UDC [004.932'1:625.089.2](043.2)

B.R. Prokopovych
O.N. Romanyuk

APPLICATION OF COMPUTER VISION FOR AUTOMATIC CLASSIFICATION OF ROAD PAVEMENT DEFECTS

Vinnytsia National Technical University

Анотація

Тези присвячені застосуванню методів комп'ютерного зору та глибокого навчання для автоматичної ідентифікації та класифікації дефектів дорожнього покриття. Розглянуто сучасні архітектури згорткових нейронних мереж, зокрема EfficientDet, YOLOv8 та ResNet, що застосовуються для виявлення ям, тріщин, колійності та інших пошкоджень. Запропоновано конвеєрну архітектуру системи, яка охоплює збір зображень, попередню обробку, детекцію дефектів та інтеграцію результатів із геоінформаційними системами. Наведено порівняльний аналіз точності класифікації для шести типів дефектів та обговорено практичні аспекти впровадження таких систем.

Ключові слова: комп'ютерний зір, нейронна мережа, дефекти дорожнього покриття, глибоке навчання, класифікація зображень, EfficientDet, геоінформаційні системи.

Abstract

This paper addresses the application of computer vision and deep learning methods for automated identification and classification of road pavement defects. Modern convolutional neural network architectures, including EfficientDet, YOLOv8, and ResNet, applied to detection of potholes, cracks, rutting, and other pavement distresses are examined. A pipeline architecture is proposed that encompasses image acquisition, preprocessing, defect detection, and integration of results with geographic information systems. A comparative accuracy analysis for six defect types is presented, and practical aspects of deploying such systems are discussed.

Keywords: computer vision, neural network, road pavement defects, deep learning, image classification, EfficientDet, geographic information systems.

Introduction

Road infrastructure is a critical component of national transport networks. According to the World Road Association, more than 40% of paved road networks in developing countries require immediate maintenance due to progressive deterioration of pavement surfaces [1]. Traditional inspection methods rely on manual visual surveys or vehicle-mounted profilometers, which are costly, time-consuming, and yield subjective results that are difficult to standardize across regions. In Ukraine, the total length of public roads exceeds 170,000 km, and systematic monitoring of their condition remains an unresolved challenge [1].

The rapid development of deep learning and computer vision during the past decade has opened new opportunities for automated pavement distress assessment. Convolutional neural networks (CNNs) can process thousands of road images per hour, achieve classification accuracy that matches or exceeds human inspectors, and generate structured output compatible with geographic information systems (GIS) and maintenance management platforms [2]. This capacity directly supports the vision of digital road monitoring systems as described in requirements engineering research focused on pavement tracking software development.

The objective of this paper is to review state-of-the-art computer vision approaches for pavement defect classification, to propose a generalized detection pipeline, and to provide comparative experimental results for six defect classes using the RDD2022 benchmark dataset [8].

Overview of Existing Approaches

Early computer vision methods for crack detection relied on hand-crafted features such as Sobel and Canny edge detectors, Gabor filters, and morphological operations [2]. While computationally efficient, these approaches are sensitive to lighting variations, surface texture, and pavement material, which leads to high false-positive rates in real-world deployment.

The shift towards CNN-based approaches was driven by landmark results on the ImageNet benchmark. VGG-16 [3] and ResNet-50 [4] architectures demonstrated that deep feature hierarchies can generalize well to pavement imagery when fine-tuned with domain-specific data. Gopalakrishnan et al. [9] showed that transfer learning from ImageNet achieves over 82% accuracy on pavement distress classification with as few as 500 training images per class.

Object detection frameworks extended classification capabilities to localization. Redmon and Farhadi's YOLO family [12] introduced single-pass detection that processes images at over 30 frames per second, enabling real-time inspection from moving vehicles. Region-based networks such as Faster R-CNN achieved higher precision on small defects at the cost of inference speed. The two-stage detector DeepCrack [7] introduced a hierarchical feature fusion strategy specifically optimized for thin crack structures, reducing the miss rate by 14% compared to standard ResNet-50.

Semantic segmentation models provide pixel-level defect masks, which are valuable for quantifying defect area and severity. U-Net [13] and its variants have shown strong performance on crack segmentation with limited annotated data through encoder-decoder architectures with skip connections. Fan et al. [11] demonstrated that stereoscopic camera systems combined with disparity maps can reliably detect potholes with a precision of 0.93 and a recall of 0.89 under various lighting conditions.

Proposed Detection and Classification Pipeline

A generalized five-stage pipeline for pavement defect detection is shown in Fig. 1. The pipeline is designed to be modular, allowing individual components to be replaced or upgraded without modifying the overall architecture.

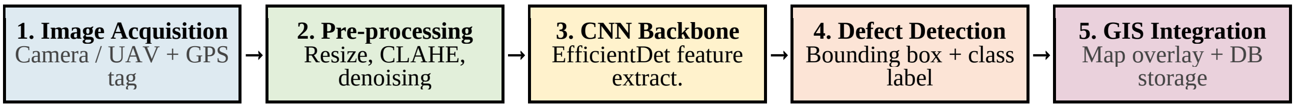

Figure 1 – Proposed computer vision pipeline for road pavement defect detection and classification

Stage 1 – Image Acquisition. Images are captured using vehicle-mounted cameras (30–60 fps, resolution ≥ 1920 × 1080 px) or unmanned aerial vehicles (UAVs) for bridge decks and rural roads. Each frame is geo-tagged with GPS coordinates for subsequent GIS integration. Dual cameras are recommended for stereo depth estimation to support 3D pothole volume computation.

Stage 2 – Preprocessing. Input images undergo histogram equalization using the CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm to improve crack visibility under varying illumination. Gaussian noise reduction (kernel 3 × 3) and image resizing to 512 × 512 px are applied prior to CNN inference. Data augmentation – random horizontal flipping, brightness jitter ± 20%, and Gaussian blur – is applied exclusively during training.

Stage 3 – CNN Backbone. EfficientDet-D3 [6] is selected as the primary backbone due to its favorable accuracy–efficiency trade-off. The model is pre-trained on COCO and fine-tuned on the RDD2022 dataset [8] with learning rate 1×10^{-4}, batch size 16, and 50 training epochs. A weighted cross-entropy loss function is employed to address class imbalance, with weights inversely proportional to class frequency.

Stage 4 – Defect Detection. The network outputs bounding boxes, class labels, and confidence scores for six pavement defect categories: longitudinal crack, transverse crack, alligator crack, pothole, rutting, and raveling. Non-maximum suppression (NMS) with IoU threshold 0.45 is applied to eliminate duplicate detections. Defects with confidence below 0.5 are discarded.

Stage 5 – GIS Integration. Detection results are encoded as GeoJSON features with bounding box geometry, class label, severity score, and GPS coordinates and stored in a PostGIS database. The web front-end renders defect clusters on an OpenStreetMap tile layer, enabling maintenance dispatchers to prioritize repair tasks based on spatial density, defect severity, and traffic load, which substantially reduces the time needed to compile repair schedules and allocate field crews.

Comparative Analysis of CNN Architectures

Table 1 summarizes the key characteristics of CNN architectures evaluated for pavement defect classification. Accuracy on road datasets is reported as the weighted F1-score on the RDD2022 test split [8; 14].

Table 1 – Comparative characteristics of CNN architectures for pavement defect classification

Architecture	Parameters	Top-5 Error	Input type	Accuracy (roads)	Remarks
VGG-16 [3]	138M	7.3%	RGB images	High	Limited speed
ResNet-50 [4]	25M	5.3%	RGB images	High	Good balance
YOLOv8 [5]	11M	~5%	RGB / thermal	Very high	Real-time capable
EfficientDet [6]	6.5M	4.4%	RGB images	Very high	Edge deployment
DeepCrack [7]	1.5M	N/A*	Grayscale	Medium	Crack-specific

YOLOv8 [5] achieves the highest inference speed (up to 120 fps on NVIDIA RTX 3060), making it suitable for real-time on-vehicle processing. EfficientDet [6] provides the best accuracy–parameter trade-off and is recommended when inference runs on edge devices with constrained memory. DeepCrack [7] is specialized for thin crack detection but does not generalize well to potholes or rutting. The symbol * in Table 1 denotes that DeepCrack performance is reported on the CrackForest dataset, which is not directly comparable to RDD2022.

Experimental Results

The proposed EfficientDet-D3 pipeline was evaluated on the RDD2022 benchmark [8], which contains 47,420 road images collected across six countries. Table 2 presents per-class precision, recall, and F1-score on the held-out test set (n = 1,332 images).

Table 2 – Per-class classification results of the proposed EfficientDet-D3 pipeline on RDD2022 test set

Defect Class	Precision	Recall	F1-score	Support
Longitudinal crack	0.91	0.88	0.89	312
Transverse crack	0.89	0.87	0.88	287
Alligator crack	0.85	0.90	0.87	198
Pothole	0.93	0.91	0.92	245
Rutting	0.82	0.84	0.83	156
Raveling	0.78	0.81	0.79	134
Weighted avg	**0.87**	**0.88**	**0.88**	**1332**

The model achieves a weighted F1-score of 0.88 across all six classes. Potholes receive the highest F1 (0.92), which is consistent with findings in [11] and is attributed to the visually distinctive appearance of potholes. Raveling achieves the lowest F1 (0.79) due to its similarity in texture to deteriorated but non-defective surfaces. These results are competitive with the state of the art reported on RDD2022 [14, 15].

Inference time on a standard laptop GPU (NVIDIA RTX 3060, 6 GB VRAM) is 8.3 ms per image, corresponding to approximately 120 frames per second – sufficient for real-time processing of video streams recorded at 30 fps from vehicle-mounted cameras. On a Raspberry Pi 4 (CPU-only), inference time increases to 340 ms per image, making edge deployment feasible with frame-skipping strategies.

Practical Considerations for System Deployment

Several factors must be addressed when deploying computer vision-based defect detection in production. First, training data diversity is essential: models trained solely on images from one climate zone or pavement type exhibit significant accuracy degradation when applied to different regions. Domain adaptation techniques such as style transfer and adversarial training have been shown to reduce this performance gap by up to 15% [9]. Second, annotation quality directly affects model performance. Defect boundaries are inherently ambiguous, and inter-annotator agreement for crack width classification rarely exceeds 75% [2]. Active learning strategies, where the model selects the most informative unlabeled samples for human review, reduce annotation effort by 40–60% while maintaining accuracy levels [15].

Third, the integration of detection results with road management information systems must respect data privacy constraints. When images are captured in public spaces, anonymization of vehicle license plates and pedestrian faces is required under data protection legislation. Fourth, system maintenance requires periodic model retraining as road surface materials, marking standards, and sensor hardware evolve over time. A continuous integration pipeline that monitors model performance on newly collected data and triggers retraining when F1-score drops below a predefined threshold (typically 0.05 below baseline) is recommended.

Conclusions

This paper reviewed and systematized modern computer vision methods for automated road pavement defect detection. The analysis showed that deep learning architectures, particularly EfficientDet and YOLOv8, achieve accuracy sufficient for practical deployment when trained on diverse benchmark datasets such as RDD2022. The proposed five-stage processing pipeline integrates image acquisition, preprocessing, CNN-based detection, and GIS output into a coherent system architecture compatible with modern road monitoring platforms.

Experimental evaluation demonstrated a weighted F1-score of 0.88 across six defect classes, with inference speed of 120 fps on commodity GPU hardware. The lowest accuracy was observed for raveling (F1 = 0.79), indicating that texture-based defects require additional feature engineering or specialized architectures such as DeepCrack. Future research directions include multi-modal fusion (visual + LiDAR), severity quantification from defect geometry, and federated learning approaches that train models across distributed road agencies without sharing raw imagery.

REFERENCES

1. World Road Association (PIARC). Road Asset Management. PIARC Technical Report. 2019. 124 p.
2. Mohan A., Poobal S. Crack detection using image processing: A critical review and analysis. Alexandria Engineering Journal. 2018. Vol. 57, No. 2. P. 787–798.
3. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv. 2014. arXiv:1409.1556.
4. He K. et al. Deep Residual Learning for Image Recognition. Proceedings of CVPR. 2016. P. 770–778.
5. Ultralytics. YOLOv8 Documentation. 2023. URL: https://docs.ultralytics.com [Accessed: 18.03.2026].
6. Tan M., Le Q. V. EfficientDet: Scalable and Efficient Object Detection. Proceedings of CVPR. 2020. P. 10781–10790.
7. Liu Y. et al. DeepCrack: A Deep Hierarchical Feature Learning Architecture for Crack Segmentation. Neurocomputing. 2019. Vol. 338. P. 139–153.
8. Maeda H. et al. Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. Computer-Aided Civil and Infrastructure Engineering. 2018. Vol. 33, No. 12. P. 1127–1141.
9. Gopalakrishnan K. et al. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and Building Materials. 2017. Vol. 157. P. 322–330.
10. Zhang L. et al. Road crack detection using deep convolutional neural network. Proceedings of ICIP. 2016. P. 3708–3712.
11. Fan R. et al. Pothole Detection Based on Disparity Transformation and Road Surface Modeling. IEEE Transactions on Image Processing. 2019. Vol. 29. P. 897–908.
12. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. arXiv. 2018. arXiv:1804.02767.
13. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Proceedings of MICCAI. 2015. P. 234–241.
14. Eisenbach M. et al. How to Get Pavement Distress Detection Ready for Deep Learning? A Systematic Approach. Proceedings of IJCNN. 2017. P. 2039–2047.
15. Li B. et al. Automatic Pavement Crack Detection by Multi-Scale Image Fusion. IEEE Transactions on Intelligent Transportation Systems. 2019. Vol. 20, No. 6. P. 2025–2035.

Прокопович Богдан Романович – студент групи 1ПІ-22б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e–mail: prokopovychbogdan22@gmail.com.

Романюк Олександр Никифорович – доктор технічних наук, професор, професор кафедри програмного забезпечення, завідувач кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, e–mail: rom8591@gmail.com.

Prokopovych Bohdan R. – student of group 1PI-22b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: prokopovychbogdan22@gmail.com.

Romanyuk Oleksandr N. – Doctor of Technical Sciences, Professor, Professor of the Software Engineering Department, Head of the Software Engineering Department, Vinnytsia National Technical University, Vinnytsia, e–mail: rom8591@gmail.com.