ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Unmanned Aerial Vehicle Enabled Road Damage Detection using Advanced Deep Learning Techniques

[1]KOLAMTI SURYA VARA PRAKASH, [2]MANUKONDA VAJRAM, [3]KODI SHANKARA NARAYANA, [4]MOYYA PAVAN KUMAR, [5]Dr. A. RAMA MURTHY,

[1234]Student, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5] Professor, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

ABSTRACT

In this research, we provide a new method for automatically detecting road damage using deep learning and photos taken by Unmanned Aerial Vehicles (UAVs). For a transportation system that is both safe and sustainable, road infrastructure maintenance is vital. But gathering information on road deterioration by hand may be dangerous and time-consuming. So, to greatly enhance the efficacy and precision of road damage identification, we advise using UAVs and AI technology. We provide a method that detects and localizes objects in UAV photos using three algorithms: YOLOv4, YOLOv5, and YOLOv7. We put these algorithms through their paces using two datasets: one from Spain and one from China, the RDD2022. The experimental findings show that our method is effective, with a mean average accuracy of 59.9% for the YOLOv7 version and 73.20% for the Transformer Prediction Head model. These findings open the door to further study into the use of unmanned aerial vehicles (UAVs) and deep learning for automated road damage identification. Term Index: unmanned aerial vehicle, object identification, deep learning, road damage detection. I.

INTRODUCTION

For a nation's economy to grow, its road maintenance infrastructure must be well-managed. To keep roads in good repair and safe for drivers, regular inspections are required. Historically, this task has been handled manually by public or commercial organizations using cars equipped with a variety of sensors to identify road degradation. On the other hand, human operators run the risk of injury, expense, and length of time while using this approach. To overcome these obstacles, scientists and engineers are automating the process using Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence (AI). Halil Ersin Soken was the associate editor who oversaw the manuscript's evaluation and gave final approval for publishing. end of detecting road damage. There has been a recent uptick in research into developing efficient and cost-effective solutions for road damage identification utilizing UAVs and deep learning-based techniques. Aerial inspections of things and settings in urban areas are only one of many uses for the adaptable unmanned aerial vehicles. Due to their many benefits over more conventional approaches, they have found widespread application in road inspections. These vehicles can survey the road surface from all angles and heights thanks to their high-resolution cameras and other sensors, which provide a complete picture of the road's state. There is less need for manual inspections—which may be risky for human operators—because UAVs can cover a lot of ground fast. Therefore, engineers and academics have taken a keen interest in the possibility of using UAVs for road inspections. An efficient and cost-effective method for detecting road damage may be developed by combining unmanned aerial vehicles (UAVs) with artificial intelligence methods like deep learning. It is often said to be used for urban inspections of roofs [2], vegetation [3], urban settings [4], and swimming pools [5]. Manual road condition checks are still the norm in Spain, with inspectors physically walking the roads in search of problems. The method's hefty price tag reflects the fact that it requires human work as well as task-specific cameras and sensors. An expert is responsible for making the decisions on fixing road problems. China, on the other hand, has an extensive system of roads and highways that are vulnerable to surface cracks and precipitation infiltration. These factors might hasten the roads' degradation and endanger drivers and passengers. Vehicles are more likely to experience excessive wear and tear and an increase in the probability of traffic accidents without prompt identification and the quick availability of data on road problems, which may result in additional financial losses. Consequently, several academic institutions are working together to discover efficient answers to the growing problem of developing

automated methods for identifying road damage. Researchers are actively working on automatic road damage detection systems that use a variety of approaches, including vibration sensors, Light Detection and Ranging (LiDAR) sensors, and image-based algorithms, to identify and map different kinds of road damage [6]. The accuracy of damage detection is typically enhanced by combining these strategies. Several kinds of road deterioration may be identified using image-based methods that use machine learning algorithms, such deep learning. A dataset of images is usually necessary for these methods. These images can come from a variety of sources, such as top-down photos, images taken by drones, mobile devices, satellite image platforms, thermal images, and even 3D or stereo vision of the asphalt surface. Drone footage, in-car cameras, and satellite photos are just a few of the many sources of data used to train the model in recent research. Various forms of road damage, including as potholes, cracks, and rutting, are often labeled in these datasets to aid in the learning process. By adding labels to these photos, the algorithm can train itself to correctly identify and categorize different kinds of road damage. To better detect and repair various forms of road damage, researchers may improve the accuracy and reliability of their models by using a big and varied dataset.

THE ROAD DAMAGE DETECTION DATASET

The IEEE BigData Cup 2022 included the Crowdsensing-based Road Damage Detection Challenge (CRDDC) [13], an event that aimed to encourage the development of automated methods for detecting road damage. Japan, India, Norway, the Czech Republic, the US, and China are all participating in this global competition using a released dataset of 47,420 road pictures. Road damage, such as potholes, alligator cracks, transverse cracks, and longitudinal cracks, totaling more than 55,000 incidences, has been marked with. The mission of CRDDC is to promote the research and development of automated road damage detection and classification systems that use deep learning. The RDD2022 dataset may be used by road organizations and municipalities to automatically monitor road conditions at a cheap cost. Researchers in the fields of computer vision and machine learning may also use the dataset to compare how various algorithms perform in similar image-based tasks, such object recognition and classification. While some groups omitted the China Drone data from their models, others relied on the RDD2022 dataset. Organizations like these rely on YOLOv5, YOLOv7, YPLNet, and Faster RCNN-series models as their primary algorithms. Image patch schemes, attention modules, individualized anchor boxes, and ensemble models trained with several layers of augmentations were some of the methods utilized by many businesses to improve the accuracy of their models. Image enhancements, label smoothing, coordinate attentions, cutting Norway photos to isolate road regions, and training country-specific models utilizing data from all nations are some of the other strategies. Section B: The Hey-Lo-Series You Only Look Once (YOLO) is supposedly one of the most popular object identification algorithms according to what's in the books. There have been many releases of this object detecting method, which has gained popularity. There has been a noticeable shift in detection time when comparing the development of all the YOLO series. In the first publication [14] Devices with limited computing power can execute the YOLO since it just requires a single back-propagation neural network to produce a forecast. This approach has been refined several times since its foundational version was built on AlexNET. The YOLO algorithm's history includes both YOLOv3 [15] and YOLOv4 [16]. Overall, both YOLOv3 and YOLOv4 rely on deep learning to identify objects, but YOLOv4 outshines YOLOv3.

To enhance its accuracy, YOLOv4 has been trained on a huge dataset of photos and videos and tuned for real-time object recognition. To further improve its speed, YOLOv4 incorporates new methods including DropBlock and Mosaic data augmentation. When it comes to YOLO, YOLOv4 is the most up-to-date and accurate version that will be available until 2021. To identify objects in photos and videos, it employs a custom-designed neural network architecture that combines convolutional and transposed convolutional layers. After being trained on a massive dataset of photos and videos, YOLOv4 was fine-tuned for object recognition in real-time. Later on, YOLOv5 [17]—the algorithm's fifth version—was revealed. Although it still has a ways to go before reaching the level of refinement shown in the 5th major update, this algorithm has proven to be an ideal model, expanding our possibilities in terms of picture segmentation highlight. There was a lot of effort put into YOLOv4, and all the subtleties were considered, and the outcomes are very comparable. After YOLOv4, YOLOv5 is a huge upgrade. Its foundation is the recently-introduced SPADE architecture, which enhances object identification accuracy by combining spatial and semantic data. Mosaic Data Augmentation is a novel training approach that YOLOv5 employs to

ISSN 2454-9940

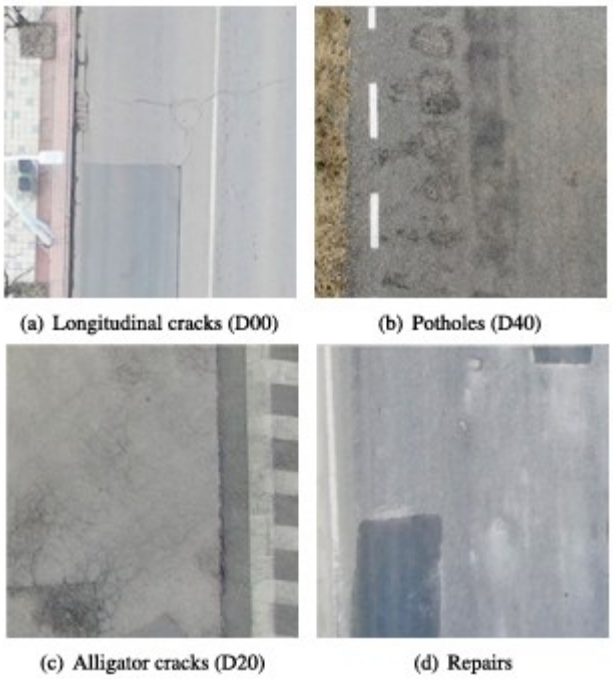
www.ijasem.org

Vol 19, Issue 2, 2025

improve the model's generalizability. The most current version of the algorithm, which represents the seventh iteration in the life cycle of YOLO models, was published very recently [18]. Compared to its predecessor, YOLOv5, YOLOv7 infers with more speed and accuracy. The most recent iteration of YOLO is YOLOv7. The foundational network of this new design, Efficient-YOLO, is EfficientNet. A huge dataset was used to train YOLOv7, which has been fine-tuned for object recognition in real-time. It outperforms earlier YOLO versions in terms of accuracy and speed. Finally, YOLOv4 is the best version of YOLO for real-time object recognition and is expected to be the most accurate version until 2021. The most recent iteration of YOLO, YOLOv7, is built on a whole new architecture known as Efficient-YOLO, which is both quicker and more accurate than its predecessors.

OBJECTIVES AND STRUCTURE

This study expands upon an earlier effort that suggested a framework for a pavement monitoring system that might detect potholes in photographs taken by unmanned aerial vehicles [7]. Here, we take the prior work and improve it by comparing it to new techniques and datasets, adding additional types of dam age, and using data augmentation during training to better respond to objects' drastically changing sizes in photos. Lastly, this study compares YOLOv5 with YOLOv7, and it uses the Transformer Prediction Head to enhance the YOLOv5 model for the UAV application. In this study, we used a combined dataset consisting of both prior research and Crowdsensing-based Road Damage. Detec tion Challenge, which includes additional pavement damage lessons for a more thorough comprehension of the issue. Our suggested technique is efficient and effective, as shown by experimental findings that achieve higher accuracy on the test dataset. Using drone-captured photos and cutting-edge AI and vision algorithms, this research aims to enhance the autonomous road monitoring system. One feature of the proposed system is the ability to transmit messages with the geographical coordinates of the damages observed. This would allow the maintenance business to be notified whenever road damage is detected.



(a) Longitudinal cracks (D00) (b) Potholes (D40)

(c) Alligator cracks (D20) (d) Repairs

The dataset includes road damage types (FIGURE 1). Among the many things our group has accomplished, we have: • Added an additional prediction head to deal with the problem of huge scale fluctuations in objects. • YOLOv5's better object localization in dense situations is a direct outcome of adding Transformer Prediction Heads (TPH) to the model. In order to recognize objects in drone footage, it is necessary to provide a variety of successful methods while excluding those that do not. • Using a self-trained classifier to improve the classification accuracy of certain ambiguous categories. Several new types of pavement degradation have been created by the project, as shown in Figure 1. Alligator cracks, longitudinal cracks, potholes, bumps, and repairs are all part of this category. By including these extra lessons, the initiative provides a more thorough understanding of pavement deterioration and allows for more accurate and efficient monitoring of road infrastructure. As a whole, the project makes use of convolutional neural networks to detect asphalt flaws, with the added capability of enabling operator overrides or suggestions to gradually increase accuracy. In addition, we will include a function that uses PIX4D to automatically generate routes that cover the full road, doing away with the need for the pilot to manually operate the system. This paper is organized in the following way: In Section II, we review the literature on damage detection techniques and UAVs in great detail. We explore the architectural design, dataset, and implementation of the proposed system in Section III. Section IV delves

into the conducted experiments and their outcomes. Lastly, the report is concluded in Section V with a brief overview of the results and an outline of possible future research. Section Two: Works Connected When evaluating the state of a road or highway for the first time, imagery capturing is essential. A unmanned aerial vehicle (UAV), and more especially a drone, can take comprehensive pictures of the road surface from different angles quickly and cheaply. The DJI Mavic Air 2S, a more modern drone model that was specifically allocated for this purpose, was used in this investigation. This drone can take very accurate pictures of road surfaces thanks to its high-resolution camera, GPS, and obstacle avoidance sensors. More thorough road surface coverage, particularly in inaccessible regions, may be achieved safely and rapidly with the use of a UAV. Deep learning and UAV algorithm improvements are the main topics of related publications. Using deep learning techniques and geotagged ultrasonic beacons, autonomous UAVs have been used for structural health monitoring and real-time damage mapping, for instance [19], [20].

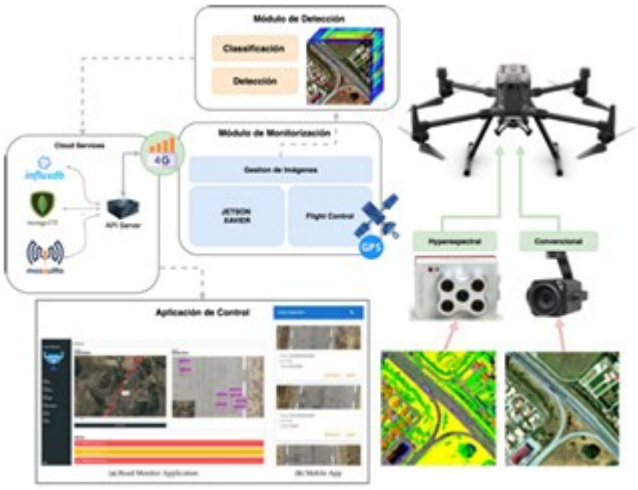
In a number of fields, including electric component detection[25], wind generator inspection[24], animal identification[23], vehicle traffic monitoring[21], and huge population monitoring[22], deep learning approaches like CNNs have shown encouraging results. Automated road damage identification is made easier using these approaches, which may also be used to video or still photos captured by onboard cameras to identify potholes. Road damage detection is one area that stands to benefit from the fast development and widespread use of deep learning technologies, which are applicable across many industries including transportation. It is feasible to identify road potholes by analyzing footage or stills captured by cameras placed on automobiles using deep learning algorithms like convolutional neural networks (CNNs). Using deep learning algorithms is a basic method to automated road damage identification. These algorithms are quite good at detecting damage and other various items. Convolutional Neural Networks (CNNs) are often used in this field for deep learning. A deep convolutional neural network (CNN) was suggested by the authors of the article [26] as a means of detecting road damage from UAV photos. After being trained and evaluated on a dataset consisting of UAV photos, the suggested CNN demonstrated its ability to reliably identify road damage. Without using image processing methods (IPTs) to extract fault information, suggests a new method for identifying concrete fractures using a deep CNN architecture in [27]. The CNN trains on a massive dataset consisting of 40,000 pictures and attains an accuracy level of about 98%. When compared to more conventional approaches to edge identification, such as the Canny and Sobel techniques, the suggested method outperforms them over a range of test structures and environmental variables. Another study that came out recently [28] suggested a way to automatically identify road damage using UAV photos that relies on deep learning-based object recognition. For object detection, they turned to the Faster R-CNN method. The results showed that compared to other road damage detecting systems, the suggested one is the best. In addition, the authors of [29] and [30] suggest using R-CNN and its improved version, Faster R-CNN, for structural visual inspection. This method can identify various damages, such as concrete cracks, steel corrosion, bolt corrosion, and steel delamination. With the suggested approach, we may get an average precision rating of 87.8 percent. With a test time of just 0.03 seconds per picture, the suggested technique is lightning quick and has applications in areas such as trained network-based quasi-real-time damage detection in video. Using Faster R-CNN in conjunction with tweaks to TuFF and DTM, a fracture detection and quantification approach is finally proposed in [31].

With an average precision of 95%, an intersection over union of 83%, and a crack length accuracy of 93%, the suggested approach attained very high levels of accuracy. Using a deep learning-based image processing algorithm with super-resolution and semi-supervised learning techniques based on GAN, the authors of [32] created a novel sensor technology for road damage identification. On 400 road photos, the suggested technique achieved an average recognition performance of 81.54% by mean junction over union and 79.228% by F1-score. In the future, the research says, the suggested strategy may be employed for effective road management. These days, it's not enough to only identify damage in structural photos. Quantifying the damage by measuring the magnitude of the observed faults is crucial for a comprehensive understanding and assessment of its extent. In order to precisely define the bounds of the damaged regions in the picture, a more sophisticated method called pixel-level segmentation is needed. In order to achieve high performance and rapid processing speed in crack segmentation in complicated sceneries, Kang [33] suggests a new semantic transformer representation network (STRNet). A high-performance deep learning network for real-time pixel-level segmentation of internal damages in concrete members using active thermography was proposed in

[34], and the network was evaluated and compared with other advanced networks, demonstrating superior performance and processing speed compared with other networks. With a mean intersection over the union of 0.900, a positive predictive value of 0.952, an F1-score of 0.941, and a sensitivity of 0.942, the attention-based IDSNet much beats state-of-the-art networks. The alternative perspective is known as Single-Shot Detection (SSD), and it is designed to identify damage to roads or concrete. Using a manually generated dataset, the SDDNet, a deep learning network for real-time crack segmentation in photographs of concrete, achieved good accuracy (as reported in [35]). The model processes pictures at 36 frames per second, which is a substantial improvement over earlier work, and it beats more contemporary models. Arya et al. [36] detailed a collection of cutting-edge methods for classifying and detecting road damage on a worldwide scale. As an example, Pham et al. [37] conducted an experiment using Detectron2 and Faster R-CNN. From what we can tell from the research we looked at, the Faster R-CNN model outperforms the YOLO model in terms of accuracy, but at the cost of 8 frames per second more prediction time. When it comes to time and accuracy of predictions, SSD strikes a compromise between the two.

YOLO IMPLEMENTATIONS

In order to enhance road maintenance and decrease accidents, [38] proposes a deep learning method for detecting potholes on Indian roadways by using the YOLO technique. We train YOLOv3, YOLOv2, and YOLOv3-tiny on a new dataset consisting of 1500 photos of roads in India, and then we compare the results in terms of accuracy. Alternatively, in [39], the authors introduce the M-YOLO, a lightweight network architecture that improves the detection efficiency of pavement oil repair using UAV photos. It is based on MobileNet V3 and YOLOv5S. Experimental findings showed that compared to YOLOv3, the M-YOLO algorithm outperformed it in every respect: accuracy (98.3%), speed (96.6% fps), and number of parameters (95.5% average). Furthermore, a unique automatic pavement distress detection framework using deep learning and stereo vision is presented by the authors in [12]. Various situations are tested on asphalt roads using the suggested approach. Compared to existing models, the improved 3D crack segmentation model outperforms them in inference speed and accuracy, reaching millimeter-level precision in crack and pothole segmentation. Accurate volume measurements of potholes are also achieved by use of the high-resolution pothole segmentation map. Using multi spectral pictures to identify road defects is another method connected to the literature [40]. Unmanned Aerial Vehicle (UAV) multispectral imaging is an effective method for finding and analyzing road damage. Alternatively, hyperspectral pictures might be used to identify pavement fractures on roads. The research [41] presents an asphalt crack index that is effective for crack identification. It outperforms the current metric in literature by an average of 21.37% in terms of F1-score. Hyperspectral image categorization may also make use of Transformer and Convolutional Neural Networks (CNNs) [42]. While convolutional neural networks (CNNs) learn spectral spatial patterns to extract features from hyperspectral data, transform-formers model long-range relationships to capture global contextual information. When applied to hyperspectral image classification tasks, both methods have shown encouraging outcomes. There is a wealth of literature on this rapidly evolving topic; more research is needed to determine the most effective method for this particular situation. We opted for YOLO in this project since we believe it to be the most effective method. The most current version was YOLOv4 when we authored the first article. Extensive testing of the most recent version of this technique, YOLOv7, is done here. At this time, the recognized authority is the study of Wang et al. [18].



The suggested method's design is shown in Figure 2. Some of the authors are the same as in the YOLOv4 version [16]. Five to one hundred and sixty frames per second are YOLOv7's speed and accuracy range. Using the available freeware and experimenting with different hyperparameters and models, this project

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

trained models to identify and classify road dam ages. Section III.

DESIGN AND IMPLEMENTATION

A. IMPLEMENTATION

Preventing defects on roadways, streets, highways, and other surfaces used by vehicles is the primary goal of the concept. Figure 2 shows the original project concept, which called for a commercial drone outfitted with a high-resolution camera—specifically, a multispectral camera. Just as its name implies, the multispectral camera can capture many light spectra. This article's dataset only makes use of high-resolution camera pictures and does not include any data acquired with a multispectral camera. B. Image Dataset from UAVs At the beginning of our work, we sought for a dataset of asphalt potholes and cracks by conducting a literature search. The databases, however, were distinct from the present recommendation of taking pictures from a safe distance from the road using an unmanned aerial vehicle. For this reason, an updated dataset was necessary for a realistic representation of the road conditions in Spain. The total number of photographs shot was 600, and their resolution was 3840×2160 pixels. The photos, captured from a DJI Air 2S drone flying 50 meters above Spanish roadways, only included two categories: potholes (D40) and cracks (D00). There were 568 labeled pictures that were retrieved after creating the dataset and tagging all of the shots. During the pre-processing step, the orientation of the photographs was changed and they were given a new size of 640 × 640. The augmentation methods were used to produce different versions of every photograph in the collection.

TABLE 1. Spain roads dataset

Class	Spain Annotations
D00 (Crack)	327
D40 (Pothole)	3480

The photos have zoom settings between zero and fifteen percent. There are a total of 1,362 photos in the collection. The training model used 70% of the pictures, the validation model 20%, and the effectiveness test 10%. Previous work [7] made use of this dataset, and you can find it in the repository. 1 The organization of the categorization is shown in Table 1. In order to automatically identify road damage from the gathered films, we used the prior datasets (Spain) as a reference while continuing to construct the dataset. We also included the dataset from the CRDDC2022. Road damage in many nations is documented in this dataset [43]. Automated pavement distress detection machine learning models are trained and tested on this benchmark dataset. A total of 47,420 road images across five nations (India, China, Japan, the Czech Republic, and Norway) make up the collection. The four kinds of pavement damage—alligator cracks (D20), transverse cracks (D10), longitudinal cracks (D00), and pothole cracks (D40)—are identified using these photographs in the training and testing of models. For the purpose of training machine learning models to identify the four distinct kinds of pavement damage, this dataset is used. Using the training set of images, the models learn to distinguish between different kinds of damage and their defining traits. To check how well the trained models did, we utilize the testing set of this dataset. The accuracy of the models is assessed by applying them to the testing set of photographs and comparing their predictions to the actual labels. Because it offers a big and varied collection of photos, this dataset is helpful for academics and engineers working on automated pavement distress identification. Models trained using this dataset will be able to generalize well to varied road conditions and environments if it includes photos from multiple nations. Images from satellites, high-resolution cameras, and cellphones were used to compile these sets. We use automobiles, motorbikes, and drones to get all of it. Table 2 clearly shows the distribution of damage kinds (of the four significant damage types) across nations. Two datasets pertaining to China were made available: one for mobile phone pictures (Ch_M) and another for drone photographs (Ch_UAV). As noted in the table above, we used the first dataset of Spanish roads plus a tiny portion of the photographs captured by drones in China (Ch_UAV) to compile the dataset for this article. There are two supplementary classes in this dataset as well. Fix, which can mean both roadwork and Block Crack.

	JPN	India	CZ	NW	US	Ch_M	Ch_UAV
D00	4049	1555	988	8570	6750	2678	1426
D10	3979	68	399	1730	3295	1096	1263
D20	6199	2021	161	468	834	641	293
D40	2243	3187	197	461	135	235	86

TABLE 2. Damage category-based data statistics for RDD2022.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

FIGURE 3. UAV used to obtain images for the dataset.

TABLE 3. Damage category-based data statistics for the merged dataset

Class	China_Drone	Spain	Total
D00	1426	327	1753
D10	1263	0	1263
D20	293	0	293
D40	86	3480	3566
Repair	769	0	769
Block Crack	3	0	3

TABLE 4. Dataset split.

	Number of Images	Percentage
Training Set	4000	82%
Validation Set	584	12%
Testing Set	289	6%
TOTAL	4873	100%

The proposed training set became more heterogeneous when we included the China Drone data. Their goals were congruent with those of this project and RDD2020, both of which aim to provide practical, low-cost solutions for autonomous road damage identification. Fig. 3(a) and Fig. 3(b) show the two nations' UAV-obtained photos, which make up the final dataset of 2893 images. The distribution of classes and annotations for the China Drone and Spain datasets, as well as an overview of the damage category-based data statistics for the combined dataset, are shown in Table 3. The dataset has been scaled to 640 × 640 and auto-oriented for enhancement and preprocessing. In order to artificially boost the quantity and variety of the dataset, the photos were enhanced. In this situation, two outputs have been generated from each training example. In order to make the model more resistant to various object orientations, the photos have been randomly rotated between -15◦ and +15◦ using the rotation. Table displayed the final dataset description.

DATA PREPARATION

The two datasets were combined and split into three variants corresponding to YOLOv4, YOLOv5, and YOLOv7, as shown in the preceding table. Two directories are required: one for validation and one for training. You also need to include the picture and label directories in these two files. Each labelled picture would have its own text file with the image annotation, while the images themselves would have the photographs. The name of the picture file should correspond to the name of the text file. The folder's structure mirrors that of the YOLO dataset after fresh YOLO annotations have been generated. The file "data.yaml" contains information about the data set, such as the names and the number of classes. All datasets are kept on the platform Roboflow, which allowed all of this to be done. Section D: Competence and Modeling This study follows the coordinated prediction approach similar to YOLOv2 and YOLOv3, and it was initially based on the YOLOv4-tiny model. Unlike previous versions, which only allowed for single-class classification, this one allows for multi-class categorization as well. Two classes were intended to be detected by this first network from the 568 photos. The number of classes was later raised to six with the introduction of YOLOv5, YOLOv5-Tranformer, and YOLOv7. Following that, fresh training was carried out using the four thousand photos and the six courses. We prepped the data and entered it into our YOLO models so that they could be trained. Alligator cracks (D20), pothole cracks (D40), transverse cracks (D10), repair and block cracks (D00), and longitudinal cracks (D00) are all detectable by the trained model. In order to train the model, we used 4873 pictures from the dataset that was created by the roboflow platform. Because of the abundance of affordable GPUs, the models detailed in this study were trained and validated on a computer with an Intel(R) Core(TM) i9-10940X CPU@3.30GHz, 128GB of RAM, and an RTX3090 GPU with 24GB of integrated memory.

EXPERIMENTAL RESULTS AND DISCUSSION

We compare the final photographs recognized by the algorithm to the labeled dataset photos, paying attention to quantitative aspects, in order to evaluate

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

our system. In this scenario, various tests provide distinct models, and while using three distinct models, it is crucial to closely observe the assessment procedure. Hence, a robust measure is required to choose the most appropriate models from all the trials. Two commonly used metrics for evaluation are available in this domain. The Mean Accuracy (mAP) at the 0.5 Intersection Overlap (IoU) threshold (mAP@0.5) is the principal one. The F1 score is the second one. The MAP was

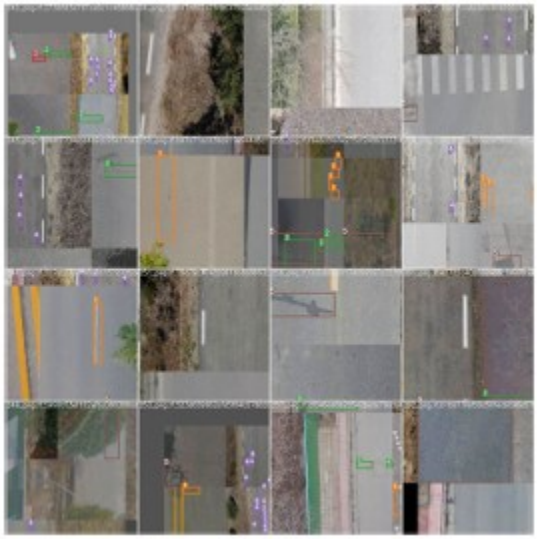


Figure 5 shows the visual distortion and inconsistencies that might result from using YOLOv7's default settings on enhanced images. a useful metric to use for checking that the model remains consistent across several confidence levels (robust), since the F1 score is calculated for a given confidence level. By default, when working with validation data, mAP@0.5 is used to choose the best model, and when reporting model performance on test data, the F1 score is used. Similar to other projects, this one uses mAP@0.5 to find the top models and then reports the F1 scores from the test sets. We will use the following metrics for the quantitative assessment: the accuracy, which may be expressed as the ratio of positive results (TP) to total results (TP + FP) Basic formula 1. The second equation integrates the recall, the probability of an image being classified as positive, and the ratio of TP to TP + FN. Combining the first two measurements, the third and last one is the F1 metric. In addition, there is the identification of overlap (IoU), which is the area where the detected and imaged regions overlap, the mean average precision (mAP), and the classification speed, which is measured in frames per second (FPS).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

We tested the model training procedure in three stages, varying the number of iterations and the resolution of the images. During training, the best models are selected using the validation data and the metrics mAP@0.5. Along with the primary assessment procedure, we also included four new approaches to assess the comparison analysis and the

TABLE 5. Performance metric for YOLOv4.

Label	TP	FP	AP (%)
Block crack	0	0	0.00%
D00	87	106	27.14%
D10	149	232	35.17%
D20	4	14	10.39%
D40	162	42	44.25%
Repair	63	69	44.24%

carry out. Through the use of ensemble approaches, error analysis, transfer learning, and hyperparameter tweaking (aka evolve). • Hyperparameter tuning: This is the process of optimizing a model's or algorithm's setup for a specific job by modifying its settings and parameters. Researchers may find the optimal settings and learn which elements are most significant by regularly changing the values of various parameters. • Analyzing model or algorithm mistakes for trends or patterns is known as error analysis. In order to learn about a system's limits and find ways to make it better, researchers look at the particular instances when it fails. • Transfer learning: this method entails tailoring an algorithm or model that has already been trained to do a certain job. It is common for researchers to get better results with less data and training time when they use pre-trained models, which include the expertise and experience of other researchers. To increase overall performance, ensemble approaches combine the predictions of numerous models or algorithms. Ensemble approaches are able to outperform individual models by combining their strengths and addressing their faults. Section A. YOLOv4 Experiments. We executed the first processing phase using YOLOv4. By using a pre-trained weight model, the convolutional layers were fine-tuned as needed. This training's outcomes may have been better compared

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

to the previous work, suggesting that the previous work was overfitting or overtraining. Here, we got an F1score of0.39, a recall of0.32, and an accuracy of 0.50. In under three seconds of detection time, they picked a mean average precision (mAP@0.50) of 0.268638 percent for this training. Table 5 provides a breakdown of the performance by class. B. FUN YOLOv5 TESTS Table 6 shows that as compared to YOLOv4, using YOLOv5v7.0 for road damage identification produced considerable gains. Using an IoU threshold of 0.5 (mAP@.5), the model YOLOv5x significantly improved its capacity to reliably identify road damage, going from 26.86% to 59.90%.

TABLE 6. Performance metric for YOLOv5.

Class	Imgs	Labels	P	R	mAP@.5	mAP@.5
all	584	1447	0.787	0.561	0.599	0.344
Block	584	1	1	0	0.0203	0.0142
D00	584	367	0.68	0.557	0.594	0.294
D10	584	248	0.778	0.742	0.829	0.46
D20	584	65	0.702	0.507	0.566	0.299
D40	584	623	0.814	0.785	0.817	0.419
Repair	584	143	0.745	0.776	0.766	0.578

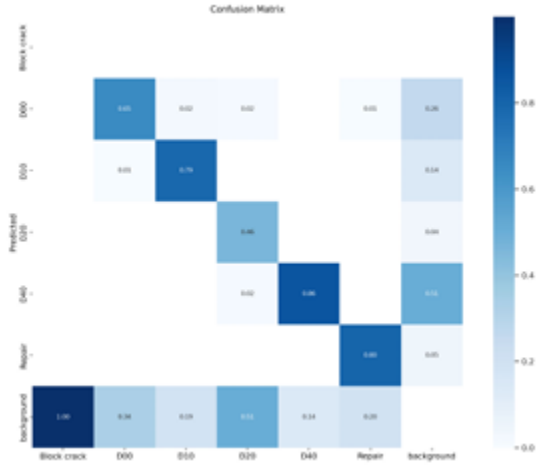


FIGURE 6. Confusion matrix for the YOLOv5 model.

There was a noticeable improvement of around 27% in the mAP when the IoU criterion was 0.5 and the recall threshold was 0.95 (mAP@.5:.95). The percentage of accurately recognized real positive events, known as the recall percentage, likewise rose from 32% to 56.10%. The remarkable accuracy and recall of YOLOv5 in identifying road damage is shown by these data. In addition, at shape (1, 3, 640, 640), YOLOv5's inference time is 17.2 ms, comprising 0.9 ms for pre-processing and 17.2 ms for inference plus 6.8 ms for Non-maximal Suppression (NMS). These results demonstrate the accuracy and processing efficiency of YOLOv5. Figure 6 shows the confusion matrix for six groups that were identified using test data and YOLOv5. The model achieved a high rate of accurate classifications. The ground truth is shown on the horizontal axis, while the predicted classes are shown on the vertical axis. A high degree of accuracy is shown by the fact that the diagonal elements, which stand for the properly categorized classes, are the greatest among all matrix elements. Some misclassifications, however, are also seen in classes D10 and D40. The model may have been more adept at identifying these classes if they had been more prevalent in the sample. On the whole, nonetheless, the data show that YOLOv5 is effective. Figure 7 also shows the F1 scores we computed for each class. Being accurate is more important than this.

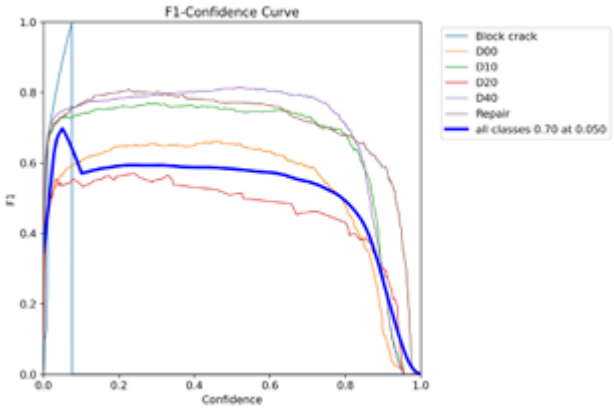
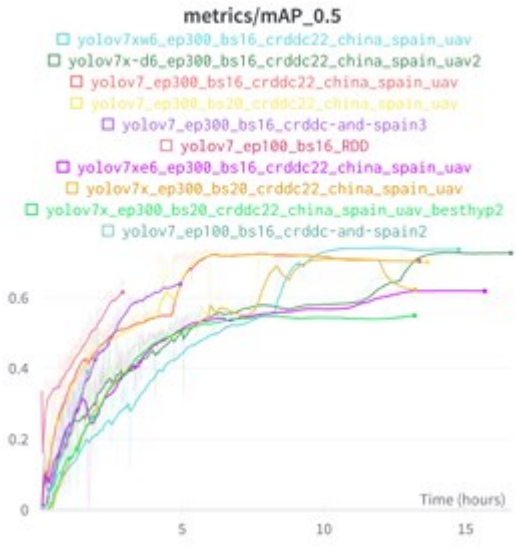


FIGURE 7. F1-Confidence curve



INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

useful indicator. The difficulty level rises due to the visual and structural differences across the six categories. The overall F1 score achieved by the system was 70%. Class allocations of F1 scores, however, range from 50% to 80%. The F1 score is highest in the D40 class and lowest in the D20 class because of the D20 class's worse recall. A result that is corroborated by the Confusion Matrix that was previously shown. C. Experiments with YOLOv7 The YOLOv7x-W6 model with 300 epochs was used to train the best YOLOv7 experiment. After 16 batches were completed, the grid study proved that the other YOLO models (YOLOv7-W6, YOLOv7-E6, YOLOv7-D6, and YOLOv7-E6E) were accurate in prediction and classification. Altering the batch size and hyperparameters resulted in 10 completed trains using these models, as shown in Figure 8. With an overall mAP of 0.737, YOLOv7 did a respectable job at detecting the various Table 7. classifications of pavement deterioration using the YOLOv7 algorithm. The Repair class achieved the greatest mean absolute precision (mAP) with 0.827. Due to its limited prevalence across all datasets, the Block crack class was removed, although achieving 100% recall and accuracy, demonstrating that the model could reliably recognize this form of damage. Nevertheless, the model had difficulty correctly detecting this damage in some classes, such D20, which had lower precision and recall levels. Nevertheless, the model was able to successfully identify pavement degradation in general, with an overall accuracy of 0.788 and recall of 0.714. All of this is shown by the YOLOv7's faster inference time of only 11.4 ms. At batch size 1, the speed for each 640 × 640 picture is 11.4/6.7/18.1 ms for inference, NMS, and totality. All types of pavement damage are broken down into their respective Precision (P), Recall (R), and mean Average Precision (mAP) ratings in Table 7. Table data shows that compared to other classes, D00 and D20 have a weaker detection impact. One possible explanation is that the model has a harder time differentiating between classes D00 (potholes) and D20 (alligator cracks) due to their shared visual characteristics. To make YOLOv7 a better detector on D00 and D20, we might use a few targeted tactics. The training data may be more varied, incorporating variations in illumination, angle, and distance, for instance, by using data augmentation methods. Before fine-tuning the individual classes, the model might be pre-trained using a large dataset of comparable items, such pictures of road surfaces, using transfer learning. Additionally, in order to better manage the complexity and variances of potholes and alligator cracks, the model architecture was altered in the YOLOv5 with Transformer Head. The next sections will elaborate on this enhancement. Figure 9 displays the confusion matrix for six groups that were classified using test data and YOLOv7. The results demonstrate that the model properly classifies the majority of the classes. The predicted classes are shown on the vertical axis, while the ground truth is shown on the horizontal axis. A high degree of accuracy is shown by the fact that the diagonal elements, which stand for the properly categorized classes, are the greatest among all matrix elements. The overrepresented D10 and D40 classes are more prominent in this matrix when contrasted with the YOLOv5 matrix. This could be because YOLOv7 is better able to identify these classes, despite their overrepresentation, which causes it to

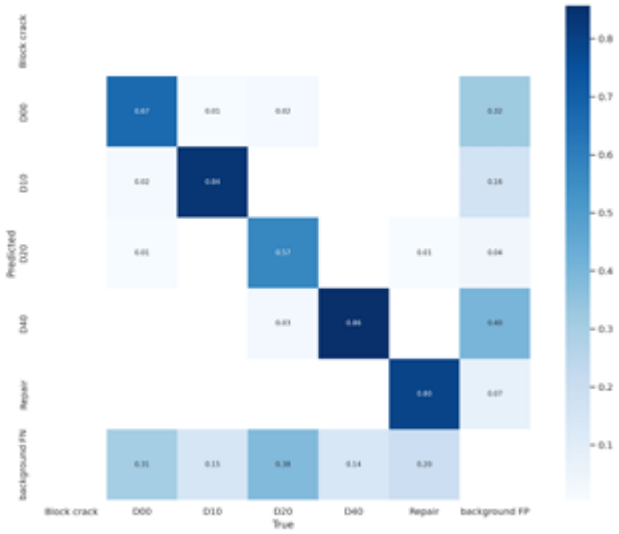


FIGURE 9. Confusion matrix for the YOLOv7 model

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

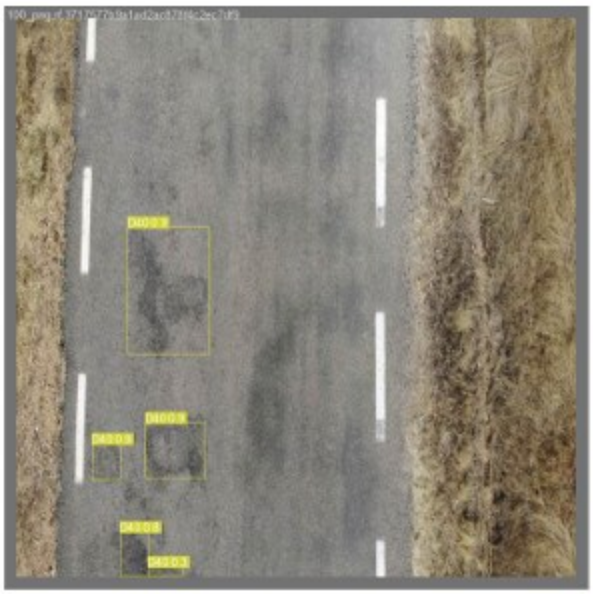
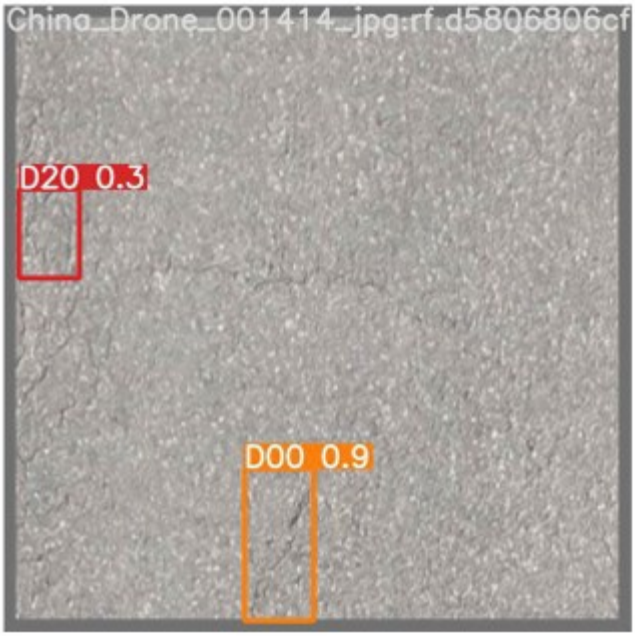


FIGURE 14. Validation YOLOv5.

TABLE 9. Performance comparison for YOLOv4, YOLOv5 and YOLOv7

	Imgs	P	R	@.5	.5:.95	Time
YOLOv4	584	0.50	0.32	0.26	-	3s
YOLOv5x	584	0.78	0.56	0.59	0.895	17.2n
YOLOv5 + TPH	584	0.71	0.67	0.65	0.35	6.1ms
YOLOv7	584	0.65	0.78	0.73	0.289	11.4n



CONCLUSION AND FUTURE WORKS

Finally, 62929 VOLUME 11, 2023 is a part of this study's comparison of the YOLOv4 from previous work with the YOLOv5 and YOLOv7 designs. Automated Road Damage Detection Using UAV

photos and Deep Learning Techniques is a work by L. A. Silva et al. that uses YOLOv5 with Transformer to identify road damage from UAV photos. The study accomplished its aim of developing an architecture that can identify road damage and shown that newer versions of the design, including YOLOv5 and YOLOv7, may enhance earlier efforts. This work made a big deal by creating a UAVimagedatabase that was optimized for training the YOLOversions and then improving it by combining it with the RDD2022 dataset. This helped level the playing field when it came to detecting potholes and alligator cracks, among other types of road deterioration, and it worked especially well for roads in China and Spain. The study's results represent an important step toward further investigation in this area and a significant addition to the field as a whole. In the results section, we showed that our implementation obtained mAP.5 of 26.8% with YOLOv4, 59.9% with YOLOv5, and 73.20% with YOLOv7. Lastly, the implemented Transformer achieved 65.7%. Our work still has room for improvement. To further improve performance, future study may investigate other picture kinds, including multispectral images and LIDAR sensors. Using an embedded computer, it may be feasible to fuse such data in order to get superior results. Also, fixed-wing UAVs are another option for this kind of job.

REFERENCES

[1]. H. S. S. Blas, A. C. Balea, A. S. Mendes, L. A. Silva, and G. V. González, ‘‘A platform for swimming pool detection and legal verification using a multi-agent system andremoteimagesensing,’’Int.J. Interact. Multimedia Artif. Intell., vol. 2023, pp. 1–13, Jan. 2023.

[2]. V.J. Hodge, R.Hawkins, andR.Alexander, ‘‘Deep reinforcement learning for drone navigation using sensor data,’’ Neural Comput. Appl., vol. 33, no. 6, pp. 2015–2033, Jun. 2020, doi: 10.1007/s00521-020-05097-x.

[3]. A. Safonova, Y. Hamad, A. Alekhina, and D. Kaplun, ‘‘Detection of Norway spruce trees (Picea abies) infested by bark beetle in UAV images using YOLOs architectures,’’ IEEE Access, vol. 10, pp. 10384–10392, 2022.

[4]. D. Gallacher, ‘‘Drones to manage the urban environment: Risks, rewards, alternatives,’’ J. Unmanned Vehicle Syst., vol. 4, no. 2, pp. 115–124, Jun. 2016.

[5]. L. A. Silva, A. S. Mendes, H. S. S. Blas, L. C. Bastos, A. L. Gonçalves, and A. F. de Moraes, ‘‘Active actions in the extraction of urban objects for information quality and knowledge recommendation with machine learning,’’ Sensors, vol. 23, no. 1, p. 138, Dec. 2022, doi: 10.3390/ s23010138.

[6]. L. Melendy, S. C. Hagen, F. B. Sullivan, T. R. H. Pearson, S. M. Walker, P. Ellis, A. K. Sambodo, O. Roswintiarti, M. A. Hanson, A. W. Klassen, M. W. Palace, B. H. Braswell, and G. M. Delgado, ‘‘Automated method for measuring the extent of selective logging damage with airborne LiDARdata,’’ISPRSJ.Photogramm.RemoteSens.,vol.139,pp. 228–240, May 2018, doi: 10.1016/j.isprsjprs.2018.02.022.

[7]. L. A. Silva, H. S. S. Blas, D. P. García, A. S. Mendes, and G. V. González, ‘‘An architectural multi-agent system for a pavement monitoring system withpotholerecognitioninUAVimages,’’Sensors,vol.20,no.21,p. 6205, Oct. 2020, doi: 10.3390/s20216205.

[8]. M. Guerrieri and G. Parla, ‘‘Flexible and stone pavements distress detection and measurement by deep learning and low-cost detection devices,’’ Eng. Failure Anal., vol. 141, Nov. 2022, Art. no. 106714, doi: 10.1016/j.engfailanal.2022.106714.

[9]. D. Jeong, ‘‘Road damage detection using YOLO with smartphone images,’’ in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2020, pp. 5559–5562, doi: 10.1109/BIGDATA50022.2020.9377847. 62930

[10]. M. Izadi, A. Mohammadzadeh, and A. Haghighattalab, ‘‘A new neuro-fuzzy approach for post-earthquake road damage assessment using GA and SVM classification from QuickBird satellite images,’’ J. Indian Soc. Remote Sens., vol. 45, no. 6, pp. 965–977, Mar. 2017.

[11]. Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, and A. Akula, ‘‘Convolutional neural networks based potholes detection using thermal imaging,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 34, no. 3, pp. 578–588, Mar. 2022, doi: 10.1016/j.jksuci.2019.02.004.

[12]. J. Guan, X. Yang, L. Ding, X. Cheng, V. C. Lee, and C. Jin, ‘‘Auto mated pixel-level pavement distress detection based on stereo vision and deep learning,’’ Automat. Constr., vol. 129, p. 103788, Sep. 2021, doi: 10.1016/j.autcon.2021.103788.

[13]. D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, ‘‘RDD2022: A multi-national image dataset for automatic road damage detection,’’ 2022, arXiv:2209.08538.

[14]. J. RedmonandA.Farhadi, ‘‘YOLO9000: Better, faster, stronger,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),

Honolulu, HI, USA, 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[15]. [15] J. Redmon and A. Farhadi. YOLOv3: An Incremental Improvement. [Online]. Available: https://pjreddie.com/yolo/

[16]. A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, ‘‘YOLOv4: Optimal speed and accuracy of object detection,’’ 2020, arXiv:2004.10934.

[17]. G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, J. Fang, C. Wong, D. Montes, Z. Wang, C. Fati, J. Nadar, V. Sonck, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, ‘‘Ultralytics/YOLOv5: V7.0—YOLOv5 SOTA realtime instance segmentation,’’ Zenodo, Tech. Rep., Nov. 2022. [Online]. Available: https://zenodo.org/record/7347926

[18]. C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, ‘‘YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,’’ 2022, arXiv:2207.02696.

[19]. R. Ali, D. Kang, G. Suh, and Y.-J. Cha, ‘‘Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural net works for GPS-denied structures,’’ Autom. Construct., vol. 130, Oct. 2021, Art. no. 103831. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092658052100282X

[20]. D. Kang and Y.-J. Cha, ‘‘Autonomous UAVs for structural health monitor ing using deep learning and an ultrasonic beacon system with geo-tagging: Autonomous UAVs for SHM,’’ Comput.-Aided Civil Infrastruct. Eng., vol. 33, no. 10, pp. 885–902, Oct. 2018.

[21]. Z. Xu, H. Shi, N. Li, C. Xiang, and H. Zhou, ‘‘Vehicle detection under UAVbased on optimal dense YOLO method,’’ in Proc. 5th Int. Conf. Syst. Informat. (ICSAI), Nov. 2018, pp. 407–411, doi: 10.1109/ICSAI.2018. 8599403.

[22]. P. Kannadaguli, ‘‘YOLO v4 based human detection system using aerial thermal imaging for UAV based surveillance applications,’’ in Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA), Nov. 2020, pp. 1213–1219, doi: 10.1109/DASA51403.2020.9317198.

[23]. T. Petso, R. S. Jamisola, D. Mpoeleng, and W. Mmereki, ‘‘Individual animal and herd identification using custom YOLO v3 and v4 with images taken from a UAV camera at different altitudes,’’ in Proc. IEEE 6th Int. Conf. Signal Image Process. (ICSIP), Oct. 2021, pp. 33–39, doi: 10.1109/ICSIP52628.2021.9688827.