

Visualization of Learned:



Failure Cases:


pred
gt