e-Journal of Nondestructive Testing - ISSN 1435-4934 - www.ndt.net

NDT-CE 2025 - The International Symposium
on Nondestructive Testing in Civil Engineering
Izmir, Türkiye, September 24-26, 2025

Automated Road Damage Detection Using UAVs and Deep Learning: A Scalable Solution for Infrastructure Maintenance

Tuğba YILDIZLI[1], Merve BAYRAKTAR[2], Burcu GÜLDÜR ERKAL[2]

[1]Civil Engineering & Geosciences, Water Systems Engineering, Delft University of Technology, Delft, Netherlands,
[2]Digital Design and Production, Technical University of Applied Sciences Augsburg, Augsburg, Germany,
T.Yildizli@tudelft.nl, merve.bayraktar@hacettepe.edu.tr, burcu.guldur.erkal@tha.de

Abstract
This study presents an automated pavement surface inspection approach using UAV imagery and the YOLOv7 deep learning model. The aim is to detect common surface defects such as longitudinal cracks, reflective cracks, alligator cracks, and potholes with high accuracy and efficiency. The system was trained on publicly available datasets and tested with aerial images captured at Hacettepe University's Beytepe Campus. Evaluation results demonstrated a precision of 0.51, recall of 0.45, and mean Average Precision (mAP@0.5) of 0.42. These findings confirm the feasibility of integrating UAV platforms and deep learning for road defect detection. Unlike traditional methods, which are time-consuming and labor-intensive, this approach enables faster, scalable, and cost-effective inspections. The proposed framework contributes to safer and more sustainable road infrastructure maintenance by facilitating proactive monitoring and reducing operational burdens. Its ability to analyze large areas in real time makes it particularly suitable for modern transportation networks and smart city applications.

Keywords: UAV imagery, pavement distress, object detection;

1. Introduction

Roads are an important aspect of the transportation system and are needed for economic growth and for people to go about. The more people use roads and the more traffic there is, the more likely it is that the structure of pavement systems will break down. Cracks, potholes, and ruts are all problems that can happen on the surface because of weather, aged materials, and heavy loads. These issues not only make the trip less comfortable, but they also make the road less safe and less long-lasting. These issues could cause reinforced concrete layers to rust more quickly, which would limit their useful life and make them exceedingly dangerous for cars, especially when vision is poor or when they are moving rapidly [1, 2]. People who are trained or special vehicles with laser scanners, infrared sensors, or ground-penetrating radar have traditionally checked the state of a pavement by hand. These methods do offer useful results, but they have certain disadvantages, including being costly to run, needing a lot of labor, not covering a vast area, and being subjective in their judgment [3, 4]. Because of this, infrastructure organizations are feeling more and more pressure to utilize inspection methods that are more automated, scalable, and objective. Drones, which are also called unmanned aerial vehicles (UAVs), are becoming more popular because they can rapidly and efficiently get high-resolution images of large areas. They are great for keeping an eye on civil infrastructure since they can fly over areas that are hard to get to or dangerous without putting anyone in danger [5]. Recent advances in computer vision and artificial intelligence have made UAV-based systems excellent for automated road surface inspection. Thresholding, edge detection, morphological operations, and wavelet transforms were among of the first standard Digital Image Processing (DIP) methods used to automatically check the condition of pavement [6 - 8]. These algorithms look for binary features in photographs of pavement, like cracks or holes,

 https://doi.org/10.58286/31691

that are different from the background. On the other hand, DIP methods are usually impacted by noise, shadows, and variations in lighting, which makes them less trustworthy in the field. Researchers investigated into using Machine Learning (ML) to get past these difficulties. Support Vector Machines (SVM), Random Forests (RF), and Decision Trees are all supervised learning algorithms that assisted sort flaws by learning from samples that had been tagged [9,10]. These models were more flexible than rule-based algorithms, but they still depended a lot on handcrafted features and couldn't be changed to work with other types of pavement. Two types of unsupervised learning that became popular, especially when it was challenging to find labeled data, are K-means clustering and Minimum Path Selection (MPS) [11,12]. Deep Learning (DL) has made a lot of progress, and its performance and capacity to generalize have both gotten a lot better. CNNs can learn hierarchical features directly from the pixels in a picture. They do a better job of classifying and locating things than traditional ML. Some early uses include CrackNet, which divides up pictures of pavement into pixels [13], and models like VGG16 and ResNet, which were changed through transfer learning to locate cracks [14,15]. Not only did these models make it easier to find objects, but they also cut down on the need for pre-processing. Faster R-CNN [16], SSD [17], and the YOLO family [18] are all object identification algorithms that can locate and sort a lot of defects at once. YOLOv3 and YOLOv7, in particular, look like they could be useful for finding road hazards in real time because they are so fast and precise [19,20]. These algorithms can discover all kinds of surface problems in photographs obtained by UAVs, such as alligator cracks, reflecting cracks, potholes, and lengthy cracks. Semantic segmentation models, on the other hand, sort pixels, which helps you quantify fault areas quite accurately. People first suggested using U-Net for biological purposes [21], but it has subsequently become quite popular for mapping road issues since it is good at modeling spatial hierarchies. A two-stage pipeline has been made from detection and segmentation in recent studies (for example, YOLO + U-Net). This combines the speed of object detection with the detail of segmentation [22]. This integration has worked well to highlight not only that damage exists, but also how severe it is, how awful it is, and how bad it is. Even with these changes, it's still impossible to utilize deep learning models on different types of roads because the pavement is made of different materials, the illumination is varied, and the flaws are shaped differently. Another thing that makes it hard is that there aren't many labeled datasets. Researchers have made large annotated datasets like CRACK500, PID, and GAPs in response. They have also employed augmentation methods to improve generalization [19,23]. In this scenario, the present study advances the field by proposing a unified framework that combines UAV imagery with the YOLOv7 deep learning model to autonomously detect and classify pavement defects. All experiments are conducted on aerial data collected from Hacettepe University's Beytepe Campus, and performance is evaluated using standard detection metrics. Although many deep learning models—such as VGG16, ResNet, Faster-R-CNN, and SSD—have been investigated in prior work, this research focuses exclusively on YOLOv7 due to its optimal balance between speed and accuracy for UAV-based inspections. The overarching goal is to deliver a scalable, rapid, and cost-effective solution for contemporary pavement-condition monitoring.

2. Datasets

To create strong and generalizable deep learning models for detecting pavement degradation, you need a large and varied dataset. This part presents the original UAV-based dataset that was collected at Hacettepe University and a few publicly accessible annotated datasets that were utilized for training and validation.

2.1 UAV-Based Data Collection at Hacettepe Campus

To evaluate the model's generalization to real-world environments, an original dataset was created using unmanned aerial vehicles (UAVs) over Hacettepe University's Beytepe Campus. A DJI Mavic 2 Pro drone equipped with a 1-inch CMOS sensor (20 MP resolution, 5472 × 3648 pixels) was employed for data acquisition. Four manual missions were conducted, covering a total of approximately 1.4 kilometers of roadway including straight paths, intersections, and parking areas with asphalt and concrete surfaces. Images were taken under varying lighting conditions (midday and late afternoon), which allowed the capture of different illumination and shadow patterns that are typical in urban scenes. A total of 102 high-resolution RGB images were acquired with nadir view angles and altitudes between 15 and 25 meters. While the dataset includes both damaged and undamaged surfaces, only 66 images containing visible defects were labeled for detection and segmentation tasks. The remaining 36 images, although not manually annotated, were retained as background-only samples and used during testing to validate model robustness in false positive scenarios. Importantly, none of the Hacettepe UAV images were used in training or validation. They were reserved exclusively as an independent test set to assess the trained model's adaptability to a new domain. This approach ensures that model performance reflects true generalization capability rather than memorization of training patterns.

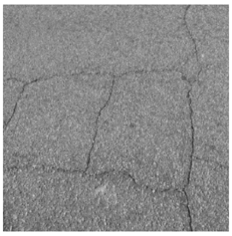
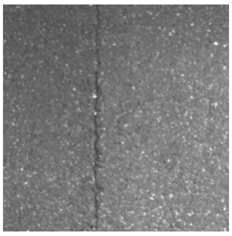
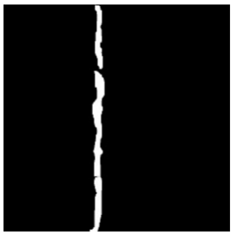
2.2 Publicly Available Datasets

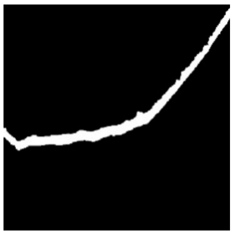
For the training and validation of deep learning models, five publicly available datasets were utilized. These datasets are widely accepted in the literature and contain rich, labeled samples of pavement surface defects under diverse conditions.

- Crack500: Contains 500 RGB images with manually segmented binary masks [24]. It includes a balanced mix of concrete and asphalt roads.
- GAPs: The German Asphalt Pavement Distress dataset includes 1,969 annotated pavement images under varying lighting conditions [25].
- CFD (CrackForest Dataset): Consists of 118 grayscale images annotated with crack masks [26].
- CrackTree: Offers 206 grayscale images with annotated crack skeletons, suitable for evaluating edge detection and connectivity [27].
- DeepCrack: Contains 537 grayscale images with fine pixel-wise crack annotations, intended for segmentation-based tasks [28].

To illustrate the annotation quality and visual diversity of these datasets, **Table** 1 presents representative image–mask pairs. Each row shows a raw pavement image alongside its binary or skeletonized ground-truth mask, highlighting differences in appearance and defect geometry. All images were normalized and resized to 416 × 416 for object-detection models and to 512 × 512 for segmentation models, preserving aspect ratios where feasible. This multi-source training strategy enhances model robustness and mitigates overfitting to any single acquisition modality.

Table 1 The images and their masks in the crack datasets

Dataset	Raw Image	Mask
Crack Forest Dataset (CFD)		
Crack500		
CrackTree		
GAP v1		

DeepCrack		

2.3 Data Augmentation

To enhance the diversity of the training set and reduce the risk of overfitting, various real-time data augmentation strategies were employed during model training. These included horizontal flipping of images, moderate adjustments to brightness and contrast (up to ±30%), and the injection of low-level Gaussian noise to simulate sensor variability. Additionally, random cropping and resizing operations were applied to mimic different framing perspectives and scale variations. A limited proportion of the dataset was also converted to grayscale, thereby encouraging the model to generalize across different color modalities and lighting conditions. Since the distribution of crack types in the datasets was imbalanced—particularly with reflective cracks and potholes appearing less frequently—targeted oversampling was performed to mitigate this issue. This ensured that the model did not become biased toward more dominant classes such as longitudinal or alligator cracks. To further improve robustness, non-defective pavement tiles containing shadows, patches, or vegetation were included during training. These samples helped the model learn to differentiate true surface distresses from natural background artifacts or lighting inconsistencies, which are often misclassified in real-world applications.

2.4 Dataset Preperation

Labeling was required only for the Hacettepe UAV dataset. Manual annotation was conducted using LabelImg [29], a lightweight annotation tool that allows bounding box drawing and export in multiple formats. Four distinct damage categories were defined based on visual morphology and literature standards:(i) longitudinal cracks, (ii) reflective cracks, (iii) alligator cracks, and (iv) potholes. From the original images, a sliding window approach was used to crop 512×512 pixel tiles with 50% overlap, resulting in 1,930 annotated tiles. Labels were reviewed by two independent annotators to ensure consistency. Public datasets already included verified annotations and were used directly. Some minor format conversions were performed to harmonize the data input for model training (e.g., converting binary masks to YOLO-compatible label files for object detection training). The five public datasets were merged and divided into 80% training and 20% validation splits in a stratified manner, preserving the class balance. This combination created a diverse and challenging dataset of approximately 1,500 training and 400 validation images. No portion of the Hacettepe UAV dataset was used in the training or validation stages to maintain an unbiased testing environment. The entire 102-image UAV set (both annotated and unannotated) was reserved for final evaluation. This strategy enables assessment of how well the model generalizes to a domain-shifted, real-world dataset—captured using different equipment and under uncontrolled conditions.

3. Object Detection

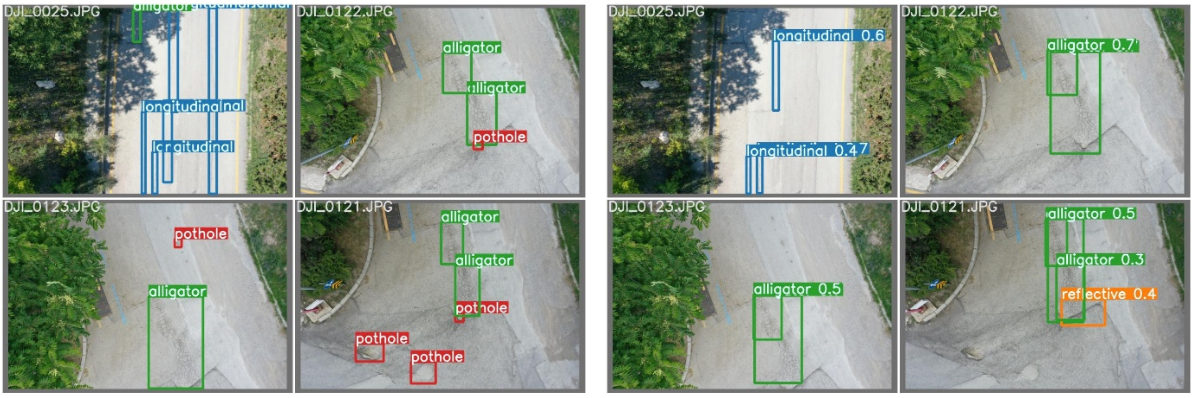
Finding and identifying several things in an image is a basic problem in computer vision called object detection. In the context of monitoring pavement distress, it helps find problems like cracks, potholes, and surface deformations by giving both class names and geographical coordinates through bounding boxes. Recent progress in deep learning has made object detection much better, especially with the arrival of one-stage detectors like the YOLO (You Only Look Once) family. Because they can anticipate object classes and bounding boxes in a single forward pass, these models are quite popular for real-time applications. This efficiency is especially helpful for surveys done with UAVs, which may make high-resolution photos with hundreds of frames.

2.4 YOLO – YOLOv7

Among the various deep learning-based approaches, the YOLO family has stood out for combining high detection accuracy with real-time inference capability. Originally introduced by Redmon et al. [18], YOLO redefined object detection as a single regression problem, predicting object bounding boxes and class probabilities directly from full images in one network pass. This one-stage design marked a shift from two-stage detectors like Faster R-CNN [16], which separate region proposal and classification tasks. The simplicity of the YOLO design contributed to faster inference and suitability for time-sensitive applications such as UAV-based road inspection. Since its inception, YOLO has undergone multiple revisions, each improving the backbone architecture, training procedures, and feature fusion mechanisms. YOLOv3 leveraged the Darknet-53 backbone and multi-scale predictions through feature pyramid networks [20]. YOLOv5, though unofficial, introduced practical enhancements such as auto-anchor learning, mosaic augmentation, and export-friendly deployment options. More recently, advanced versions up to YOLOv12 have been proposed, continuing to push the limits of detection performance with transformer-based modules, better generalization, and multi-resolution adaptability. In this study, YOLOv7 [30] was selected as the detection framework due to its favorable balance between accuracy, model size, and inference speed. YOLOv7 features the Extended Efficient Layer Aggregation Network (E-ELAN), which deepens the network without degradation, and a compound model scaling approach that tunes depth, width, and resolution concurrently. It also introduces coarse-to-fine label assignment strategies that help stabilize the learning process and improve precision, especially for small and overlapping objects. Thanks to its flexible and optimized architecture, YOLOv7 can effectively handle high-resolution UAV imagery, detecting thin and irregular road surface defects under varying illumination and environmental conditions. Its performance in identifying common pavement anomalies—such as longitudinal cracks, reflective cracks, alligator cracks, and potholes—demonstrates its suitability for aerial road monitoring tasks. Although newer versions like YOLOv8 through YOLOv12 offer continued improvements, YOLOv7 remains a widely adopted and rigorously evaluated benchmark in the literature. Its modular design, reproducibility, and training efficiency make it a solid choice for the problem addressed in this study.

4. Results

The performance of the YOLOv7-based object detection system was evaluated on UAV-acquired test images from Hacettepe University's Beytepe Campus. These images were annotated for four primary classes of pavement surface anomalies: longitudinal cracks, reflective cracks, alligator cracks, and potholes. The model's predictions were assessed against the ground-truth annotations to evaluate accuracy and robustness. **Figure 1** presents an example of the annotated test imagery, and corresponding predictions made by the YOLOv7 model. The model demonstrated a strong ability to localize and classify extended crack patterns (e.g., longitudinal and alligator cracks), although performance was slightly reduced in more subtle or irregular defect types such as potholes and reflective cracks.



(a) (a)

Figure 1 Labels (a) and predictions (b) of drone test image with YOLOv7

Quantitative evaluation metrics are summarized in **Table 2**. The overall test precision was 0.51, recall was 0.45, and the mean Average Precision (mAP@0.5) score reached 0.42. While these values are moderate, they highlight the model's effectiveness under realistic and challenging aerial conditions. Factors affecting performance included shadow interference, vegetation overlaps, and differences in altitude and image resolution compared to training data.

Table 2 YOLOv7 test results for each class

Class	Precision	Recall	mAP
All Classes	0.51	0.45	0.42
Longitudinal Crack	0.60	0.43	0.44
Reflective Crack	0.37	0.43	0.34
Alligator Crack	0.55	0.47	0.48
Pothole	0.53	0.45	0.43

Figure 2 provides a confusion matrix, visualizing the classification performance across all four defect categories. Based on the confusion matrix, the YOLOv7 model demonstrated higher accuracy in identifying longitudinal and alligator cracks, with true positive rates of 46% and 50%, respectively. However, reflective cracks and potholes were more frequently misclassified. Reflective cracks were correctly identified 47% of the time but were also misclassified as potholes in 29% of the cases. Similarly, while 46% of potholes were correctly detected, 15% were incorrectly classified as background. High misclassification rates between defect classes and the background are evident, particularly for reflective cracks (51%) and potholes (54%). This suggests that the model struggles to distinguish subtle damage features from the background, especially in low-resolution or shadowed areas. Overall, the model performs better on defect types with distinct geometric patterns, while ambiguous or less pronounced defects pose greater challenges.

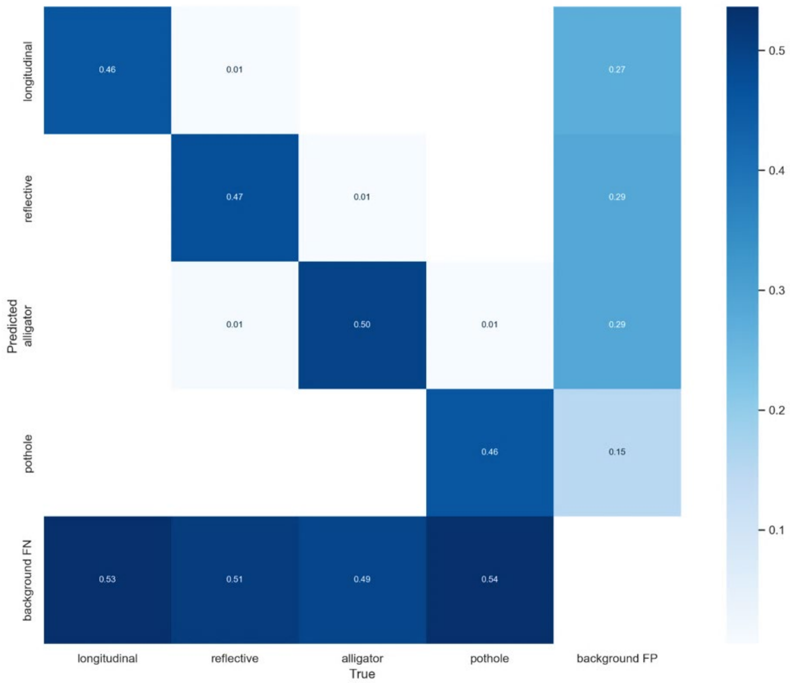


Figure 2 Confusion Matrix as a result of the YOLOv7 model

These results affirm YOLOv7's capability for multi-class pavement distress detection from aerial platforms. Despite certain limitations, the system offers a practical, automated solution for preliminary road surface assessment, reducing the reliance on labor-intensive manual inspections.

5. Discussion

The results of this investigation show both the pros and downsides of using a YOLOv7-based UAV system to automatically check the condition of roads. One of the best things about it was that it could find different types of defects in real time without needing specific scanning equipment. Combining drone photography with deep learning is a step toward infrastructure assessment that is both scalable and affordable. But the differences in performance that were seen, especially when it came to finding smaller problems like reflecting fissures and potholes, imply that there are limits to generalization. These differences could be because the training datasets were mostly made up of ground-level images, while the test set had both oblique and aerial views. In the future, using domain adaptation approaches or training on datasets that were fully recorded by UAVs could help fix these problems. It is also still hard to find little or low-contrast features. YOLOv7 is designed to be fast and accurate, however adding attention methods or transformer-based improvements from subsequent YOLO versions (like YOLOv12) may make it even more accurate. Lastly, although though this study only looked at four types of defects, adding additional complicated situations, including multi-layered degradation, surface water presence, or changing lighting, will make the framework even more reliable. These things are necessary for using the model in real-world traffic situations.

6. Conclusion

The YOLOv7 object recognition framework was used in this study to show how to monitor the surface of a road using a UAV from start to finish. The system showed that it could find common road surface flaws by being trained on a lot of public datasets and tested on aerial images from Hacettepe University's Beytepe Campus. The test findings revealed that the system was good at finding longitudinal and alligator cracks, but it had trouble with more irregular types of damage, such potholes. Even though the YOLOv7 model's precision and recall scores aren't very high, it is a good starting point for future work that will involve real-time drone-based road assessment. As infrastructure organizations get more and more requests for quick and scalable ways to evaluate things, the combination of aerial platforms and deep learning architectures shown in this work will be very important for the growth of smart transportation systems. In the future, we will concentrate on making the model more generalizable, adding more data to the training sets, and looking into hybrid detection-segmentation pipelines to get a better picture of how healthy the roads are.

References

[1] Zou, Q., Cao, Y., Li, Q., Mao, Q., & Wang, S. (2012). "CrackTree: Automatic crack detection from pavement images." Pattern Recognition Letters, 33(3), 227–238.
[2] Gopalakrishnan, K., et al. (2017). "Deep learning for computer vision-based crack detection." Construction and Building Materials, 142, 402–414.
[3] Huidrom, L., et al. (2013). "Video-based detection and quantification of pavement defects." International Journal of Pavement Engineering, 14(1), 1–9.
[4] Tanaka, N., & Uematsu, K. (1998). "Automatic crack detection using morphological image processing." Automation in Construction, 7(6), 687–699.
[5] Zhu, J., et al. (2022). "UAV-based pavement inspection: Recent advances and future prospects." Sensors, 22(8), 2955.
[6] Koch, C., & Brilakis, I. (2011). "Pothole detection in asphalt pavement images." Advanced Engineering Informatics, 25(3), 507–515.
[7] Zhao, X., et al. (2010). "Edge detection methods for pavement inspection." Construction and Building Materials, 24(8), 1506–1513.
[8] Subirats, P., et al. (2006). "Wavelet-based crack detection in pavements." Pattern Recognition Letters, 27(12), 1648–1656.

[9] Gavilán, M., et al. (2011). "Adaptive road crack detection using SVM." Expert Systems with Applications, 38(3), 2857–2866.
[10] Shi, Y., et al. (2016). "CrackForest: Road crack detection using random structured forests." IEEE Trans. on ITS, 17(12), 3434–3445.
[11] Amhaz, R., et al. (2016). "Automatic crack detection using minimum path selection." Automation in Construction, 69, 66–76.
[12] Akagic, A., et al. (2018). "Crack detection using Otsu and histogram thresholding." Measurement, 123, 223–231.
[13] Zhang, L., et al. (2017). "CrackNet: A deep CNN for pavement crack detection." Computers in Civil Engineering, 31(5), 04017036.
[14] Zhang, L., et al. (2016). "Crack detection using pre-trained VGG16 CNN." Journal of Computing in Civil Engineering, 30(5), 04016003.
[15] Cubero-Fernandez, A., et al. (2017). "Decision tree-based classification of road defects." Automation in Construction, 81, 58–66.
[16] Ren, S., He, K., Girshick, R., & Sun, J. (2017). "Faster R-CNN: Towards real-time object detection." IEEE TPAMI, 39(6), 1137–1149.
[17] Liu, W., et al. (2016). "SSD: Single shot multibox detector." ECCV, 21–37.
[18] Redmon, J., et al. (2016). "YOLO: Real-time object detection." CVPR, 779–788.
[19] Majidifard, H., et al. (2020). "Pavement image dataset for deep learning." ISPRS J. of Photogrammetry, 167, 189–200.
[20] Du, Y., et al. (2021). "YOLOv3-based distress detection." Automation in Construction, 129, 103847.
[21] Ronneberger, O., et al. (2015). "U-Net: Convolutional networks for biomedical image segmentation." MICCAI, 234–241.
[22] Liu, Y., et al. (2020). "YOLO-U-Net hybrid for pavement crack segmentation." Applied Sciences, 10(23), 8382.
[23] Yang, F., et al. (2020). "CRACK500: Benchmark dataset and deep network." IEEE Access, 8, 142876–142890.
[24] Zhang, L., Yang, F., Zhang, Y.D., & Zhu, Y.J. (2016). Road crack detection using deep convolutional neural network. Image and Vision Computing, 69, 118–125.
[25] Eisenbach, M., et al. (2017). How to get pavement distress detection ready for deep learning? A systematic approach. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences.
[26] Shi, Y., Cui, L., Qi, Z., Meng, F., & Chen, Z. (2016). Automatic road crack detection using random structured forests. IEEE Transactions on Intelligent Transportation Systems, 17(12), 3434–3445.
[27] Zou, Q., Cao, Y., Li, Q., Mao, Q., & Wang, S. (2012). CrackTree: Automatic crack detection from pavement images. Pattern Recognition Letters, 33(3), 227–238.
[28] Yang, F., Zhang, L., Yu, S., & Prokhorov, D. (2019). DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing, 338, 139–153.
[29] Tzutalin. LabelImg. Git code (2015). https://github.com/tzutalin/labelImg.
[30] Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv:2207.0269