

Article | Received 10 March 2026; Revised 12 May 2026; Accepted 28 May 2026; Published 27 August 2026
<https://doi.org/10.55092/pcs20260012>

ELW–YOLO: A next–generation lightweight model for pavement defects detection

Yufeng Xu*, Shiling Jin, Chongwei Yang and Bo Tong

School of Artificial Intelligence, Mianyang City College, Mianyang, Sichuan, China

E–mail: 3320240441@mycc.edu.cn.

Abstract: Reliable pavement–distress detection remains difficult because defects vary greatly in size, geometry, and visual appearance. We present ELW–YOLO, a lightweight detection framework built on YOLOv2026n with three targeted additions: (1) an Enhanced Multi-Scale Attention (EMA) block at P3/P4 that captures spatial–channel cues at several scales and suppresses background interference; (2) a Localization Quality Estimation (LQE) head that predicts localization quality alongside class scores to improve NMS ranking; and (3) a Wise–IoU loss with attention-based outlier suppression for high aspect–ratio variance. On RDD2022 China (26,620 images, four defect classes), ELW–YOLO reaches 86.0% mAP@0.5 and 55.9% mAP@0.5:0.95, beating the baseline by 3.6 and 4.4 points while using only 2.7M parameters, 6.5 GFLOPs, and 74 FPS.

Keywords: Deep learning; pavement defect detection; YOLOv2026; multi–scale feature enhancement; lightweight network.

1. Introduction

Monitoring pavement condition is essential for both road upkeep and the safety of road users. The most common distress types—longitudinal, transverse, and alligator cracks and potholes—have become more frequent as traffic loads grow, creating safety risks and shortening pavement service life. Traditional manual inspection is too slow to meet today’s maintenance demands, which motivates the development of automated detection approaches [1].

Yet building such systems is challenging. Defects differ widely in type, scale, and shape; road scenes contain distracting background clutter; and there is an inherent tension between detection accuracy and the limited compute on embedded devices. Heavy networks deliver accuracy at high resource cost, whereas compact ones often lose robustness. Closing this accuracy–efficiency gap is therefore a key research aim [2].

The main contributions of this work are:



Copyright©2026 by the authors. Published by ELSP. This work is licensed under Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

- 1) We embed an EMA module in the backbone so that high-level semantic features respond more strongly to defect regions, effectively suppressing complex background interference. An LQE branch is added to the detection head, linking classification confidence with localization quality to refine the NMS ranking and improve detection precision.
- 2) We replace the default regression loss with Wise-IoU, whose dynamic non-monotonic focusing stabilizes bounding-box regression for irregular cracks with large aspect-ratio changes and fuzzy edges.

2. Related Work

Early work combined manual inspection with classical image processing on cameras and laser scanners, using edge detection and texture analysis. Guo *et al.* [3] applied ground-penetrating radar for crack identification; Ha *et al.* [4] built ITVCrack to extract cracks from point clouds. These methods still weaken under changing illumination and diverse road conditions.

Machine-learning methods such as SVM and random forests raised accuracy by learning complex feature relations. Sari *et al.* [5] fused SVM with Otsu for crack segmentation; Peng *et al.* [6] used random structured forests with a triple-threshold scheme; Liu *et al.* [7] reached 95.8% accuracy via improved multi-scale Retinex. Yet these rely on large labeled sets and hand-crafted features, limiting generalization [8].

More recently, deep CNNs enabled solutions that learn high-level representations directly from data [9]. Zhang *et al.* [10] enhanced YOLOv3 with an MLAB module, lifting mAP to 68.75%; Deng *et al.* [11] paired YOLOv5 with Res-UNet for accurate crack measurement. Even so, such deep models demand heavy computation, hindering real-time deployment [12].

3. Methods

3.1. Overall architecture

ELW-YOLO unifies the three proposed components—EMA attention inserted in the neck, an LQE branch attached to the detection head, and the Wise-IoU regression loss. The resulting compact structure is well suited to mobile or edge devices that require real-time inference.

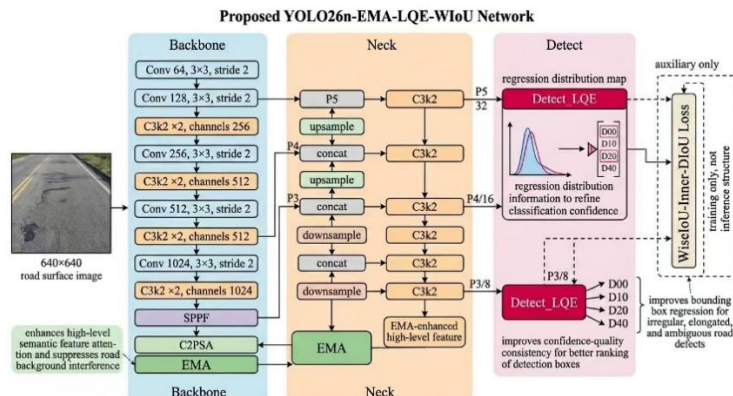


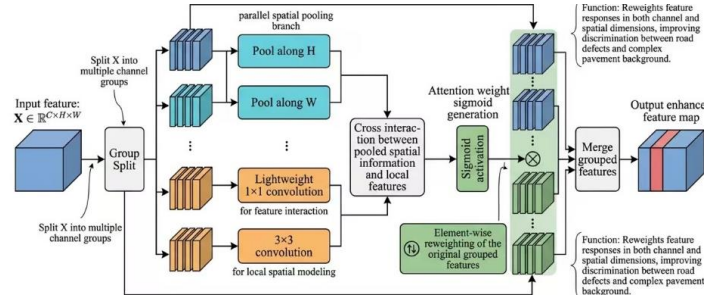
Figure 1. Overall architecture of ELW-YOLO.

3.2. YOLOv2026 network

Our framework adopts YOLOv2026n, the newest compact member of the YOLO family. By casting detection as a single regression task, YOLO achieves an attractive trade-off between accuracy and speed [13].

3.3. EMA: Enhanced multi-scale attention module

Unlike ordinary channel-attention operators that act at one scale only, EMA captures both inter-channel and inter-spatial relations through parallel branches. It partitions the channel dimension into groups of sub-features and applies cross-space learning so that local detail and global context are both retained, amplifying defect-relevant features while damping background. By replacing costly non-local attention with this lightweight cross-dimensional interaction, EMA cuts computational overhead. We place EMA inside the C3 modules at the P3 and P4 levels.

**Figure 2.** EMA module structure.

3.4. LQE: Localization quality estimation

One-stage detectors commonly suffer a mismatch: high classification confidence does not guarantee accurate localization when NMS is applied. To address this, we add an LQE branch that estimates an IoU quality value for every candidate box. Following the IACS idea in VarifocalNet, the final score is the product of predicted IoU quality and class confidence, so better-localized boxes rank first; the branch is jointly optimized with Varifocal Loss. Drawing on GFL [14], we also model each box as a general distribution rather than a single point, which further steadies training on elongated cracks.

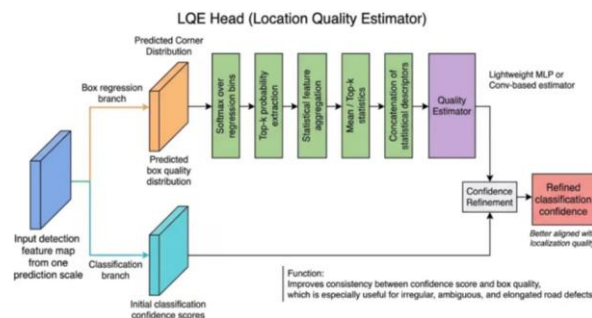


Figure 3. LQE detection head structure.

3.5. Wise-IoU: Dynamic bounding box regression loss

Standard IoU-based losses give every training sample the same gradient weight, which is ill-suited to objects of differing localization difficulty. We therefore use Wise-IoU v3, whose dynamic non-monotonic focusing grades samples by outlier degree rather than IoU alone. Boxes with moderate overlap receive the largest gain for fast convergence, while extreme outliers are down-weighted to avoid harmful gradients. This helps irregular cracks with very large aspect ratios and blurry edges.

4. Model Training and Evaluation

4.1. Experimental environment

All experiments ran on Ubuntu 22.04 with PyTorch 2.2, an NVIDIA 4090D GPU (24 GB), an 18-core Intel CPU, CUDA 12.8, and Python 3.11, using YOLOv26n as the baseline.

4.2. Evaluation metrics

We report seven metrics: Precision (P), Recall (R), mAP@0.5, mAP@0.5:0.95, FLOPs, parameter count, and FPS. Precision is the fraction of positive predictions that are correct, while Recall reflects how many actual objects are found. mAP@0.5 averages per-class AP at IoU 0.5, and mAP@0.5:0.95 averages AP over IoU 0.5–0.95 (step 0.05). FLOPs and parameters indicate computational and storage cost, and FPS reflects inference speed.

5. Experimental Results and Analysis

5.1. Dataset

For training and testing we use the RDD2022 China portion of CRDDC2022 [14], containing 3,853 images from the China_Drone subset (aerial views over Nanjing) and the China_MotorBike subset (phone-captured images from Southeast University). Each image is a 512×512 asphalt patch covering four defect classes: longitudinal crack (D00), transverse crack (D10), alligator crack (D20), and pothole (D40) [15], split 8:1:1 for training, validation, and testing. Three difficulties stand out: (1) cluttered backgrounds with shadows and wet surfaces; (2) an imbalance favoring D00 and D10; and (3) many small or elongated boxes that are hard to localize.

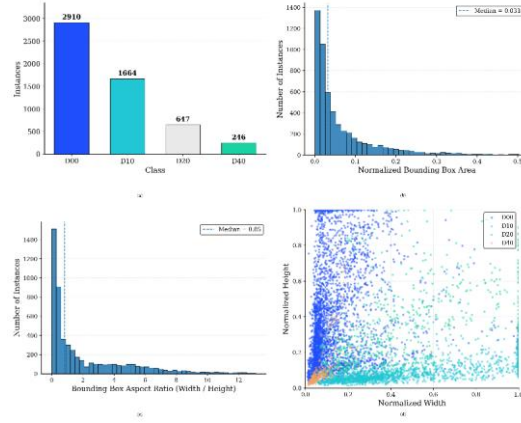


Figure 4. Dataset distribution: (a) split ratio; (b) defect type distribution.

5.2. Ablation study

Table 1 presents an incremental study in which EMA2, LQE, and Wise-IoU are added one by one to the YOLOv2026n baseline. The complete model reaches 86.0% mAP@0.5 and 55.9% mAP@0.5:0.95, gains of 3.6 and 4.4 points, with only 1.2 G and 0.4 M added to FLOPs and parameters.

Table 1. Ablation experiment results.

Model	P	R	mAP50	mAP50-95	FLOPs	FPS	Para(M)
YOLOv26n	0.811	0.776	0.824	0.515	5.3	85	2.3
YOLOv26n +EMA2	0.854	0.779	0.843	0.525	5.2	84	2.3
YOLOv26n +EMA2+LQE	0.868	0.795	0.858	0.558	6.5	73	2.7
ELW-YOLO	0.870	0.796	0.860	0.559	6.5	74	2.7

5.3. Effectiveness of EMA2

Adding EMA2 lowers GFLOPs by 0.1 G (5.3→5.2 G) and increases parameters by merely 0.01 M, yet lifts mAP@0.5:0.95 by 0.88%. This shows that EMA’s gain comes from more efficient feature use, not from extra parameters.

5.4. Effectiveness of LQE

LQE mainly boosts the high-IoU measures: mAP@0.5 and mAP@0.5:0.95 each rise by 2.5%, and Recall goes up by 1.2%. The gain is larger at stricter IoU thresholds, confirming that LQE helps surface well-localized predictions during ranking.

5.5. Effectiveness of Wise-IoU

Wise-IoU v3 with its non-monotonic focusing module yields the strongest results on every metric, especially mAP@0.5 and mAP@0.5:0.95. Its dynamic focusing redistributes gradient effort by outlier degree, which is vital for steady regression on irregular cracks.

5.6. Comparison with other methods

Table 2. Comparison with state-of-the-art methods.

Model	P	R	mAP50	mAP50-95	FLOPs(G)	FPS	Para(M)
YOLOv3-tiny	0.785	0.705	0.770	0.425	18.9	107	12.1
YOLOv5	0.808	0.725	0.794	0.442	7.3	55	2.5
YOLOv7-tiny	0.783	0.713	0.774	0.380	12.0	54	4.8
YOLOv8n	0.825	0.712	0.807	0.502	8.3	65	3.05
Rtdetr-l	0.830	0.715	0.800	0.465	108	19	32.7
Wang's	0.710	0.680	0.812	0.472	6.5	33	2.25
Zhou's	0.690	0.706	0.712	0.390	5.1	50	2.1
ELW-YOLO	0.870	0.796	0.860	0.559	6.5	74	2.7

Table 2 sets ELW-YOLO against recent methods on the RDD2022 test set under the same protocol. Our model records the top mAP@0.5 (0.860) and mAP@0.5:0.95 (0.559) while staying efficient at 6.5 GFLOPs and 2.7 M parameters. It clearly surpasses YOLOv3-tiny (12.1 M, 0.770) and YOLOv7-tiny (4.8 M, 0.774), and beats RTDETR-l on mAP50-95 (0.559 vs 0.465) with 2.7 M vs 32.7 M parameters. Against Wang's and Zhou's similar-size models, ELW-YOLO is superior on every reported metric.

6. Conclusion

In this work we propose ELW-YOLO, a compact real-time detector for pavement distress that runs on portable and edge devices. With three focused upgrades—EMA2 for richer feature fusion, an LQE branch for better detection ranking, and Wise-IoU for stable box regression—the model exceeds the YOLOv2026n baseline by 3.6 and 4.4 points in mAP50 and mAP50-95. Running at 74 FPS with just 6.5 G FLOPs and 2.7 M parameters, it is well suited to mobile use. Future efforts will test it on larger, more diverse road scenes and keep refining it for real deployment in smart road-maintenance systems.

Acknowledgments

The authors would like to thank Prof. Tianrui Li (School of Computing and Artificial Intelligence, Southwest Jiaotong University) and Prof. Haoxi Zhang (School of Cyberspace Security, Chengdu University of Information Technology) for their invaluable guidance and mentorship throughout the study.

Authors' contribution

Y.X. and B.T. conceived the study; Y.X. and C.Y. designed the method; S.J. and C.Y. carried out validation; Y.X. performed formal analysis; S.J. conducted the investigation; C.Y. and S.J. managed the data; Y.X. wrote the first draft; B.T. reviewed and edited the manuscript. All authors have read and approved the final version.

Conflicts of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Yu X, Marinov M. A Study on Recent Developments and Issues with Obstacle Detection Systems for Automated Vehicles. *Sustainability*.2020, 12(8): 3281.
- [2] Guo K, He C, Yang M, Wang S. A pavement distresses identification method optimized for YOLOv5s. *Sci. Rep.* 2022, 12: 3542.
- [3] Ha J, Kim D, Kim M. Assessing severity of road cracks using deep learning-based segmentation and detection. *J. Supercomput.* 2022, 78: 17721–17735.
- [4] Zhong M, Sui L, Wang Z, Hu D. Pavement crack detection from mobile laser scanning point clouds using a time grid. *Sensors.* 2020, 20(15): 4198.
- [5] Sari Y, Prakoso P, Baskara A. Road crack detection using SVM and Otsu algorithm. In *Proceedings of 2019 6th International Conference on Electric Vehicular Technology (ICEVT)*, 2019, pp. 349–354.
- [6] Peng C, Yang M, Zheng Q, Zhang J, Wang D, *et al.* A triple-thresholds pavement crack detection method leveraging random structured forest. *Constr. Build. Mater.* 2020, 263.
- [7] Liu S, Han Y, Xu L. Recognition of road cracks based on multi-scale Retinex fused with wavelet transform. *Array.* 2022, 15.
- [8] Abagiu M, Cojocaru D, Manta F, Mariniuc A. Detecting Machining Defects inside Engine Piston Chamber with Computer Vision and Machine Learning. *Sensors.* 2023, 23(2): 785.
- [9] Jia Y, Yang C, Gao Y. LL-YOLO: A lightweight pavement disease detection algorithm in low-light environments. *Measurement.* 2026, 258: 119494
- [10] Zhang Y, Ma Z, Song X, Wu J, Liu S, *et al.* Road surface defects detection based on IMU sensor. *IEEE Sens. J.* 2022, 22: 2711–2721.
- [11] Deng L, Zhang A, Cuo J, Liu Y. An Integrated Method for Road Crack Segmentation and Surface Feature Quantification under Complex Backgrounds. *Remote Sens.* 2023, 15(6): 1530.
- [12] Hsieh C, Jia H, Huang W, Hsieh M. Deep Learning-Based Road Pavement Inspection by Integrating Visual Information and IMU. *Information.* 2024, 15, 239.
- [13] Sapkota R, Cheppally RH, Sharda A, Karkee M. YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection. *arXiv preprint arXiv:2509.25164*. 2025 Sep 29.
- [14] Arya D, Maeda H, Ghosh S, Toshniwal D, Sekimoto Y. RDD2022: A multi-national image dataset for automatic road damage detection. *Geosci. Data J.* 2024, 11(4): 846–862.

-
- [15] Arya D, Maeda H, Ghosh S, Toshniwal D, Omata H, *et al.* Crowdsensing-based road damage detection challenge. In *Proceedings of the 2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 6378–6386.
- [16] Huang P, Wang S, Chen J, Li W, Peng X. Lightweight Model for Pavement Defect Detection Based on Improved YOLOv7. *Sensors*. 2023, 23(16): 7112.