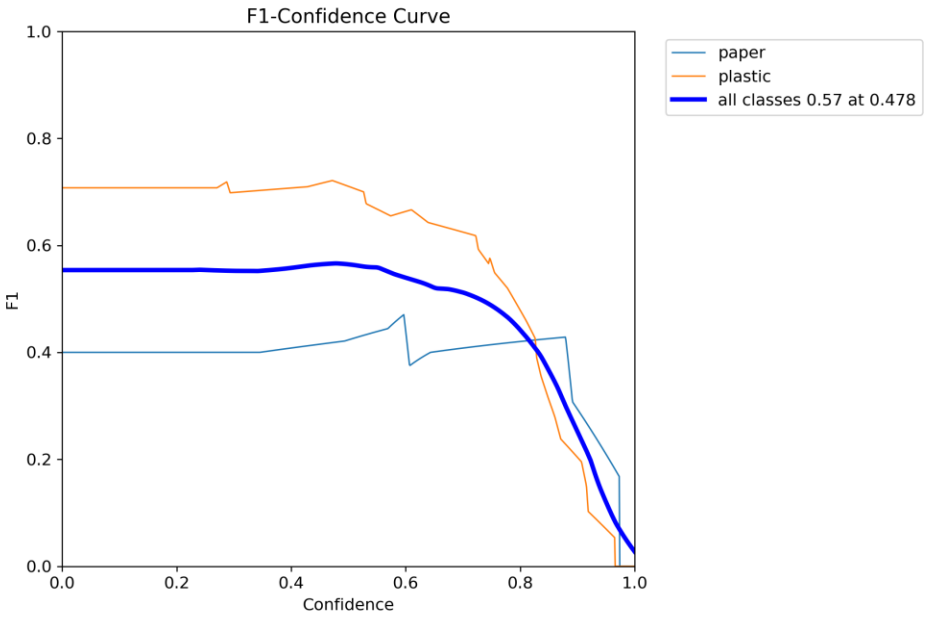
THE METRICS CURVES

F1–Confidence Curve (BoxF1_curve)



What this plot shows

- X-axis: Confidence threshold
- Y-axis: F1 score (harmonic mean of Precision & Recall)
- Curves shown:
    - paper
    - plastic
    - all classes (bold blue)

Key readings

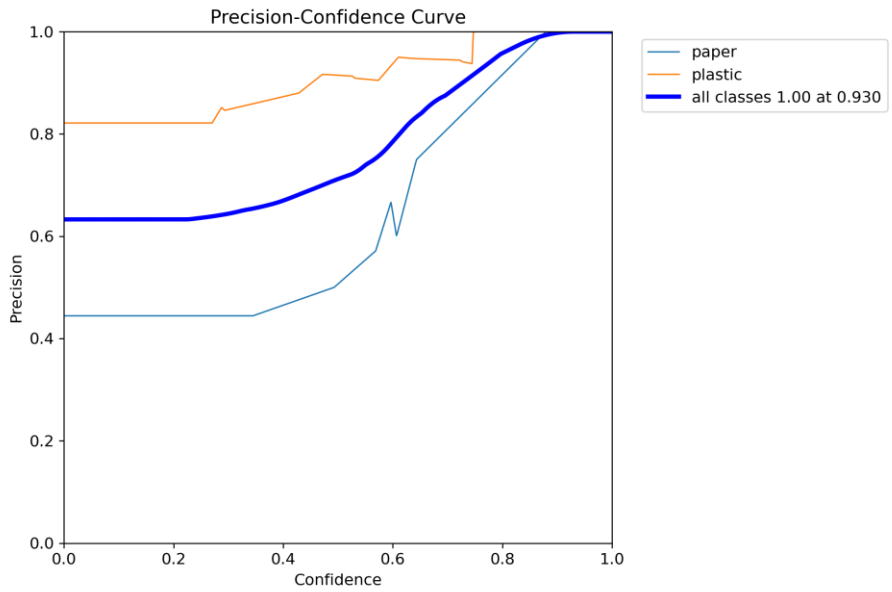
- Best overall F1 ≈ 0.57 at confidence ≈ 0.48
- Plastic consistently has higher F1 than paper
- Paper F1 remains relatively low and unstable

Interpretation

- A confidence threshold around **0.45–0.50** gives the **best balance** between false positives and false negatives.
- Paper class struggles due to:
    - Fewer samples
    - Visual similarity to background

- Plastic objects are easier to detect (distinct shape & texture).

Precision–Confidence Curve (BoxP_curve)



What this plot shows

- X-axis: Confidence threshold
- Y-axis: Precision
- Precision increases as confidence increases

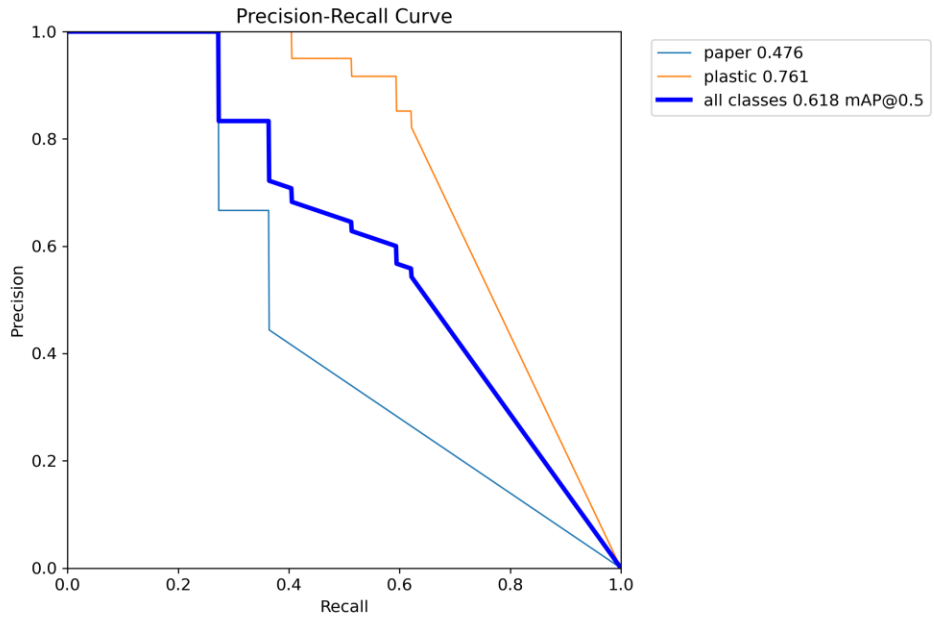
Key readings

- Plastic precision ≈ 0.95–1.00 at high confidence
- Paper precision improves slowly
- Overall precision reaches 1.00 at confidence ≈ 0.93

Interpretation

- At high confidence thresholds:
    - Predictions are very reliable
    - But many objects are missed (low recall)
- Plastic predictions are highly reliable even at moderate thresholds
- Paper predictions need higher confidence to be trustworthy

Precision–Recall Curve (BoxPR_curve)



What this plot shows

- X-axis: Recall
- Y-axis: Precision
- Area under the curve ≈ mAP@0.5

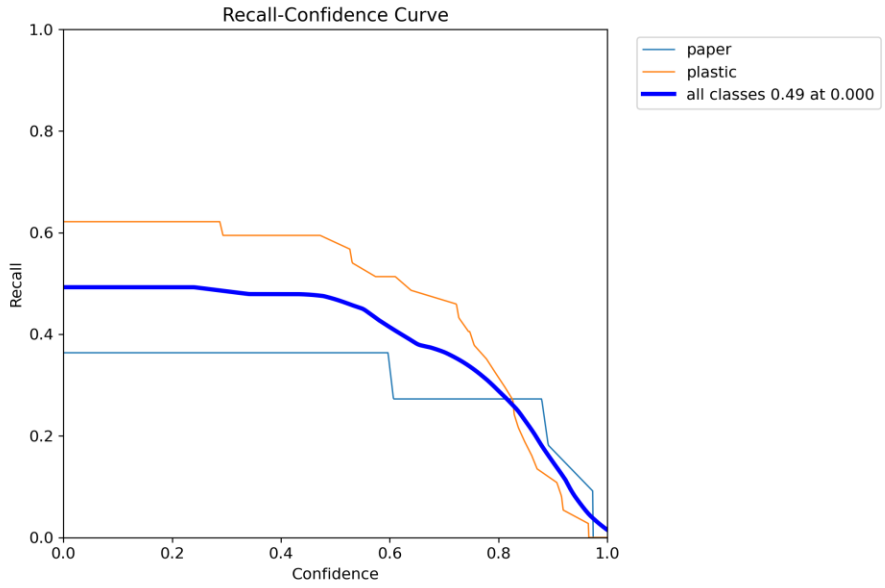
Key readings

- Plastic mAP@0.5 ≈ 0.761
- Paper mAP@0.5 ≈ 0.476
- Overall mAP@0.5 ≈ 0.618

Interpretation

- Plastic has strong performance across recall levels
- Paper performance drops quickly as recall increases
- Confirms class imbalance problem

Recall–Confidence Curve (BoxR_curve)



What this plot shows

- X-axis: Confidence threshold
- Y-axis: Recall
- Recall decreases as confidence increases

Key readings

- Recall starts around:
    - Plastic ≈ 0.62
    - Paper ≈ 0.36
- Recall drops sharply after confidence > 0.7

Interpretation

- High confidence thresholds:
    - Reduce false positives
    - Increase missed detections
- This is why general waste fallback is important