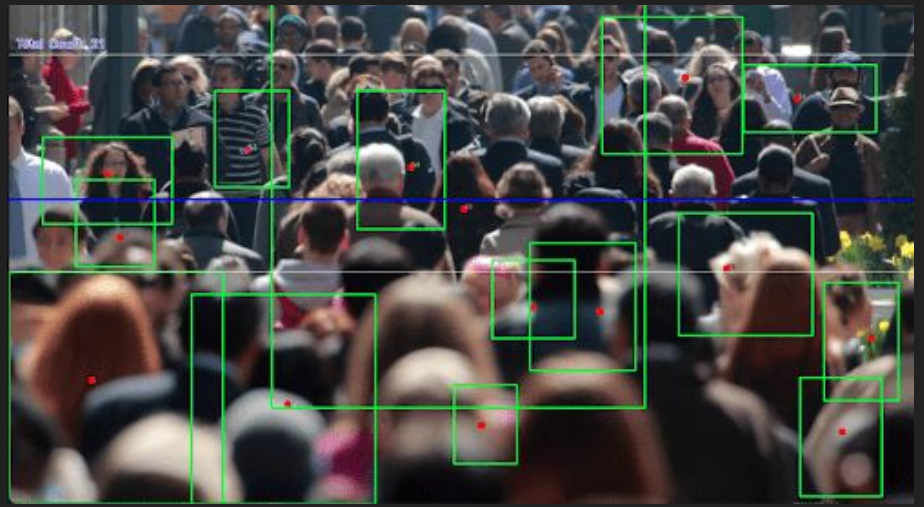
Crowd Searching

Ethan Peterson

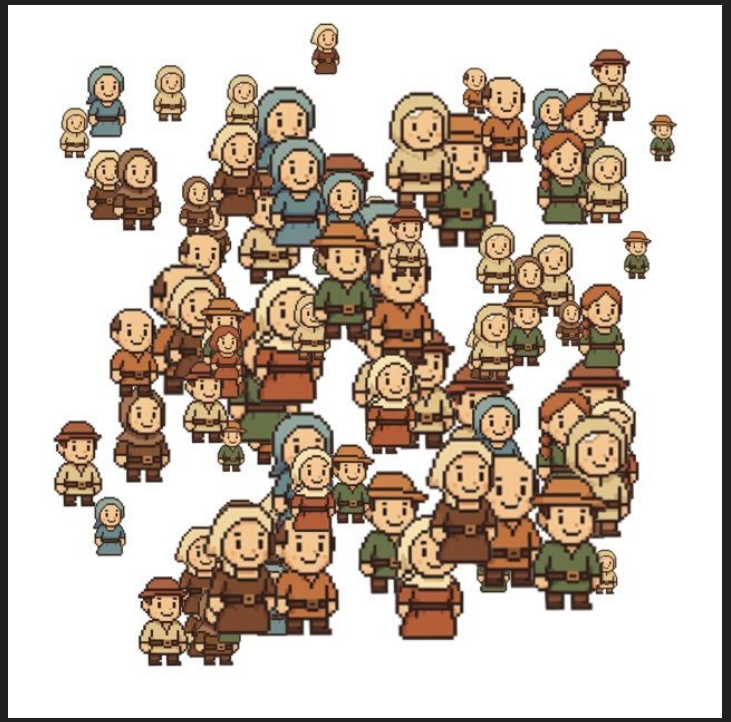
Purpose

The goal of this project was to investigate object detection in dense crowd environments, a setting where model performance typically degrades significantly due to factors such as:

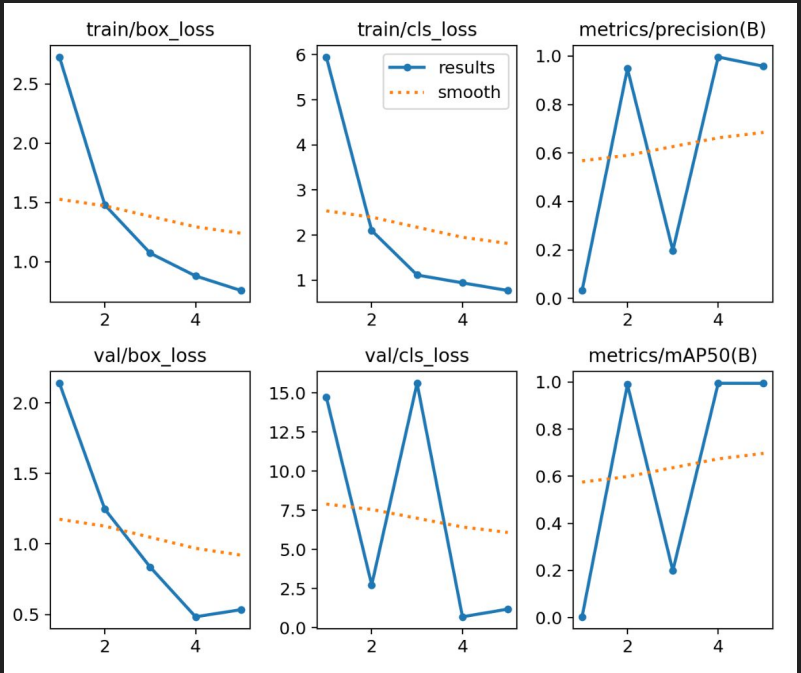
- Severe scaling
- Occlusion
- Noise
- Bulk processing

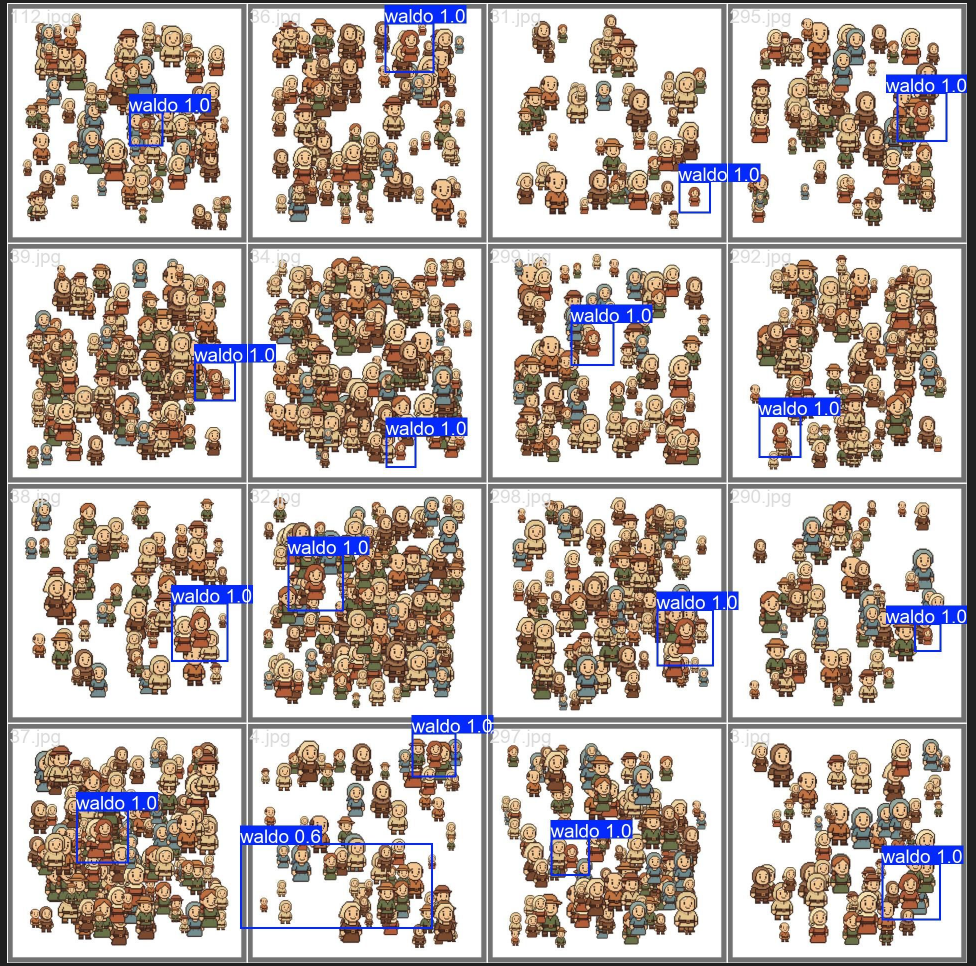


Synthetic Data & Simple Puzzles



Basic Puzzles - 99% mAP



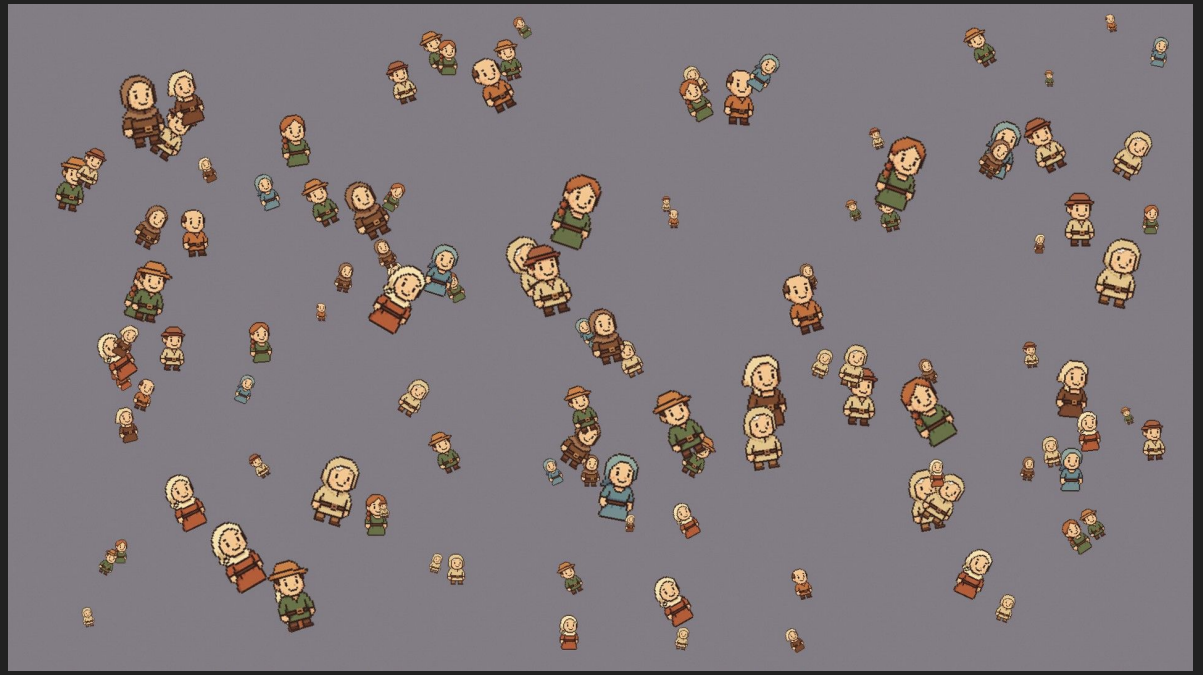


2. Medium Puzzles

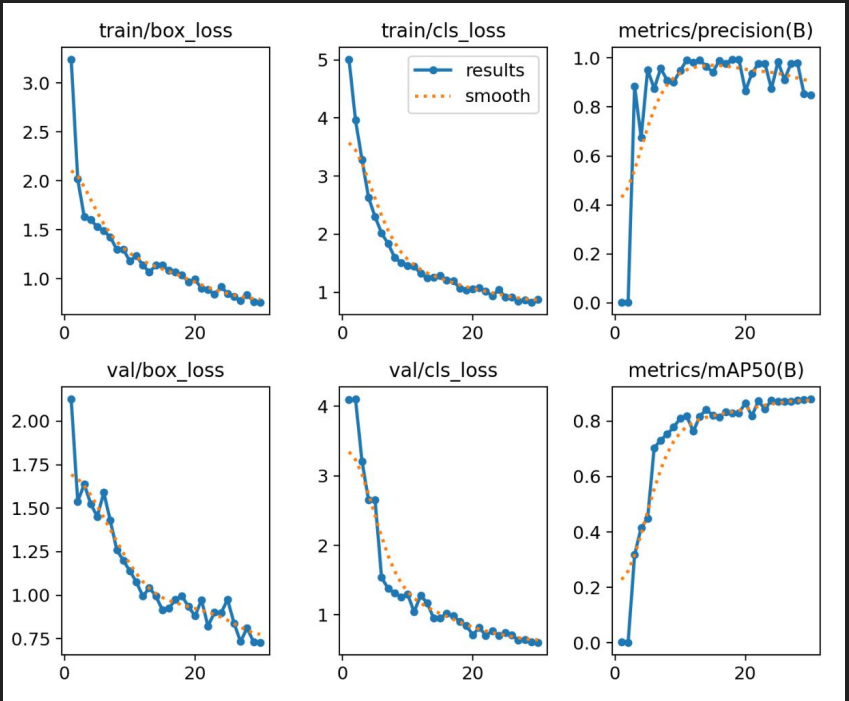
Changes:

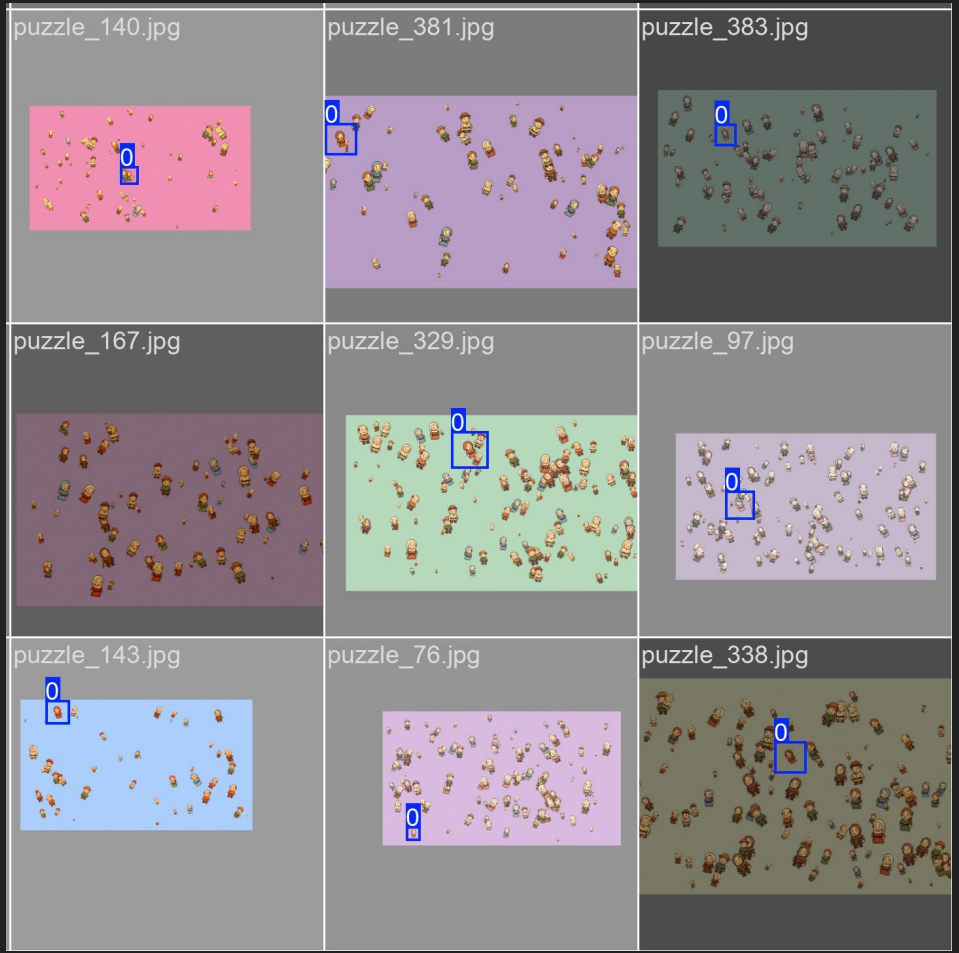
- 30-150% scaling
- -35° to 35° rotations
- Random occlusion
- Random backgrounds
- Waldo can be blocked

Additionally: 400 images & 30 epochs.



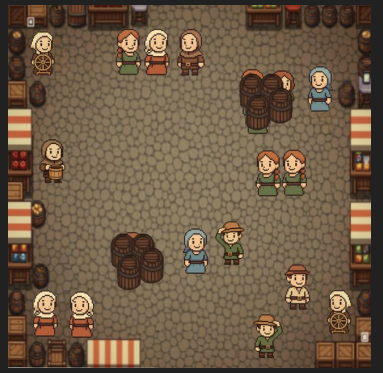
Results - 97% mAP



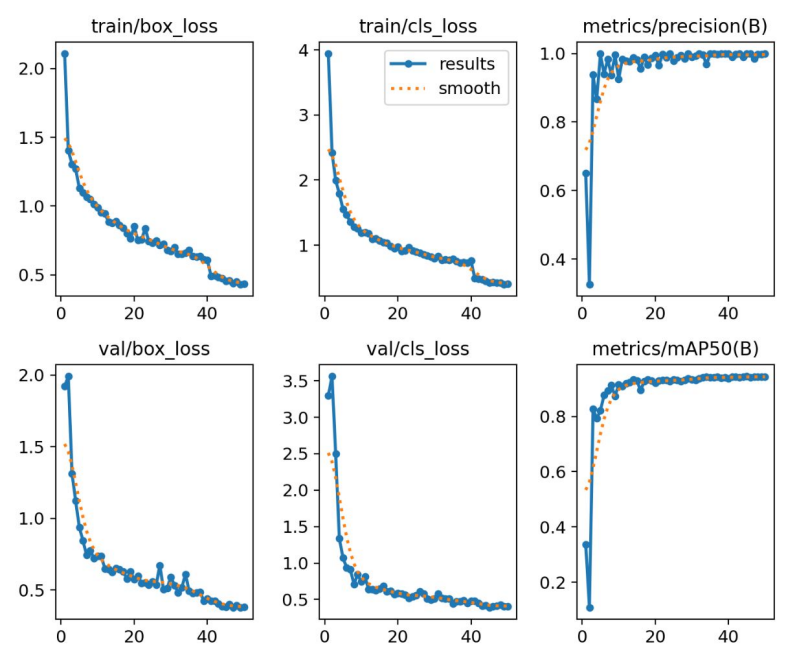


Harder Puzzles

- 1200 training images
- 3 different backgrounds
- Scene layers for villagers to hide behind
- Reflections, blocking, slight scaling and rotation
- Secondary poses for each villager
- 50 epochs

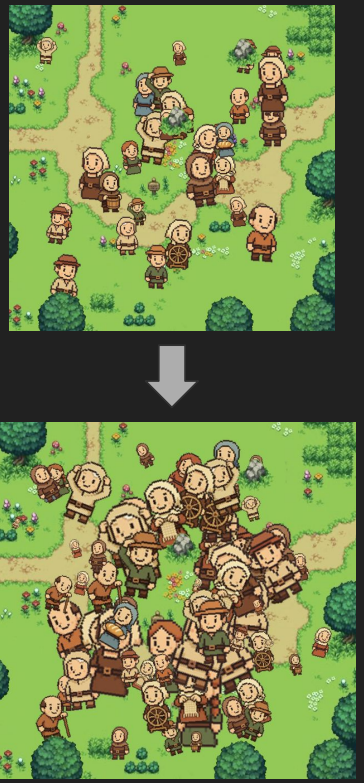


Results -95% mAP



Severe Augmentation + Handmade Validation Tests

Test Set	mAP50	mAP50-95	Precision	Recall
A-baseline	0.96	0.91	0.99	0.93
B-clutter	0.81	0.693	0.99	0.67
C-rotation	0.90	0.79	0.983	0.74
D-second pose	0.96	0.91	0.98	0.93
E-scaling	0.96	0.90	0.98	0.93
F-reflection	0.96	0.87	0.99	0.95
Handmade	0	0	0	0



Let’s Use Real People Now (Research Reimplementation)

“Rethinking Counting and Localization in Crowds: A Purely Point-Based Framework” (2021)

Objective:

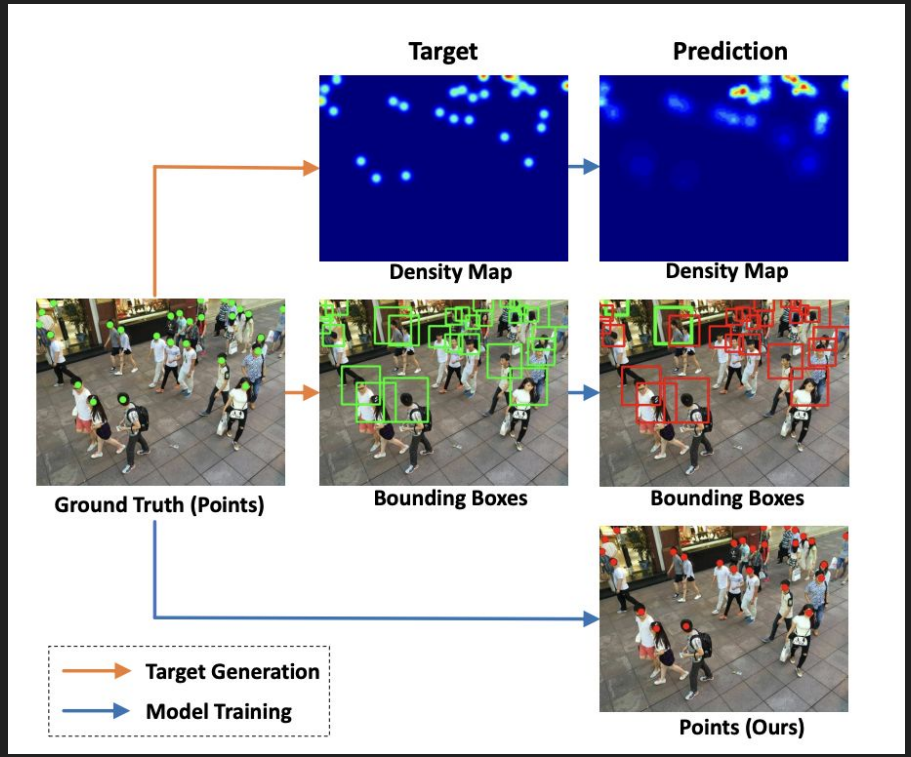
- Understand the research
- Reproduce the results and on a different dataset

P2P Idea & Motivation

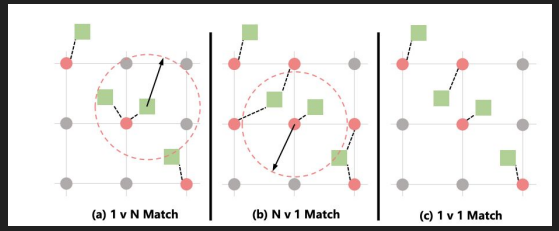
Point predictions >> bounding boxes and density maps.

Benefits:

- More precise performance evaluation
- Less annotation time
- Reduced matching ambiguity



Proposal: P2PNet Framework



Input Image → P2PNet → Point Predictions (x, y coordinates + confidence)

Training

1) Generate Predictions
2) One to One Matching: Cost(i,j) = τ × distance(pred_j, gt_i) - confidence_j
3) Loss Calculation:
4) Back prop.

$$\mathcal{L}_{cls} = -\frac{1}{M}\left\{\sum_{i=1}^{N}\log \hat{c}_{\xi(i)} + \lambda_1 \sum_{i=N+1}^{M}\log\left(1-\hat{c}_{\xi(i)}\right)\right\}, \tag{4}$$

$$\mathcal{L}_{loc} = \frac{1}{N}\sum_{i=1}^{N}\left|\left|p_i - \hat{p}_{\xi(i)}\right|\right|_2^2, \tag{5}$$

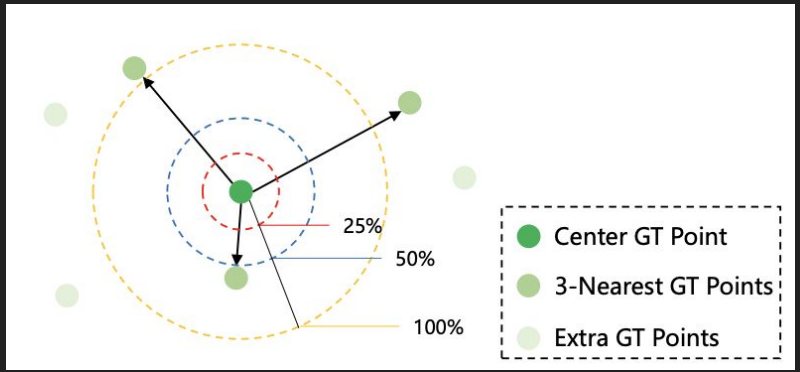
$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_2 \mathcal{L}_{loc}, \tag{6}$$

Evaluation: Density nAP

$$\mathbb{1}(\hat{p}_j, p_i) = \begin{cases} 1, & \text{if } d(\hat{p}_j, p_i)/d_{k\text{NN}}(p_i) < \delta, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

Allows us to create a local thresholding distance to avoid duplication errors

1) Filter out predictions with confidences below some threshold
2) Take predictions and sort by confidence
3) Match ground truths (closest GT to pred)
4) Create PR Curve
5) Calculate nAP



=== Final Evaluation ===
MAE : 57.24175824175824
RMSE: 66.15258021144665

- Close to performance on datasets from the paper! (e.g. ShanghaiTech Part_A)

RESULTS


CENTURYLINK FIELD
SEAHAWKS

References

- Varanasi, K. *Building a Waldo Finder*. 2020.
  Available at: https://keshav.is/building/waldo-finder/

- Lin et al. *Feature Pyramid Networks for Object Detection* (2017).

- Ultralytics. *YOLOv8 Documentation*.

- Song, Qi, et al. "Rethinking Counting and Localization in Crowds:
  A Purely Point-Based Framework." arXiv:2107.12746, 2021.