Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.39099

# *A Review of Vision-Based Road Distress Recognition across Diverse Scenarios*

**Huiliang Pan[1], Yan Jin[2], Xuanyi Pan[3], Yaocheng Mo[4], Ping Zhu[1,*]**

[1]*School of Engineering, Hangzhou City University, Hangzhou, China*
[2]*School of Journalism and Communication, Hangzhou City University, Hangzhou, China*
[3]*School of Computer and Computing Science, Hangzhou City University, Hangzhou, China*
[4]*Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China*
**Corresponding Author. Email: zhup@hzcu.edu.cn*

***Abstract.*** Rather than treating road distress detection as a single, uniform task, this review considers how inspection needs change between carriageways and sidewalks. It first follows the development of recognition methods, from field surveys and rule-based image processing to handcrafted descriptors, deep feature learning, and real-time detection. Particular attention is given to the YOLO family, including YOLOv5, YOLOv8, YOLOv11, YOLOv12, and YOLOv26. Their feature extraction designs, detection heads, and deployment requirements are discussed alongside published benchmark and road-distress results. The reported evidence shows that a larger network is not necessarily the better choice: designs addressing the visual properties of pavement damage can offer accuracy gains with lower computational demands. Such comparisons must, however, be interpreted in light of the dataset, evaluation protocol, and hardware used. The review draws out the implications for choosing models in practical road inspection.

***Keywords:*** Road distress recognition, Computer vision, Deep learning, YOLO

## 1. Introduction

Pavement defects develop under repeated traffic loads, environmental changes in temperature and moisture, and the gradual aging of road materials. Cracks, potholes, surface raveling, and deformation are among the visible consequences. Besides affecting ride quality, they may create hazards for road users. Maintenance decisions consequently depend on knowing what damage is present, where it occurs, and how severe it is [1]. These requirements are not identical for every paved surface. Carriageways and sidewalks use different materials, exhibit different patterns of deterioration, and are usually photographed under different conditions [2-4]. In-person surveys allow inspectors to exercise engineering judgment, but their results may vary between individuals; they are also time-consuming and can require personnel to work close to moving traffic [5]. The need to inspect larger networks more frequently has therefore motivated research into automated, image-based assessment methods.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.39099

The earliest image-based approaches identified pavement defects through operations such as intensity thresholding, edge detection, and morphological filtering. Although relatively straightforward to implement, these techniques relied on prescribed thresholds or manually selected visual cues. A shadow or a change in aggregate texture could consequently be mistaken for a defect [2]. Research subsequently moved toward data-driven feature learning as image collection became more systematic and deep learning methods matured. Learned representations have been used not only to identify damage, but also to locate and segment it and to support condition assessment [6-8]. Where geometric measurements are needed, photogrammetric reconstruction and three-dimensional imaging provide information that an ordinary image cannot supply [9, 10]. Within the object detection literature, YOLO is especially relevant to mobile inspection because classification and localization take place in a single-stage pipeline [11]. Road-specific variants have explored multiscale fusion, preservation of fine crack details, and computationally economical designs; some also seek to connect detections with quantitative assessments of damage [12-14].

The discussion is organized around the requirements of different inspection settings rather than around algorithms alone. Section 2 identifies the pavement defects encountered on carriageways and sidewalks and explains why their assessment needs differ. Section 3 then examines the successive approaches used for visual recognition, including their principal limitations. Section 4 considers the design choices of several YOLO families and compares results reported for general object detection and road distress. The closing section draws together the implications for model selection and field deployment.

## 2. Road distress and the need for inspection

### 2.1. Distress types of carriageways

Asphalt and cement concrete are both widely used on carriageways, but the damage patterns they produce are not the same. On asphalt, longitudinal, transverse, and alligator cracks are common [2]. Potholes arise where surface material has been lost, often leaving an uneven outline and exposed texture. In photographs, however, some potholes resemble repaired patches, pooled water, or darker stretches of pavement. Their appearance illustrates why a detector cannot rely on color contrast alone [3]. Rutting and other local deformations are chiefly changes in surface shape. Raveling and bleeding, by comparison, can affect broader areas and may be recognized through their surface appearance [5].

For cement concrete carriageways, the relevant defects include slab cracks, damaged joint edges, faulting, and failure of joint seals. The challenge is not always to find a conspicuous mark: a sound joint and a deteriorated one can look much alike. Joint width, continuity, and spalling along the edges therefore need to be considered together. Faulting presents a different problem because it involves an elevation difference that a two-dimensional image cannot reliably quantify. Depth or three-dimensional profile measurements are required when its severity is assessed [8, 10]. Photogrammetry and three-dimensional sensing can supply such measurements, although this benefit comes with additional equipment and processing demands [9, 10].

### 2.2. Distress types of sidewalks

Sidewalk surfaces often consist of precast concrete units, brick, or stone rather than continuous pavement. Cracks and broken or missing units are visible forms of deterioration, while displacement and differential settlement change the level of the walking surface. Roots growing beneath the

DOI: 10.54254/2755-2721/2026.39099

paving can lift individual units and create a tripping hazard. For this reason, a sidewalk survey must consider passability and the distribution of obstacles, not just damage to the material itself [4]. Some defects are easy to recognize in a photograph; a missing slab is one example. A slab that rocks when stepped on is more difficult, since looseness may not be apparent without observing its movement or measuring a response to loading. Similarly, misaligned joints may suggest settlement, but the extent of settlement still needs an elevation measurement [4, 9].

Sidewalk photographs are generally captured from shorter distances than carriageway views, which can make small defects easier to see. That advantage is offset by changes in camera angle and by frequent occlusions from pedestrians, bicycles, or street furniture. Regular patterns of paving joints create another source of confusion because they may resemble cracks in an image [4].

## 3. Evolution of road distress recognition technologies

### 3.1. Manual inspection

In a manual survey, trained inspectors examine roads on foot or from vehicles traveling at low speed. They note visible damage and use their experience of pavement condition and the surrounding site to judge its extent [5, 15]. Such contextual judgment is a strength of the method. Its cost becomes apparent, however, when the same work must be repeated across a large network: staffing and survey time grow with both road length and inspection frequency [15]. There are operational consequences as well. Personnel may need to work beside active traffic, and temporary closures or slowly moving survey vehicles can interrupt normal traffic flow. Even when the same inspection procedure is followed, differences in experience, weather, and fatigue can affect the recorded assessment [5]. Manual surveys thus remain useful, but their workload, exposure to traffic, and between-inspector variability limit their suitability for frequent, network-wide monitoring.

### 3.2. Traditional image processing

A conventional image-processing pipeline typically begins by converting images to grayscale or reducing noise, then identifies possible defects with thresholding, edge operators, and morphological operations. When a crack is clearly separated from its surroundings, this approach is inexpensive to run and its output is relatively easy to interpret [2]. Edges highlight local changes in intensity; morphological filtering can join broken crack traces and suppress isolated responses [2, 16]. The difficulty lies in the road surface itself. Shadows, standing water, white markings, exposure differences, and aggregate patterns may all create strong intensity changes without indicating damage. Conversely, a faint crack may produce little contrast. Thresholds chosen for one road or lighting condition therefore transfer poorly to another, making false detections a persistent problem for rule-based methods in complex scenes [2, 15, 16].

### 3.3. Handcrafted feature extraction

Handcrafted feature approaches introduce a separate representation step between the image and the classifier. Texture, shape, gradient, and transform-domain measurements are computed first; methods such as support vector machines, random forests, or AdaBoost then use those measurements to distinguish damage from the background. Examples include gray-level co-occurrence matrices, local binary patterns, histograms of oriented gradients, wavelet responses, and geometric descriptors such as crack curvature and aspect ratio [5, 16]. Some studies further model

DOI: 10.54254/2755-2721/2026.39099

relationships among adjacent image regions so that discontinuous crack segments can be interpreted as parts of one structure. Because the features are explicitly defined, their contribution to a decision can be examined, and the approach can be practical with limited labels or computing resources. The same design choice is also its weakness. A descriptor developed for one pavement texture may be much less informative on another, and different defects can share similar local patterns. Regular joints between sidewalk slabs, for instance, may be classified as cracks. Handcrafted representations improve on simple pixel thresholds, but their range of application is ultimately shaped by what their designers anticipated [5, 16].

### 3.4. Deep feature learning

Deep learning changes where the visual representation comes from. Instead of specifying texture or edge descriptors in advance, convolutional neural networks learn features at several spatial scales, combining local detail with surrounding context [2]. The form of the prediction depends on the task: a network may assign an image-level class, locate damaged regions with bounding boxes, or produce a pixel-wise segmentation mask [7]. Learned features reduce the need for manual descriptor design, but they do not eliminate dependence on the training data. Annotation consistency, the variety and balance of examples, augmentation choices, and the way data are divided for evaluation can all affect the reported outcome. A strong result on one road dataset should therefore not be assumed to hold for different materials or acquisition conditions [2, 8, 10].

### 3.5. YOLO-based real-time object detection

YOLO performs localization and classification in a single-stage detection pipeline, making it an attractive option when road images must be processed quickly. Changes to its backbone, feature aggregation, and prediction heads have progressively altered the trade-off between accuracy and inference time; the actual balance still depends on model scale and the available hardware [11]. Road-oriented modifications illustrate the importance of matching the architecture to the target defects. YOLO-RD combines feature fusion with wavelet-based processing to address small or irregular damage [12]. YOLOv8-PD instead emphasizes feature enhancement and lightweight modules, seeking better recognition without a comparable increase in computational demand [13]. A parameter count alone cannot establish whether either design is suitable for an inspection vehicle or portable device: memory consumption, throughput, and end-to-end latency also need to be measured on the intended hardware [11, 13]. Detection is only one stage of an inspection workflow. A bounding box does not specify the physical dimensions of a pothole or crack [9], and consecutive video frames may repeatedly report the same defect. Practical systems also require confidence-threshold management and links between detections, geographic positions, and maintenance records [15]. Environmental robustness is equally important. In results reviewed by Abdelwahed et al. [1], YOLOv3 obtained an F1 score of 0.774 in sunny conditions on Slovakia_potholes, compared with 0.219 at night. Such a difference cautions against choosing a model from favorable-condition tests alone. Lighting, weather, and pavement setting must also be represented in evaluation.

DOI: 10.54254/2755-2721/2026.39099

## 4. Comparative capabilities of YOLO model families

### 4.1. Technical emphasis of different YOLO models

Road distress places several demands on a detector at once: thin or small defects may occupy few pixels, pavement textures can be visually confusing, and inspection may need to run in real time. The YOLO families address these demands through different architectural choices. Table 1 sets out the feature extraction, detection-head design, and deployment considerations for YOLOv5, YOLOv8, YOLOv11, YOLOv12, and YOLOv26 [17-19].

Table 1. Main technical characteristics and design priorities of YOLO model families [17-19]

| Model family | Main technical characteristics | Considerations for road distress recognition |
|---|---|---|
| YOLOv5 | Uses an anchor-based detection head with C3 modules and multiscale feature fusion. | Can serve as a baseline for road distress detection, with a focus on anchor configuration for target objects. |
| YOLOv8 | Uses C2f modules and an anchor-free, decoupled detection head with separate classification and localization branches. | Provides a basis for feature enhancement and lightweight model development for road distress detection. |
| YOLOv11 | C3k2 and C2PSA modules are used to refine feature representation while keeping computational requirements in view. | An option to consider when background clutter and hardware constraints must be handled together. |
| YOLOv12 | Area Attention and R-ELAN place greater emphasis on modeling contextual relationships. | Useful for examining whether contextual information helps trace cracks and separate defects from pavement texture. |
| YOLOv26 | Changes the regression design by removing DFL-related components, supports an optional NMS-free inference path, and revises training procedures. | Highlights the effects of head complexity and post-processing on latency when deployment hardware is limited. |

### 4.2. Comparison of detection performance across YOLO models

Table 2 brings together published COCO results for representative YOLO variants, reporting mAP@50-95 alongside parameter count, computational cost, and inference latency [18].

Table 2. Reported performance of representative YOLO models on COCO [18]

| Model | mAP@50-95 (%) | Parameters (M) | GFLOPs | Latency (ms) |
|---|---|---|---|---|
| YOLOv8-N | 37.4 | 3.2 | 8.7 | 1.77 |
| YOLOv11-N | 39.4 | 2.6 | 6.5 | 1.50 |
| YOLOv12-N | 40.5 | 2.6 | 6.0 | 1.62 |

These results do not point to a single best model on every measure. YOLOv12-N reports the highest mAP@50-95 among the three variants shown, whereas YOLOv11-N has the shorter reported inference time. Model choice consequently depends on which constraint matters most for the application.

A similar caution applies when comparing road-distress studies. Youwai et al. [20] reported that model rankings changed with the evaluation metric, model scale, and data split. Results calculated on a split also used to select checkpoints should not be interpreted in the same way as tests on an

DOI: 10.54254/2755-2721/2026.39099

independent set. Zeng and Zhong [13] provide a more controlled comparison of YOLOv5 and YOLOv8 variants. They sampled 10,000 images from RDD2022, using 8,000 for training and 2,000 for validation. Four defect classes were represented: longitudinal, transverse, and alligator cracks, together with potholes. Their experiments used 640 × 640-pixel inputs, 300 training epochs, and an NVIDIA RTX 2060 GPU. Table 3 reports the validation measurements from that study.

Table 3. Validation results for road distress detection [13]

| Model | mAP@50 (%) | mAP@50–95 (%) | Parameters (M) | GFLOPs |
|---|---|---|---|---|
| YOLOv5n | 59.5 | 25.4 | 1.7 | 4.2 |
| YOLOv5s | 67.2 | 35.5 | 7.03 | 16.0 |
| YOLOv8n | 69.2 | 38.0 | 3.1 | 8.2 |
| YOLOv8-PD | 70.6 | 39.5 | 2.3 | 6.1 |

On the RDD2022 validation split, YOLOv8n outperformed YOLOv5s in detection accuracy despite using fewer parameters and less computation. The modified YOLOv8-PD produced a further gain of 1.4 percentage points in mAP@50 and 1.5 points in mAP@50-95 relative to YOLOv8n, while reducing model complexity [13]. For the conditions tested, these results favor changes designed for pavement defects over an indiscriminate increase in network size.

## 5. Conclusion

This review examined visual road distress recognition for carriageways and sidewalks, tracing its development from manual inspection and traditional image processing to deep learning and YOLO-based detection. The reviewed studies suggest that YOLO supports rapid distress localization and that architectural improvements tailored to distress characteristics can improve accuracy while reducing computational cost. However, performance depends on pavement materials, acquisition conditions, and evaluation protocols, limiting the applicability of benchmark rankings across road scenarios. Model selection should therefore consider detection accuracy, resource requirements, and measured efficiency on target hardware. Future research should prioritize independent evaluation across diverse road environments and incorporate geometric or mechanical measurements where necessary to connect visual recognition with distress severity assessment.

## References

[1] Abdelwahed, S. H., Sharobim, B. K., Wasfey, B., & Said, L. A. (2025). Advancements in real-time road damage detection: A comprehensive survey of methodologies and datasets. *Journal of Real-Time Image Processing, 22*(4), Article 137.

[2] Manoni, L., Orcioni, S., & Conti, M. (2024). Recent advancements in deep learning techniques for road condition monitoring: A comprehensive review. *IEEE Access, 12*, 154271-154293.

[3] Safyari, Y., Mahdianpari, M., & Shiri, H. (2024). A review of vision-based pothole detection methods using computer vision and machine learning. *Sensors, 24*(17), Article 5652.

[4] Yussif, A.-M., Taiwo, R., Zayed, T., Mohandes, S. R., & Abdulai, S. F. (2025). Enhancing sustainable urban mobility: Computer vision applications in pedestrian path studies. *Journal of Traffic and Transportation Engineering (English Edition), 12*(5), 1415-1451.

[5] Kothai, R., Prabakaran, N., Srinivasa Murthy, Y. V., Cenkeramaddi, L. R., & Kakani, V. (2024). Pavement distress detection, classification, and analysis using machine learning algorithms: A survey. *IEEE Access, 12*, 126943-126960.

[6] Shang, J., Zhang, A. A., Dong, Z., Zhang, H. & He, A. (2024). Automated pavement detection and artificial intelligence pavement image data processing technology. *Automation in Construction, 168*, Article 105797.

DOI: 10.54254/2755-2721/2026.39099

[7] El Hakea, A. H. & Fakhr, M. W. (2023). Recent computer vision applications for pavement distress and condition assessment. *Automation in Construction, 146*, Article 104664.
[8] Wang, Z., Huang, Z., Zhao, Q., Xu, R. & Yan, H. (2026). A comprehensive survey on road damage detection: Datasets, methods, and platforms. *IEEE Transactions on Intelligent Transportation Systems, 27*(10), 11949-11965.
[9] Tafida, A., Alaloul, W. S., Zawawi, N. A. B. W., Musarat, M. A. & Sani, A. A. (2025). Advancing smart transportation: A review of computer vision and photogrammetry in learning-based dimensional road pavement defect detection. *Computer Science Review, 56*, Article 100729.
[10] Abu Dabous, S., Ait Gacem, M., Zeiada, W., Hamad, K. & Al-Ruzouq, R. (2025). Artificial intelligence applications in pavement infrastructure damage detection with automated three-dimensional imaging – A systematic review. *Alexandria Engineering Journal, 117*, 510-533.
[11] Sapkota, R., Flores-Calero, M., Qureshi, R., Badgujar, C., Nepal, U., Poulose, A., Zeno, P., Vaddevolu, U. B. P., Khan, S., Shoman, M., Yan, H., & Karkee, M. (2025). YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series. *Artificial Intelligence Review, 58*(9), Article 274.
[12] Wang, W., Yu, X., Jing, B., Tang, Z., Zhang, W., Wang, S., Xiao, Y., Li, S., & Yang, L. (2025). YOLO-RD: A road damage detection method for effective pavement maintenance. *Sensors, 25*(5), Article 1442.
[13] Zeng, J., & Zhong, H. (2024). YOLOv8-PD: An improved road damage detection algorithm based on YOLOv8n model. *Scientific Reports, 14*(1), Article 12052.
[14] Cano-Ortiz, S., Lloret Iglesias, L., Martinez Ruiz del Árbol, P., Lastra-González, P., & Castro-Fresno, D. (2024). An end-to-end computer vision system based on deep learning for pavement distress detection and quantification. *Construction and Building Materials, 416*, Article 135036.
[15] Yang, X., Zhang, J., Liu, W., Jing, J., Zheng, H., & Xu, W. (2024). Automation in road distress detection, diagnosis and treatment. *Journal of Road Engineering, 4*(1), 1-26.
[16] Huang, S., Chen, H., Yan, L., Zou, X., Li, B., & Bi, Y. (2025). A review of the progress in machine vision-based crack detection and identification technology for asphalt pavements. *Digital Transportation and Safety, 4*(1), 65–79.
[17] Ultralytics. (n.d.). YOLO architecture explained: From YOLOv3 to YOLO26. *Ultralytics Documentation*. Retrieved from https://docs.ultralytics.com/guides/yolo-architecture (accessed 16 October 2026).
[18] Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems, 38*, 78433–78457.
[19] Ultralytics. (n.d.). Ultralytics YOLO26. *Ultralytics Documentation*. Retrieved from https://docs.ultralytics.com/models/yolo26 (accessed 16 October 2026).
[20] Youwai, S., Chaipetch, P., Samaikul, H., & Yonseng, T. (2026). YOLO26-RD: An end-to-end road damage detection network with learnable contrast enhancement and edge-guided downsampling. *arXiv, 2608.15713v2* (preprint). https://arxiv.org/abs/2608.15713v2.