Deep Learning Enhanced Feature Extraction of Potholes Using Vision and LiDAR Data for Road Maintenance

by Karukayil, A., Quail, C. and Cheein, F.A.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 September 2024, accepted 11 November 2024, date of publication 9 December 2024,
date of current version 17 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3512783

RESEARCH ARTICLE

Deep Learning Enhanced Feature Extraction of Potholes Using Vision and LiDAR Data for Road Maintenance

ABHIRAM KARUKAYIL[1], CHRISTOPHER QUAIL[1], AND FERNANDO AUAT CHEEIN[2,3], (Senior Member, IEEE)

[1]School of Engineering and Physical Sciences, The National Robotarium, Heriot-Watt University, EH14 4AS Edinburgh, U.K.
[2]Department of Engineering, Harper-Adams University, Newport, TF10 8NB England, U.K.
[3]Department of Electronic Engineering, Advanced Center of Electrical and Electronic Engineering, Federico Santa Maria Technical University, Valparaiso 1680, Chile

Corresponding authors: Abhiram Karukayil (ak5001@hw.ac.uk) and Fernando Auat Cheein (fauat@harper-adams.ac.uk)

This work was supported in part by The National Robotarium–the Institute of Sensors, Signals and Systems, Heriot-Watt University; and in part by the Advanced Centre of Electrical and Electronic Engineering (ANID), Universidad Tecnica Federico Santa Maria (UTFSM), under Grant AFB240002.

ABSTRACT As the global population increases, so does the number of vehicles on our roads, which makes maintenance of the road infrastructure critical for safe and efficient transportation. A significant challenge in road maintenance is to address surface defects, such as potholes, which pose the risk of accidents and vehicle damage. This work proposes an automated solution to improve the detection and aid in the repair of potholes, thus reducing the reliance on manual inspections and reducing the overall maintenance time. Our methodology integrates LiDAR (Light Detection and Ranging) with RGB (Red, Green, and Blue) camera data to enhance depth information for accurate pothole characterisation. Geo-positioning using the GNSS (Global Navigation Satellite System) allows for precise mapping of detected potholes. An RGB image dataset created by aggregating publicly available pothole image datasets was used to train the object detection model YOLO (You Only Look Once) implemented in this work. Using this data, the models YOLOv5, YOLOv6, YOLOv7, and YOLOv8 were trained and their performance analysed. Remarkably, YOLOv5 showed the best implementation performance during the training phase, and it was lately selected for real time deployment. The data provided by the LiDAR sensor were used to compute the area, volume and depth of the detected pothole using the Convex Hull approaches. During deployment on Edinburgh City roads, our work was able to effectively detect and characterise 52 potholes of different volume and area. The implementation of this technology has the potential to significantly reduce inspection time, and our findings offer promising directions for future developments in automated road maintenance systems.

INDEX TERMS Machine learning, machine vision, pothole detection, road interpretation.

I. INTRODUCTION

ThE infrastructure of road transportation serves a major role in connecting various locations and maintaining the mobility of goods and people [1], [2]. Therefore, maintaining the infrastructure of the road is essential, as the condition of the road surfaces directly impacts the effectiveness of the transportation system. Potholes are a significant challenge, affecting the quality of the road surface and affecting travel safety and comfort [3]. According to the National Accident Helpline in the UK, 1,766 road accidents in the second quarter of 2020 were a result of potholes [4]. Between April 2018 and June 2021, more than 1.5 million potholes were reported, with 43,947 claims for damages, including over 7,000 from Scottish councils. Of these, 13,187 claims were successful in securing compensation [5]. Potholes not only damage vehicles, but also pose serious risks to cyclists and pedestrians, with these road defects that cause 15% of bicycle accidents [5]. In 2016, a cyclist in San Diego suffered serious injuries from a pothole, resulting in a $1 million

The associate editor coordinating the review of this manuscript and approving it for publication was Razi Iqbal.

settlement [6]. This review underscores the need for effective pothole detection and management strategies to address this global problem that affects road transport networks.

Contributions from researchers around the world have significantly advanced and improved the automatic pothole detection method. These methods can be broadly classified into vibration-based, vision-based, and depth-based approaches, each with their own advantages and limitations [7], [8]. In addition to detection and feature extraction, the geo-positioning of potholes on a map is essential for understanding the areas affected by potholes within road transportation networks. The following section outlines the key contributions made by each three methods.

Vibration-based detection uses sensors such as accelerometers and gyroscopes in vehicles to measure changes in acceleration, identifying potholes based on variations in road surface geometry [9]. However, it is often misclassified due to factors such as speed bumps, terrain transitions, along others, as these affect the variation in acceleration values [7], [8]. Although it is cost-effective [7], it lacks detailed extraction of the features of the pothole. Vision and depth based methods offer potential solutions by providing more comprehensive data on pothole characteristics. Integrating these methods enhances accuracy and informs targeted road maintenance strategies. In a study [9], researchers used three different algorithms within a vibration-based approach to detect potholes, relying on acceleration data as a key metric. Several factors can impair the accuracy of the acceleration data, causing false positives in the detection of potholes. To mitigate this, the authors precisely segmented acceleration data based on predicted factors such as calls, door slams, messages, potholes, short humps, and gutters, using advanced algorithms such as convolutional neural networks (CNN), reservoir computing (RC) and long-short-term memory (LSTM) networks to differentiate between terrains based on specific vibration thresholds. In [10], researchers introduced an innovative Dynamic Sliding Window (DSW) technique to analyse accelerometer data for pothole detection. This method segments sensor data into windows that are dynamically adjusted based on contextual factors such as vehicle speed, aligning window sizes with expected road anomaly sizes to improve accuracy while minimising interference from normal road conditions. To verify their methodology, the researchers collected data on various road anomalies such as speed bumps, metal bumps, and potholes across thirty different virtual road scenarios. They then assessed this data using seven heuristic algorithms such as Z-Thres, Z-Diff, STDEV (Z), G-Zero, Swarm, Pothole Patrol, and Nericell resulting in F1 scores of 0.725, 0.856, 0.853, 0.600, 0.852, 0.728, and 0.715, respectively. Furthermore, the study delved into the capabilities of machine learning classifiers by examining the performance of seven algorithms: Decision Trees (DT), Gradient Boosting (GB), Random Forest (RF), Naive Bayes (NB), Neural Networks (NN), k-Nearest Neighbor (kNN), and Support Vector Machine (SVM). The F1 scores obtained from these classifiers were 0.967, 0.917, 0.922, 0.888, 0.904, 0.973, and 0.922, respectively.

The implementation of vision-based pothole detection marks a significant advancement in road inspection techniques. These systems use cameras or image data, often integrated with GNSS for geo-positioning. Object detection architectures such as YOLO [3], [11], SSD [12], [13], RetinaNet [14], and ResNet [12] are some techniques used to detect potholes. Image enhancement improves detection under varying light conditions. In post-detection, image segmentation can measure pothole dimensions, aiding in damage severity categorisation for repair prioritisation. However, vision-based systems lack depth data, which highlights the need for advancements such as LiDAR [2]. The authors of [3] evaluated custom trained neural networks, YOLOv4, YOLOv4-tiny, and YOLOv5s, for real-time pothole detection. They used a dataset of 655 images that was split into 70 training, 20 validation, and 10 testing subsets. Then, they compared these models using the mean average precision (mAP) at 50% intersection over union. The results show mAP@0.5 values of 74.8% for YOLOv5S, 77.7% for YOLOv4, and 78.4% for YOLOv4-tiny, making YOLOv4-tiny the preferred model due to its balance of accuracy and computational efficiency. The study in [11] compared YOLO-v3, YOLO-v3 with Spatial Pyramid Pooling (SPP) and Sparse R-CNN. The models were tested under various lighting conditions, revealing that larger image input sizes improve detection accuracy in low light but increase inference time. YOLOv3's mAP@0.5 dropped from 74.7% on a clear day to 7.01% at night, while YOLO-v3 with SPP achieved 79.1% on clear days. Sparse R-CNN performed best in low light and rainy conditions. In [15], the authors used Mask R-CNN for pothole detection and risk assessment, achieving over 95% accuracy after 2000 training iterations. Their model segmented potholes within images, calculating the area in pixel values, and assessing risk based on the percentage of the pavement affected.

Depth-based pothole detection methods provide critical depth information, allowing precise estimations of volume, area, and shape. These methods use sensors like LiDAR, kinetic sensors, and stereo vision cameras, and some researchers used custom multi-camera setups. LiDAR-based methods less affected by lighting conditions that offer an advantage over vision-based methods, but can struggle with depth estimation when potholes are filled with water [16]. In [16], researchers used a Mobile Mapping System (MMS) with a SICK LMS 511 laser scanner, panoramic camera, GPS (Global Positioning System), and IMU (Inertial measurement unit). Their system used DeepLabv3+ for the initial detection of potholes, achieving 81.2% accuracy. The laser scanner then collected point cloud data to capture geometric features such as depth, area, and number of potholes, which were geographically positioned on a map. In [17], the HiPRoSS (High Performance Road Survey System) from the University of Messina used 3D cameras and linear lasers emitting

810nm infrared light to capture 3D data. After converting the data into 2D images, they were subjected to semantic segmentation, thresholding, and post-processing to remove false positives and refine pothole contours. This system achieved a precision of 89.75% on a dataset with 44 pothole images out of 50, calculating depth, perimeter, and area. In [18], researchers combined data from a 3D LiDAR (Velodyne VLP32) and a camera (Alvium 1800 U-158) to obtain accurate depth data. Through extrinsic calibration, a transformation matrix integrated the data from both devices. Using YOLOv7 for pothole detection, the system transferred bounding box coordinates to LiDAR data for point cloud analysis. The pothole area and volume were calculated using the k-means clustering and the convex hull algorithm, while depth was determined by comparing pavement and pothole point cloud data. This system, tested at vehicle speeds of 30, 40, and 50 mph, showed a deviation less than 8% in depth estimation and an error margin less than 20% compared to manual inspections. Each detected pothole was geo-positioned, providing valuable information for road maintenance.

To improve road safety, maintaining the health of the road network is a key element. Systematic inspections, necessary for detailed damage assessments, traditionally rely on manual labour, which is time-consuming and labour intensive [3]. Potholes, if not immediately addressed, escalate into more significant problems, increasing repair costs, and disrupting transportation [19]. To overcome these limitations, modern technologies such as cameras, LiDAR, and GNSS can automate the inspection process, allowing for precise pothole detection and 3D reconstruction. This enables accurate volume and area estimation, optimising repair material calculations, and reducing manpower requirements [13]. These advances offer a rapid, efficient, and cost-effective solution for road maintenance. In our system the LiDAR and camera has been calibrated after mounting on a custom made 3D printed mount, which keeps the relative position of the camera and LiDAR fixed. By this, it can be used as a standalone device in various inspection vehicles without the need for a recalibration. This helps ensure consistent performance and easy adaptability across different platforms.

II. SYSTEM OVERVIEW

Figure 1 outlines the system architecture designed for pothole detection, showcasing an inspection vehicle equipped with a modern range of data collection technologies. This setup includes an onboard computer for data recording and processing, accompanied by a camera for visual data capture, LiDAR to collect 3D point cloud data, and a GNSS module for geo-positioning the pothole location. The core objective of this experiment is to detect and extract detailed geometric data about the pothole along with its geographic coordinates. The process begins with the camera, which collects visual data to detect potholes. Although the camera provides valuable 2D data, it lacks the ability to capture depth data. To overcome this limitation and gain

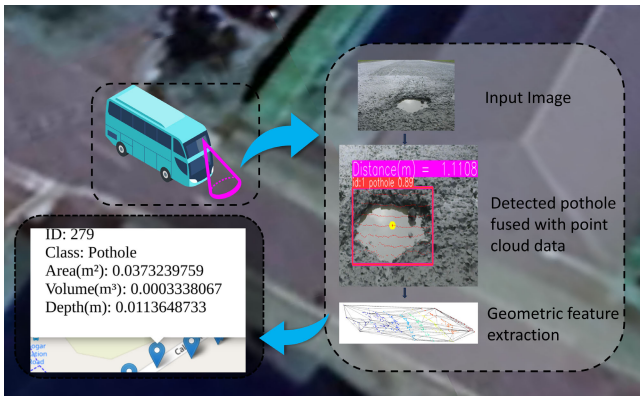

FIGURE 1. System architecture illustrating the processing cycle from capturing input image to feature extraction. The input image is processed through various stages, including pre-processing, detection, 3D reconstruction, and feature extraction, where key pothole attributes such as area, volume, and depth are computed.

a comprehensive understanding of the depth, volume, and area of each pothole, the camera and 3D LiDAR data were fused. This sensor fusion technique enriches the 2D pixel coordinate system with depth information, creating a more detailed representation of the pothole's geometry.

Upon activation of the system, the camera starts collecting 2D data. This visual information was fed into the YOLO (You Only Look Once) object detection algorithm. This cutting-edge algorithm detects potholes and locates them with bounding boxes. The coordinates of these bounding boxes are then passed to the LiDAR data processing algorithm to extract point cloud data that lies within the detected pothole's area. The point cloud data are further analysed using the Convex Hull algorithm, which calculates the pothole's area, volume, and depth. By combining the data with the GNSS coordinates, the system can accurately map each pothole. Additionally, the geometric characteristics of the pothole can be visualised on the map, alongside the location. This integrated approach improves the efficiency of feature extraction and pothole detection. By this, the system can provide comprehensive data that can be useful in road maintenance and repair strategies, potentially revolutionising the way infrastructure is monitored and maintained.

III. DATA ACQUISITION

This section outlines the process of collecting pothole images from online resources and preparing the data to train the custom pothole detection model using YOLO. In this section, we discuss the calibration of the camera and LiDAR.

A. TRAINING DATA ACQUISITION FOR POTHOLE DETECTION

In terms of performing custom object detection such as pothole detection, one of the initial steps is to find a dataset of pothole images. For this project, the pothole data were collected from various publicly available online datasets [20], [21], [22]. As the images were collected from different

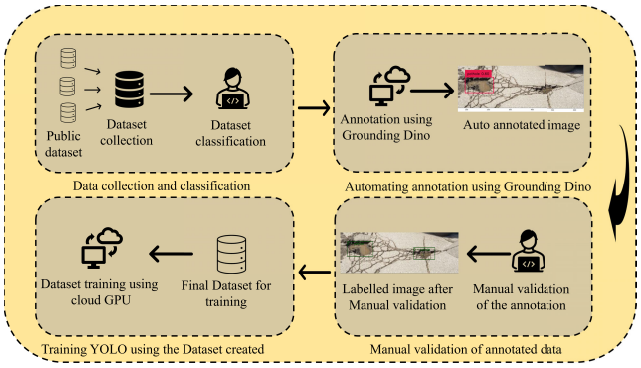

FIGURE 2. Pictorial representation of the process of collecting and preparing pothole images for training. The top left part shows how the images were collected from online datasets and classified. The upper right part illustrates how the classified images were annotated using Grounding Dino. The bottom left part shows how the annotated images were manually validated. Finally, the bottom right part displays the dataset with annotated images, ready for training.

datasets, their sizes varied. Therefore, they were resized to 640 × 480 pixels. This step also helps in testing various object detection architectures, as some require images of the same size for training. After analysing the dataset, high-quality images were selected to create a custom dataset for training and testing the object detection architecture. This custom made datasets were passed to the Grounding Dino [23] for automatic annotation of the images. The result from Grounding Dino was not highly reliable for detecting potholes. Therefore, after the completion of the annotation, a manual validation was performed and the annotations were edited to meet the standards. However, using automatic labelling saves a huge amount of time spent on the manual annotation method. The YOLO labelling data were made in the COCO (Common Objects in Context) format as it is the accepted input format for the YOLO training architecture. By this process, a dataset was created with 800 images for training, 200 images for validation, and 80 images for testing. Figure 2 shows the visual representation of the whole process of making the custom dataset for training the object detection model.

B. CAMERA LiDAR CALIBRATION

Extraction of the geometric features of the pothole is one of the primary objectives of this project. To achieve this, LiDAR technology plays an essential role. LiDAR provides detailed point cloud data surrounding the robot environment, which is essential for depth determination. For effective depth extraction of potholes, it is critical to synchronise the data captured by both the camera and the LiDAR. This involves aligning the camera's visual detection of potholes with the depth data collected by LiDAR. The key to this integration is establishing a common field of view between the camera and the LiDAR system. Once this common field of view is achieved, the depth information from the LiDAR can be precisely aligned with the pixel data from the camera. By fusing these two data sources, the system can improve the accuracy of the depth information associated with each detected pothole. This integrated approach ensures that the geometric features of the potholes are captured, facilitating better assessments and actions for road maintenance. For the purposes of our work, we have followed the guidelines previously published in [24].

To perform camera and LiDAR calibration, the camera calibration matrix was first calculated using the ROS (Robot Operating System) camera calibration package [25]. The calibration of the camera and LiDAR was performed using the ROS package developed by [26]. The authors introduced a method to enhance camera LiDAR calibration, focusing on reducing user errors and overfitting through a new metric, the Variable of Quality (VOQ). This approach highlights the role of quality sample selection in ensuring accurate and reliable calibration. By performing the LiDAR camera calibration, the rotation matrix (R) and the translation matrix (T) between the camera and LiDAR were obtained.

IV. DATA PROCESSING

This part of this article will talk about object detection models, camera and LiDAR data sensor fusion, feature extraction techniques, pothole classification approaches, and geo-positioning techniques for mapping the potholes. These components form an integrated system that enables efficient and reliable detection, classification, and localisation of potholes.

A. YOLO

YOLO (You Only Look Once) follows single-stage detection, because of this it has a higher detection speed [27]. For custom training a class outside the coco dataset or to create a new custom detection model, the image training is done using the pre-trained weight to create a new weight file. That is, by using transfer learning. The pre-trained weights are used for training. The selection of weights depends on the application. The YOLO architecture generally starts with a backbone responsible for extracting features from the input image. This is complemented by the neck, which enhances the image and facilitates further feature fusion. The output from the head of the YOLO architecture includes essential information such as a bounding box, bounding box coordinates, detected object class, confidence value, and more. In this experiment, the YOLOv5, YOLOv6, YOLOv7, YOLOv8 models have been chosen to test the custom dataset created to find a suitable model to deploy on field trails. Performance evaluation in this project involves the analysis of precision, recall and mAP@0.5. The mAP@0.5 indicates the mean average precision at a 50% IoU threshold. These adjustments and evaluations provide information on the accuracy and effectiveness of the model in detecting potholes.

B. SENSOR FUSION

From the camera and LiDAR calibration, the relative position of the camera and LiDAR in the real world was calculated.

As a result of this, the values of t (translational matrix), roll, pitch, and yaw were obtained. The R (rotation matrix) was calculated by multiplying row, pitch, and roll matrices respectively from the calibration results obtained. Using these results, depth values that correspond to the pixel value of the camera can be computed.

C. GEOMETRIC FEATURE EXTRACTION

From the YOLO output, a bounding box is created around the detected pothole. Using the pixel coordinates of this bounding box, the corresponding point cloud data inside the bounding box is extracted from the rest of the point cloud data from LiDAR. This data is then passed on to the Convex Hull algorithm. The convex hull algorithm calculates the volume, area, and depth of the pothole. The resulting geometric data, along with the GNSS coordinates captured during the process, are saved on the computer.

D. GEO-POSITIONING

Using the latitude and longitude data collected during detection, the pothole is mapped with its calculated volume, area, and depth. This comprehensive geometric and location data is saved on the computer. The mapping is performed using the Folium package in Python, which facilitates the visualisation of the data on an interactive map. The saved data are in HTML (HyperText Markup Language) format and the results can be viewed using a web browser.

V. EXPERIMENTAL RESULTS

This section covers the equipment used for the experiment and the results obtained from training, validation and that geo-positioning of the detected pothole with features obtained.

A. EXPERIMENTAL SETUP

The figure 3 shows the experimental setup used to perform the experiment. For performing the field testing of the pothole detection, the Agile X Hunter 2.0 mobile robot, equipped with a camera, LiDAR, and GNSS, was used. The 3D printed mounted camera lidar can be seen in fig. 3 which makes the camera and lidar work as a standalone device keeping their relative position fixed. Sensor communication and data collection were performed using ROS Noetic. The data processing onboard system is using an Intel i7 12th generation CPU, and postprocessing system having Intel i7 11th generation CPU an Nvidia RTX 3060 graphics card, and 16GB RAM. The LiDAR used for data collection was the Ouster OS1-32 having a resolution of 2048 × 32, and the camera was the Logitech 505e. The GNSS system used were made by Swift Navigation, which runs on a frequency of 10Hz. The experiment was carried out on the E Hermiston Road (55.91888133259224, -3.3134189608152917), Edinburgh, United Kingdom.

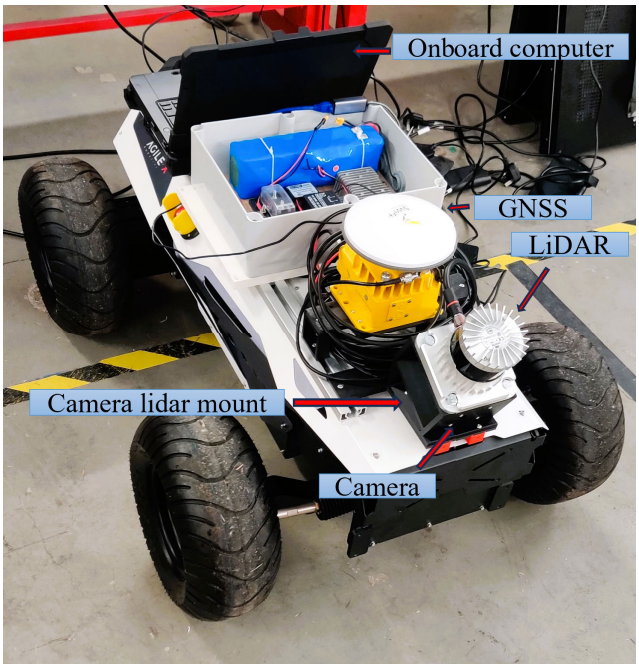

FIGURE 3. Agile X Hunter 2.0 mobile robot car mounted with LIDAR for depth data, camera for vision, GNSS for geo-positioning and onbord computer for sensor communication and ROS bag recording.

TABLE 1. YOLO architectures training performance comparison.

YOLO	Precision	Recall	mAP@0.5	Training time (h)
v5	86.0%	67.2%	76.0%	1.848
v6	72.4%	56.2%	72.0%	1.470
v7	75.7%	73.2%	76.0%	2.124
v8	77.2%	67.2%	72.5%	1.933

B. YOLO TRAINING RESULTS

To better examine how well four YOLO models work in finding potholes, a custom dataset was used to train the four networks (YOLOv5, YOLOv6, YOLOv7, and YOLOv8). These models have been evaluated across several metrics to determine their efficiency. In the table 1 shows the YOLO architecture used for training and the result of the training, such as precision, recall, mAP@0.5 and training time. All four models operate with the same input resolution of 640 × 480 pixels, a batch size of 8 and undergone 250 epochs of training. In terms of detection accuracy, it is evaluated through precision, recall, and mean average precision. Precision measures the accuracy of the positive predictions, and YOLOv5 leads with 86.0%, indicating fewer false positives and higher reliability in its detections. YOLOv8 follows with a precision of with 77.2%, YOLOv7 with 75.7% and YOLOv6 with 72.4%. The YOLOv7 shows highest recall of 73.2% and YOLOv6 with the lowest 56.2%. The mAP@0.5 evaluates the overall accuracy when IoU (intersection over union) thresholds are at 50% between the ground truth and the prediction, balancing both precision and recall. YOLOv5 and YOLOv7 shares higher mAP@0.5 with

76.0% followed by YOLOv8 and YOLOv6 with 72.5% and 72.0% respectively. In the following subsection shows the trained architectures performance in the test set and why it is important.

C. YOLO VALIDATION RESULT ON TEST SET

Table 2 shows the results of the performance evaluation metrics in the test dataset with 80 images. Upon evaluating the precision, YOLOv5 leads with a precision of 76.8%. YOLOv7 and YOLOv8 closely followed each other with a precision of 73.7% and 72.3% and lowest with YOLOv6 with 69.1%. Similarly to precision, YOLOv5 has the highest recall with 73.1% which was then followed by YOLOv6, YOLOv7, YOLOv8 with 71%, 70.9% and 68.3% respectively. Analysing the mAP metric provides a comprehensive overview of overall model performance, which shows that YOLOv5 stands out with the highest mAP@0.5 at 75.3%. By comparing both the training and the testing phases, YOLOv5 is the one with the highest mAP.

TABLE 2. YOLO architectures performance comparison on test dataset.

YOLO	Precision	Recall	mAP@0.5
v5	76.8%	73.1%	75.3%
v6	69.1%	71.0%	72.0%
v7	73.7%	70.9%	74.7%
v8	72.3%	68.3%	70.0%

D. RESULTS FROM FIELD TESTING

Analysis of pothole data collected from the field trial reveals a significant variation in pothole dimensions, which is useful for developing effective road maintenance strategies. From this experiment 52 potholes were extracted from the inspected road. The average volume of the potholes is 0.0033 cubic metres, inferring potholes are of small sizes, with volumes ranging from 0.0001 cubic metres to 0.0244 cubic metres. This suggests the presence of very small and significantly larger potholes within the inspected area. The surface area of the potholes also varies widely, from 0.0103 square metres to 0.6853 square metres. In terms of depth, the potholes have an average depth of 0.0270 metres, with the shallowest at 0.0104 metres and the deepest at 0.1549 metres. The smaller potholes are more frequent and the less impact to vehicles, whereas larger and deeper potholes require more extensive maintenance. In general, this detailed analysis highlights the diverse impact of potholes on road surfaces and the potential of modern measurement techniques in road management practices.

The pie chart in Fig. 4 represents the classification of potholes based on their volume into three categories, small, medium and large. From the pie chart, it is evident that the inspected road has a higher number of potholes having a smaller volume, which is 51.9.% indicating that almost half of the potholes have a volume of 0.001 m^3 or less. The medium-sized potholes account for around 32.7%. Finally,

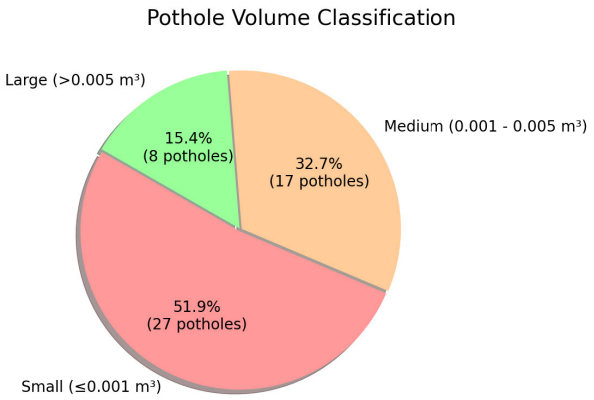

FIGURE 4. Pothole classification based on volume which is calculated from 3D reconstructed model.

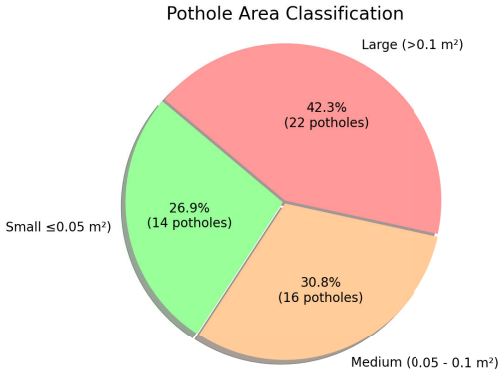

FIGURE 5. Pothole classification based on area which is calculated from 3D reconstructed model.

TABLE 3. Automated v/s manual inspection results.

ID	Depth (m)		Area (m^2)		Volume ($10^{-3} m^3$)	
	Auto	Manual	Auto	Manual	Auto	Manual
9	0.023	0.020	0.094	0.060	0.916	0.800
46	0.016	0.015	0.019	0.017	0.158	0.150
97	0.018	0.017	0.0412	0.038	0.321	0.300
111	0.012	0.010	0.063	0.061	0.455	0.400
218	0.052	0.045	0.018	0.016	0.166	0.150

the large potholes amount to about 15.4%., being the least frequent but larger in volume, exceeding 0.005 m^3 and require quicker repairs. Similarly in Fig. 5 shows the classification of potholes based on their surface area. The results reveal a significant dominance of pothole with smaller area which is 42.3% of the total. The potholes with a medium and smaller area were 30.8% and 26.9% respectively. This data shows that most of the potholes in the dataset are small and could require less extensive repairs. By time these smaller potholes will become larger potholes that may pose more significant repair challenges if road maintenance was not performed in the initial stage and it will cost more to repair these road defects.

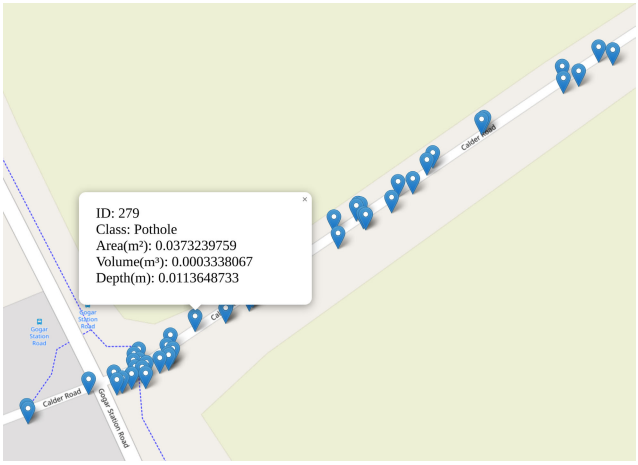

FIGURE 6. Map displaying detected potholes. Selecting a pothole opens a pop-up window that shows its unique ID, along with the area, volume, and maximum depth, as determined from 3D reconstructed models.

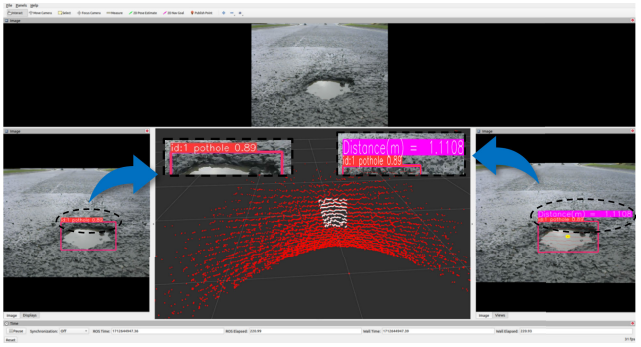

FIGURE 7. RVIZ (Robot Visualization tool) visualisation of the pothole detection and sensor fusion process. The top sub-window displays the input image, while the left sub-window shows the detected pothole using the YOLO algorithm. The central sub-window visualises the LIDAR data in red, and the LiDAR points corresponding to the detected pothole are highlighted in white. The right sub-window presents the sensor-fused data, including the depth information of the detected pothole.

Figure 6 presents a geo-positioned map of the detected potholes, allowing a visual differentiation between individual potholes according to their unique IDs. In addition, the map displays the volume, depth, and area of each pothole. This comprehensive visualisation aids in assessing the intensity and potential impact of each pothole on the road. By providing detailed information on the size and severity of each pothole, this map enables a more informed understanding of the specific challenges posed by different potholes, facilitating targeted maintenance efforts and more efficient resource allocation to address the most critical areas first. The spatial distribution and the varying characteristics of the potholes highlighted in the map underscore the importance of a strategic approach to road maintenance, taking into account the specific dimensions and locations of each pothole to effectively mitigate its impact.

E. SENSOR FUSION AND GEOMETRIC FEATURE EXTRACTION

The feature extraction process is a major step after the detection of potholes. Figure 7 displays this entire process result, from pothole detection to point cloud data extraction. The ROS visualisation tool RVIZ illustrates the results of the process in a single window, combining all the data. In this figure, the video feed from the camera, which serves as the system's input, is displayed in the top image window. Simultaneously, the centre window visualises the point cloud data collected from the LiDAR, illustrated as red dots, which occur simultaneously with the camera feed. This camera feed is subsequently processed through the YOLO architecture, where potholes are detected. The results from the YOLO architecture are presented in the leftmost window of the figure, where each detected pothole is encased within a bounding box. These bounding boxes have an ID that makes it possible to distinguish between variously detected potholes, and they also show a confidence value that denotes how certain each detection is. The coordinates of these bounding boxes, derived from YOLO, provide the base line for further calculations. Utilising these coordinates, the pixel coordinates inside the bounding boxes are accurately determined. These bounding box coordinates are then communicated to the LiDAR data processing node, where the integration of data from both the camera and LiDAR occurs. This sensor fusion is conducted using the calibration matrices to enhance the precision of the data. It's not necessary to look at the full depth data from the whole camera frame in order to evaluate the depth data of potholes. Instead, by focusing only on the pixel coordinates within the bounding box, the relevant depth data associated specifically with the detected pothole is isolated and analysed. In the image window on the right side of Fig. 7, the output from the LiDAR node showcases the reprojected depth data beneath the bounding boxes, marked in red dots. This visualisation also includes the calculated distance to the pothole, which is specifically measured by considering the depth data located at the centre of the bounding box. Given that the LiDAR utilised possesses 32 rings and not all individual pixels may have corresponding depth data due to gaps between these rings, interpolation is necessary to estimate the depth more accurately. The interpolated point nearest to the centre of the bounding box is highlighted with a yellow dot, representing the most accurate estimation of the distance to the centre of the pothole. Additionally, the middle section of Fig. 7 features the reprojected LiDAR point cloud data, displayed as white points, providing the data inside the bounding box, and the whole point cloud data before the extraction is in red dots, as discussed above. As from the sensor fusion, the point cloud data belonging to the pothole is extracted. As the next step, by analysing the point cloud data, feature extraction needs to be done. To perform the feature extraction, the convex hull algorithm was used. From the reconstructed model of the pothole, the geometric features such as area, volume, and depth were estimated. Figure 8 demonstrates the

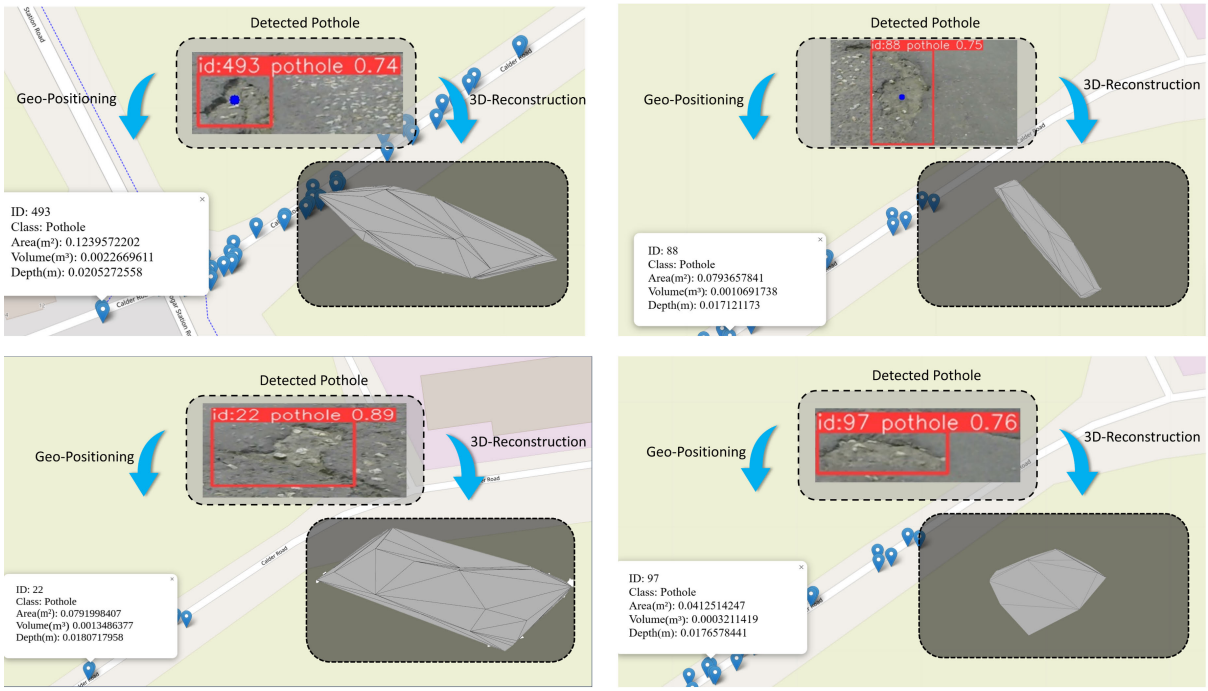

FIGURE 8. Examples of various detected potholes and their 3D reconstructed models using the Convex Hull approach. The corresponding positions of the potholes are mapped using GNSS data, illustrating their locations on the map.

results from our experiment and it shows various potholes detected. In this figure the detected pothole with its unique ID and confidence value of detection can be seen and the corresponding 3D reconstructed model and the extracted feature has been plotted on the map. This examples shows how our system handles potholes detection to geo-positioning process.

VI. CONCLUSION

Automatic pothole inspection methods offer a promising solution by providing real-time health status updates of road networks to relevant authorities with minimal labour requirements. Our system, utilising the YOLOv5 object detection architecture combined with LiDAR data, successfully detected the potholes on Edinburgh City roads and estimated both area, volume, and depth from the 3D reconstructed Convex Hull model. The integration of 2D and 3D data with geo-positioning information facilitates a comprehensive analysis and view of potholes on roads.

REFERENCES

[1] J. J. Yebes, D. Montero, and I. Arriola, "Learning to automatically catch potholes in worldwide road scene images," *IEEE Intell. Transp. Syst. Mag.*, vol. 13, no. 3, pp. 192–205, 2020.

[2] E. M. Thompson, A. Ranieri, S. Biasotti, M. Chicchon, I. Sipiran, M.-K. Pham, T.-L. Nguyen-Ho, H.-D. Nguyen, and M.-T. Tran, "SHREC 2022: Pothole and crack detection in the road pavement using images and RGB-D data," *Comput. Graph.*, vol. 107, pp. 161–171, Oct. 2022.

[3] S.-S. Park, V.-T. Tran, and D.-E. Lee, "Application of various Yolo models for computer vision-based real-time pothole detection," *Appl. Sci.*, vol. 11, no. 23, p. 11229, Nov. 2021.

[4] National Accident Helpline. (May 2023). *Potholes in the Uk: A Serious and Dangerous Problem*. [Online]. Available: https://www.national-accident-helpline.co.uk/news/post/potholes-uk-serious-and-dangerous-problem

[5] The AA. (Jun. 2022). *Potholes on Uk Roads*. [Online]. Available: https://www.theaa.com/breakdown-cover/advice/potholes-on-uk-roads

[6] The San Diego Union-Tribune. (Feb. 2019). *San Diego Paying Out $1m in Another Cycling Injury Settlement*. [Online]. Available: https://www.sandiegouniontribune.com/2019/02/11/san-diego-paying-out-1m-in-another-cycling-injury-settlement/

[7] Y.-M. Kim, Y.-G. Kim, S.-Y. Son, S.-Y. Lim, B.-Y. Choi, and D.-H. Choi, "Review of recent automated pothole-detection methods," *Appl. Sci.*, vol. 12, no. 11, p. 5320, May 2022.

[8] V. Kaushik and B. S. Kalyan, "Pothole detection system: A review of different methods used for detection," in *Proc. 2nd Int. Conf. Comput. Sci., Eng. Appl. (ICCSEA)*, Sep. 2022, pp. 1–4.

[9] B. Varona, A. Monteserin, and A. Teyseyre, "A deep learning approach to automatic road surface monitoring and pothole detection," *Pers. Ubiquitous Comput.*, vol. 24, no. 4, pp. 519–534, Aug. 2020.

[10] N. Chibani, F. Sebbak, W. Cherifi, and K. Belmessous, "Road anomaly detection using a dynamic sliding window technique," *Neural Comput. Appl.*, vol. 34, no. 21, pp. 19015–19033, Nov. 2022.

[11] B. Buč ko, E. Lieskovská, K. Zábovská, and M. Zábovský, "Computer vision based pothole detection under challenging conditions," *Sensors*, vol. 22, no. 22, p. 8878, Nov. 2022.

[12] S. Gupta, P. Sharma, D. Sharma, V. Gupta, and N. Sambyal, "Detection and localization of potholes in thermal images using deep neural networks," *Multimedia Tools Appl.*, vol. 79, nos. 35–36, pp. 26265–26284, Sep. 2020.

[13] E. Salcedo, M. Jaber, and J. R. Carrión, "A novel road maintenance prioritisation system based on computer vision and crowdsourced reporting," *J. Sensor Actuator Netw.*, vol. 11, no. 1, p. 15, Feb. 2022.

[14] G. Ochoa-Ruiz, A. A. Angulo-Murillo, A. Ochoa-Zezzatti, L. M. Aguilar-Lobo, J. A. Vega-Fernández, and S. Natraj, "An asphalt damage dataset and detection system based on RetinaNet for road conditions assessment," *Appl. Sci.*, vol. 10, no. 11, p. 3974, Jun. 2020.

[15] S.-Y. Lee, T. H. M. Le, and Y.-M. Kim, "Prediction and detection of potholes in urban roads: Machine learning and deep learning based image segmentation approaches," *Develop. Built Environ.*, vol. 13, Mar. 2023, Art. no. 100109.

[16] H. Wu, L. Yao, Z. Xu, Y. Li, X. Ao, Q. Chen, Z. Li, and B. Meng, "Road pothole extraction and safety evaluation by integration of point cloud and images derived from mobile mapping sensors," *Adv. Eng. Informat.*, vol. 42, Oct. 2019, Art. no. 100936.

[17] G. Bosurgi, M. Modica, O. Pellegrino, and G. Sollazzo, "An automatic pothole detection algorithm using pavement 3D data," *Int. J. Pavement Eng.*, vol. 24, no. 2, Jan. 2023, Art. no. 2057978.
[18] S. A. Talha, M. A. Karasneh, D. Manasreh, A. Al Oide, and M. D. Nazzal, "A LiDAR-camera fusion approach for automated detection and assessment of potholes using an autonomous vehicle platform," *Innov. Infrastruct. Solutions*, vol. 8, no. 10, p. 274, Oct. 2023.
[19] S. N. M. Rum and N. Rajaratinam, "A pothole detection using VGG16," *Civil Eng. Archit.*, vol. 10, no. 3, pp. 1222–1232, May 2022.
[20] S. Rath. (2022). *Fine Tuning Yolov7 on Custom Dataset*. Accessed: Nov. 2023. [Online]. Available: https://learnopencv.com/fine-tuning-yolov7-on-custom-dataset/
[21] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 12, pp. 1127–1141, Dec. 2018, doi: 10.1111/mice.12387.
[22] A. R. Chitholian. (2020). *Pothole Object Detection Dataset*. Accessed: Nov. 2023. [Online]. Available: https://public.roboflow.com/object-detection/pothole/
[23] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, Q. Jiang, C. Li, J. Yang, H. Su, J. Zhu, and L. Zhang, "Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection," 2023, *arXiv:2303.05499*.
[24] M. Vel'as, M. Španěl, Z. Materna, and A. Herout, "Calibration of RGB camera with velodyne lidar," in *Proc. 22nd Int. Conf. Central Europeon Comput. Graph., Vis. Comput. Visionin Co-operation EUROGRAPHICS Assoc.*, 2014, pp. 135–144.
[25] R. Perception. *Image*. Accessed: Jul. 28, 2024. [Online]. Available: https://github.com/ros-perception/image
[26] D. Tsai, S. Worrall, M. Shan, A. Lohr, and E. Nebot, "Optimising the selection of samples for robust lidar camera calibration," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 2631–2638.
[27] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023.

ABHIRAM KARUKAYIL received the bachelor's degree in mechatronics engineering with a specialization in robotics from the SRM Institute of Science and Technology, India, in 2022, and the M.Sc. degree in robotics with industrial application from Heriot-Watt University, U.K., in 2024, where he is currently pursuing the Ph.D. degree in electrical engineering, focusing on agricultural robotics. He has gained valuable experience by working with The National Robotarium, Edinburgh, U.K., as an Engineer, where he successfully bridged the gap between academic research and industry projects. Throughout his academic journey, he has contributed to various innovative projects, including underwater robotics, large language models (LLM), and robotic automation.

CHRISTOPHER QUAIL received the bachelor's degree in mathematics from the University of Exeter, U.K., in 2018, and the M.Sc. degree in robotics from Heriot-Watt University, U.K., in 2022, where he is currently pursuing the Ph.D. degree with the School of Engineering and Physical Sciences. He is part of The National Robotarium, Edinburgh, U.K. His research is focused on instantaneous power consumption of autonomous field robots in agriculture.

FERNANDO AUAT CHEEIN (Senior Member, IEEE) received the bachelor's degree in electronic engineering from the Universidad Nacional de Tucuman, Argentina, in 2002, and the M.Sc. and Ph.D. degrees in engineering from the Universidad Nacional de San Juan, Argentina, in 2005 and 2009, respectively. He is currently a Full Professor with the Department of Engineering, Harper-Adams University, U.K.; and an Associate Researcher with the Advanced Center for Electrical and Electronic Engineering, Federico Santa Maria Technical University, Valparaiso, Chile. His research is focused on autonomous and intelligent vehicles, robotics and perception in agriculture, motion planning and control, and efficient navigation strategies. He is an associate editor of several journals.