Structured Training Ablation and GSD-Adaptive Orthophoto Inference for UAV-Based Pavement Defect Detection with YOLO26

Ahsan Mustafa,[1,2,*] Mikhail Yu. Kataev,[3] and Dmitry S. Efremenko[2]

[1] *National University of Science and Technology MISIS, Moscow, 119049, Russia*

[2] *HSE University, Moscow, 101000, Russia*

[3] *Tomsk State University of Control Systems and Radioelectronics, Tomsk, 634050, Russia*

(Received xx.xx.2026; Revised xx.xx.2026; Accepted xx.xx.2026)

Abstract

Automated detection of pavement surface defects from unmanned aerial vehicle (UAV) imagery can substantially reduce the cost and labour of road condition surveys. Two questions that remain largely unanswered in current literature are which training configuration choices actually drive accuracy gains, and how well a model trained on low-altitude UAV images generalises to the higher-altitude orthophotos used in real-world deployment. This paper addresses both. We train YOLO26 on the UAV-PDD2023 benchmark dataset (2,439 images, six defect classes) and run a controlled five-configuration ablation study that independently varies hyperparameter settings, preprocessing, and data augmentation. The best combination reaches $\mathrm{mAP}_{50} = 0.859$ and $\mathrm{mAP}_{50\text{–}95} = 0.631$ on the held-out test split, a 57% and 133% improvement over the default baseline respectively. Raising input resolution from 640 px to 1280 px is the single largest contributor; preprocessing and augmentation each add complementary gains. We then deploy the best model on three real-world GeoTIFF orthophotos through a ground-sampling-distance (GSD) adaptive tiling pipeline. The pipeline selects tile dimensions to match the ground coverage seen during training, applies cross-tile non-maximum suppression, and converts bounding boxes to WGS84 coordinates. Despite a measurable domain gap, the pipeline identifies all six defect classes across scenes spanning up to 322×843 m, with 2,201 detections on the largest mosaic. Exported GeoJSON and annotated GeoTIFF outputs are validated in QGIS against source rasters.

PACS numbers: 07.05.Mh, 42.30.Va

Keywords: pavement defect detection UAV imagery YOLO26 ablation study GSD-adaptive tiling geo-referenced detection

1. INTRODUCTION

Surface cracks, potholes, and alligator fractures are the early warning signs of pavement failure. Left unaddressed, they compound under traffic and weather into repairs that cost orders of magnitude more than timely maintenance would. The World Bank estimates that more than half of paved road surfaces worldwide show measurable deterioration [1]. At the same time, traditional inspection, where engineers walk or drive road sections and log defects by hand, cannot scale to the size of modern road networks.

UAVs change the economics. A single flight covers kilometres of road in minutes, producing high-resolution RGB imagery without traffic disruption. Processed into georeferenced or-

* E-mail: iahsanmustafa@gmail.com

thophoto mosaics, these images achieve centimetre-level ground sampling distances over large areas, which is fine enough to resolve cracks a few millimetres wide. Combined with a trained object detector, this creates a scalable inspection pipeline.

Detection methods based on YOLO variants [2–4], convolutional segmentation networks [5], and transformer architectures [6] have all produced good benchmark numbers on pavement datasets. But the published work has three consistent omissions that motivate this paper.

First, controlled ablation studies are rare. Most papers report a single optimised configuration without isolating the contributions of hyperparameter choices, preprocessing pipelines, and augmentation strategies separately. Practitioners can't determine from the literature which axis of improvement is worth the effort.

Second, evaluation almost always stops at the benchmark test split. Deploying a model trained at a GSD of 0.7 cm/px onto a survey orthophoto at 2.5 cm/px is a fundamentally different problem, and that transition is seldom studied.

Third, geographic output is rarely addressed. Infrastructure engineers use GIS platforms such as QGIS; they need bounding boxes that carry real-world coordinates and physical dimensions, not pixel offsets.

This paper addresses all three. Our contributions are: (i) a five-configuration ablation study on UAV-PDD2023 that isolates the effect of hyperparameter tuning, preprocessing, and augmentation, with YOLO26 achieving $\mathrm{mAP}_{50} = 0.859$; (ii) a GSD-adaptive tiling and georeferencing pipeline for deploying the trained model on full-resolution survey orthophotos; and (iii) field deployment and QGIS validation on three real-world orthophotos not present in the training set.

2. RELATED WORK

Early automated pavement inspection relied on edge detection [7], Gabor filter banks [8], and entropy-based thresholding [9]. These methods require hand-crafted feature design and are brittle across road surface types and illumination conditions. Deep CNN-based approaches replaced manual feature engineering with learned representations. Zhang et al. [10] showed that a CNN trained end-to-end on smartphone pavement images surpassed SVM baselines by a clear margin. Encoder-decoder networks such as DeepCrack [5] handle the thin, elongated geometry of pavement cracks at multiple scales.

In the YOLO family, several task-specific variants have been proposed for road inspection. YOLOv5-PD [11] replaced standard convolutions with large-kernel versions to capture wide crack patterns. Sun [2] and Wang [3] each introduced structural changes to YOLOv8 achieving mAP_{50} values of 0.774 and a 5.8% improvement over baseline respectively. RCYOLO [4] added dynamic snake convolution and SimAM attention, outperforming YOLOv8 by 5.9% in mAP_{50} on UAV crack imagery. YOLOv12 [12], the predecessor of YOLO26, introduced area attention for efficient long-range spatial dependency modelling in detection tasks. An earlier study by the present authors [13] used YOLOv12 to evaluate RGB versus single-channel inference robustness on UAV-PDD2023; the present paper uses YOLO26 and focuses on training configuration and orthophoto deployment.

For high-resolution inference, SAHI [14] proposed slicing-aided hyper inference, showing AP gains of 5–7% on aerial datasets. The critical gap not addressed by SAHI or its adaptive variant ASAHI [15] is that fixed tile sizes ignore the GSD ratio between training and deployment imagery. Our approach selects tile dimensions to match ground coverage, not pixel count.

3. DATASET

All experiments use the UAV Pavement Distress Detection 2023 dataset (UAV-PDD2023) [16], which contains 2,439 georeferenced RGB images captured by UAV at altitudes producing a GSD of approximately 0.7 cm/px over road surfaces in China. Annotations cover six defect classes using axis-aligned bounding boxes: alligator crack, longitudinal crack, oblique crack, transverse crack, pothole, and repair patch. Table 1 summarises the split statistics.

Table 1: UAV-PDD2023 dataset statistics.

Split	Images	Instances
Train	1,709	7,689
Validation	480	2,283
Test	250	1,182
Total	2,439	11,154

The dataset is class-imbalanced: longitudinal and transverse cracks each contribute several thousand instances, while potholes and oblique cracks together account for fewer than 1,500. Addressing this imbalance is one motivation for the augmentation axis in the ablation study.

For the ablation, three Roboflow dataset versions are used. Version v2 is the raw dataset with no modification. Version v7 adds auto-orientation, stretch resize to 1280×1280 px, and contrast stretching, but no augmentation. Version v4 applies all of v7's preprocessing and additionally generates three augmented outputs per training example using rotation ($\pm 5°$), saturation ($\pm 15\%$), brightness ($\pm 15\%$), blur (up to 1 px), and pixel noise (up to 1.49%).

4. YOLO26

YOLO26 is the latest major release in Ultralytics' YOLO series. Its small variant, YOLO26s, is used throughout this work. It extends the attention-centric design of YOLOv12 [12], retaining Area Attention and Residual ELAN feature aggregation while adding architectural improvements to the neck and detection head that improve multi-scale feature fusion.

Area Attention divides feature maps into fixed-size regions and applies self-attention within each region, keeping computational cost sub-quadratic in image area. This matters for pavement defects, which span a wide range of spatial extents from compact potholes to longitudinal cracks hundreds of pixels long. Compared to YOLOv12s, YOLO26s achieves higher mAP on COCO benchmarks at a similar parameter count, and that improved multi-scale handling is the most relevant architectural difference for this task.

All training used an NVIDIA A100-SXM4-80GB GPU via HPC resources at HSE University [17].

5. ABLATION STUDY

5.1. Experimental design

Three independent training axes are studied. The hyperparameter axis compares Ultralytics default settings (DP) against a tuned configuration (BP) that raises input resolution from 640 px to 1280 px and extends training to 300 epochs. The preprocessing axis compares no preprocessing (NPP) against best preprocessing (BPP), which applies auto-orientation, 1280×1280 stretch, and contrast stretching via Roboflow. The augmentation axis compares no augmentation (NAUG) against the best augmentation strategy (BAUG) described in Section 3. Five configurations are evaluated; Table 2 lists the axis settings for each.

5.2. Quantitative results

Table 3 reports three mAP metrics on the held-out test split for all five configurations.

The best configuration (BP,BPP,BAUG) reaches $\mathrm{mAP}_{50} = 0.859$ and $\mathrm{mAP}_{50\text{–}95} = 0.631$, gains of +0.312 and +0.360 over the baseline. Hyperparameter tuning alone (BP,NPP,NAUG) already

Table 2: Five ablation configurations and their axis settings.

Configuration	HP	Preprocessing	Augmentation
Baseline (DP,NPP,NAUG)	DP	None	None
DP,BPP,NAUG	DP	BPP	None
BP,NPP,NAUG	BP	None	None
DP,BPP,BAUG	DP	BPP	BAUG
BP,BPP,BAUG (Best)	BP	BPP	BAUG

Table 3: Ablation results on the UAV-PDD2023 test split, ranked by $mAP_{50–95}$.

Configuration	mAP_{50}	$mAP_{50–95}$	mAP_{75}
BP,BPP,BAUG (Best)	**0.859**	**0.631**	**0.740**
DP,BPP,BAUG	0.769	0.480	0.539
BP,NPP,NAUG	0.763	0.477	0.508
DP,BPP,NAUG	0.645	0.350	0.347
Baseline (DP,NPP,NAUG)	0.547	0.271	0.218

achieves $mAP_{50} = 0.763$, a 39% improvement without any preprocessing or augmentation. The jump from 640 px to 1280 px input resolution is the dominant single change: at higher resolution, oblique cracks become distinguishable from longitudinal cracks by their angular geometry, and fine crack structures that were sub-pixel at 640 px become resolvable. Augmentation applied with default hyperparameters (DP,BPP,BAUG, $mAP_{50} = 0.769$) slightly outperforms best hyperparameters alone, confirming that both axes are independently valuable. The full combination outperforms every partial configuration by a clear margin.

5.3. Per-class analysis

Table 4 shows per-class AP_{50} for the baseline and the best configuration.

Oblique crack records the largest absolute gain (+0.435) but remains the second-weakest class. This suggests that baseline failure on oblique cracks was primarily a resolution problem rather than a genuine class ambiguity: once the input size preserves diagonal crack structure at 1280 px, the model can distinguish it from longitudinal instances. Pothole's gain (+0.328) is more attributable to augmentation-driven sample diversity, consistent with its status as the

Table 4: Per-class AP_{50}: Baseline vs. Best (BP,BPP,BAUG).

Class	Baseline	Best	Gain
Alligator crack	0.623	0.949	+0.326
Longitudinal crack	0.625	0.901	+0.276
Oblique crack	0.362	0.797	+0.435
Pothole	0.446	0.774	+0.328
Repair	0.624	0.877	+0.253
Transverse crack	0.605	0.857	+0.252
mAP_{50}	0.547	0.859	+0.312

most underrepresented class. Repair and transverse crack show the smallest gains, likely because their visual signals were already learnable at lower resolution. Figure 1 shows the per-class comparison across all five configurations.

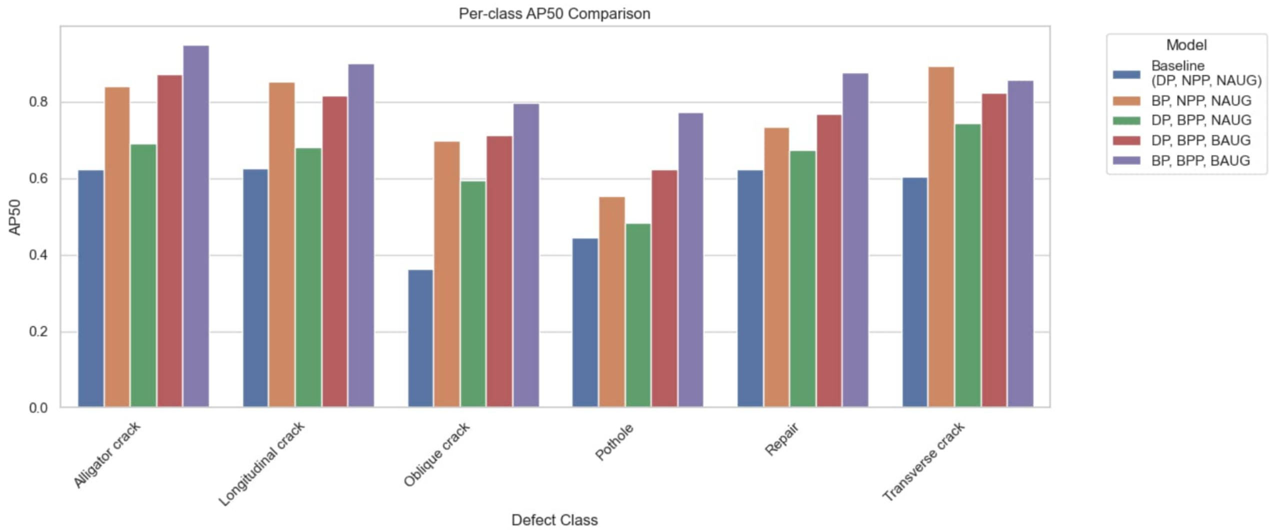
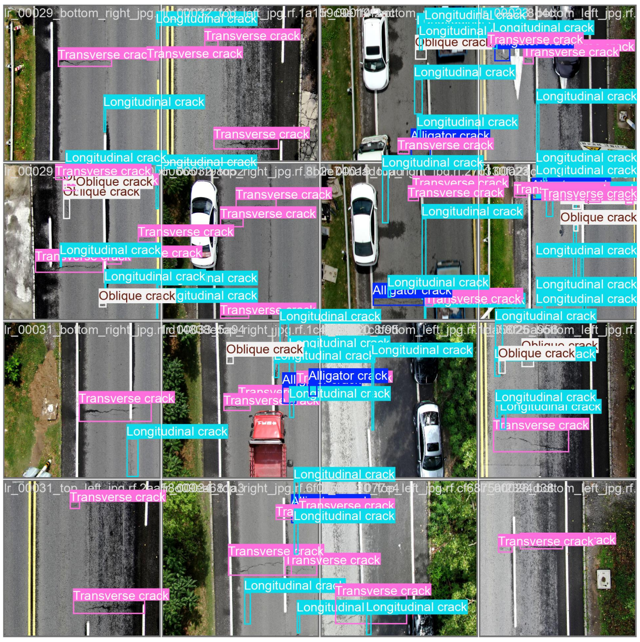


Figure 1: Per-class AP_{50} across all five ablation configurations. Oblique crack and pothole are the most sensitive to configuration choice.

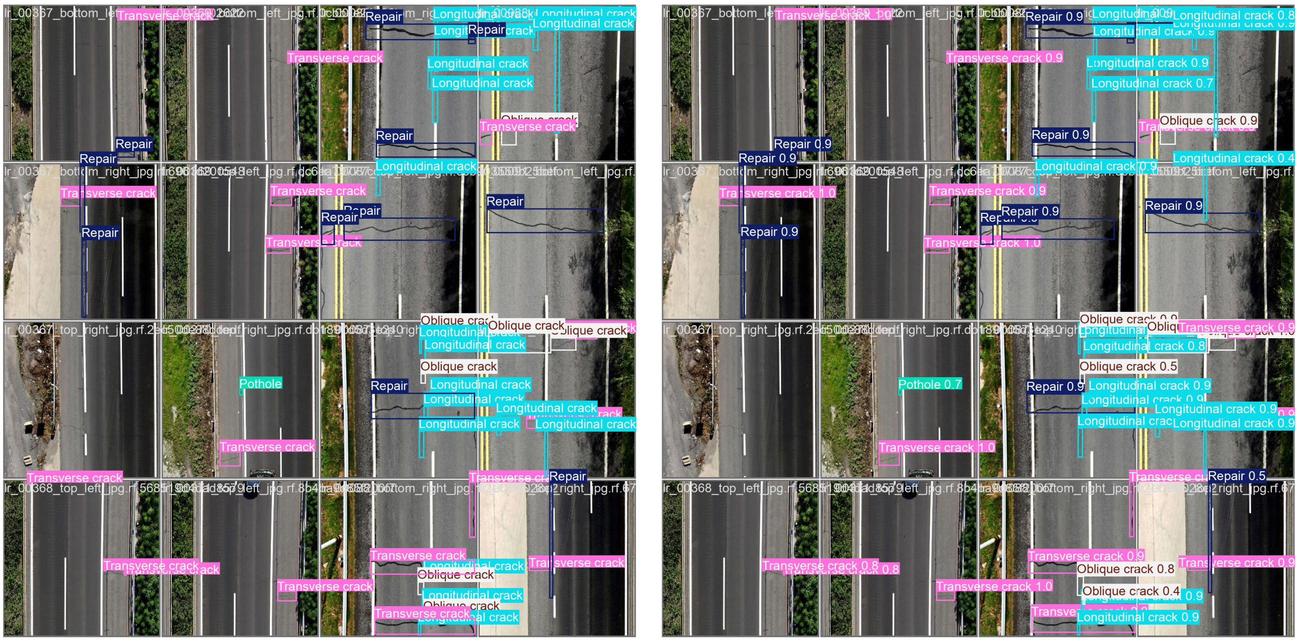
Figures 2 and 3 provide qualitative support for these numbers. The best configuration (BP,BPP,BAUG) correctly localises the dominant crack types at high confidence on both urban road sections and highway surfaces, while pothole instances, consistent with their lower AP_{50}, occasionally receive confidence scores in the 0.4–0.7 range.



(a) Ground-truth annotations, urban road tiles.

(b) Model predictions with confidence scores (BP,BPP,BAUG).

Figure 2: Validation batch on urban road sections. Each panel is a 4×4 mosaic of held-out UAV-PDD2023 tiles. Longitudinal, transverse, and oblique cracks are correctly localised at confidence ≥ 0.9.



(a) Ground-truth annotations, highway tiles.

(b) Model predictions with confidence scores (BP,BPP,BAUG).

Figure 3: Validation batch on highway sections containing repair patches and potholes. Repair patches are reliably identified; pothole confidence is lower (0.4–0.7), consistent with class imbalance in the training data.

5.4. Training convergence

The best configuration continues to improve through epoch 280 of the 300-epoch budget, confirming that a shorter training run would underestimate what the combined pipeline achieves. In contrast, the baseline converges rapidly but plateaus with residual validation loss indicative of overfitting. BP,NPP,NAUG (best hyperparameters but no preprocessing or augmentation) shows an unstable optimisation period between epochs 80 and 150, suggesting that high-resolution training without augmentation creates an irregular loss landscape until the model settles into a stable feature representation. The divergence between BP,BPP,BAUG and BP,NPP,NAUG grows continuously after epoch 50, reaching a gap of approximately 0.14 in $\mathrm{mAP}_{50\text{–}95}$ by epoch 300. This confirms that preprocessing and augmentation provide cumulative benefit rather than a one-time boost.

6. GSD-ADAPTIVE ORTHOPHOTO INFERENCE PIPELINE

6.1. Scale mismatch

UAV-PDD2023 images have a GSD of approximately 0.7 cm/px. Real survey orthophotos, stitched from higher-altitude flights, typically have GSDs of 1–3 cm/px. This creates a direct scale mismatch: a 1280 px tile at 2.5 cm/px covers $1280 \times 0.025 = 32$ m of road, whereas the same tile at the training GSD covers only $1280 \times 0.007 \approx 9$ m. A pothole that occupies 50 px in a training image appears at roughly 14 px in a field orthophoto tile of the same physical area, which is below the reliable detection threshold.

6.2. Adaptive tiling

The fix is to match ground coverage rather than pixel count. Given training GSD g_{tr} and field GSD g_{f}, the field tile size T_{f} is set as

$$T_{\mathrm{f}} = T_{\mathrm{tr}} \cdot \frac{g_{\mathrm{tr}}}{g_{\mathrm{f}}}, \tag{1}$$

where $T_{\mathrm{tr}} = 1280$ px is the training input size. For field orthophotos at 2.5 cm/px: $T_{\mathrm{f}} = 1280 \times 0.7/2.5 \approx 358$ px, rounded to 400 px. A 400 px tile at 2.5 cm/px covers 10 m of road, which matches the 9 m coverage of a 1280 px training tile. Ultralytics' internal upsampling then presents each 400 px tile to the model at 1280 px, preserving the spatial context the model

Table 5: Field orthophotos used in deployment evaluation.

Scene	Dimensions (px)	GSD	Coverage
rychlewskiego	10,769×14,461	2.48 cm/px	267 × 359 m
JP2	3,795×15,040	2.54 cm/px	97 × 383 m
road2	28,576×74,778	1.13 cm/px	322 × 843 m

expects. Tiles are extracted with a 25% overlap (stride 300 px) so that any defect near a boundary appears fully in at least one neighbouring tile.

6.3. Cross-tile NMS and georeferencing

After inference, detections from all tiles are expressed in full-image pixel coordinates. Per-class non-maximum suppression is then applied globally across the entire orthophoto using OpenCV's `dnn.NMSBoxes` with an IoU threshold of 0.45. This removes duplicate detections that arise where tiles overlap.

Each surviving bounding box centre is converted to geographic coordinates via the six-parameter affine transform embedded in the GeoTIFF:

$$\begin{pmatrix} x_{\text{geo}} \\ y_{\text{geo}} \end{pmatrix} = \begin{pmatrix} c + a \cdot \text{col} + b \cdot \text{row} \\ f + d \cdot \text{col} + e \cdot \text{row} \end{pmatrix}, \tag{2}$$

extracted via rasterio. For UTM-projected rasters the result is subsequently converted to WGS84 (EPSG:4326) using pyproj. Physical bounding box dimensions follow as $w = (x_2 - x_1) \cdot g_{\text{f}}$ and $h = (y_2 - y_1) \cdot g_{\text{f}}$.

The pipeline produces four output artefacts per orthophoto: an annotated GeoTIFF with colour-coded bounding boxes; a detection CSV with class label, confidence, latitude, longitude, width and height in metres, and area in m^2; a GeoJSON polygon file loadable in QGIS; and a confidence score histogram.

7. FIELD DEPLOYMENT EVALUATION

7.1. Test orthophotos

Three georeferenced orthophotos not present in the training set are used to evaluate the pipeline under genuine domain shift. Their properties are summarised in Table 5.

All three scenes were captured in a different geographic region from UAV-PDD2023 and at higher altitude. GSD ranges from 1.13 to 2.54 cm/px, well above the training GSD of 0.7 cm/px.

7.2. Detection results

Table 6 reports total detections at a confidence threshold of 0.10. Per-class counts are in Table 7.

Table 6: Detection summary across field orthophotos (confidence ≥ 0.10, Best model BP,BPP,BAUG).

Scene	Total	Mean conf.	Dominant class
rychlewskiego	104	0.249	Oblique crack (62)
JP2	49	0.302	Oblique crack (16)
road2	2201	0.297	Oblique crack (1065)

Table 7: Per-class detection counts across field orthophotos. Long. = longitudinal; Trans. = transverse.

Scene	Allig.	Long.	Oblique	Pot.	Repair	Trans.
rychlewskiego	11	3	62	9	0	19
JP2	8	14	16	5	0	6
road2	147	438	1065	59	18	474

All six defect classes are identified in the road2 mosaic. The dominance of oblique cracks is consistent with the turning stress patterns expected in the surveyed road sections; the road2 scene includes a roundabout with a notably high concentration of oblique and transverse cracks near the turning lanes. Repair patches are absent in the two smaller scenes, which likely reflects the actual condition of those road segments.

Mean detection confidence (0.25 to 0.30) is substantially lower than the 0.9+ values seen on the held-out test set. Confidence score histograms for all three scenes are right-skewed, with modal bins between 0.10 and 0.15. This is the expected signature of domain shift from lower-GSD training data to higher-GSD field imagery. The conservative threshold of 0.10 is adopted to maximise recall for a survey use case where a human operator reviews results in

GIS. Raising it to 0.25 would discard roughly 40% of detections in the rychlewskiego scene without a verifiable precision gain in the absence of ground-truth labels.

7.3. GIS validation

The exported GeoJSON and annotated GeoTIFF files were loaded as separate layers in QGIS for all three orthophotos. Detection polygon outlines aligned correctly with visible pavement features in the source raster, and the coordinate reference system reported by QGIS matched the CRS embedded in each source file. This confirms that the affine transform in Eq. (2) is correctly applied and that the exported vector outputs are suitable for direct use in maintenance scheduling workflows.

8. DISCUSSION

8.1. What the ablation reveals

The clearest finding from the ablation is that input resolution dominates the other axes. Going from 640 px to 1280 px alone recovers more than two-thirds of the total mAP_{50} gain achievable by the full pipeline. This is not surprising for a task that involves thin, elongated objects, but it is not always reported quantitatively in the literature. Preprocessing and augmentation each add further gains and are complementary to each other and to the resolution change.

8.2. Comparison with related work

Table 8 places the best configuration alongside related published results. Direct numerical comparison is imprecise because each study uses a different dataset and defect taxonomy, but the positioning is instructive. Our mAP_{50} of 0.859 on a six-class UAV dataset compares favourably with Sun's 0.774 on a four-class ground-level benchmark [2]. Zhang's 93.9% classification accuracy with ViT on UAV imagery [18] concerns image-level classification rather than bounding-box detection, so the tasks are not directly comparable. The consistency between the strong ViT result in that study and YOLO26's area attention mechanism both point to the value of long-range spatial features for crack detection.

Table 8: Comparison with related work. Datasets and defect taxonomies differ across rows.

Study	Architecture	Metric	Value
Sun [2]	YOLOv8+SPD-Conv	mAP_{50}	0.774
Wang [3]	Improved YOLOv8s	mAP_{50}	+5.8% vs base
Zhang [18]	ViT	Accuracy	93.9%
This work	YOLO26s	mAP_{50}	**0.859**

8.3. Limitations

No manually annotated ground-truth labels exist for the field orthophotos, so quantitative metrics such as precision and recall cannot be computed for the deployment evaluation. The ablation is conducted on a single architecture (YOLO26s) trained on one dataset (UAV-PDD2023); whether the relative ranking of the three training axes holds for other architectures or geographies is untested. Tile boundary effects can split large defects that span a non-overlapping tile edge into partial bounding boxes that cross-tile NMS cannot merge. In mixed urban orthophotos, the model occasionally fires on non-road surfaces such as rooftops whose texture resembles weathered asphalt, as the model has no explicit scene-semantic context. Masking inference to a road network polygon from OpenStreetMap before tiling is a practical mitigation that requires no model changes.

9. CONCLUSION

We presented a systematic study of training configuration choices and a GSD-aware deployment pipeline for UAV-based pavement defect detection. The five-configuration ablation shows that input resolution is the dominant axis: moving from 640 px to 1280 px training input, combined with Roboflow preprocessing and moderate augmentation, raises mAP_{50} from 0.547 to 0.859 on UAV-PDD2023 with YOLO26s. The GSD-adaptive tiling pipeline then bridges the gap between the benchmark dataset and real survey orthophotos by selecting tile dimensions to match ground coverage rather than pixel count. Applied to three field orthophotos covering up to 322 × 843 m of road, the pipeline produces georeferenced GeoJSON and GeoTIFF outputs validated in QGIS, identifying all six target defect classes. Future work will target the residual domain gap through fine-tuning on field-annotated tiles and multi-GSD dataset construction, and will quantify recall on manually labelled field orthophotos.

ACKNOWLEDGMENTS

This research was supported in part through computational resources of the HPC facilities at HSE University [17].

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

---

[1] S. Muzira, *Paving the Way to Better Roads*, World Bank Feature Article (2021). https://www.worldbank.org/en/news/feature/2021/04/22/paving-the-way-to-better-roads

[2] Z. Sun, L. Zhu, S. Qin, Y. Yu, R. Ju, and Q. Li, Electronics **13**, 2413 (2024). https://doi.org/10.3390/electronics13122413

[3] J. Wang, R. Meng, Y. Huang, L. Zhou, L. Huo, Z. Qiao, and C. Niu, Sci. Rep. **14**, 16758 (2024). https://doi.org/10.1038/s41598-024-67953-3

[4] J. Dang and L. Wang, arXiv:2407.06013 (2024).

[5] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li, Neurocomputing **338**, 139 (2019). https://doi.org/10.1016/j.neucom.2019.01.036

[6] J. Sun, A. Cao, and H. Duan, Appl. Sci. **13**, 2227 (2023). https://doi.org/10.3390/app13042227

[7] I. Abdel-Qader, O. Abudayyeh, and M. E. Kelly, J. Comput. Civ. Eng. **17**, 255 (2003). https://doi.org/10.1061/(ASCE)0887-3801(2003)17:4(255)

[8] E. Zalama, J. Gomez-Garcia-Bermejo, R. Medina, and J. Llamas, Comput. Aided Civ. Infrastruct. Eng. **29**, 342 (2014). https://doi.org/10.1111/mice.12042

[9] H. Oliveira and P. L. Correia, in *Proc. Eur. Signal Process. Conf. (EUSIPCO)* (2009), p. 1.

[10] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, in *Proc. IEEE Int. Conf. Image Process. (ICIP)* (2016), pp. 3708–3712. https://doi.org/10.1109/ICIP.2016.7533052

[11] Y. Li, Z. Ma, X. Liu, and L. Ma, Appl. Sci. **12**, 12085 (2022). https://doi.org/10.3390/app122312085

[12] Y. Tian, Q. Ye, and D. Doermann, arXiv:2502.12524 (2025).

[13] A. Mustafa, M. Yu. Kataev, and D. S. Efremenko // Light & Enigineering. 2026 [accepted].

[14] F. C. Akyon, S. O. Altinuc, and A. Temizel, in *Proc. IEEE Int. Conf. Image Process. (ICIP)* (2022), pp. 966–970. https://doi.org/10.1109/ICIP46576.2022.9897990

[15] W. Pang, Y. Li, and D. Guo, arXiv:2309.10093 (2023).

[16] H. Yan and J. Zhang, Data Brief **51**, 109692 (2023). https://doi.org/10.1016/j.dib.2023.109692

[17] P. S. Kostenetskiy, R. A. Chulkevich, and V. I. Kozyrev, J. Phys.: Conf. Ser. **1740**, 012050 (2021). https://doi.org/10.1088/1742-6596/1740/1/012050

[18] Y. Zhang, J. Chen, Z. Wu, X. Guo, and S. Jia, KSCE J. Civ. Eng. **29**, 100095 (2025). https://doi.org/10.1016/j.kscej.2024.100095