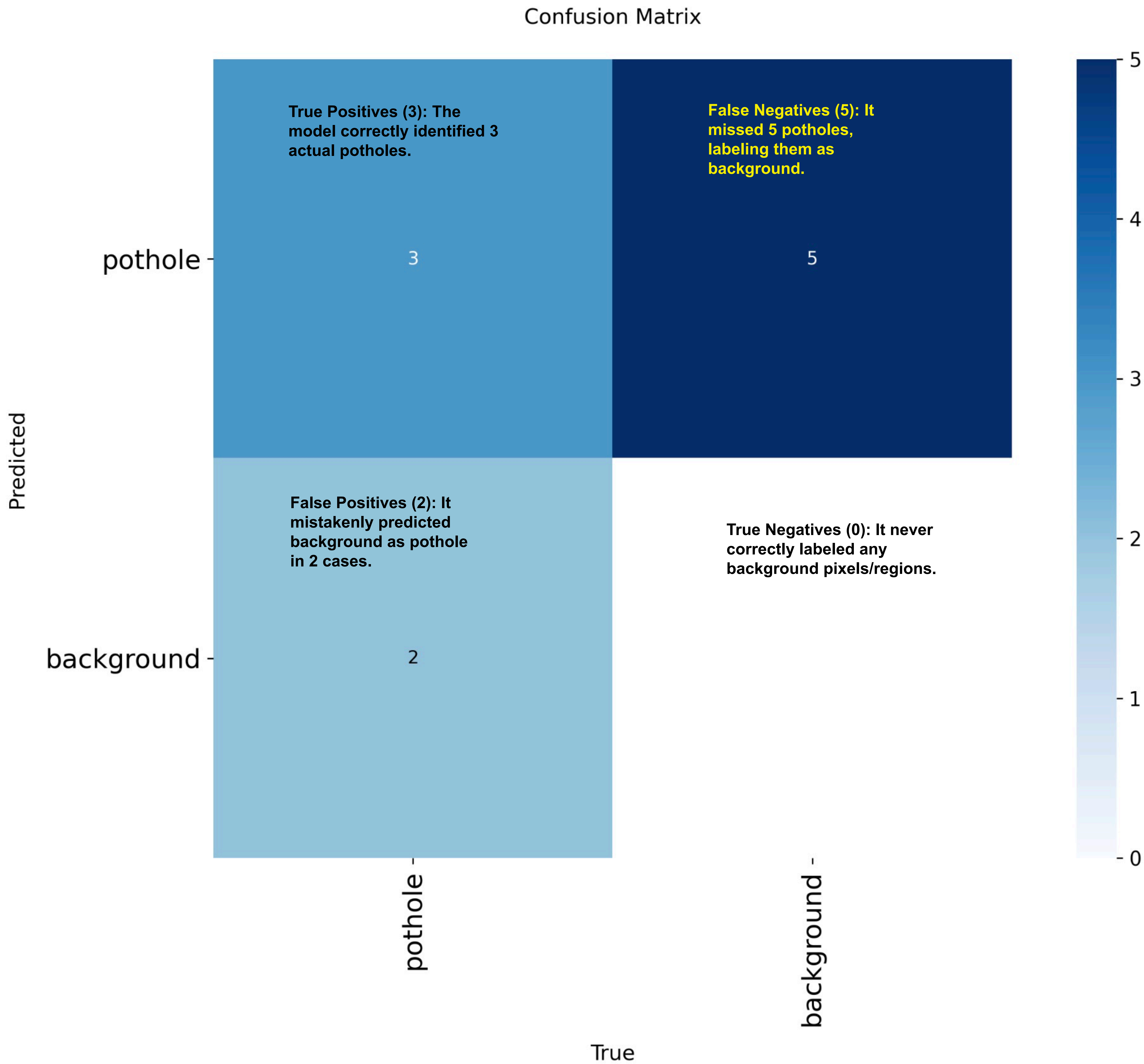
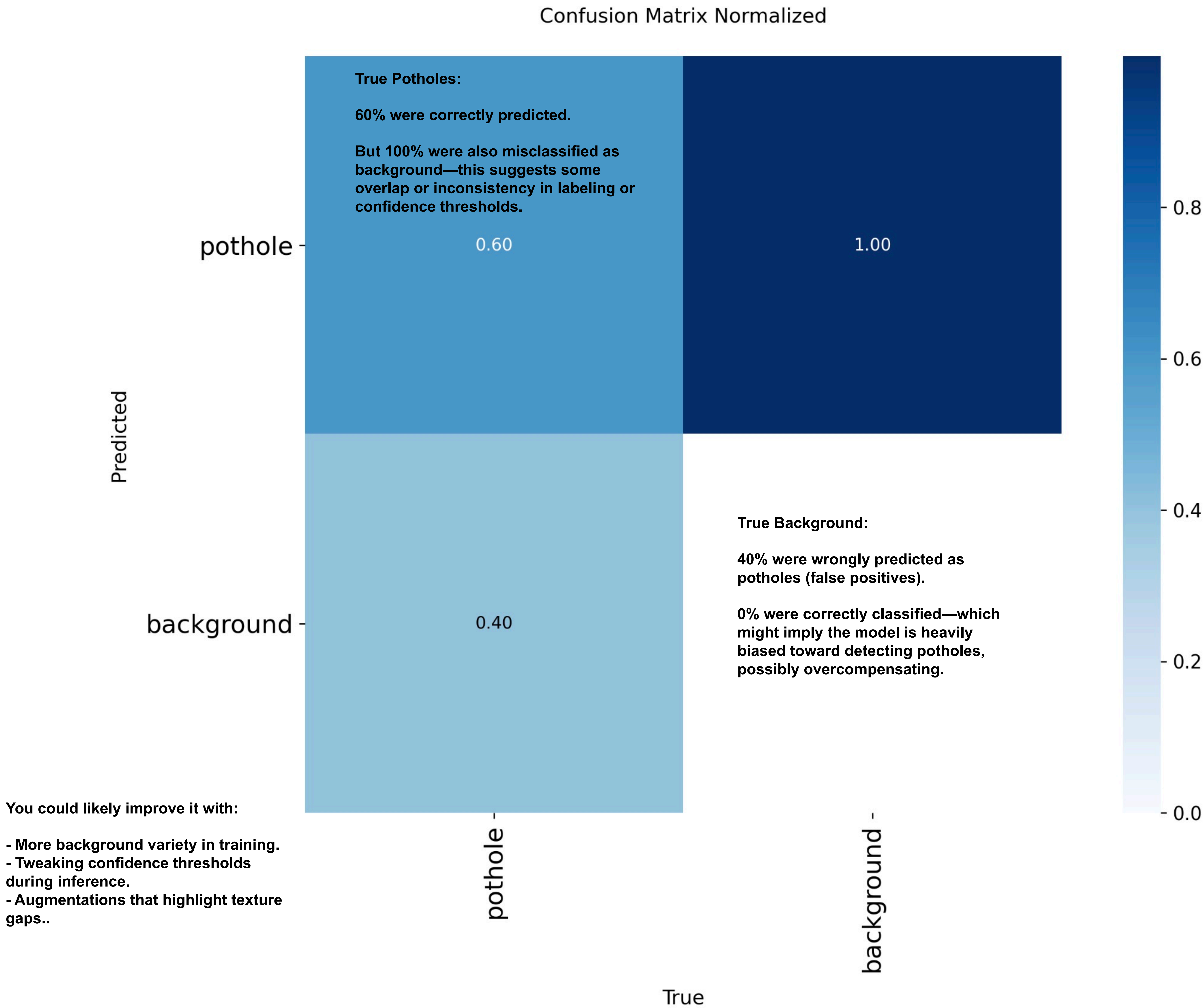
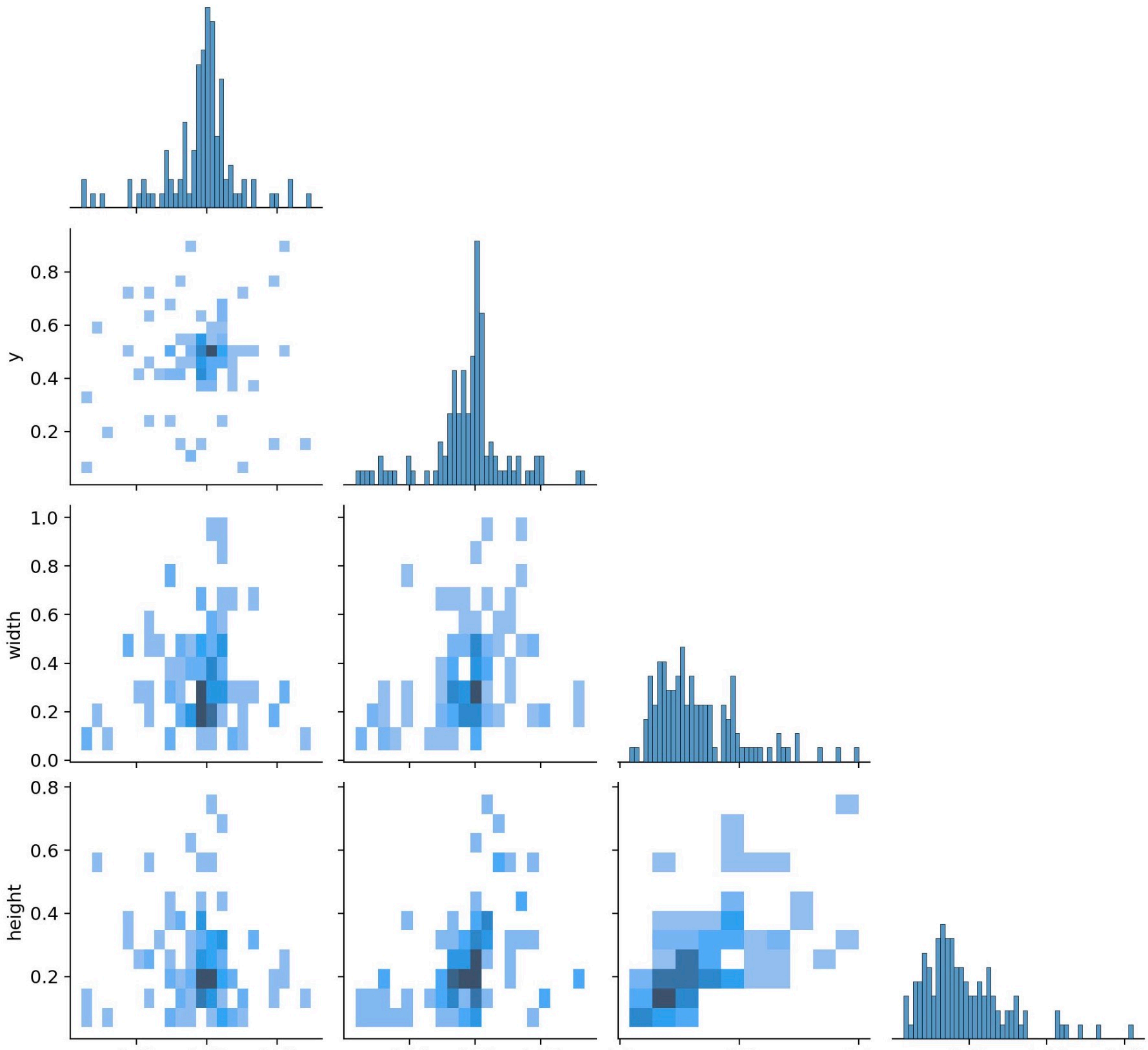
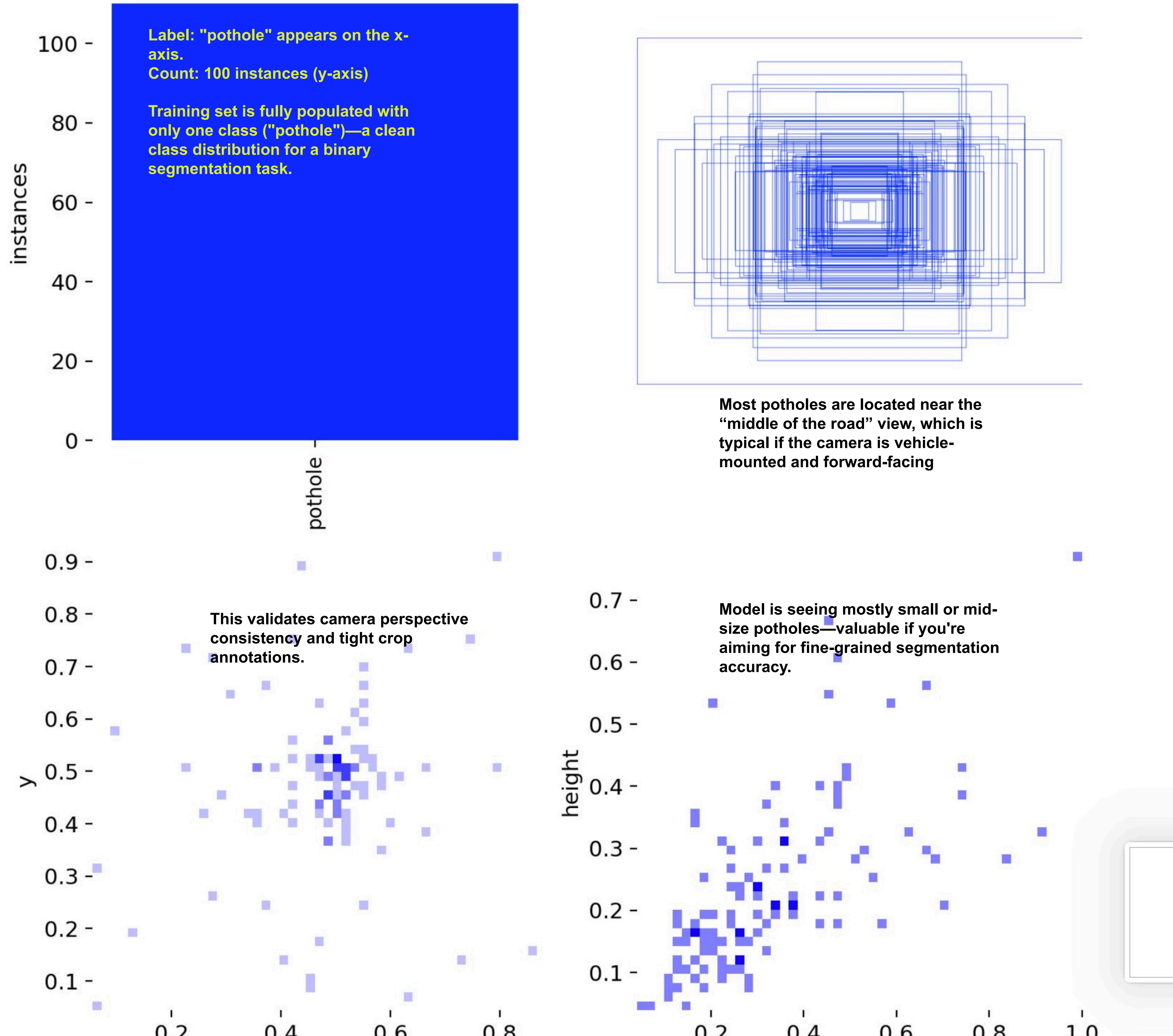

Confusion Matrix
True Positives (3): The model correctly identified 3 actual potholes.
3
False Negatives (5): It missed 5 potholes, labeling them as background.
5
False Positives (2): It mistakenly predicted background as pothole in 2 cases.
2
True Negatives (0): It never correctly labeled any background pixels/regions.
pothole
background
Predicted
pothole
background
True
5
4
3
2
1
0


Confusion Matrix Normalized
True Potholes:
60% were correctly predicted.
But 100% were also misclassified as background—this suggests some overlap or inconsistency in labeling or confidence thresholds.
0.60
1.00
0.40
True Background:
40% were wrongly predicted as potholes (false positives).
0% were correctly classified—which might imply the model is heavily biased toward detecting potholes, possibly overcompensating.
pothole
background
Predicted
True
0.8
0.6
0.4
0.2
0.0
You could likely improve it with:
- More background variety in training.
- Tweaking confidence thresholds during inference.
- Augmentations that highlight texture gaps..



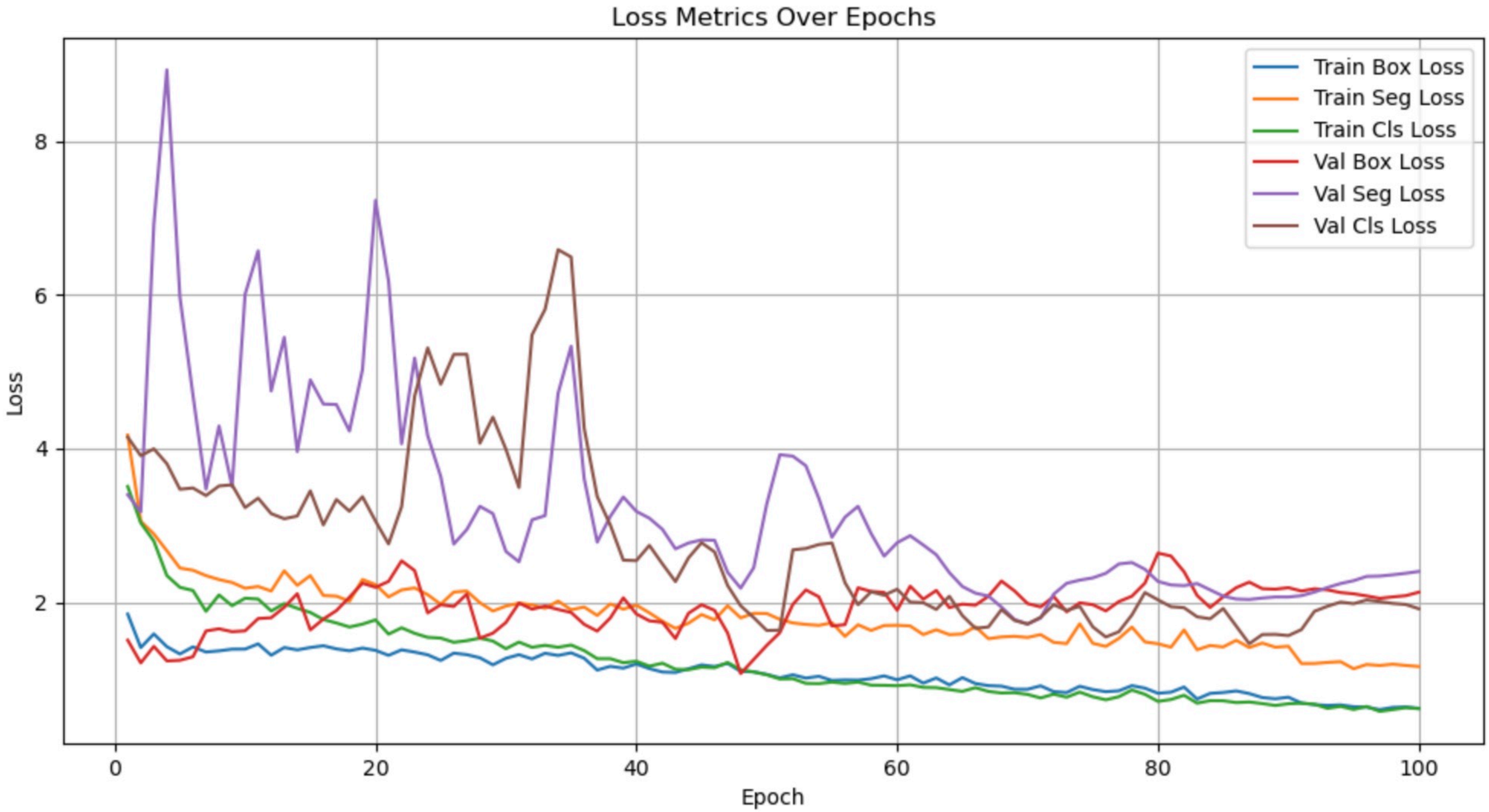
y
width
height
0.2
0.4
0.6
0.8
0.0
1.0



Label: "pothole" appears on the x-axis.
Count: 100 instances (y-axis)
Training set is fully populated with only one class ("pothole")—a clean class distribution for a binary segmentation task.
instances
100
80
60
40
20
0
pothole
Most potholes are located near the "middle of the road" view, which is typical if the camera is vehicle-mounted and forward-facing
This validates camera perspective consistency and tight crop annotations.
y
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.2
0.4
0.6
0.8
Model is seeing mostly small or mid-size potholes—valuable if you're aiming for fine-grained segmentation accuracy.
height
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.2
0.4
0.6
0.8
1.0

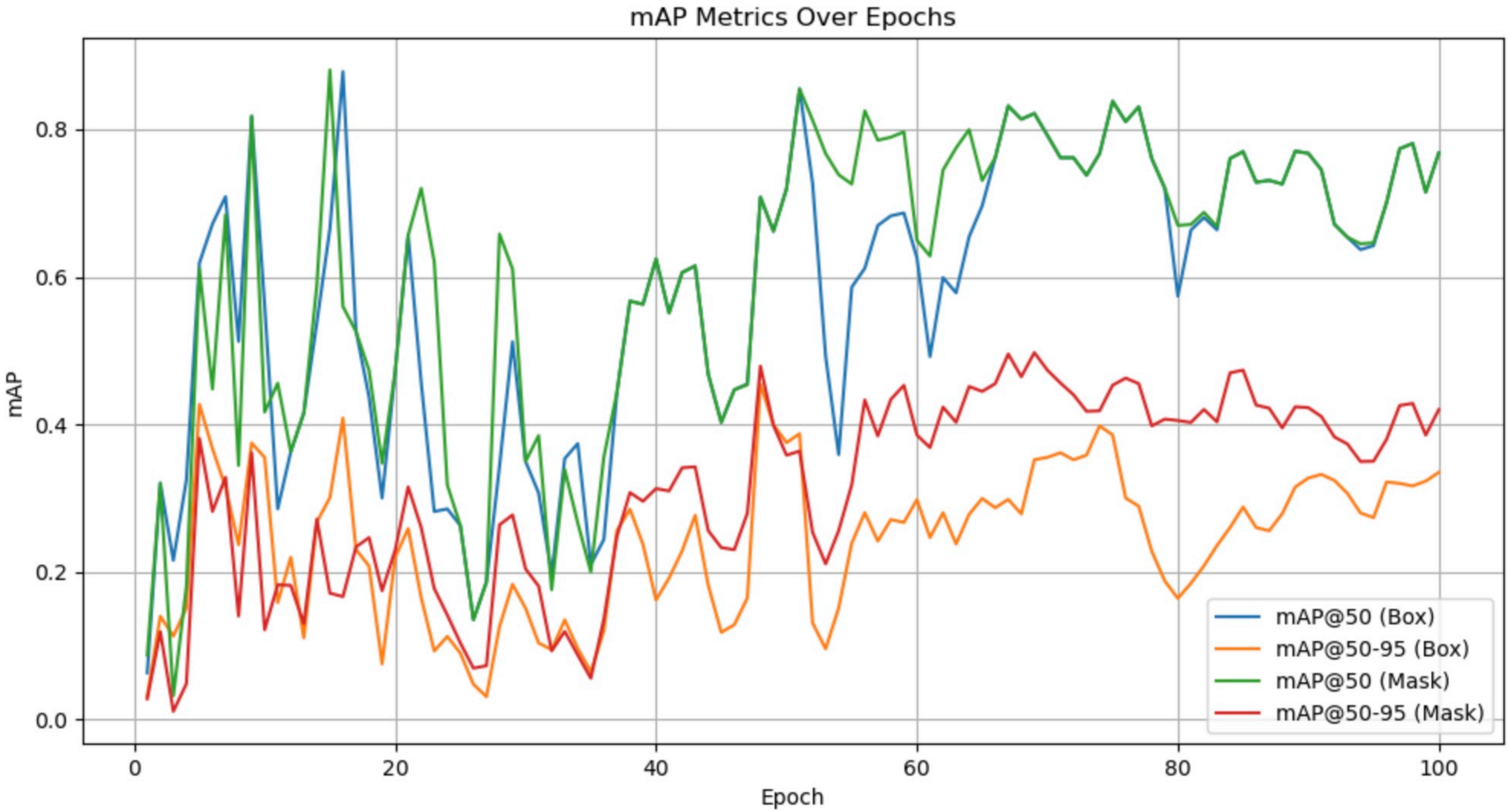
Training Stability
All major losses—box, segmentation, and class—declined smoothly across 100 epochs. The model showed strong learning behavior, especially in bounding box accuracy, suggesting a well-prepared dataset and effective convergence.

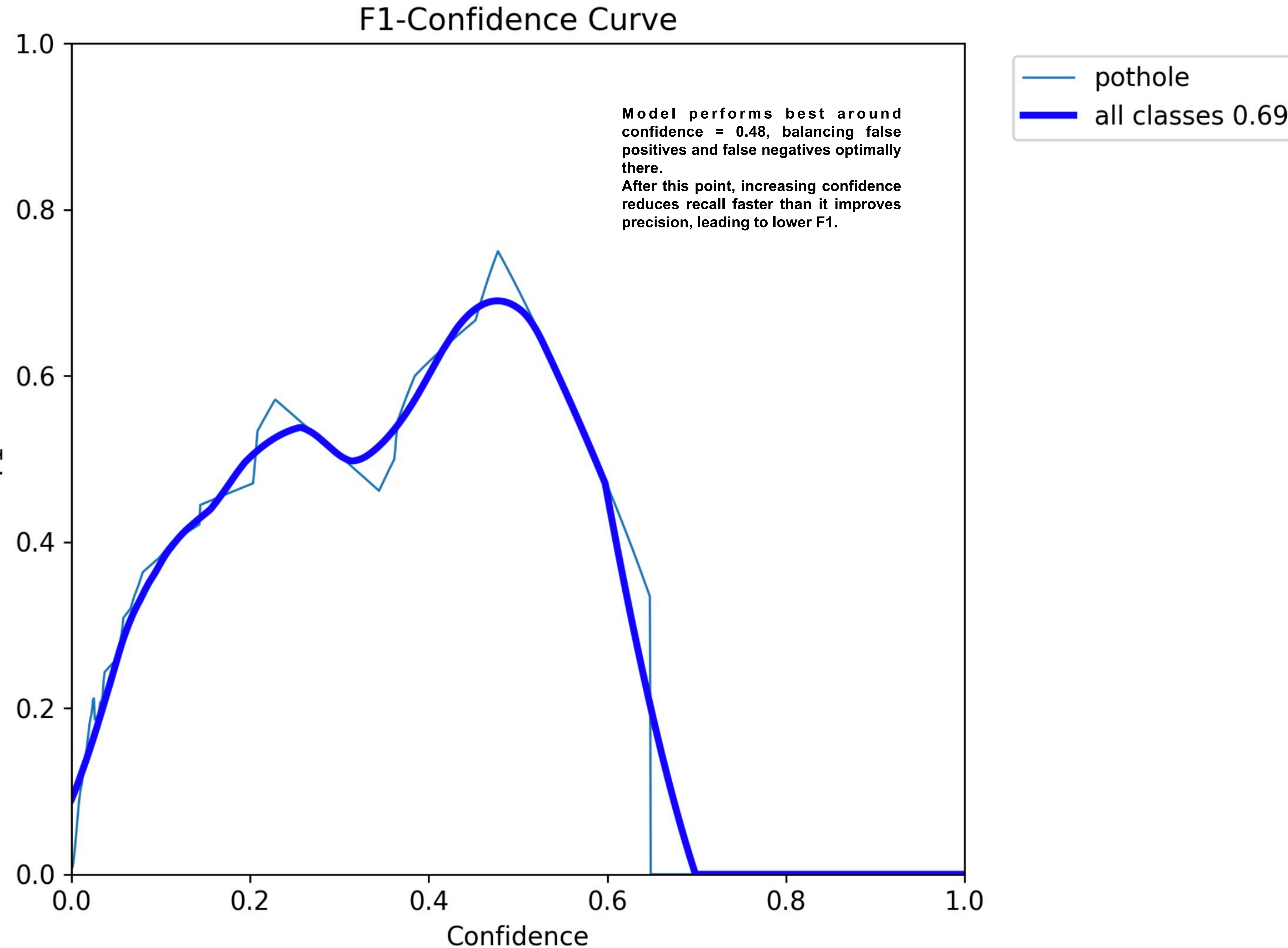
Validation Trend
Validation losses followed training curves early on but flattened later, hinting at mild overfitting. Segmentation loss could benefit from more background variety or size-balanced masks to refine model generalization.

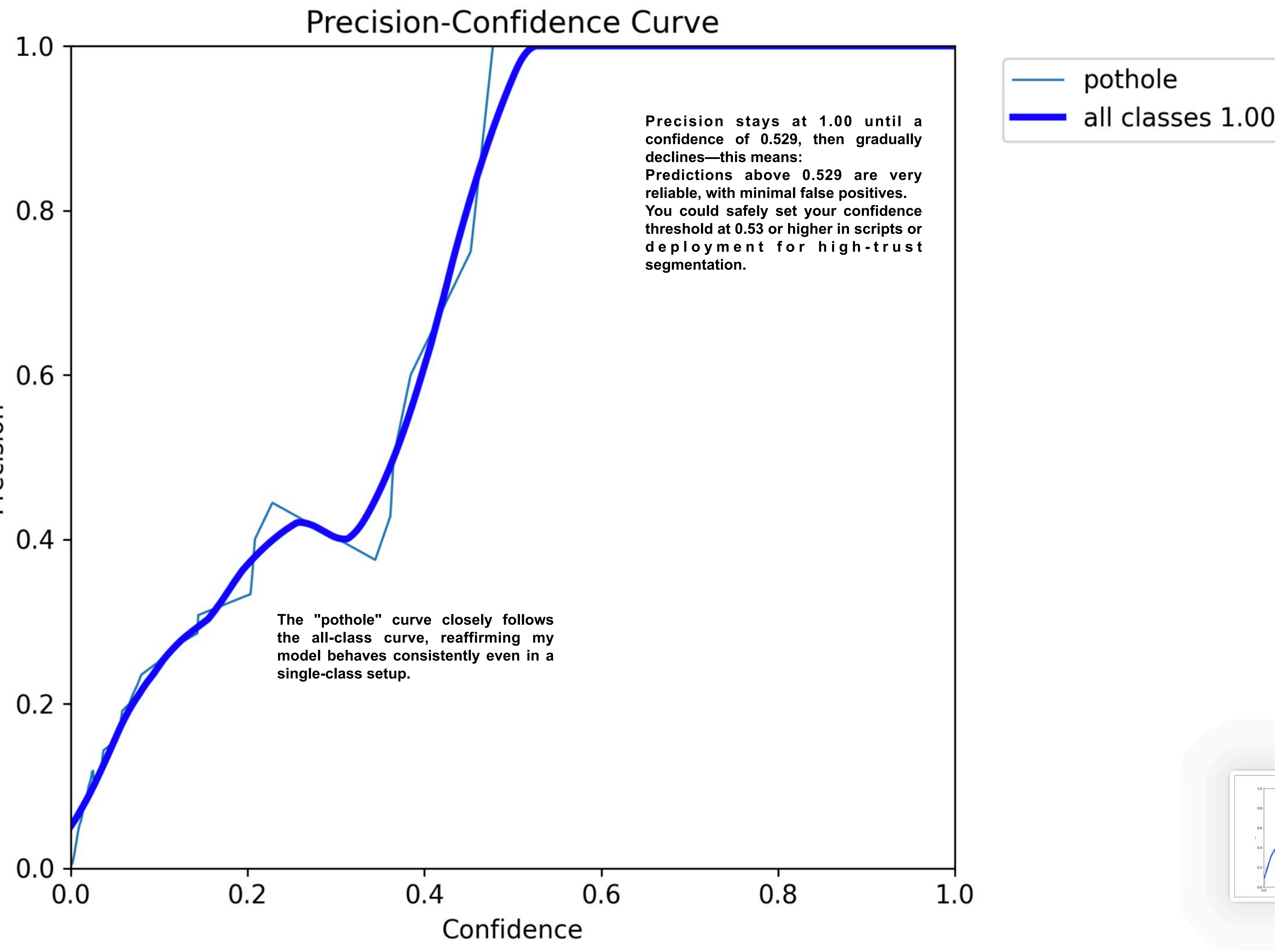


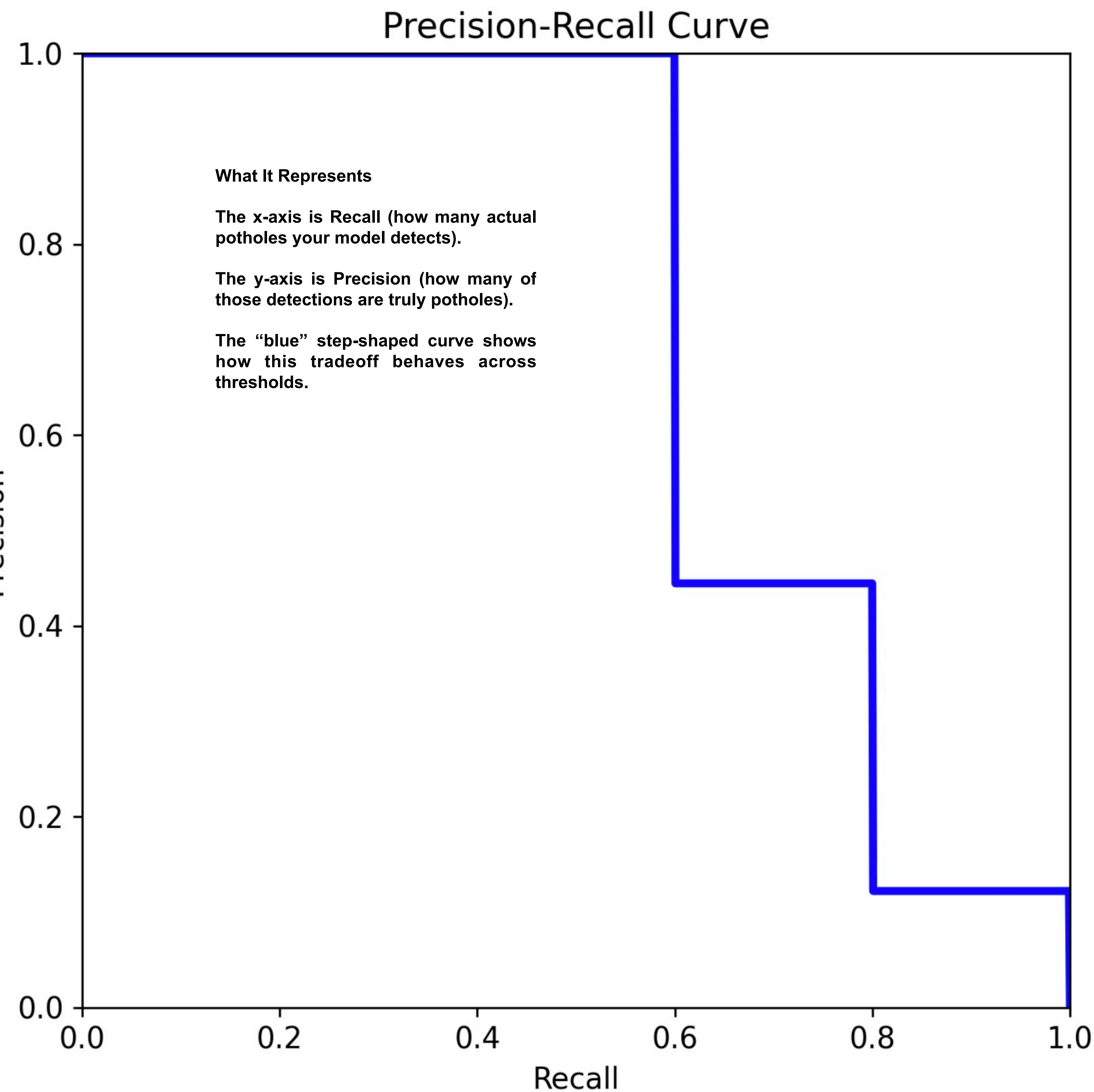
Object Detection
Performance mAP@50 (Box) peaked at 0.70 with a steady learning curve, showing accurate and consistent bounding box predictions. Lower mAP@50-95 suggests room to improve precision across IoU thresholds

Segmentation Metrics
Mask mAP@50 reached ~0.62, with mAP@50-95 stabilizing near 0.40. Performance was solid but could benefit from refined annotations or diverse mask shapes to improve robustness.




F1-Confidence Curve
Model performs best around confidence = 0.48, balancing false positives and false negatives optimally there.
After this point, increasing confidence reduces recall faster than it improves precision, leading to lower F1.
pothole
all classes 0.69
Confidence
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Confidence Curve
pothole
all classes 1.00
Precision stays at 1.00 until a confidence of 0.529, then gradually declines—this means:
Predictions above 0.529 are very reliable, with minimal false positives.
You could safely set your confidence threshold at 0.53 or higher in scripts or deployment for high-trust segmentation.
The "pothole" curve closely follows the all-class curve, reaffirming my model behaves consistently even in a single-class setup.
0.0
0.2
0.4
0.6
0.8
1.0
Confidence



What It Represents

The x-axis is Recall (how many actual potholes your model detects).

The y-axis is Precision (how many of those detections are truly potholes).

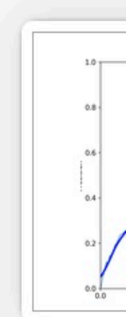
The “blue” step-shaped curve shows how this tradeoff behaves across thresholds.

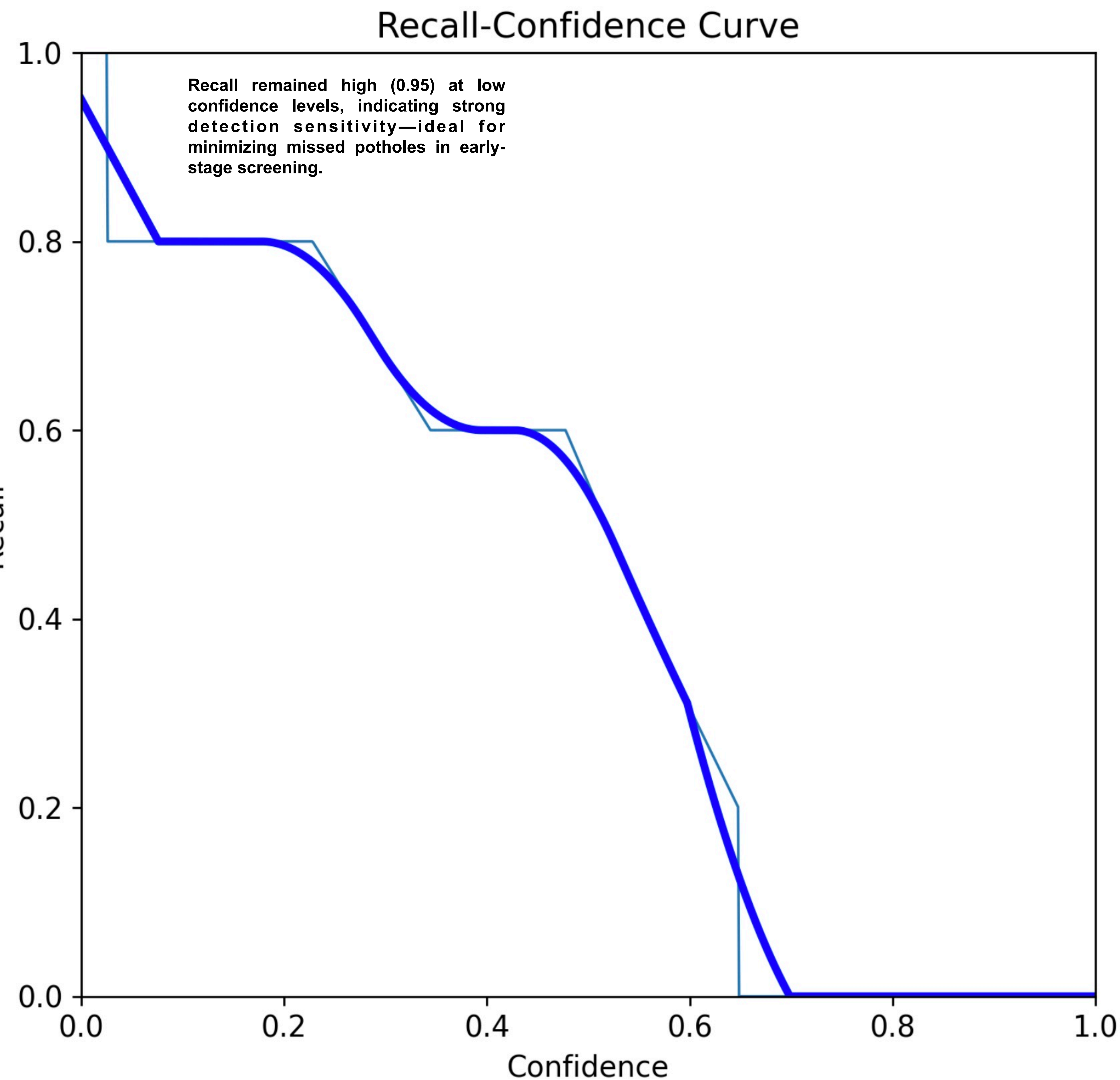
Key Metrics from the Graph

mAP@0.5 for both "pothole" and "all classes" is 0.708 — solid performance.

The curve shape indicates high precision at low recall, with a gradual tradeoff as recall improves.

This balance tells us your model avoids many false positives but might still miss some potholes (false negatives).



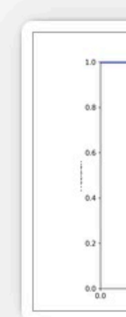


Key Observations

Recall stays at 0.95 across a broad confidence range—especially impressive when confidence is low. This means the model is very good at detecting most potholes, even when not very sure about its predictions.

As confidence increases, recall naturally begins to decline—fewer detections, but typically more precise ones.

The "all classes" and "pothole" lines almost overlap, which makes sense in your single-class setup.



Raw road image with a prominent pothole.
Great example of natural lighting and texture—ideal for validating model robustness.

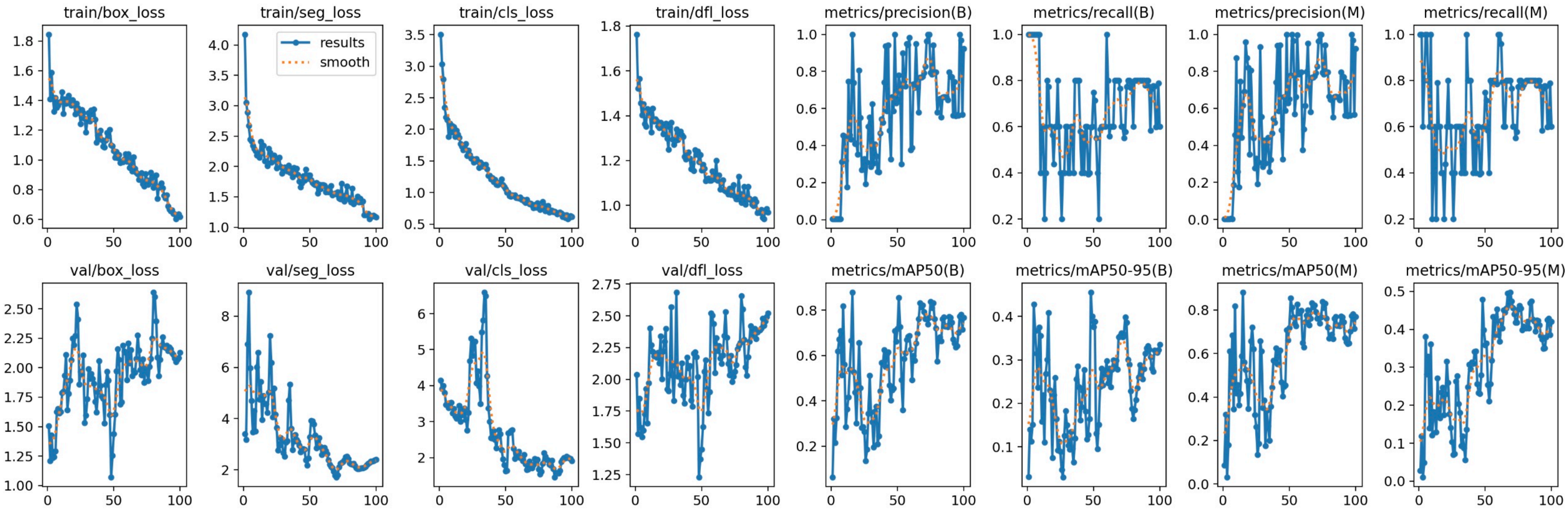
Ground Truth Mask
Manually annotated binary mask.
White area marks the true pothole region—serves as the benchmark for evaluating segmentation accuracy.

Predicted Mask
Model-generated segmentation mask.
White patch approximates the pothole shape, with some overlap and subtle spatial variance compared to the ground truth.

The model successfully identifies the general location and shape of the pothole.
Slight mismatch in boundaries suggests room for refinement—possibly from more diverse training samples or finer mask resolution.

These are strong outcomes for a lightweight segmentation model—especially impressive considering the limited class diversity.

Metrics like box_loss, seg_loss, cls_loss, and dfl_loss all show steady downward trends, indicating successful learning and convergence.

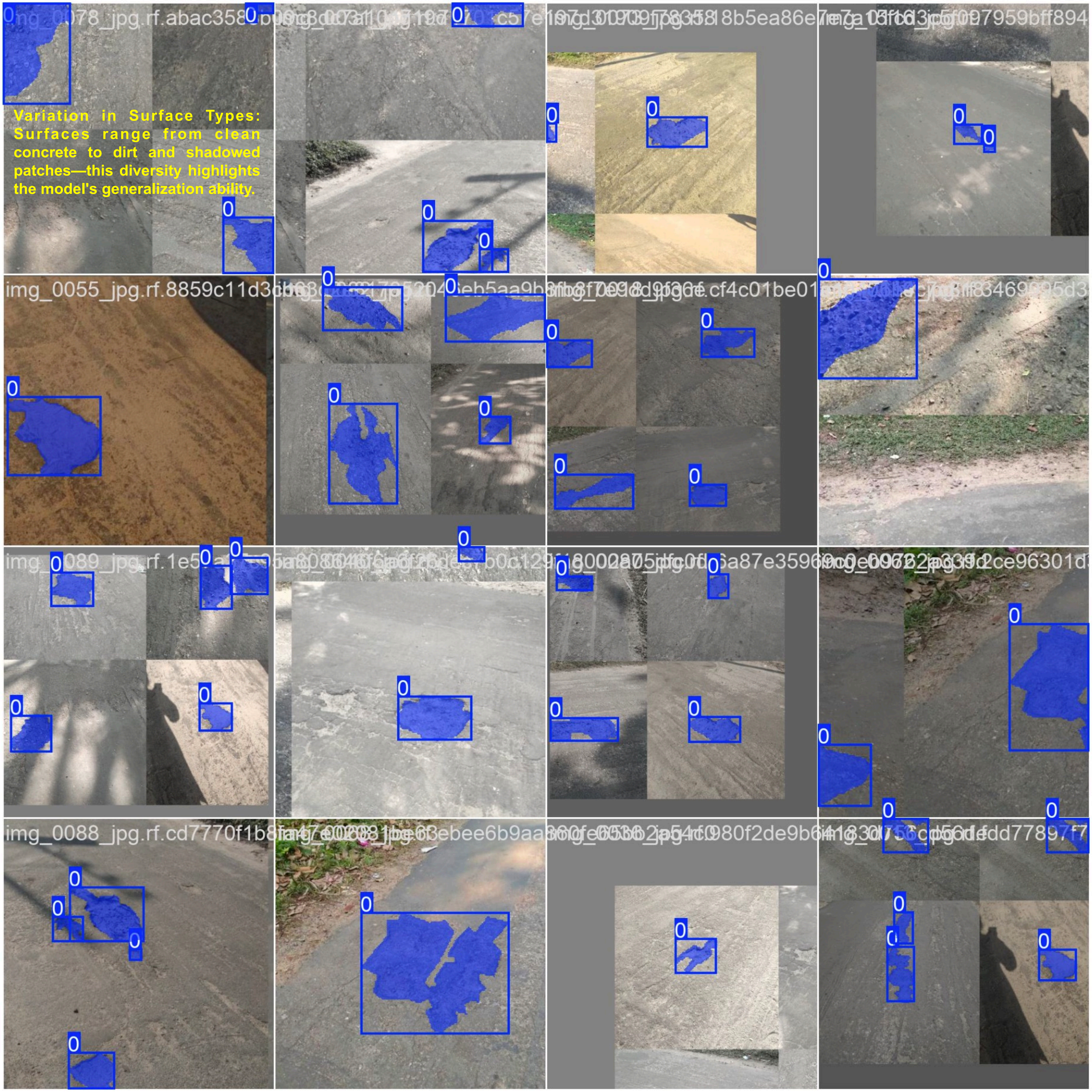


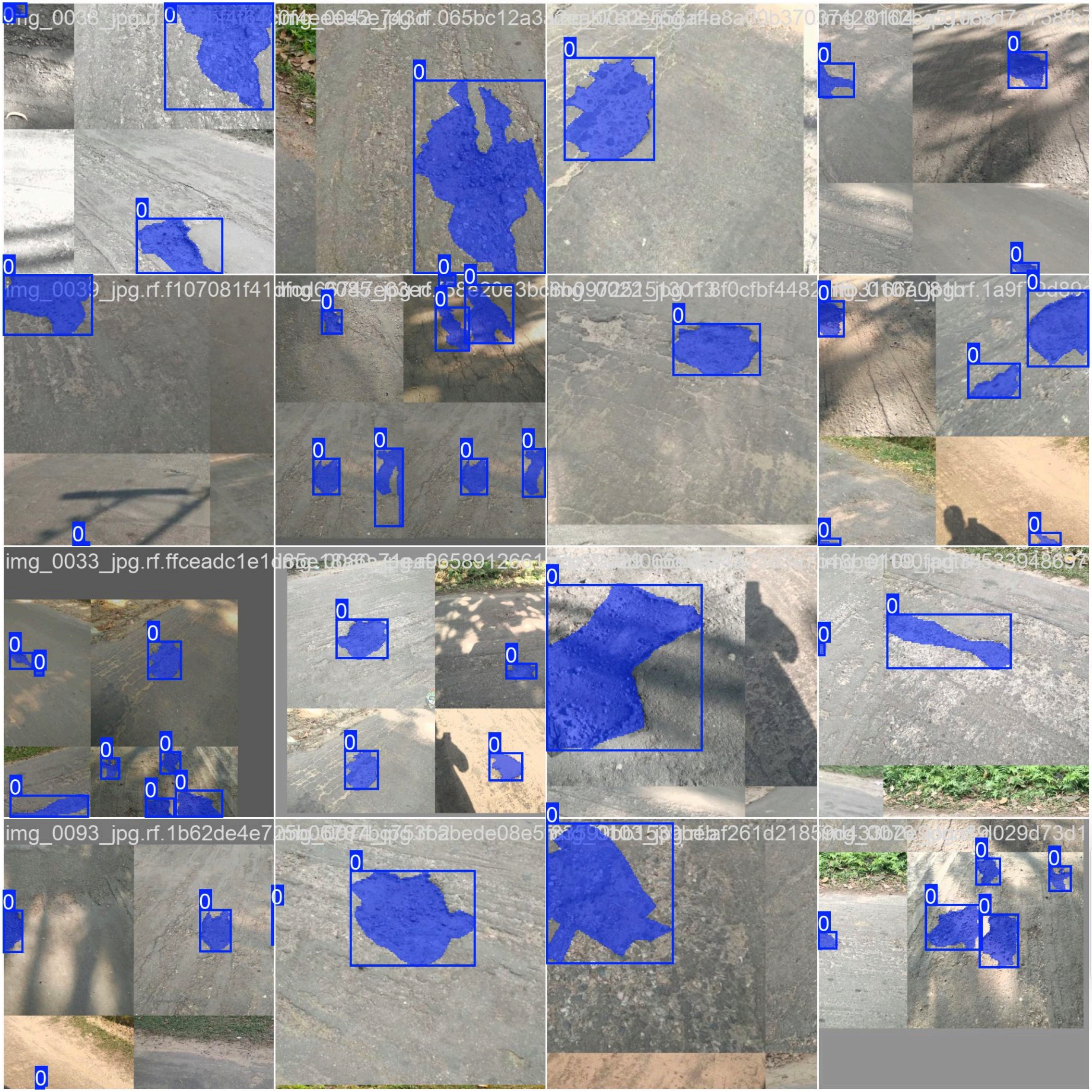
Validation losses (val/box_loss, val/seg_loss, etc.) follow similar downward trends early on, with plateaus or small fluctuations in later epochs—typical signs of reaching generalization capacity. Precision and recall also rise sharply then stabilize, suggesting solid detection and mask consistency.

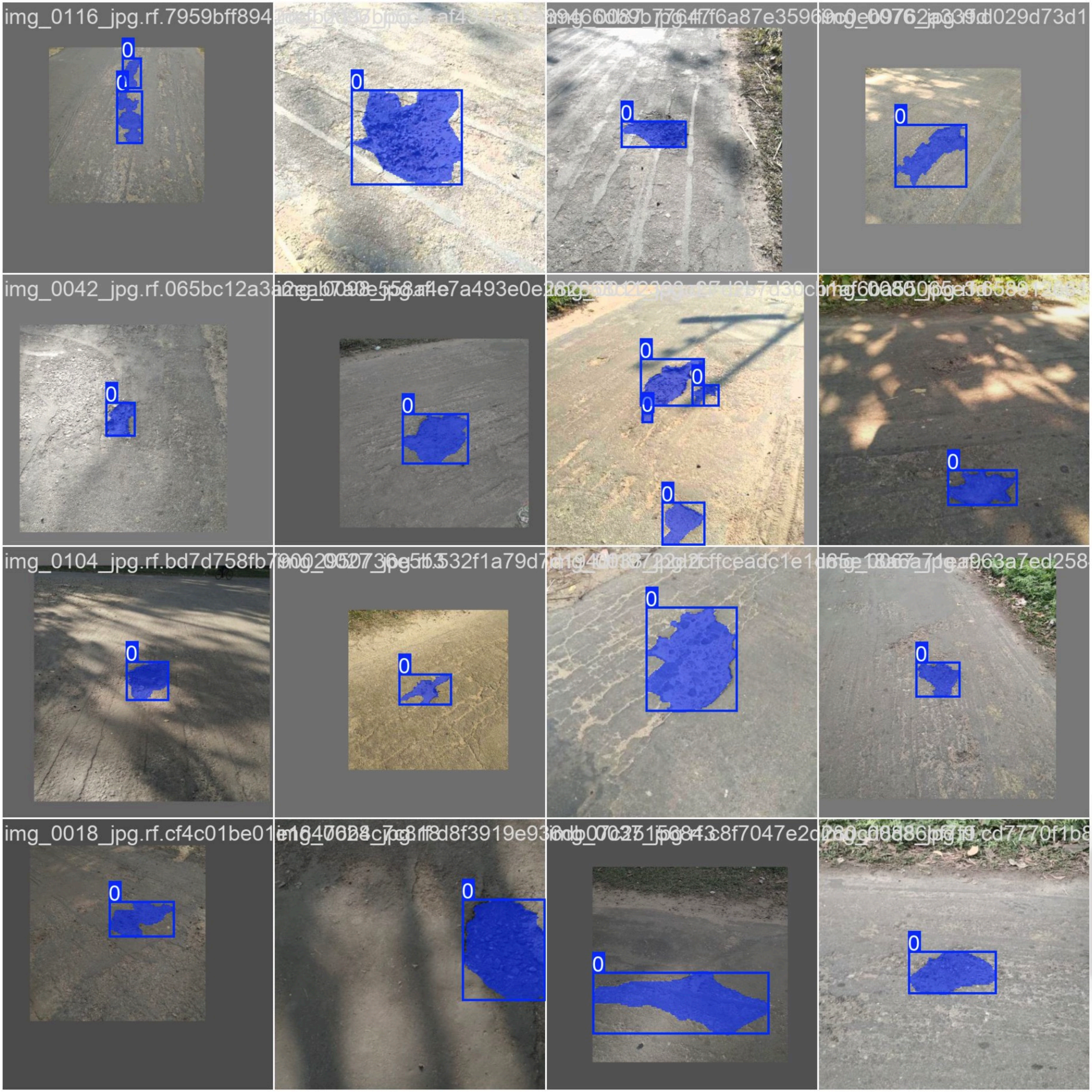
mAP50 (Box) reaches around 0.70
mAP50-95 (Box) approaches 0.46

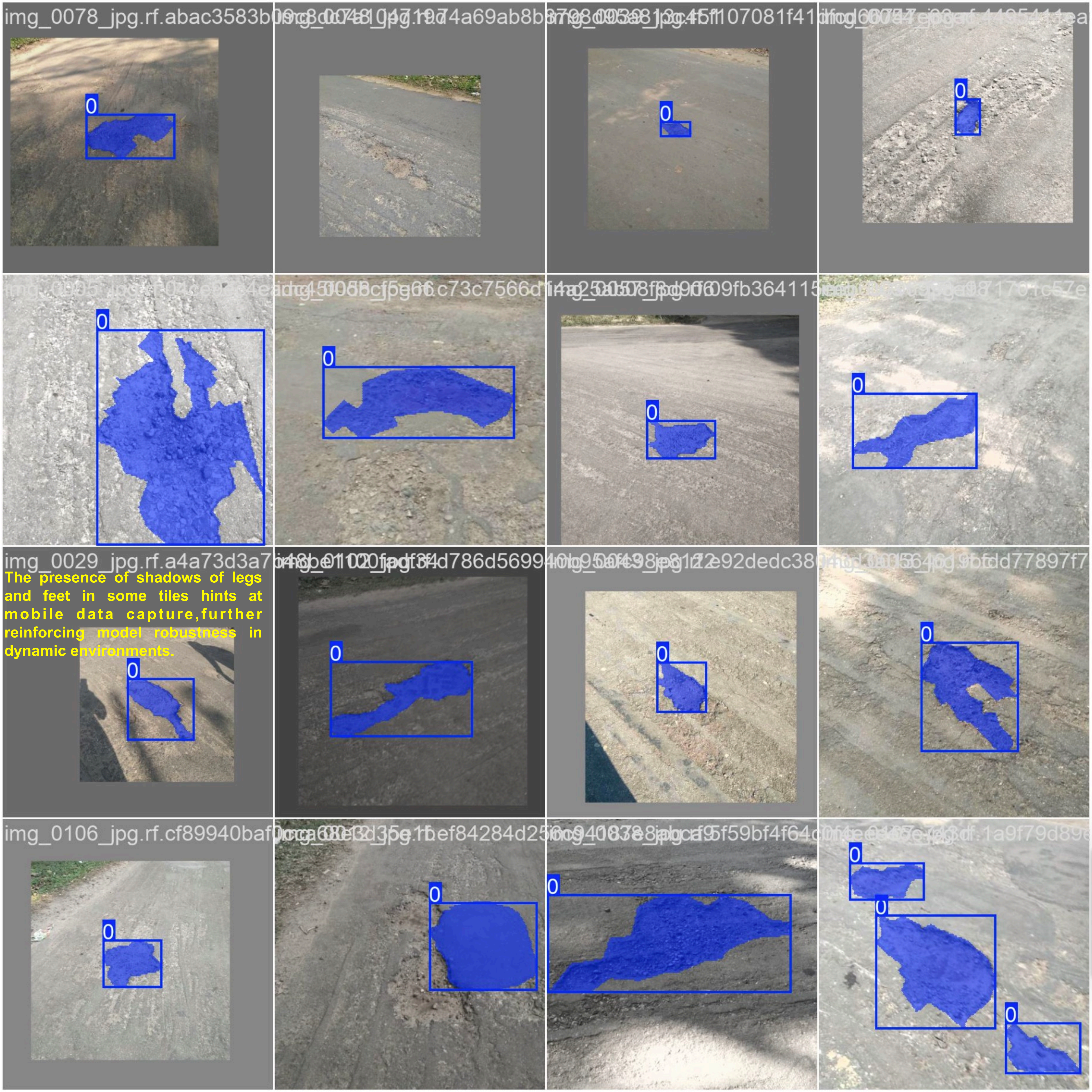
mAP50 (Mask) settles near 0.62, while
mAP50-95 (Mask) stabilizes around 0.40

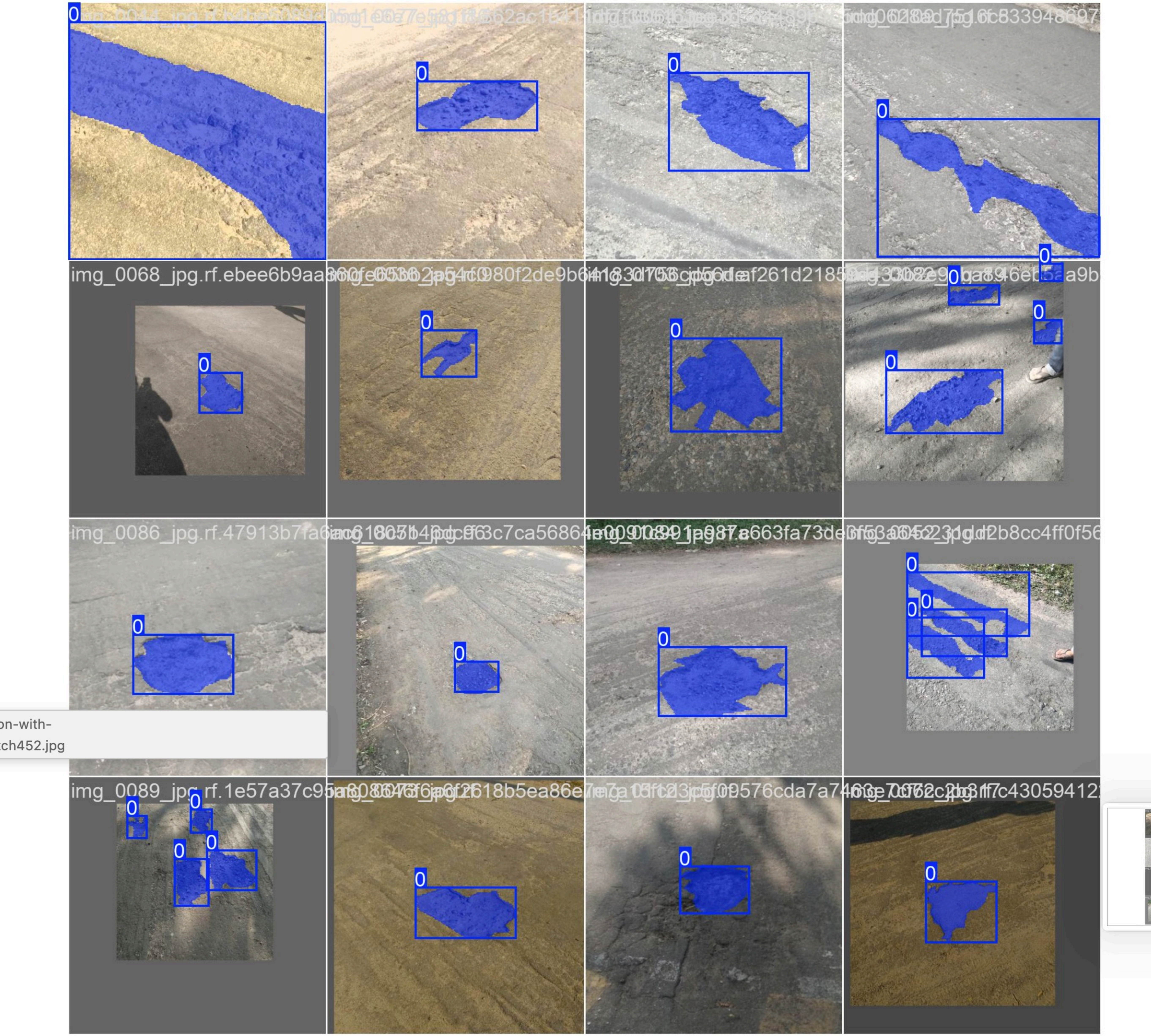

The "0" label suggests this is a single-class task (pothole).
Blue Mask Overlays:
These regions highlight the model's pixel-level segmentation of the pothole.

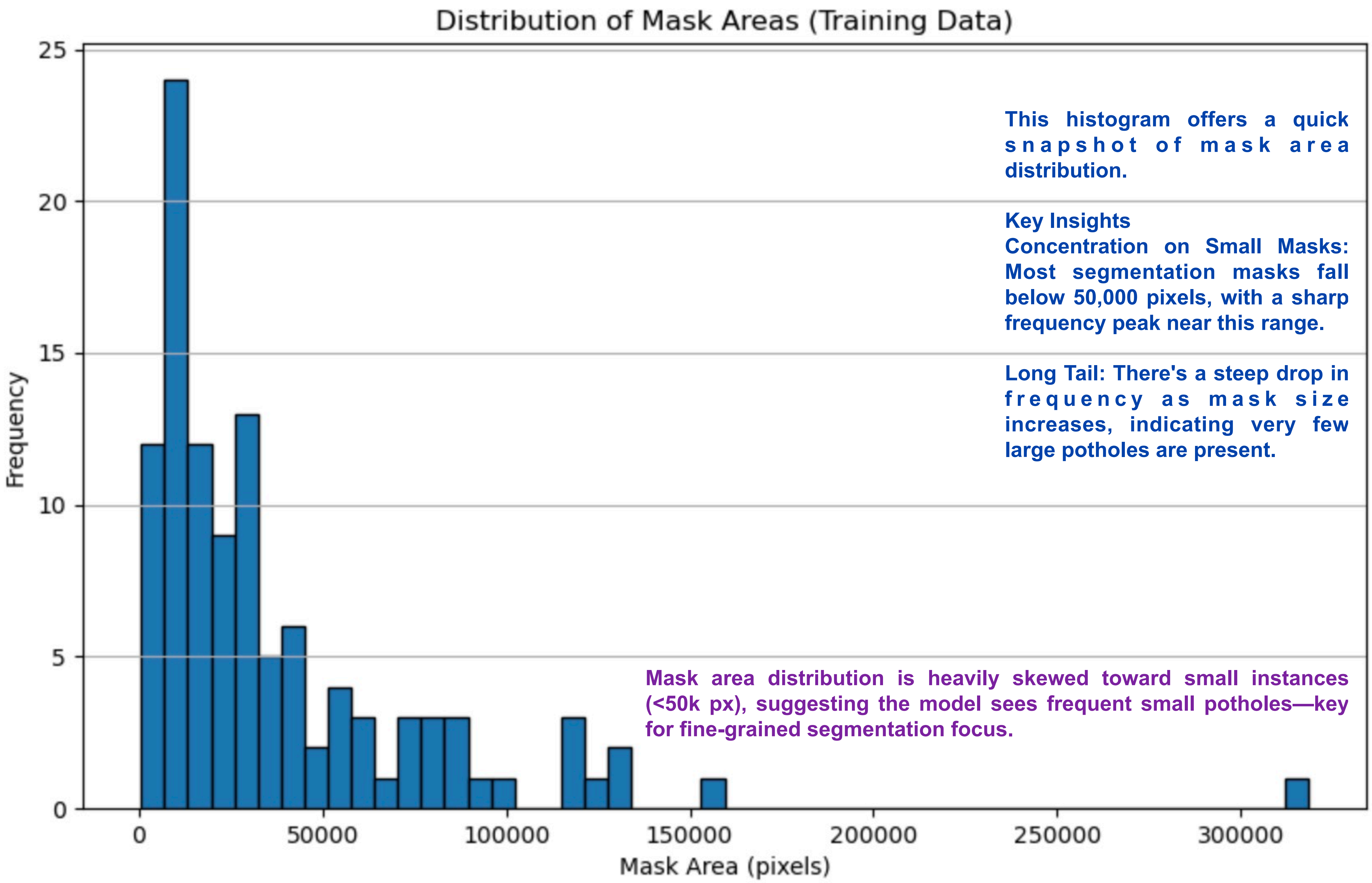

Variation in Surface Types: Surfaces range from clean concrete to dirt and shadowed patches—this diversity highlights the model's generalization ability.

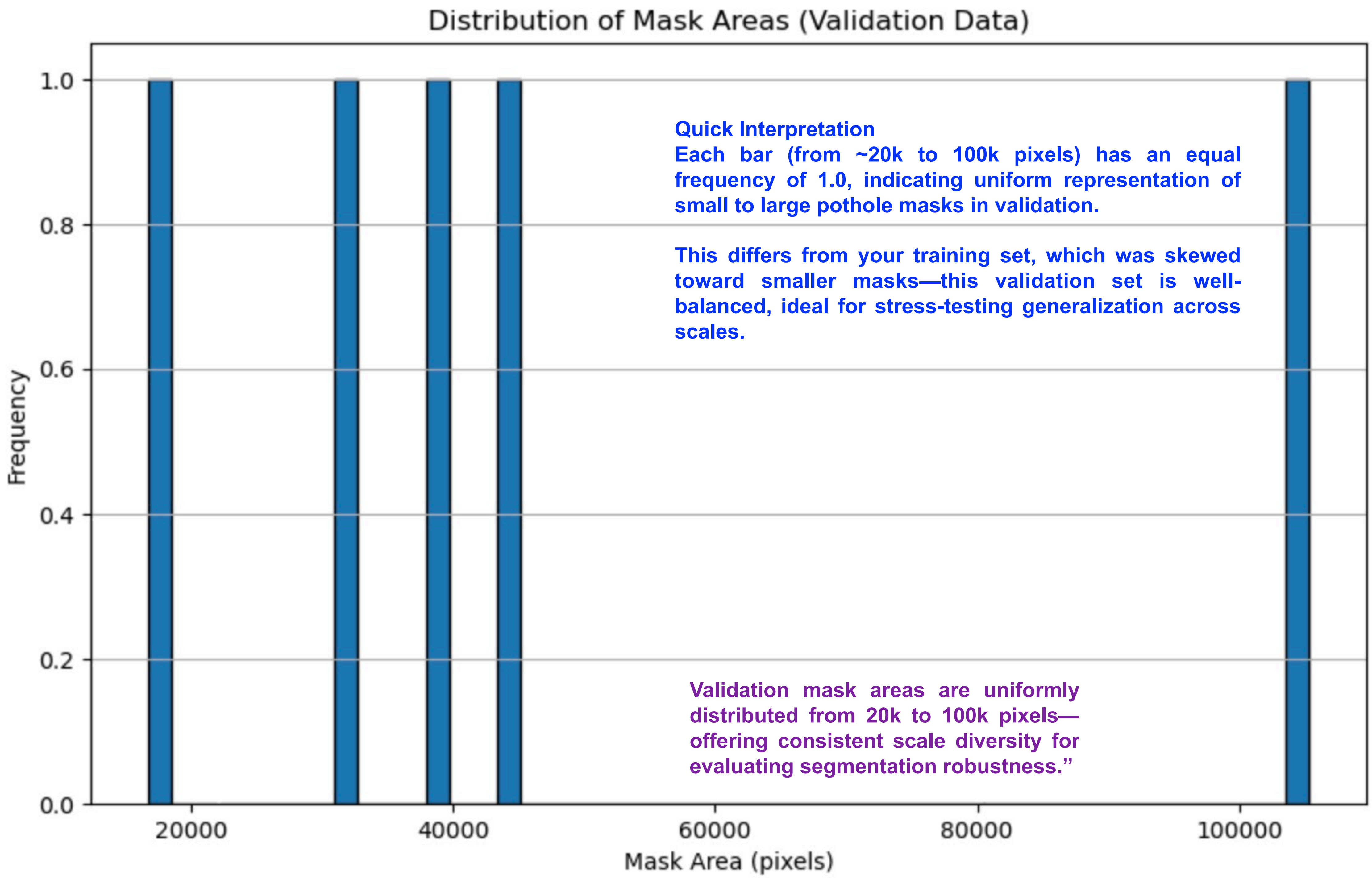

img_0039_jpg.rf.f107081f41
img_0033_jpg.rf.ffceadc1e1d
img_0093_jpg.rf.1b62de4e7

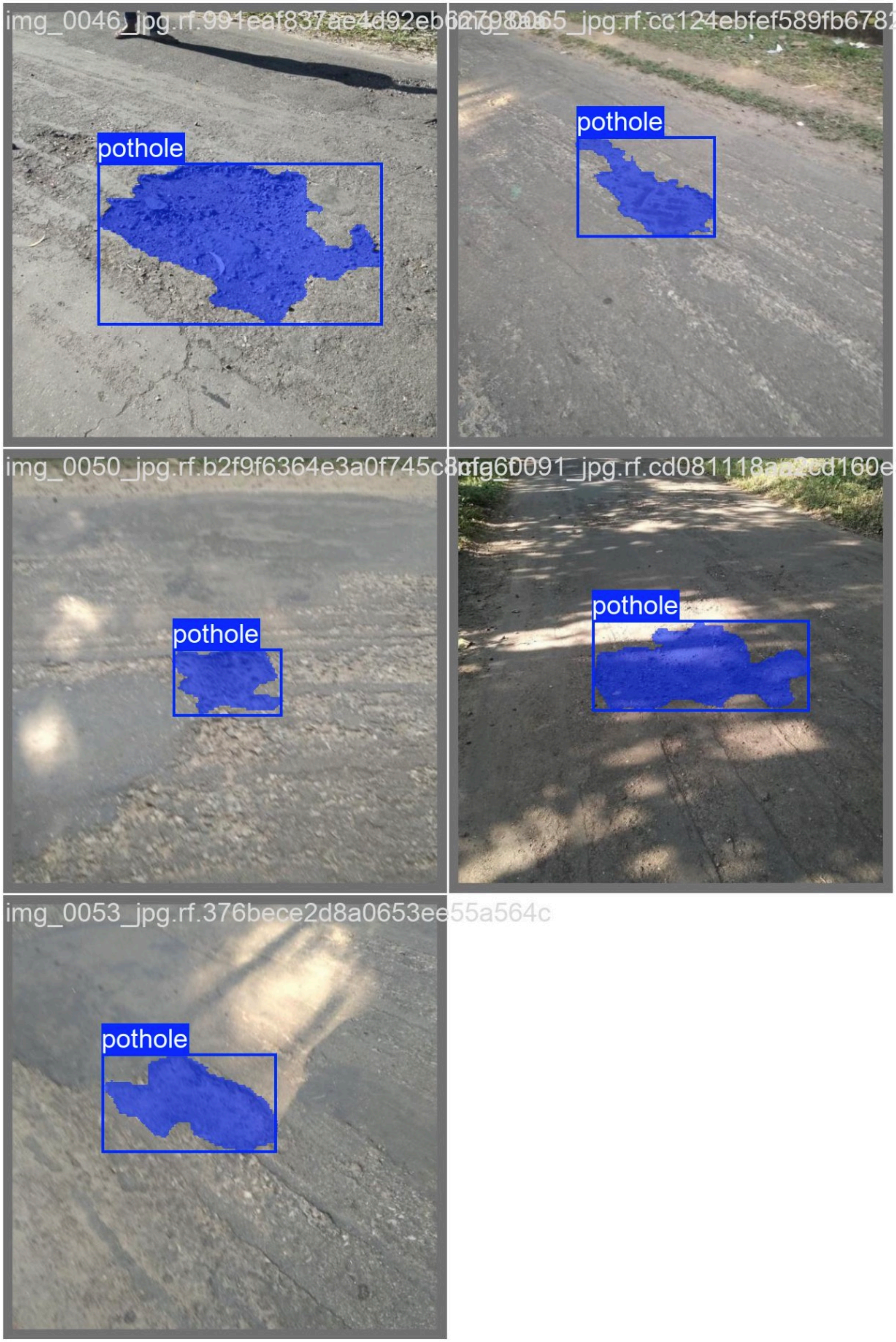

img_0116_jpg.rf.7959bff894
img_0042_jpg.rf.065bc12a3a
img_0104_jpg.rf.bd7d758fb7
img_0018_jpg.rf.cf4c01be01

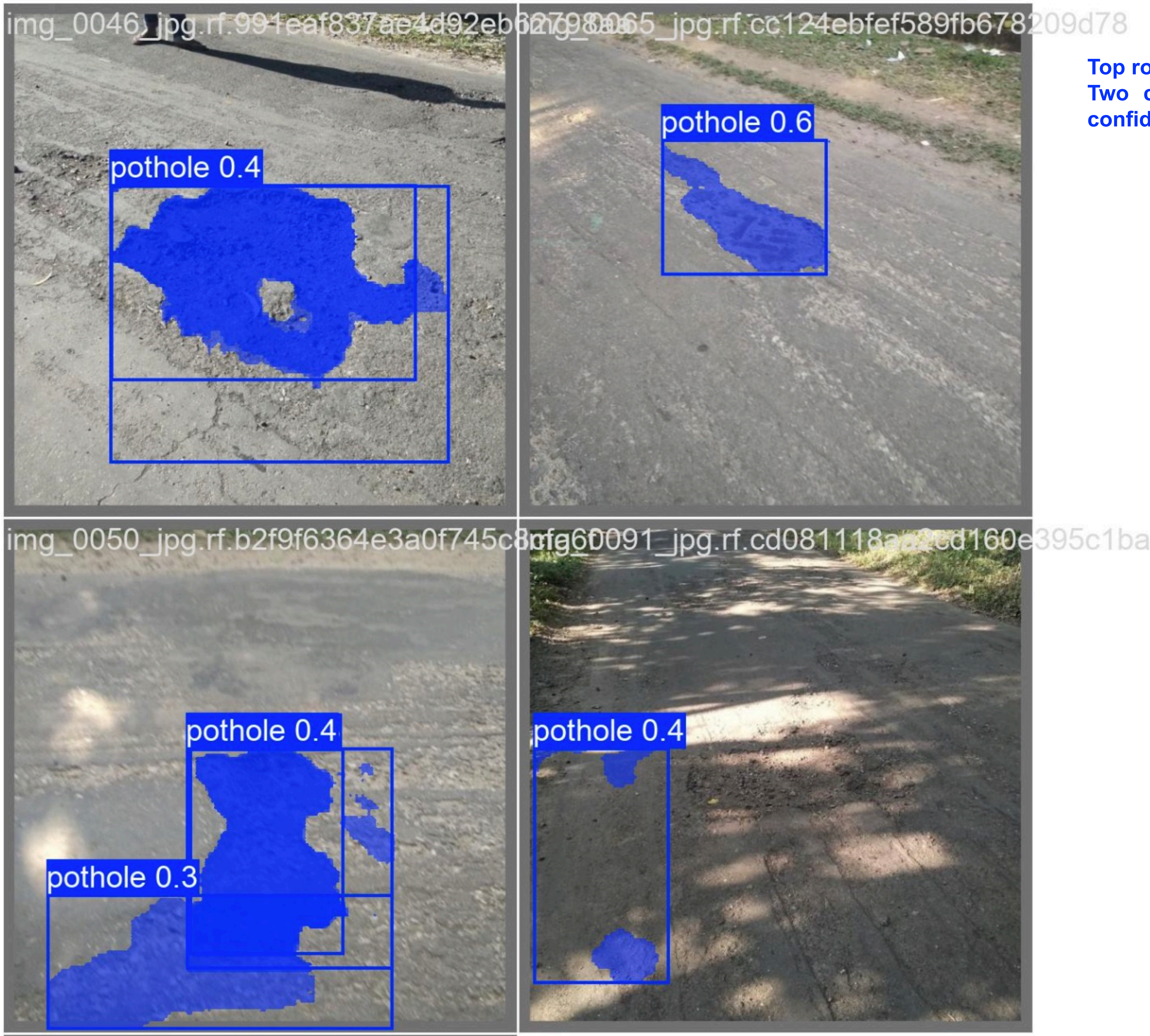

img_0078_jpg.rf.abac3583b
img_0029_jpg.rf.a4a73d3a7
The presence of shadows of legs and feet in some tiles hints at mobile data capture,further reinforcing model robustness in dynamic environments.
img_0106_jpg.rf.cf89940baf


on-with-
tch452.jpg


Distribution of Mask Areas (Training Data)
Frequency
Mask Area (pixels)
0
5
10
15
20
25
0
50000
100000
150000
200000
250000
300000
This histogram offers a quick snapshot of mask area distribution.
Key Insights
Concentration on Small Masks: Most segmentation masks fall below 50,000 pixels, with a sharp frequency peak near this range.
Long Tail: There's a steep drop in frequency as mask size increases, indicating very few large potholes are present.
Mask area distribution is heavily skewed toward small instances (<50k px), suggesting the model sees frequent small potholes—key for fine-grained segmentation focus.


Distribution of Mask Areas (Validation Data)
Frequency
Mask Area (pixels)
1.0
0.8
0.6
0.4
0.2
0.0
20000
40000
60000
80000
100000
Quick Interpretation
Each bar (from ~20k to 100k pixels) has an equal frequency of 1.0, indicating uniform representation of small to large pothole masks in validation.
This differs from your training set, which was skewed toward smaller masks—this validation set is well-balanced, ideal for stress-testing generalization across scales.
Validation mask areas are uniformly distributed from 20k to 100k pixels—offering consistent scale diversity for evaluating segmentation robustness.”


pothole
pothole
pothole
pothole
img_0053_jpg.rf.376bece2d8a0653ee55a564c
pothole



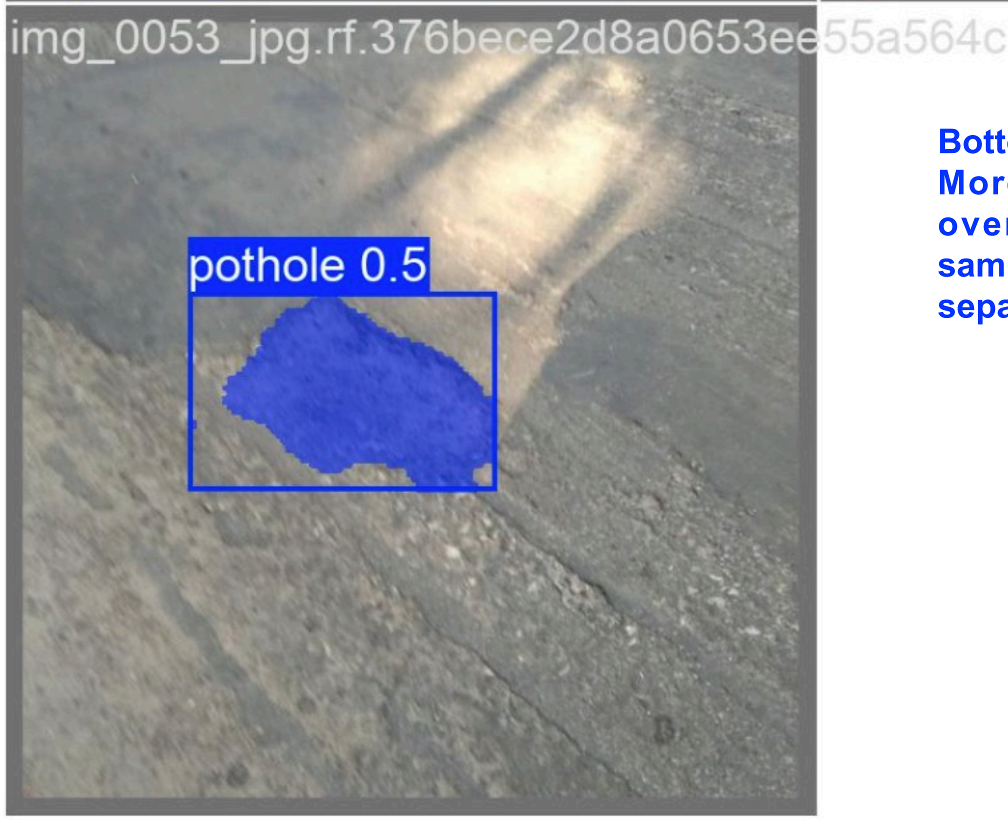
Top row:
Two clear detections with moderate confidence (~0.4–0.6).

What It Tells Us !

The model is confidently segmenting potholes even in visually noisy scenarios.

A few lower-confidence labels (0.3–0.4) suggest borderline cases where mask boundaries could benefit from finer tuning.

Overall, this is a great qualitative sample to pair with your F1/ confidence curves and precision-recall metrics.



Bottom row:
More complex visuals—including overlapping detections and one sample with multiple potholes labeled separately.