

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report Number	2. Government Association Number	3. Recipient's Catalog Number	
4. Title and Subtitle Automated Vehicle (AV) Data Sharing and Data Analytics		5. Report Date 04/30/2026	
		6. Performing Organization Code	
7. Author(s) Xinwei Yang, Peggy Wang		8. Performing Organization Report Number UCB-ITS-PRR-2026-01	
9. Performing Organization Name and Address California PATH, University of California at Berkeley 1357 South 46th Street, Richmond, CA, 94804		10. Work Unit Number	
		11. Contract or Grant Number 65A1058	
12. Sponsoring Agency Name and Address California Department of Transportation 1727 30th Street, Sacramento, CA 95816		13. Type of Report and Period Covered Final report	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract Roadway infrastructure conditions directly affect the safety, mobility, and operational costs of the transportation system. Pavement markings, traffic signs, and road surfaces degrade over time due to traffic loading, weather, and exposure to environmental conditions. Timely detection and repair of these defects, including deteriorated lane markings, damaged or obstructed traffic signs, and surface damage such as cracks and potholes, is essential for maintaining safe roadway operations and supporting the growing deployment of connected and automated vehicles (CAVs). Current roadway inspection practices in California, as in most US states, rely primarily on dedicated inspection vehicles and field surveys. While these methods have served as the foundation of asset management programs for decades, they face several limitations that constrain the frequency, coverage, and timeliness of roadway condition monitoring. This project was designed to develop detection capabilities specifically for (1) deteriorated pavement markings (including pavement signs and pavement lanes) and (2) potholes. Our approach involves collecting video data, labeling it, and building a comprehensive training dataset for roadway damages specific to the US.			
17. Key Words pavement marking, pothole detection, computer vision		18. Distribution Statement	
19. Security Classification (of this report)	20. Number of Pages 55	21. Cost of Report	

DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in alternate formats. For information, call (916) 654-8899, TTY 711, or write to California Department of Transportation, Division of Research, Innovation and System Information, MS-83, P.O. Box 942873, Sacramento, CA 94273-0001.

PARTNERS FOR ADVANCED TRANSPORTATION TECHNOLOGY
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Automated Vehicle (AV) Data Sharing and Data Analytics

Final Report

April 30, 2026



Primary Authors

Dr. Xinwei Yang

Post-doctoral Scholar

California PATH

University of California, Berkeley

Dr. Peggy Wang

Associate Research Engineer

California PATH

University of California, Berkeley

Executive Summary

Roadway infrastructure conditions directly affect the safety, mobility, and operational costs of the transportation system. Pavement markings, traffic signs, and road surfaces degrade over time due to traffic loading, weather, and exposure to environmental conditions. Timely detection and repair of these defects, including deteriorated lane markings, damaged or obstructed traffic signs, and surface damage such as cracks and potholes, is essential for maintaining safe roadway operations and supporting the growing deployment of connected and automated vehicles (CAVs).

Current roadway inspection practices in California, as in most US states, rely primarily on dedicated inspection vehicles and field surveys. While these methods have served as the foundation of asset management programs for decades, they face several limitations that constrain the frequency, coverage, and timeliness of roadway condition monitoring. This project was designed to develop detection capabilities specifically for (1) deteriorated pavement markings (including pavement signs and pavement lanes) and (2) potholes. Our approach involves collecting video data, labeling it, and building a comprehensive training dataset for roadway damages specific to the US.

The data preparation process began with manual reviewing and filtering the collected driving videos. We first inspected the video data and selected approximately 5,000 image frames containing identifiable roadway damage. This subset was divided into 3,500 training images and 1,500 validation images and was used to train a preliminary - You Only Look Once - YOLOv12m-mask segmentation model. Using this pre-trained model as a starting point, we developed a semi-automated annotation pipeline to improve labeling efficiency. The final dataset included a total of 22,588 images, which was divided into 12,653 training images and 9,935 validation images. The validation set contained 28,263 annotated instances across four roadway damage-related classes, with potholes being the most frequent category.

YOLOv12m-seg model demonstrates stable convergence during training, with minimal overfitting, and strong generalization across different environmental conditions, including low-light and shadowed scenarios. The same training and validation datasets were used for three YOLO models (YOLOv12m-seg, YOLOv11m-seg and YOLOv26m-seg). Results indicate that all three models achieved strong performance on the roadway damage dataset. The mAP50 values of the YOLOv12m-seg model ranged from 0.80 to 0.85 for both bounding-box detection and mask segmentation, indicating that the YOLOv12m-seg model is able to reliably detect and segment roadway damage instances.

Building upon the detection results, a vision-based pothole size estimation framework was developed using stereo camera depth information and geometric projection. A distance-dependent correction method was introduced to reduce estimation errors caused by perspective distortion and sensor noise. Experimental results show that the proposed correction approach significantly improved estimation accuracy, reducing length estimation errors from over 70% to below 10% in representative cases.

Further validation under varying lighting and weather conditions demonstrated that the proposed system maintains reliable performance, especially when combined with image enhancement techniques such as CLAHE. The integrated detection and size estimation system provides a

scalable and cost-effective solution for automated roadway condition monitoring, with potential real-world applications in Caltrans' infrastructure maintenance planning.

Table of Contents

Executive Summary	3
1 Introduction	7
1.1 Background: Roadway Damage Detection for Roadway Maintenance	7
1.2 Motivation: Limitations of Traditional Inspection Methods	7
1.3 Scope and Significance of This Project	8
1.4 Structure of This Report	9
2 Summary of Literature Review	10
2.1 Automated Vehicle Operation and Infrastructure Requirements	10
2.2 Industry Applications	10
2.3 Detection Methods	11
2.3.1 Pavement Marking Detection	11
2.3.2 Road Damage Detection	11
2.4 Existing Data Set.....	13
2.5 Summary of the Literature Review	13
3 Outreach and Data Platform.....	15
3.1 Outreach to Automated Vehicle Companies	15
3.2 Data Collection.....	15
3.2.1 iVision (Caltrans).....	16
3.2.2 PATH Field Data Collection	16
3.2.3 RDD2022 (Open-Source Dataset).....	16
3.2.4 AV Demonstration Project	16
3.3 Data Infrastructure and Computing Platform.....	16
4 Roadway Damage Detection.....	18
4.1 Roadway Damage Classes	18
4.2 Dataset Creation	18
4.2.1 Data Preparation	18
4.2.2 Data Annotation Pipeline	19
4.2.3 Dataset Description	20

4.3	Training Framework: YOLOv12-seg Model.....	20
4.4	Roadway Damage Detection Results	22
4.4.1	YOLO Model Training Parameters.....	23
4.4.2	Training Results of YOLOv12m-seg Model.....	23
4.4.3	Performance Comparison of YOLOv12m-seg, YOLOv11m-seg and YOLOv26m-seg	27
4.4.4	Prediction Test for the YOLOv12m-seg	30
5	Pothole Size Estimation	32
5.1	Pothole Size Estimation Process.....	32
5.2	Apparatus	36
5.3	Pothole Size Estimation Results.....	36
5.3.1	Pothole Size Estimation Before and After Correction	36
5.3.2	Pothole Size Estimation with Different Measurement Distance	38
5.3.3	Robustness Test for Pothole Size Estimation Algorithm	39
5.3.4	Statistic Evaluation for Robustness	47
6	Discussions and Conclusions.....	50
6.1	Findings	50
6.2	Areas of improvement	51
6.3	Future Research Directions.....	51
	References.....	53

1 Introduction

1.1 Background: Roadway Damage Detection for Roadway Maintenance

Roadway infrastructure conditions directly affect the safety, mobility, and operational costs of the transportation system. Pavement markings, traffic signs, and road surfaces degrade over time due to traffic loading, weather, and exposure to environmental conditions. Timely detection and repair of these defects, including deteriorated lane markings, damaged or obstructed traffic signs, and surface damage such as cracks and potholes, is essential for maintaining safe roadway operations and supporting the growing deployment of connected and automated vehicles (CAVs).

The importance of roadway infrastructure conditions extends beyond conventional road users. As automated driving systems (ADS) advance from research toward deployment, the performance and reliability of these systems is closely tied to roadway infrastructure quality. Automated Vehicles (AVs) rely on cameras, LiDAR, radar, HD maps, and GPS to sense and navigate the environment. Camera-based machine vision systems depend on visible lane markings and unobstructed traffic signs for scene understanding. LiDAR is used to detect fixed infrastructure features and support vehicle localization. Deficiencies in pavement markings or signage could degrade AV perception performance, create unsafe conditions, or trigger unnecessary system disengagements. Basic infrastructure requirements — including clear and consistent lane markings, legible traffic signs, and smooth pavement surfaces — are considered prerequisites for safe AV operation across multiple future deployment scenarios (Lu et al., 2019; Wang et al., 2022).

Roadway damage detection encompasses three primary infrastructure conditions of interest to transportation agencies: (1) deteriorated pavement markings, including faded or missing lane lines, arrows, crosswalks, and stop lines; (2) damaged or obstructed traffic signs, including signs that are defaced, missing, or partially occluded; and (3) surface damage such as cracks, potholes, and uneven pavement. These defects affect both conventional drivers and ADS performance, and their detection and repair are key responsibilities of state (such as Caltrans) and local transportation agencies.

A range of detection methods has been developed to address these three infrastructure conditions, from traditional image processing techniques to deep learning-based computer vision approaches. A detailed review of these methods and relevant industry applications is provided in Section 2.

1.2 Motivation: Limitations of Traditional Inspection Methods

Current roadway inspection practices in California, as in most US states, rely primarily on dedicated inspection vehicles and field surveys. While these methods have served as the foundation of asset management programs for decades, they face several limitations that constrain the frequency, coverage, and timeliness of roadway condition monitoring.

Traditional roadway inspection is low frequency by nature. Field inspections require trained personnel, specialized vehicles, and significant staff time, so most roadway segments are inspected once a year or every other year. Pavement markings may fade, signs may be damaged,

and surface defects may develop between inspection cycles with no efficient mechanism to detect them in the interim.

AV testing data offers a promising alternative. AV companies operating in California continuously collect large volumes of sensor data, including camera footage, LiDAR point clouds, and GPS trajectories as part of their regular testing programs on public roadways. In a prior Caltrans-funded survey of 20 plus AV companies (Wang et al., 2022), many companies expressed willingness to share their testing data with public agencies to support roadway infrastructure maintenance. This project builds on that interest to establish data-sharing partnerships, identify best practices for obtaining and processing AV testing data, and develop tools to extract infrastructure condition information from that data.

Compared to conventional inspection methods, AV testing data offers several advantages:

- High frequency: AV fleets travel the same road segments repeatedly, providing continuous data rather than periodic snapshots.
- Broad coverage: AV testing routes cover urban arterials, state highways, and interstate corridors across multiple California counties.
- Multi-modal data: Camera, LiDAR, and GPS data are collected simultaneously, providing complementary information for detecting pavement markings, traffic signs, and surface damage.
- Digital format: AV sensor data is inherently digital and compatible with computer vision and machine learning pipelines.
- No additional inspection cost: The data is collected as a byproduct of AV development and testing, so leveraging it for infrastructure monitoring does not add marginal field cost.

At the same time, AV testing data is not purpose-built for infrastructure monitoring. Extracting useful condition information requires careful data pre-processing, quality control, and appropriate analytical tools. Data volume is large and formats vary across companies, which are known challenges of this research topic.

1.3 Scope and Significance of This Project

The PATH team, based on a literature review, identified several ways that shared AV testing data and analytics could benefit Caltrans' operational needs. These include detecting poor pavement markings, identifying pavement cracking, recognizing potholes, and flagging maliciously altered or obstructed traffic signs.

To better understand current practices, the PATH team consulted with Caltrans maintenance personnel responsible for roadway condition monitoring and maintenance planning. We learned that Caltrans currently surveys all lanes of the state highway system for pavement condition biannually. The survey vehicles utilize cameras, GPS, and laser sensors, and the collected data is analyzed through a semi-automated process (combining algorithms and manual review) to identify and classify pavement distress, especially cracking. The iVision online platform allows for searching and reporting these survey results. However, Caltrans currently lacks similar reliable tools for detecting potholes or assessing pavement marking conditions.

Consequently, this project was designed to develop detection capabilities specifically for deteriorated pavement sign, pavement lane, and potholes. Our approach involves collecting video data, labeling it, and building a comprehensive training dataset for roadway damages specific to the US. We will then compare various computer vision (CV) frameworks and algorithms before applying them to train the detection models. Ultimately, the goal was to leverage data shared by third-party sources (such as AV developers and operators) to create CV algorithms that can accurately and automatically identify deteriorated lane markings and potholes.

1.4 Structure of This Report

The remainder of this report covers a summary of the literature review (Section 2), description of the outreach activities and data platform (Section 3), the methodology and analysis results for roadway damage detection (Section 4), the methodology and analysis results for pothole size estimation (Section 5), and discussion and conclusion (Section 6).

2 Summary of Literature Review

2.1 Automated Vehicle Operation and Infrastructure Requirements

AVs operate based on the classic “Sense – Plan – Act” paradigm from robotics. The Sense process draws on multiple sensors to build a picture of the surrounding environment, which then informs path planning and vehicle control. Because no single sensor can reliably handle all conditions, AVs typically use a combination of cameras, LiDAR, radar, GPS, and High Definition (HD) maps. Each sensor type has specific capabilities and limitations with respect to the infrastructure it can detect. Camera-based machine vision detects lane markings, traffic signs, and signals but is limited by line of sight, sun glare, and poor weather. LiDAR detects fixed objects and can detect lane markings through reflectivity but is costly and struggles in poor weather. Radar detects fixed infrastructure but is susceptible to radio frequency interference. HD maps support vehicle localization but become unreliable when roadway features change, such as during construction or road closure events.

A brainstorming workshop with 13 experts from the Netherlands (Farah et al., 2018) identified four high-priority research areas for AV infrastructure: (1) impacts of automated driving on existing infrastructure; (2) transitions and mixed traffic; (3) ideal infrastructure design for future AVs; and (4) business case and decision-making. These findings align with the need for better roadway condition monitoring and detection explored in this project.

For SAE Level 4 ADS, Lu et al. (2019) identified basic infrastructure requirements that must be met before roads can be opened to AVs. These include clear and harmonized road signs and lane markings, speed limit signage at exits and intersections, countdown traffic lights, warnings for vulnerable road users, HD maps, and V2I connectivity. High-cost upgrades such as dedicated lanes and intelligent traffic lights were considered feasible only under higher public expenditure scenarios. This project focuses on the most fundamental of these requirements: lane markings and road surface conditions.

2.2 Industry Applications

Several industry initiatives have demonstrated how vehicle-collected data can be used for roadway condition monitoring.

Honda Research Institute developed a pilot road condition monitoring system in collaboration with the Ohio Department of Transportation. Using GPS and onboard cameras, Honda vehicles classify lane marking conditions in real time as ideal, good, missing, or in need of repair. The data is anonymized and uploaded to a secure platform accessible to road operators. This system began generating data in early 2022 and informs Advanced Driver Assistance Systems (ADAS) in connected vehicles.

Mobileye’s Road Experience Management (REM) system constructs continuously updated HD maps by aggregating data from millions of vehicles equipped with Mobileye technology. Beyond supporting AV navigation, REM data has been used to monitor lane marking conditions —

including in low-visibility situations where markings are obscured by weather — and to assist road authorities with infrastructure maintenance and traffic management planning.

GoodRoads, a technology company specializing in road inspection and pavement management, uses high-resolution cameras and sensors mounted on vehicles to collect road surface data, which is then analyzed using machine learning to detect and classify defects such as cracks and potholes. In a project in Philadelphia, Pennsylvania, GoodRoads conducted an assessment covering approximately 1,100 miles of streets, including an inventory of pavement markings, road signs, and manholes. Studies have demonstrated potential to reduce inspection times by up to 70%.

Michelin Mobility Intelligence collects data from connected vehicles — including speed, location, tire conditions, and driver behavior — and applies machine learning to automate pavement condition assessments. Their platform provides a 0–100 pavement condition rating and presents results through an interactive online interface to support maintenance planning.

These industry examples demonstrate the feasibility and value of leveraging vehicle-collected data for roadway infrastructure monitoring. They also highlight the growing commercial interest in this space, which aligns with the data-sharing opportunity this project aims to pursue with Caltrans.

2.3 Detection Methods

The following literature review summary covers detection methods for two categories of roadway infrastructure conditions: pavement markings, and road damages. The detection methods fall into three broad groups: traditional image processing, learning-based approaches, and 3D laser scanning.

2.3.1 Pavement Marking Detection

Traditional methods for pavement marking detection rely on hand-crafted features such as histogram of oriented gradients (HOG) and support vector machines (SVM). These approaches are computationally simple but tend to perform poorly in unfamiliar or complex conditions.

Deep learning methods have substantially improved detection accuracy. Kim and Lee (2014) combined convolutional neural networks (CNNs) with random sample consensus (RANSAC) for lane detection in complex scenes. Li et al. (2017) used both a CNN and a recurrent neural network (RNN) to detect lane boundaries, with the CNN capturing geometric structure and the RNN detecting lane positions. For damaged marking detection specifically, Kawano et al. (2017) applied You Only Look Once (YOLO) to detect blurred road markings. Wei et al. (2021) used hierarchical semantic segmentation to estimate marking damage ratios from multiple images. Wang et al. (2024) introduced an approach combining pixel-based and region-based segmentation with attention-guided feature fusion to assess the degree of marking damage.

2.3.2 Road Damage Detection

Traditional road damage detection methods include accelerometer-based sensing, image thresholding, image filtering, and edge detection algorithms. Accelerometer-based methods are limited to areas where vehicle wheels pass and are rarely used in practice. Image-based traditional methods are straightforward but struggle with noise and variable lighting.

Deep learning methods have achieved strong results for road damage detection. Zhang et al. (2016) proposed one of the earliest CNN-based crack detection models. Fan et al. (2019) developed a stereo vision-based pothole detection system using a disparity transformation algorithm. Majidifard et al. (2020) used deep learning to estimate a pavement condition index (PCI) from images. Matouq et al. (2024) applied YOLOv8 to detect potholes, manholes, and patches, with camera calibration to estimate pothole area. YOLO is a single-stage deep learning detector that processes an entire image in one forward pass through a convolutional neural network, simultaneously predicting bounding boxes and class probabilities across a grid of candidate regions. It was first introduced by Redmon et al. (2016). The design makes YOLO substantially faster than two-stage detectors while maintaining competitive accuracy, which is why it has become widely adopted for real-time detection tasks in transportation and infrastructure applications. YOLOv8 is the eighth major iteration of the YOLO family of object detection models.

3D laser scanning offers high accuracy for detecting pavement texture and damage and is less affected by environmental conditions but requires expensive specialized equipment and is not suitable for lightweight or network-scale deployment.

Table 1 summarizes the pros and cons of each detection approach for the two infrastructure condition categories.

Table 1. Summary of Detection Methods

Case	Methodology	Pros	Cons
Pavement marking	Traditional (HOG, SVM)	Simple and computationally inexpensive	Poor performance in complex or unfamiliar conditions
	Machine learning (Artificial Neural Network -ANNs, Principal Component Analysis Net - PCANet)	Improved accuracy over traditional methods	Limited scalability; slower than deep learning
	Deep learning (CNN, Hybrid CNN-RNN, R-CNN, Mask R-CNN, YOLO-based)	High accuracy and robustness; real-time possible with YOLO	High computational cost; slower inference in Faster R-CNN
Road damages	Traditional (accelerometer, thresholding, edge detection)	Simple and cost-effective	Limited coverage; struggles with noise and lighting variation

	Deep learning (DCNN, Faster R-CNN, YOLO-based)	High accuracy; scalable to large datasets	Computationally expensive; requires labeled training data
	3D laser scanning	Highly accurate; resistant to environmental conditions	Expensive hardware; not suitable for lightweight devices

2.4 Existing Data Set

Several publicly available datasets support research in AV perception and roadway infrastructure condition monitoring. AV datasets such as Waymo Open (Ettinger et al., 2021), Argoverse (Wilson et al., 2023), and KITTI (Liao et al., 2022) are designed for object detection and scene understanding in complex urban environments, and include multi-modal data (LiDAR, cameras, GPS, Inertial Measurement Unit -IMU) collected across multiple US cities. However, these datasets are focused on pedestrian and vehicle detection rather than infrastructure condition monitoring.

Datasets for pavement markings, traffic signs, and road damage were primarily collected outside the US. Most came from Europe, Asia, or South Korea. The LISA Traffic Sign Dataset is one of the few US-specific traffic sign datasets. Road damage datasets such as RDD2022 (Road Damage Dataset) include some US data sourced from Google Street View but are not representative of the full range of US road conditions and damage types.

This geographic gap of the dataset is significant. Road conditions, pavement marking standards, and traffic sign conventions vary considerably across countries. Detection models trained primarily on non-US data may not generalize well to the California roadways. Developing a labeled US-specific dataset for pavement marking deterioration and pothole detection is one of the key contributions of this project.

2.5 Summary of the Literature Review

The literature identifies several challenges associated with using AV data for roadway infrastructure assessment. These include but are not limited to the algorithms and data quality.

On the algorithm side, deep learning models often struggle with false positives and negatives when detecting nuanced defects such as potholes, which vary widely in shape, size, and texture. Models trained on specific road types may not generalize well to other environments. Real-time deployment on edge devices also remains challenging due to hardware and latency constraints.

On data quality side, AV systems collect data from multiple sensors with different resolutions and sampling rates, complicating integration. Sensor failures and poor weather can result in missing or incomplete data. Noise and outliers in GPS and LiDAR readings can skew model outputs if not properly filtered during pre-processing.

These challenges informed the approach taken in this project, particularly the data pre-processing procedures, and the selection of detection algorithms.

3 Outreach and Data Platform

3.1 Outreach to Automated Vehicle Companies

One objective of this project was to establish data-sharing partnerships with AV companies operating in California. The PATH team developed a pitch document, reviewed and approved by the Caltrans technical panel, to engage AV companies on the topic of sharing their testing data with Caltrans for roadway infrastructure monitoring purposes. The pitch framed the data-sharing effort as mutually beneficial: AV companies would support better-maintained public roadways, which in turn supports safer and more reliable AV operations.

The PATH team used this pitch to engage with AV industry companies at the Automated Road Transportation Symposium (ARTS) 2024. Companies contacted through this and subsequent outreach efforts included Volkswagen, Honda, Nissan, Toyota, Waymo, and May Mobility. Several of these companies have been involved in collaborative data-sharing efforts with other state transportation agencies, which provided a basis for discussion.

The most substantive engagement was with Waymo. The PATH team held an initial discussion with Waymo in February 2025, focusing on the possibility of sharing video data for pothole detection. A follow-up discussion was held in August 2025. The PATH team presented the project background, Caltrans' current roadway monitoring practices, and the potential use cases for AV testing data, including deteriorated lane markings, potholes, cracking, and damaged or obstructed traffic signs.

Despite these active outreach and engagement efforts, no AV company has formally shared testing data for use in this project to date. The discussions with Waymo remain ongoing. Establishing formal data-sharing agreements between public agencies and private AV companies involves legal, privacy, and data governance considerations that require time to resolve.

3.2 Data Collection

The PATH team assembled a training dataset from four alternative sources: the Caltrans iVision platform, PATH's own field data collection, an existing open-source road damage dataset, and an AV demonstration project. Table 2 summarizes the data sources.

Table 2. Summary of Data Sources

Data Source	Type	Volume	Coverage
iVision (Caltrans)	Road damage images	~1,500 images	California (2022–2024)
PATH field collection	Video data (camera-mounted vehicles)	~8,000 extracted samples	East Bay area (Berkeley, Richmond, Oakland)

RDD2022 (open-source)	Road damage images	~13,000 images	Japan, Czech Republic, China
AV demonstration project	Video data (from Rosbag files)	High quality video	California (2025)

3.2.1 iVision (Caltrans)

The PATH team obtained approximately 1,500 road damage images from Caltrans' iVision platform, covering roadway conditions in California from 2022 to 2024. iVision is the online platform used by Caltrans to store and report results from its biannual pavement condition survey. The PATH team filtered and downloaded images showing degraded pavement markings and potholes. This dataset provided California specific road damage images directly relevant to Caltrans' operational context.

3.2.2 PATH Field Data Collection

The PATH team collected additional video data by mounting high-resolution cameras on PATH's own vehicles and driving in the East Bay area, including Berkeley, Richmond, and Oakland. Video analysis and frame extraction tools were used to filter frames containing relevant targets — potholes and degraded pavement markings. Approximately 8,000 samples were extracted from the collected video footage. This dataset represents road conditions specific to California urban roadways and directly supplements the iVision data.

3.2.3 RDD2022 (Open-Source Dataset)

The PATH team used the Road Damage Detection Challenge 2022 dataset (RDD2022), an open-source dataset containing approximately 13,000 road damage images collected in Japan, the Czech Republic, and China. While this dataset provides a substantial volume of labeled training data, it does not include US road conditions and has known data quality limitations. It was used primarily as a baseline dataset to support initial model training before US-specific data was available.

3.2.4 AV Demonstration Project

Video data was collected by an AV system along the demonstration routes. Some of the collected video data were made available to the PATH team via Rosbag files, which were downloaded through a batch API and decoded to extract video footage. The data was saved to the PATH local server for further processing. This dataset provided high-quality video data from an AV system operating in a California urban environment.

3.3 Data Infrastructure and Computing Platform

The PATH team used a dedicated data server to store, manage, and process the collected datasets. The selected server is a desktop equipped with an NVIDIA RTX 4090 GPU, an Intel Core i9-14900KF CPU, 64GB of RAM, and 2TB of SSD storage. The server was designed to be

expandable to accommodate larger volumes of incoming data. Remote access for Caltrans stakeholders was set up using the Secure Shell (SSH) protocol, with individual authentication credentials generated for authorized users.

The server runs Ubuntu 24.10 as the operating system. The software environment includes Python 3.12, Anaconda, PyCharm, and MATLAB. All software was installed and verified operational. The computing platform supports data storage, data pre-processing, and model training as.

4 Roadway Damage Detection

4.1 Roadway Damage Classes

The data required for this project focused on roadway condition information. The raw data was collected from three primary sources as mentioned in Section 3. In our dataset, road damage classes included potholes and degraded pavement markings (e.g., pavement signs and pavement lanes). Due to the limited availability of sufficient and representative traffic sign samples, traffic signs were not included in the target detection categories. In addition, to mitigate misclassification, particularly the confusion between potholes and visually similar objects such as manholes, we explicitly included manholes as a separate class in the dataset. However, manhole is not a focus of this project. The detailed definitions of all detection target classes are summarized in Table 3.

Table 3. Roadway Damage Classes

Damage Type		Description	Class Number
Pothole		A hole or depression in the roadway	D40
Degraded pavement markings	Pavement sign	Pavement signs lost structural integrity, functionality, or visibility	D43
	Pavement lane	Pavement lanes lost structural integrity, functionality, or visibility	D44
Manhole		Manholes, typically protected by a manhole cover	D50

4.2 Dataset Creation

4.2.1 Data Preparation

Since our approach relied on a vision-based detection framework, the data preparation process began with manually reviewing and filtering the collected driving videos. We first inspected the video data and selected approximately 5,000 image frames that contain identifiable roadway damage classes. These selected samples covered a range of environmental conditions, including daytime, low-light, and rainy scenarios, as well as different pavement surfaces such as asphalt and concrete.

After frame selection, each target category was manually annotated. In this project, we used a mask-based semantic segmentation annotation method instead of conventional bounding-box labeling. Compared with bounding boxes, mask-based annotations provide more precise object localization at the pixel level. They also better capture the boundaries of irregularly shaped targets and reduce the inclusion of irrelevant background pixels. This is especially important for roadway damage detection, because damage regions are often non-rectangular and vary considerably in size, shape, and geometry.

Therefore, all detection objects were labeled using semantic segmentation with mask annotations. An example of the annotation format is shown in Figure 1. All annotations were created and managed using the Label Studio platform.



(a) Original Detected Objects (b) Annotated with Mask

Figure 1. Illustration of Objects and Annotation

4.2.2 Data Annotation Pipeline

Because data filtering and annotation are labor-intensive and time-consuming, we first manually annotated the 5,000 images, including 3,500 images for training and 1,500 images for validation. This initial dataset was used to train a preliminary YOLO segmentation model, which served as the preliminary model for developing a semi-automated annotation pipeline.

The overall workflow of the pipeline is illustrated in Figure 2. The proposed pipeline followed an iterative human-in-the-loop process. Specifically, the preliminary model was applied to raw video frames and unlabeled image data to generate initial segmentation annotations. The research team then manually reviewed and corrected these model-generated labels to ensure annotation quality and accuracy.

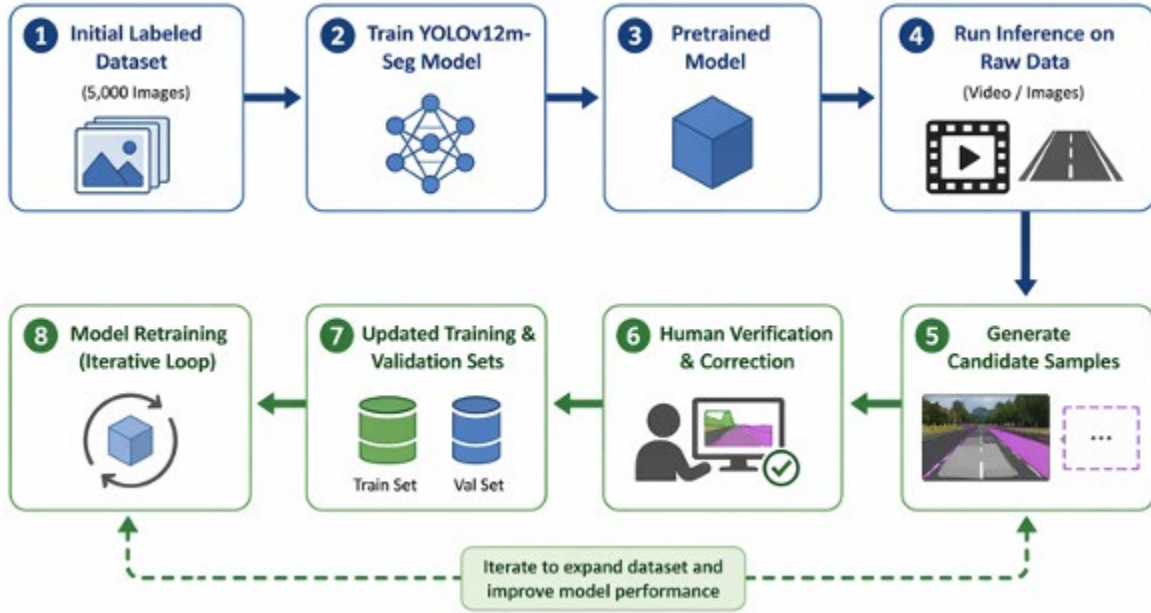


Figure 2. The Semi-Automated Annotation Pipeline

After verification, the corrected samples were added to the dataset and used to update the training and validation sets. The model was then retrained with the expanded dataset, which gradually improved its detection performance. This process was repeated over multiple rounds, allowing us to efficiently scale up the labeled dataset while reducing the amount of manual annotation work required.

4.2.3 Dataset Description

The final dataset included 22,588 samples, which was divided into 12,653 training images and 9,935 validation images. The validation set contained 28,263 annotated instances across four roadway damage-related classes, with potholes (D40) being the most frequent category. These validation data were used for model performance evaluation, while the training set followed a similar class distribution.

4.3 Training Framework: YOLOv12-seg Model

For the target detection and semantic segmentation task, we adopt a YOLO-based deep learning framework, specifically the YOLOv12 model (Tian et al., 2025). YOLOv12 model was selected because it introduced an attention-centric YOLO architecture that improved representation capability while maintaining real-time inference speed. Traditional YOLO improvements have mainly relied on CNN-based architectural designs, whereas YOLOv12 incorporated attention mechanisms to enhance feature modeling without sacrificing computational efficiency. According to Tian et al. (2025), YOLOv12 achieved higher accuracy than several popular real-time object detectors while maintaining competitive inference latency. This balance between accuracy and speed is important for our application, where roadway damage

must be detected efficiently from large-scale video data and potentially deployed on onboard vehicle computing platforms.

The Figure 3 shows that YOLOv12 consistently achieved higher accuracy (mAP) than other models at comparable latency and computational cost (FLOPs). Compared with previous YOLO versions and transformer-based detectors (e.g., RT-DETR), YOLOv12 provided a better trade-off between speed and performance, delivering higher detection accuracy without increasing inference time significantly. This makes YOLOv12 particularly suitable for real-time applications such as roadway damage detection, where both high accuracy and fast processing are required.

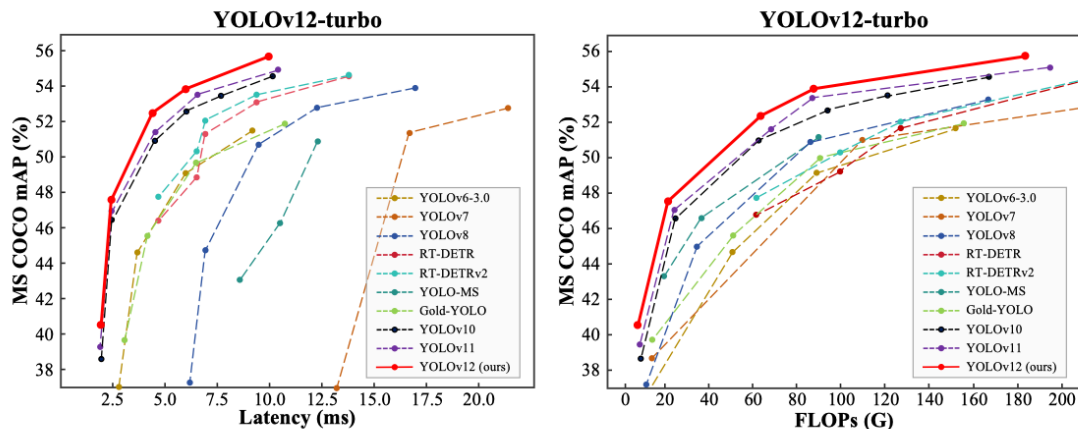


Figure 3. Comparisons with other popular methods in terms of latency-accuracy (left) and FLOPs-accuracy (right) trade-offs *Source: Tian et al., 2025 +

Building on these advantages, YOLOv12-seg further extends the framework by integrating an instance segmentation head. This enables the model to generate pixel-level masks for each detected object, which is particularly important for potholes and degraded pavement markings because these targets often have irregular shapes and ambiguous boundaries. Therefore, YOLOv12-seg provides a suitable balance of detection accuracy, segmentation capability, and computational efficiency for the proposed roadway damage analysis framework.

The overall architecture of YOLOv12-seg, as shown in Figure 4, consists of three main components: the backbone, neck, and head. The backbone extracts hierarchical feature representations from the input image. The neck aggregates multi-scale features, such as through feature pyramid structures, to improve the detection of objects at different resolutions. The head then predicts the bounding boxes, class probabilities, and segmentation masks for each detected instance.

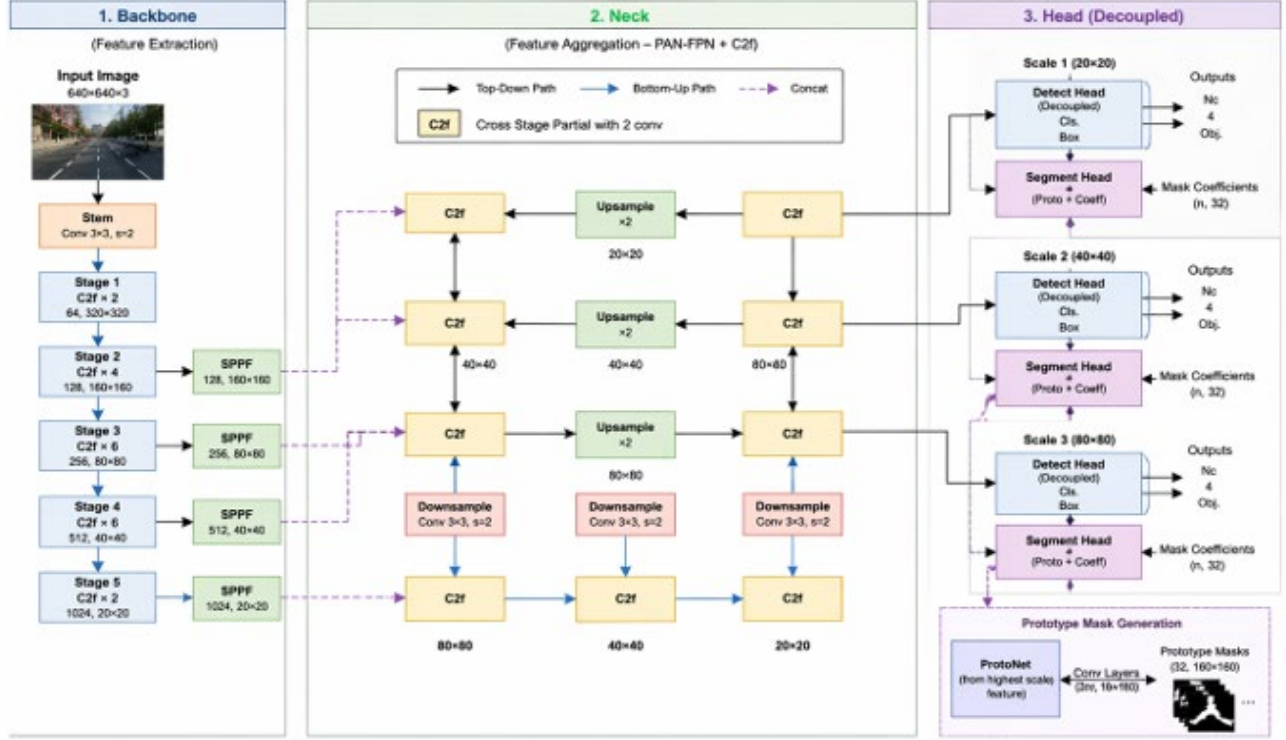


Figure 4. YOLOv12-seg Architecture

4.4 Roadway Damage Detection Results

We first evaluated the training performance, validation accuracy, and model suitability of YOLOv12m-seg. The evaluation was conducted from two perspectives. First, the convergence behavior and validation performance of YOLOv12m-seg were analyzed using training/validation loss curves, precision–recall curves, confidence-based evaluation curves, and segmentation metrics. This analysis was used to verify whether the model could provide stable bounding-box detection and reliable pixel-level mask prediction for roadway damage targets. Second, to further assess model selection, YOLOv12m-seg was compared with YOLOv11m-seg and YOLOv26m-seg using the same training and validation datasets and identical training configurations. This ensures that the observed performance differences are primarily caused by model architecture rather than data variation or training settings.

The evaluation metrics included both bounding-box detection and mask segmentation performance, reported in terms of Precision (P), Recall (R), mAP50, and mAP50–95. Precision measures the proportion of correctly detected positive samples among all predicted positive samples; while Recall measures the proportion of correctly detected positive samples among all ground-truth positive samples; mAP50 represents the mean Average Precision calculated at an IoU threshold of 0.50, indicating whether the predicted bounding boxes or segmentation masks sufficiently overlap with the ground truth; mAP50–95 represents the mean Average Precision averaged over IoU thresholds from 0.50 to 0.95, typically with a step size of 0.05. For segmentation models, these metrics are reported separately for the bounding-box branch and the mask-segmentation branch.

4.4.1 YOLO Model Training Parameters

The key training parameters are summarized in Table 4, which were applied consistently using across all models as described in the following section.

Table 4. YOLO Model Training Parameters

Parameter	Value
Epochs	600
Batch size	30
Input image size	640
Scale augmentation	0.5
Mosaic augmentation	1.0
MixUp augmentation	0.0
Copy-paste augmentation	0.1
Horizontal flip	0.5
Mask ratio	4

4.4.2 Training Results of YOLOv12m-seg Model

The training and evaluation results of the YOLOv12m-seg model demonstrated strong detection and segmentation performance with stable convergence behavior. As shown in Figure 5, all loss components, including box loss, classification loss, segmentation loss, and distribution focal loss (DFL) decreased rapidly in the early training stage and gradually converged after approximately 200–300 epochs, indicating effective optimization and model stability. Meanwhile, both training and validation losses followed similar trends, suggesting minimal overfitting and good generalization capability.

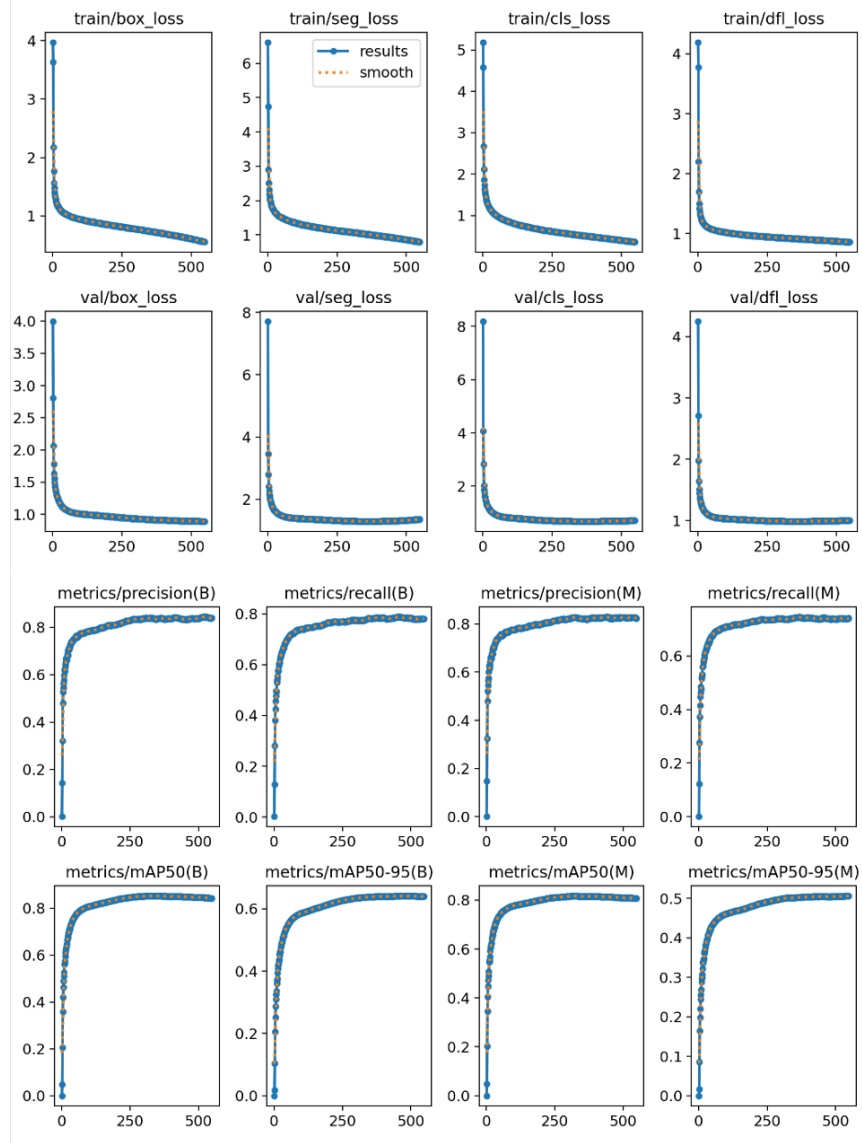


Figure 5. Training and Validation Metrics for YOLOv12m-seg Model

The evaluation metrics further confirmed the model's effectiveness. The precision and recall curves showed a smooth trade-off, with precision stabilizing above 0.83 and recall around 0.78 for bounding box detection, while segmentation metrics exhibited comparable trends. The mAP results reached approximately 0.85 (mAP50) and 0.64 (mAP50–95) for bounding boxes, and 0.81 (mAP50) and 0.50 (mAP50–95) for masks, indicating strong localization and segmentation accuracy across multiple IoU thresholds.

Figure 6 presents four standard evaluation curves for the bounding-box prediction branch across the four roadway damage classes, including D40, D43, D44, and D50. The bold blue curve represents the overall performance across all classes.

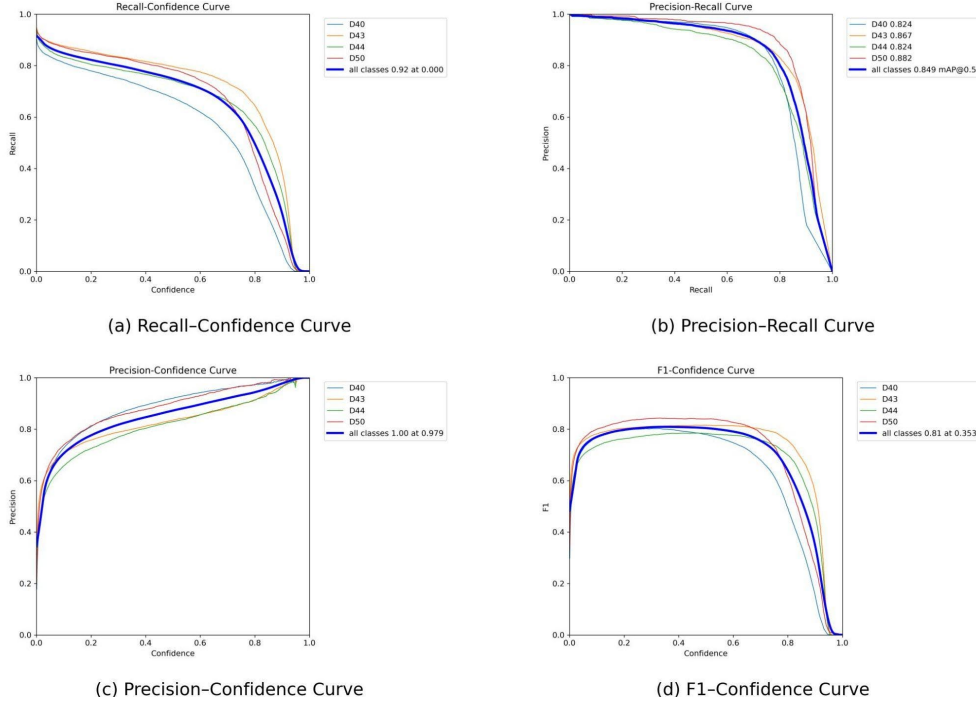


Figure 6. Detection Performance Curves of the Bounding-box Branch

In the Recall–Confidence curve of Figure 6(a), the overall recall reached 0.92 at a confidence threshold of 0. The recall remained within the range of approximately 0.75–0.85 until the confidence threshold reached around 0.70, after which it decreased sharply and approached zero near a confidence threshold of 0.95. Among the four classes, D43 achieved the highest recall, while D40 showed relatively lower recall, indicating that pothole detection was relatively more challenging.

The Precision–Recall curve of Figure 6(b) shows that the overall mAP50 for bounding-box detection was 0.849. The class-wise mAP50 values were 0.882 for D50, 0.867 for D43, 0.824 for D40, and 0.824 for D44. Precision remained above 0.90 until recall reached approximately 0.80, after which it decreased rapidly. This result indicates that the model maintained high precision over a wide recall range.

The Precision–Confidence curve of Figure 6(c) shows that precision increased from approximately 0.20 at a confidence threshold of 0 to 1.00 at a confidence threshold of 0.979. D40 and D50 generally showed slightly higher precision than D43 and D44 within the mid-confidence range, suggesting that their positive predictions were relatively more reliable when an appropriate confidence threshold was selected.

The F1 score is a harmonic meaning of precision and recall, defined as:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (1)$$

The F1–Confidence curve of Figure 6(d) indicates that how the F1 score changes with different confidence thresholds. As shown in Figure 6(d), the maximum overall F1 score was 0.81 at a confidence threshold of 0.353. Among the four classes, D50 achieved the highest peak F1 score,

approximately 0.84, followed by D43, D44, and D40. The F1 score remained above 0.75 when the confidence threshold was between approximately 0.05 and 0.70, indicating that the model was relatively robust to confidence-threshold selection within this range.

Overall, the bounding-box validation results demonstrates that YOLOv12m-seg achieved high accuracy and stable detection performance across all classes. The model provided a good balance between precision and recall, with an optimal confidence threshold around 0.35. Although pothole detection remained slightly more challenging than the other classes, the overall performance was sufficient for downstream applications such as pothole size estimation.

The segmentation performance of the trained model is shown in Figure 7. The mask evaluation curves demonstrates that YOLOv12m-seg achieved robust and stable segmentation performance across all classes. Although pothole segmentation was more difficult due to irregular shapes, unclear boundaries, and variations in pavement texture, the overall mask accuracy remained high. This indicates that the model can provide reliable pixel-level damage regions for subsequent pothole area estimation and severity assessment.

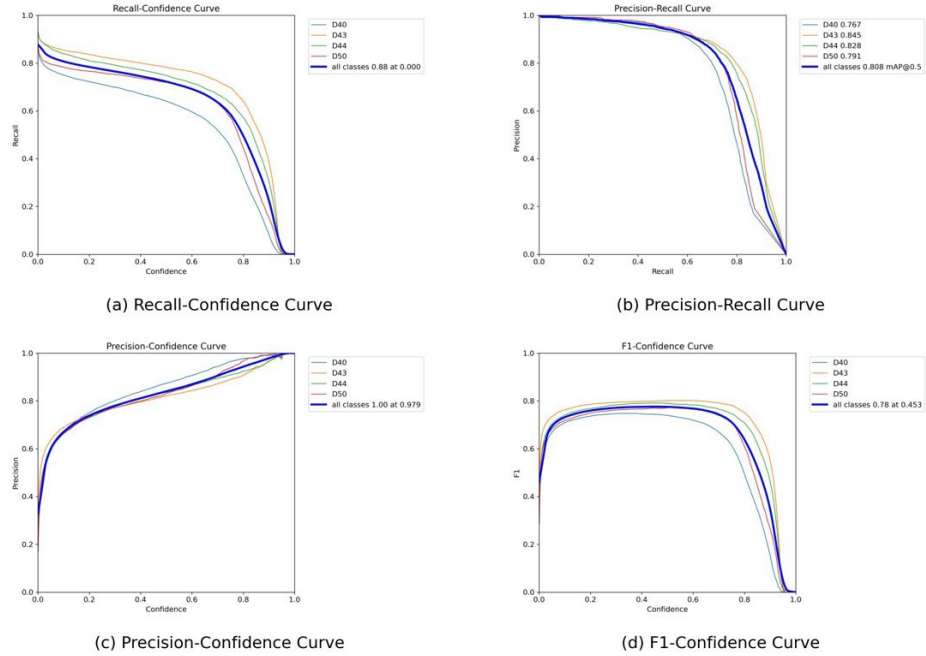


Figure 7. Detection Performance Curves of the Mask-segmentation Branch

As summarized in Table 6, the YOLOv12m-seg model achieved strong bounding-box detection performance, with a best mAP50 of 0.852 and a best mAP50–95 of 0.641. For segmentation masks, the model achieved a best mAP50 of 0.817 and a best mAP50–95 of 0.506. The final epoch results were very close to the best values, suggesting that the model training was stable and had largely converged by the end of the training process.

Table 5. YOLOv12m-seg Model Training Metrics and Loss Metric

Metrics	Best Value	Final Value
Precision	Box: 0.845, Mask: 0.830	Box: 0.841, Mask: 0.824
Recall	Box: 0.789, Mask: 0.747	Box: 0.780, Mask: 0.742
mAP50	Box: 0.852, Mask: 0.817	Box: 0.843, Mask: 0.807
mAP@50–95	Box: 0.641, Mask: 0.506	Box: 0.640, Mask: 0.506
Train box loss	0.556	0.556
Train segmentation loss	0.793	0.793
Train classification loss	0.358	0.358
Train DFL loss	0.858	0.858
Validation box loss	0.892	0.892
Validation segmentation loss	1.292	1.3565
Validation classification loss	0.685	0.710
Validation DFL loss	0.989	1.002

4.4.3 Performance Comparison of YOLOv12m-seg, YOLOv11m-seg and YOLOv26m-seg

The same training and validation datasets were used for all three models to ensure that performance differences are primarily caused by model architecture rather than data variation.

Table 5 presents a comprehensive comparison of YOLOv12m-seg, YOLOv11m-seg (Khanam & Hussain, 2024), and YOLOv26m-seg (Sapkota et al., 2026).

Table 6. Metrics Comparison of YOLOv12m-seg, YOLOv11m-seg, and YOLOv26m-seg

Model	Class	Box (Precision)	Recall	mAP50	mAP50-95	Mask(P)	Recall	mAP50	mAP50-95
Yolov12m-seg	Pothole(D40)	0.880	0.733	0.824	0.567	0.859	0.656	0.767	0.434

	Pavement sign(D43)	0.801	0.825	0.867	0.742	0.809	0.791	0.845	0.600
	Pavement lane (D44)	0.790	0.775	0.824	0.643	0.825	0.760	0.828	0.520
	Manhole(D50)	0.866	0.820	0.882	0.613	0.816	0.728	0.791	0.437
	All	0.834	0.788	0.849	0.641	0.828	0.733	0.808	0.498
Yolov11m-seg	Pothole(D40)	0.893	0.766	0.837	0.590	0.845	0.710	0.789	0.454
	Pavement sign(D43)	0.820	0.804	0.861	0.737	0.807	0.784	0.838	0.600
	Pavement lane (D44)	0.824	0.751	0.822	0.648	0.836	0.752	0.824	0.525
	Manhole(D50)	0.886	0.818	0.888	0.620	0.810	0.736	0.798	0.448
	All	0.856	0.785	0.852	0.649	0.824	0.745	0.812	0.507
Yolov26m-seg	Pothole(D40)	0.884	0.772	0.840	0.596	0.817	0.708	0.764	0.442
	Pavement sign(D43)	0.801	0.798	0.834	0.708	0.788	0.781	0.813	0.587
	Pavement lane (D44)	0.800	0.751	0.798	0.618	0.810	0.755	0.807	0.513
	Manhole(D50)	0.875	0.836	0.890	0.629	0.786	0.748	0.784	0.445
	All	0.840	0.789	0.840	0.638	0.800	0.748	0.792	0.497

Overall, all three models achieved strong performance on the roadway damage dataset. The mAP50 values exceeded 0.80 for both bounding-box detection and mask segmentation, indicating that the models were able to reliably detect and segment roadway damage instances.

Based on the overall results for all classes, YOLOv11m-seg achieved the best bounding-box detection performance, with a box mAP50 of 0.852 and a box mAP50–95 of 0.649. It slightly outperformed YOLOv12m-seg, which achieved 0.849 and 0.641, respectively, and YOLOv26m-seg, which achieved 0.840 and 0.638. A similar trend was observed for mask segmentation. YOLOv11m-seg obtained the highest overall mask performance, with a mask mAP50 of 0.812 and a mask mAP50–95 of 0.507, followed closely by YOLOv12m-seg and YOLOv26m-seg.

For the pothole class, D40, which is one primary focus of this study, YOLOv26m-seg achieved the highest bounding-box detection accuracy, with a box mAP50 of 0.840 and a box mAP50–95 of 0.596. In contrast, YOLOv11m-seg achieved the best segmentation performance for potholes, with a mask mAP50 of 0.789 and a mask mAP50–95 of 0.454. Although YOLOv12m-seg showed slightly

lower pothole-specific performance than the other two models, it still maintained competitive accuracy, with a box mAP50 of 0.824 and a mask mAP50 of 0.767.

For the other roadway damage classes, including degraded pavement signs (D43) and pavement lane damage (D44), the three models showed generally comparable performance. YOLOv12m-seg demonstrates slightly stronger segmentation consistency for some classes, such as D43, where it achieved a mask mAP50 of 0.845. For the manhole class, D50, YOLOv26m-seg achieved the strongest box-level performance, with a box mAP50 of 0.890, while the segmentation performance remained similar across the three models.

Overall, the differences among the three models were relatively small. While YOLOv26m-seg led in pothole bounding-box accuracy and YOLOv11m-seg led in pothole segmentation, neither model maintained that advantage consistently across all damage classes. YOLOv12m-seg, by contrast, achieved competitive performance on detection accuracy and segmentation quality across all different classes. Therefore, YOLOv12m-seg was selected as a suitable and reliable model for the subsequent pothole detection and size estimation analysis.

4.4.4 Prediction Test for the YOLOv12m-seg

To further evaluate the robustness of the trained YOLOv12m-seg model, we conducted inference analysis on three groups of untrained test images, with each group containing nine representative samples. These images included potholes and pavement marking damage under different environmental conditions, including normal lighting, low-light scenarios, and shadow-affected scenes (Iparraguirre et al., 2022; Maeda et al., 2018).

The prediction results in Figure 8, Figure 9, and Figure 10 show that the trained model maintained stable detection performance across different visual conditions. In particular, the model was able to correctly identify target roadway damage objects and generate reliable segmentation masks even under challenging lighting environments. These results indicate that the trained YOLOv12m-seg had a certain level of robustness and practical potential for roadway damage detection using onboard vehicle image data.

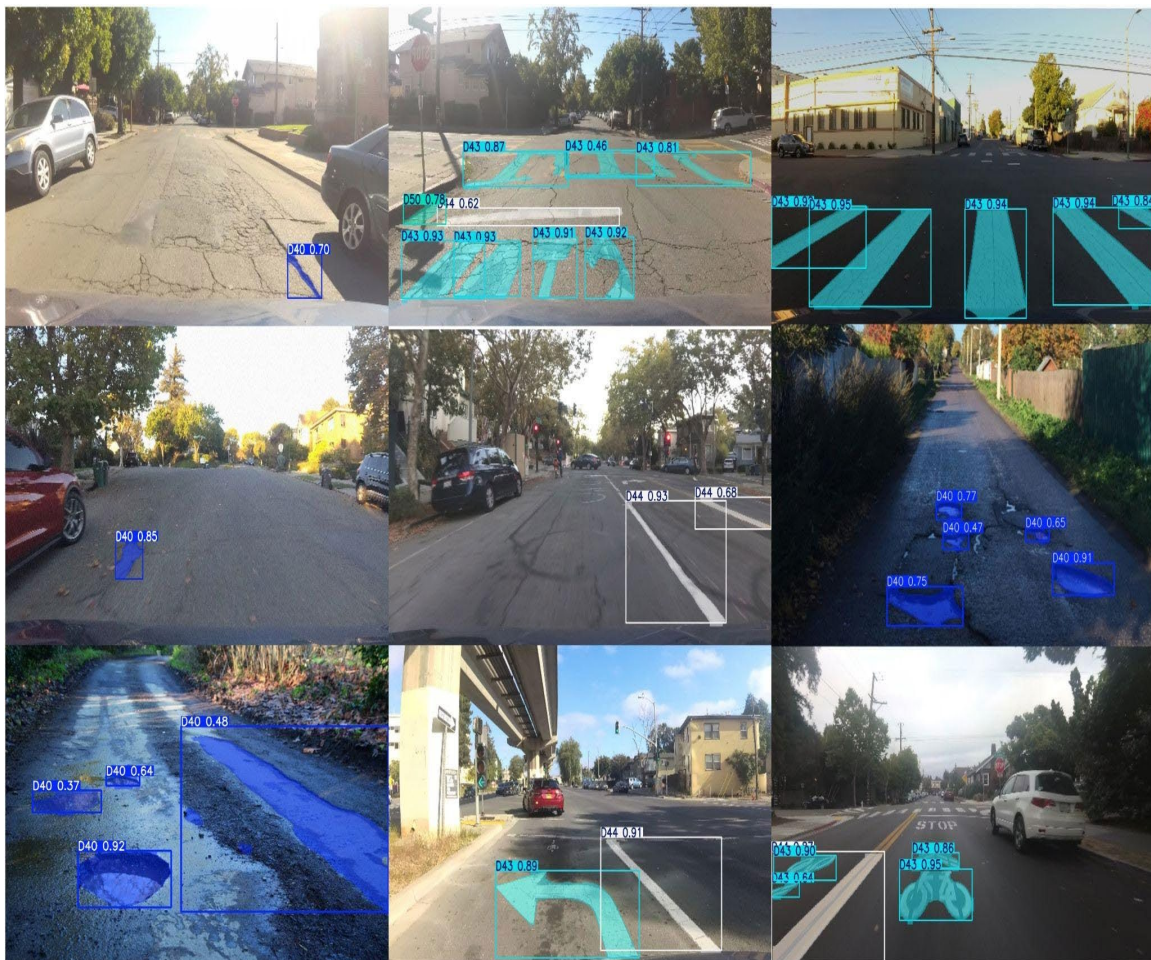


Figure 8. Prediction Examples of Group 1

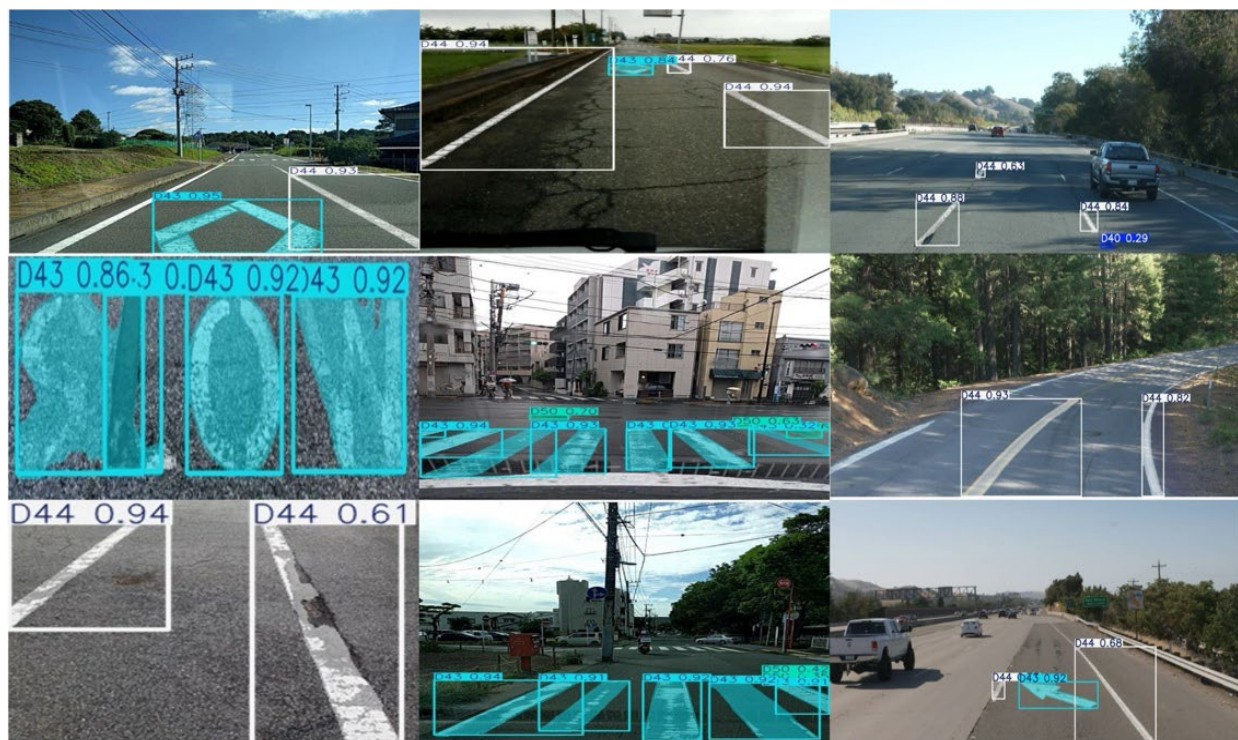


Figure 9. Prediction Examples of Group 2

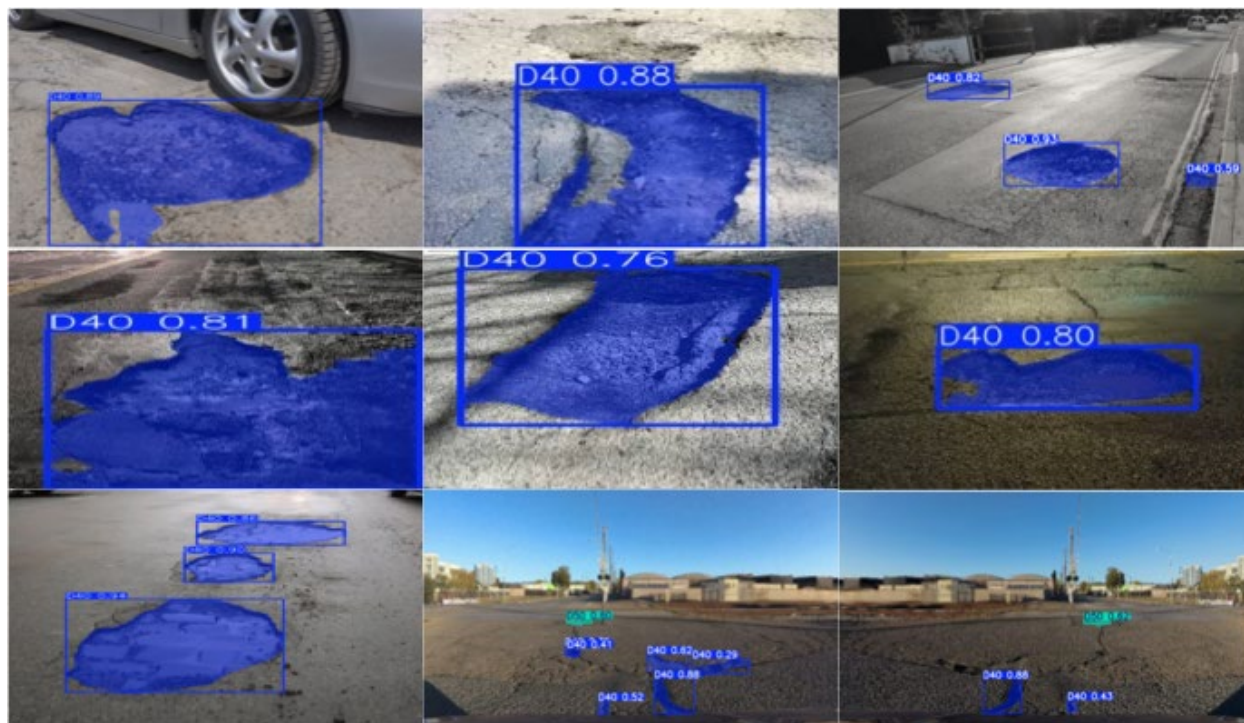


Figure 10. Prediction Examples of Group 3

5 Pothole Size Estimation

Building on the vision-based detection results, further quantitative analysis is needed to evaluate the detected roadway damage. Under the Pavement Condition Index (PCI) framework, the severity of pavement distress is largely determined by its physical dimensions, such as length, width, and affected area. Therefore, in addition to detecting potholes, estimating their real-world size is essential for assessing pavement condition and supporting maintenance decisions.

5.1 Pothole Size Estimation Process

This section presents a vision-based pothole size estimation framework that integrates instance segmentation, depth sensing, geometric projection, and calibration-based correction. The goal is to estimate the physical dimensions of potholes from image data collected by onboard cameras.

As shown in Figure 11, the framework first used the trained segmentation model to detect potholes and extract pixel-level masks. The physical dimensions of each detected pothole were then estimated using camera geometry and measured depth information. Finally, a data-driven calibration process was applied to correct systematic estimation errors and improve measurement accuracy.

Compared with traditional manual measurement methods or sensor-intensive inspection systems, the proposed framework provided a scalable, cost-effective, and automated solution for large-scale roadway monitoring. By leveraging pixel-level segmentation results, it can also capture irregular pothole shapes, which enables more accurate area estimation and severity classification under the PCI framework.

To quantify the physical dimensions of detected potholes, the overall process consisted of camera configuration, calibration, object detection, and pothole dimension estimation.

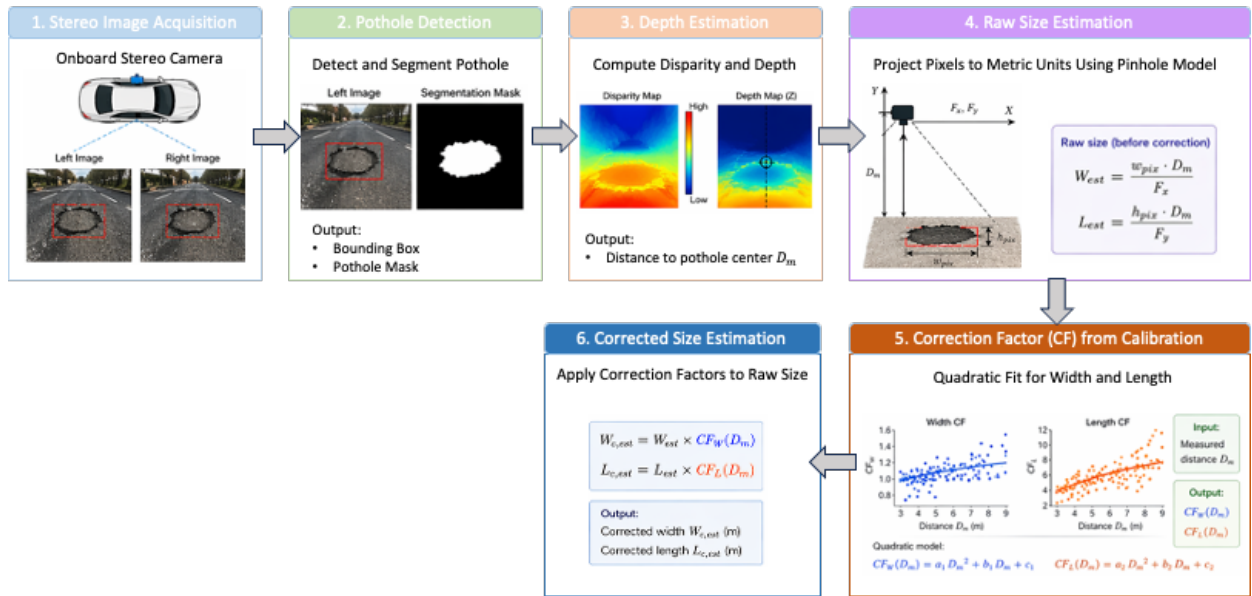


Figure 11. Pothole Size Estimation Process

1. Camera Selection and Configuration

To improve the accuracy of physical size estimation, we selected ZED Stereo Camera as shown in Figure 12, which provides direct distance measurements to detected objects. Compared with monocular cameras, depth sensing improves the reliability of scale estimation, especially when the object distance varies across different image frames (Smolyanskiy et al., 2018).



Figure 12. ZED Stereo Camera

The stereo camera was calibrated to obtain the geometric parameters required for depth-based size estimation and to correct image distortion. The calibrated camera parameters are summarized in Table 7. The intrinsic parameters include F_x and F_y , which denote the focal lengths in x and y directions, respectively, measured in pixels. C_x and C_y denote the principal point coordinates. The distortion coefficients include K_1, K_2, K_3, K_4, K_5 , and K_6 , which describe radial distortion. P_1 and P_2 denote tangential distortion. The parameter f denotes physical focal length of the camera.

Table 7. Stereo Camera Calibration Parameters

Parameters	Left camera sensor	Right camera sensor
F_x, F_y	[1911.490, 1912.430]	[1906.200, 1907.040]
C_x, C_y	[1131.790, 631.173]	[1099.980, 646.796]
K_1, K_2	[3.830, 15.438]	[-0.207, 24.220]
K_3, K_4, K_5, K_6	[-6.426, 3.936, 15.464, -4.562]	[-9.613, -0.107, 23.975, -7.266]
P_1, P_2	[-0.0002, -0.0000]	[-0.0001, -0.0002]
f	3.823 mm	3.812 mm

In this study, we assume that the camera mounting position and orientation remain fixed throughout the data collection process. This assumption ensures that the calibrated camera parameters and the derived correction factors remain valid across all collected samples. Any change in the camera mounting height, angle, or position would require recalibration before applying the proposed size estimation method.

2. Raw Perspective-Based Size Estimation

The standard pinhole camera projection model is widely used in 3D vision and photogrammetry (Sturm, 2005). Based on the pinhole camera model, the physical dimensions of a detected pothole can be approximated using the linear relationship between pixel dimensions, camera depth, and focal length:

$$W_{est} = \frac{w_{pix} * D_m}{F_x} \quad (2)$$

$$L_{est} = \frac{h_{pix} * D_m}{F_y} \quad (3)$$

where w_{pix} is the width of detected bounding box in pixels, h_{pix} is the length of detected bounding box in pixels, D_m is the measured distance from the camera to the pothole, F_x is the focal length in x-direction in pixels, F_y is the focal length in y-direction in pixels. W_{est} and L_{est} represent the estimated physical width and length of the detected pothole, respectively.

3. Correction Factor from Calibration

Although the raw perspective-based estimation provided a basic approximation of pothole size, it may introduce non-negligible errors due to lens distortion, perspective effects, and depth measurement noise. These errors grew larger as the distance between the camera and the pothole increases.

To correct these systematic errors, we collected ground-truth length and width measurements for 11 pothole samples and estimated their dimensions at multiple camera-to-object distances using the raw perspective-based method. The preliminary results showed distance-dependent deviations caused by lens distortion, perspective distortion, and depth sensing uncertainty.

Therefore, we introduced a distance-dependent correction factor, denoted as $CF(D_m)$, which was fitted using quadratic regression based on the field-collected calibration data. This correction model captured the nonlinear relationship between measurement distance and estimation error:

$$CF(D_m) = aD_m^2 + bD_m + c \quad (4)$$

where a is the quadratic coefficient, b is the linear coefficient, and c is the constant bias.

The fitted correction factor curves for all samples are presented in Figure 13. Figure 13(a) presents the width correction factor as a function of distance. The fitted curve showed a mild increasing trend as distance increases, indicating that the raw width estimation tended to require a slightly larger correction at longer distances. However, the relationship was relatively weak, as indicated by the low R^2 value of 0.1743. This suggests that width estimation errors were influenced slightly by distance. Figure 13(b) presents the length correction factor as a function of distance. Compared with width, the length correction factor shows a stronger increasing trend with distance. The fitted curve indicates that the raw length estimation became more sensitive to measurement distance, especially at longer ranges. The R^2 value of 0.4050 suggests a moderate relationship between distance and the length correction factor.

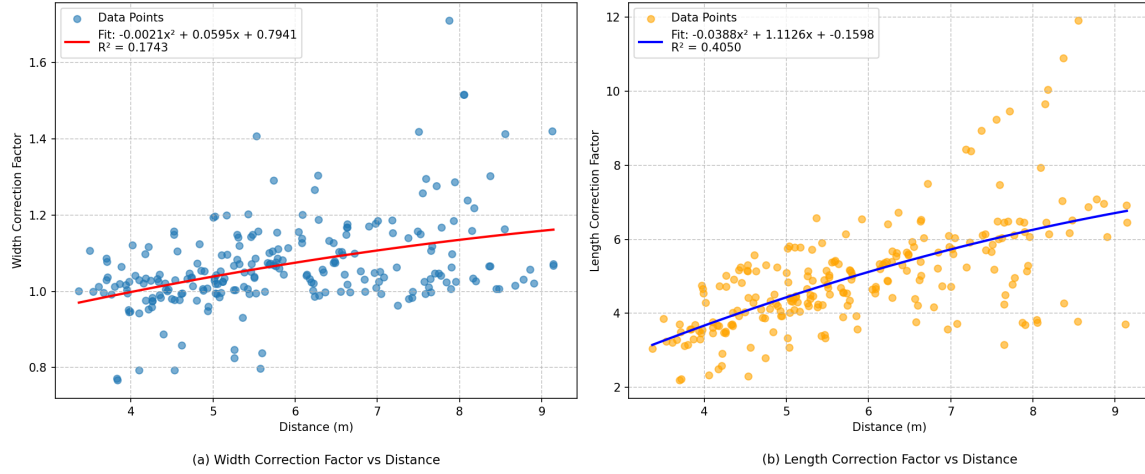


Figure 13. Distance-Dependent Correction Factor Fitting for Pothole Size Estimation

The correction factors for width and length are defined as follows:

$$CF_W = -0.0021D_m^2 + 0.0595D_m + 0.7941 \quad (5)$$

$$CF_L = -0.0388D_m^2 + 1.1126D_m - 0.1600 \quad (6)$$

where CF_W and CF_L represent the correction factors for width and length, respectively.

4. Corrected Size Estimation

The corrected pothole dimensions are then calculated by applying the corresponding correction factors to the raw perspective-based estimates:

$$W_{c,est} = W_{est} * CF_W \quad (7)$$

$$L_{c,est} = L_{est} * CF_L \quad (8)$$

where $W_{c,est}$ and $L_{c,est}$ are the corrected width and length of estimates, respectively.

After obtaining the corrected pothole dimensions, the pothole area is estimated using two approaches.

First, a rectangular approximation is computed based on the corrected width and length:

$$S_{Box} = W_{c,est} * L_{c,est} \quad (9)$$

where S_{Box} is the bounding box area.

This approach provides a simple area estimate by assuming that the pothole occupies the entire bounding box.

Second, because potholes usually have irregular shapes, the area estimation can be further refined using the segmentation mask. The mask-based area is calculated as:

$$S_{Mask} = S_{Box} * \frac{N_{Mask}}{N_{Box}} \quad (10)$$

where S_{Mask} denotes mask-based area, N_{Mask} is the number of pixels inside the segmentation mask, and N_{Box} is the number of pixels inside the bounding box.

This method adjusted the rectangular bounding box area according to the mask occupancy ratio, providing a more realistic estimate of the actual pothole area.

5.2 Apparatus

We deployed the pothole size estimation system in a field-testing environment to evaluate its performance under actual roadway conditions.

The stereo camera was mounted on the vehicle hood and connected to a portable onboard computing unit. The system was implemented on a laptop equipped with an NVIDIA RTX 3050 GPU, a multi-core Intel i7-11800 CPU, and a Linux operating system. The trained YOLOv12m-seg ONNX model and the proposed size estimation algorithm were integrated into the system, enabling real-time pothole detection and extraction of physical size information during vehicle operation.

To ensure safety and avoid interference with live traffic, all experiments were conducted in a closed-course test track at Richmond Field Station. The test site contains enough pavement damage samples, making it suitable for validating the proposed method. The apparatus setup and testing field are illustrated in Figure 14.



Figure 14. Apparatus and Testing Field

5.3 Pothole Size Estimation Results

5.3.1 Pothole Size Estimation Before and After Correction

To evaluate the effectiveness of the proposed correction method, four representative pothole samples were selected from the dataset. For each sample, the estimated width and length were calculated before and after applying the correction factors, and the results were compared with the corresponding ground-truth measurements.

Figure 15 presents a visual comparison of the pothole size estimation results before and after correction. The corrected dimensions were much closer to the ground-truth values, indicating that

the proposed calibration-based correction method can effectively reduce estimation errors caused by distance-dependent nonlinear distortion.



Figure 15. Pothole Size Estimation Before and After Correction

As shown in Table 8, the raw perspective-based estimation produced relatively large errors in length estimation, with errors reaching up to 78.25%. In contrast, the raw width estimation was generally more accurate, with most errors remaining within 10%. After applying the proposed distance-dependent correction factors, the estimation accuracy improved substantially. The length error was reduced to below 10% for all four cases, while the width error remained within a relatively small range, generally below 5%.

These results demonstrate that the proposed calibration method effectively compensates for nonlinear distortion, perspective effects, and depth-related measurement errors, particularly for length estimation. This improvement is important because accurate pothole dimensions are required for reliable pavement distress assessment and severity classification.

Table 8. Comparison of Size Estimation Before and After Correction

Case	Ground Truth (Width, Length) [m]	Before Correction (Width, Length) [m]	After Correction (Width, Length) [m]	Error of Before (%)	Error of After (%)
Case 1	0.40, 0.40	0.40, 0.42	0.40, 0.33	0.00, 5.00	0.00, -22.50
Case 2	0.40, 0.40	0.40, 0.59	0.40, 0.38	0.00, 47.50	0.00, -5.00
Case 3	0.40, 0.40	0.40, 0.51	0.40, 0.47	0.00, 27.50	0.00, -5.00
Case 4	0.40, 0.40	0.40, 0.78	0.40, 0.70	0.00, 95.00	0.00, -17.50

1	(0.433, 0.989)	(0.423, 0.317)	(0.436, 1.038)	(2.31, 67.95)	(0.69, 4.95)
2	(0.855, 1.071)	(0.845, 0.303)	(0.825, 1.027)	(1.17, 71.71)	(3.51, 4.11)
3	(0.855, 1.071)	(0.779, 0.233)	(0.802, 0.965)	(8.89, 78.25)	(6.20, 9.90)
4	(0.428, 0.510)	(0.445, 0.171)	(0.410, 0.490)	(3.97, 66.47)	(4.21, 3.92)

5.3.2 Pothole Size Estimation with Different Measurement Distance

To further evaluate the performance of the proposed pothole size estimation algorithm, experiments were conducted using three representative pothole samples at different measurement distances. For each pothole, size estimation was performed at multiple distances ranging from 4.5 m to 7.0 m. The selected test samples and estimation results are illustrated in Figure 16, and the corresponding quantitative results are summarized in Table 9.



Figure 16. Pothole Size Estimation with Different Distances

Table 9. The Summary of Pothole Size Estimation in Different Distance

Case	Ground Truth (Width, Length) [m]	Distance [m]	Estimate Value (Width, Length) [m]	Estimate Error (%)
1	(0.778, 1.114)	5.015	(0.758, 1.123)	(0.69, 0.808)
		5.670	(0.749, 1.108)	(3.728, 0.539)
		6.463	(0.695, 1.157)	(10.668, 3.860)

2	(0.335, 1.102)	4.922	(0.319, 1.1042)	(4.776, 5.445)
		5.499	(0.335, 1.135)	(0, 2.995)
		6.349	(0.331, 1.128)	(1.194, 2.174)
3	(0.445, 0.880)	4.997	(0.436, 0.871)	(2.022, 1.023)
		5.642	(0.433, 0.835)	(2.697, 5.114)
		6.584	(0.440, 0.853)	(1.124, 3.068)

As shown in Table 9, the proposed method achieved consistently high estimation accuracy across different measurement distances, although the error levels varied slightly as the camera-to-pothole distance increased.

For Case 1, the method achieved high accuracy at the shorter measurement distance of 5.015 m, with width and length errors of 0.69% and 0.81%, respectively. When the distance increased to 6.463 m, the estimation errors increased to 10.67% for width and 3.86% for length. This result suggests that measurements at longer distances are more sensitive to depth noise, perspective distortion, and small variations in segmentation or bounding-box localization.

For Case 2, the proposed method demonstrates stable and robust performance across all tested distances. The width estimation error remained within 0–4.78%, while the length estimation error stayed below 5.5%. These results indicate that the method can maintain reliable size estimation accuracy under moderate distance variations.

For Case 3, the estimation results are also reliable. The width estimation errors remained below 3% across all measurement distances. The length estimation error increased slightly at mid-range distances, reaching up to 5.11%, but still remained within an acceptable range for practical roadway damage assessment.

Overall, the experimental results demonstrate that the proposed size estimation method can provide accurate pothole dimension estimates under varying camera-to-object distances. However, the results also indicate that estimation uncertainty may increase at longer distances due to depth measurement noise, perspective effects, and segmentation sensitivity.

5.3.3 Robustness Test for Pothole Size Estimation Algorithm

Next, we evaluated the performance of the pothole size estimation algorithm under low-light and adverse weather conditions. Although the onboard camera can continuously collect raw video data, reliable detection and size estimation are generally easier under favorable illumination. In real-world vehicle operation, low-light, shadowed, and rainy conditions are unavoidable. These conditions may reduce image contrast, blur pothole boundaries, and affect both detection confidence and size estimation accuracy.

To improve system robustness, an image enhancement preprocessing step was applied before pothole detection and size estimation. Specifically, for low-light or shadowed scenes, Contrast Limited Adaptive Histogram Equalization (CLAHE) (Reza, 2004) was applied in the LAB color space. The original BGR image was first converted into LAB color space, where the L channel represents image lightness. CLAHE was then applied only to the L channel to enhance local

contrast while preserving the color information in the A and B channels. The enhanced image was subsequently converted back to BGR format and used as the input to the trained YOLOv12m-seg model and the pothole size estimation system.

The CLAHE parameters, including a clip limit of 2.0 and a tile grid size of 8×8 , were selected to balance contrast enhancement and noise suppression, thereby supporting stable performance under low-light and shadowed conditions.

To validate the robustness of the proposed pothole size estimation algorithm, its performance was evaluated under four representative environmental conditions: Case 1, low-light condition; Case 2, rainy condition with good light; Case 3, rainy condition with water-filled potholes; and Case 4, rainy condition with low lighting. For each case, pothole size estimation was conducted at three different distances to analyze the combined effects of lighting, weather, and measurement range on estimation performance.

Case 1: Low-Light Condition

The original sample for Case 1 is shown in Figure 17. The target pothole has a ground-truth width of 0.861 m and a ground-truth length of 1.148 m.

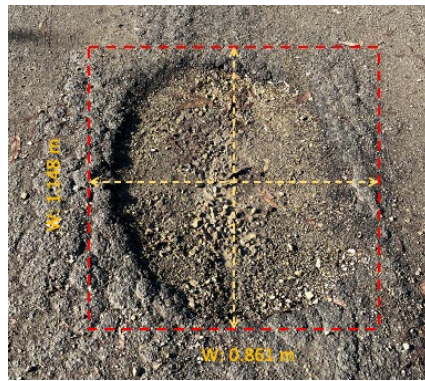


Figure 17. Original Image of Case 1 with Ground-Truth Width and Length

Figure 18 illustrates the pothole detection and size estimation results under low-light conditions at three different camera-to-pothole distances: 5.075 m, 6.550 m, and 7.266 m. For each distance, the left image shows the detection result without image enhancement, while the right image shows the corresponding result after applying image enhancement.



Figure 18. Illustration of Low-Light Test

Table 10 presents the pothole size estimation results under the low-light condition. Without image enhancement, the width estimation error increased rapidly as the measured distance increases, rising from 1.626% at 5.075 m to 5.698% at 6.550 m and 15.447% at 7.266 m. This was mainly because pothole edges become less distinguishable from the dark background as the image signal-to-noise ratio decreases. As a result, the detection confidence also decreased from 0.85 to 0.54.

Table 10. Estimate Results for Case 1

Measured Distance [m]	No Enhancement (Width, Length) [m]	No Enhancement error (%)	Confidence	With Enhancement (Width, Length) [m]	With Enhancement error (%)	Confidence
5.075	(0.847, 1.179)	(1.626, 2.700)	0.85	(0.852, 1.158)	(1.045, 0.871)	0.86
6.550	(0.811, 1.044)	(5.698, 9.059)	0.78	(0.820, 1.146)	(4.762, 0.174)	0.79
7.266	(0.728, 1.067)	(15.447, 7.056)	0.54	(0.839, 1.189)	(2.555, 3.571)	0.72

After applying image enhancement, the estimation accuracy improved substantially. The maximum width error was reduced from 15.447% to 2.555% at the farthest distance. The length estimation error also remained below 4% across all tested distances. In addition, the detection confidence remained above 0.70 even at 7.266 m, indicating that image enhancement improved both detection reliability and size estimation accuracy under low-light conditions.

Case 2: Rainy Condition with Good Lighting

The original sample for Case 2 is shown in Figure 19. The target pothole has a ground-truth width of 0.283 m and a ground-truth length of 0.885 m.

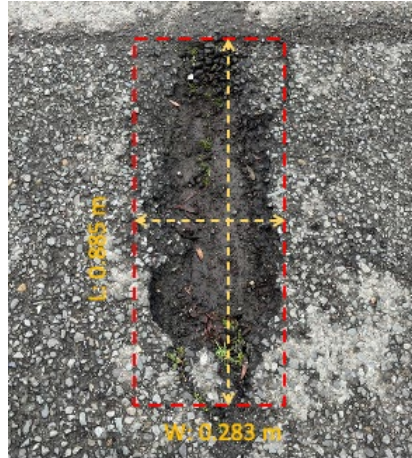


Figure 19. Original Image of Case 2 with Ground-Truth Width and Length

Figure 20 illustrates the pothole detection and size estimation results under rainy condition with good lighting at three different camera-to-pothole distances: 3.457 m, 4.401 m, and 5.635 m. For each distance, the left image shows the detection result without image enhancement, while the right image shows the corresponding result after applying image enhancement.

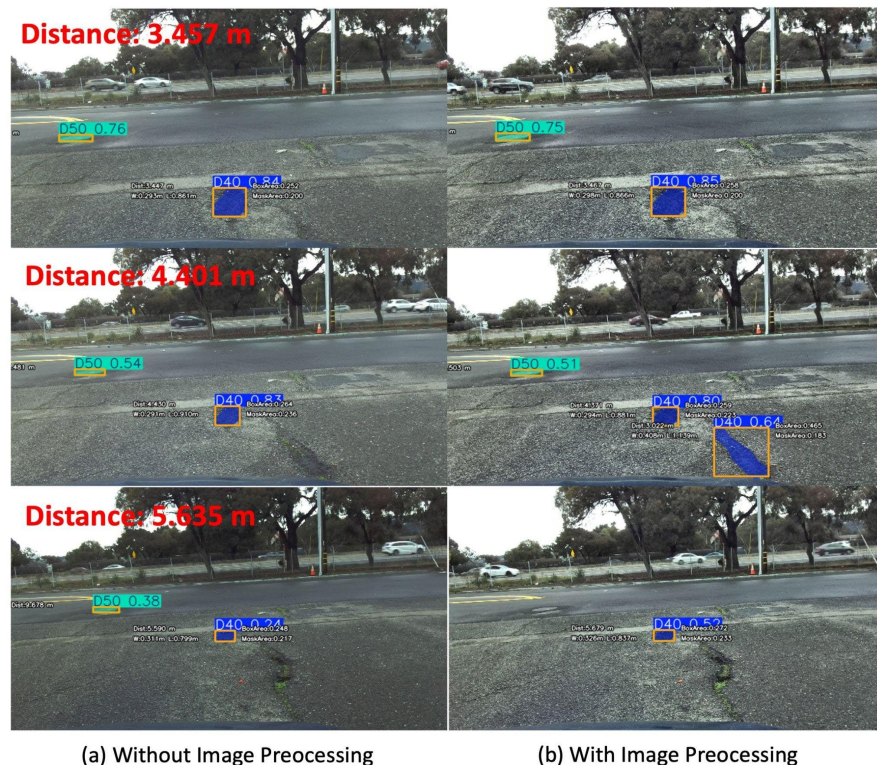


Figure 20. Illustration of Rainy Condition with Good Lighting Test

Table 11 summarizes the estimation results for Case 2 under rainy conditions with adequate lighting. Without image enhancement, the width estimation error remained relatively stable across the three tested distances, ranging from 2.658% to 3.322%. The length estimation error was low at the first two distances, with errors of 2.712% and 2.825%, but increased to 9.716% at 5.635 m. Detection confidence was high at shorter distances, 0.84 and 0.83, but decreased sharply to 0.24 at the farthest distance.

Table 11. Estimate Results for Case 2

Measured Distance [m]	No Enhancement (Width, Length) [m]	No Enhancement error (%)	Confidence	With Enhancement (Width, Length) [m]	With Enhancement error (%)	Confidence
3.457	(0.293, 0.861)	(2.658, 2.712)	0.84	(0.298, 0.866)	(0.997, 2.147)	0.85
4.401	(0.291, 0.910)	(3.322, 2.825)	0.83	(0.294, 0.881)	(2.326, 0.452)	0.80
5.635	(0.311, 0.799)	(3.322, 9.716)	0.24	(0.326, 0.837)	(8.306, 5.424)	0.52

With image enhancement, both width and length estimation accuracy improved at the first two distances. At 3.457 m, the width error decreased from 2.658% to 0.997%, and the length error decreased from 2.712% to 2.147%. At 4.401 m, the width error decreased from 3.322% to 2.326%, while the length error was reduced from 2.825% to 0.452%. At the farthest distance of 5.635 m, enhancement reduced the length error from 9.716% to 5.424% and increased the detection confidence from 0.24 to 0.52. However, the width error increased from 3.322% to 8.306%, suggesting that enhancement may improve object visibility and confidence while introducing some boundary changes for smaller potholes.

Overall, image enhancement provided clear benefits for Case 2 in terms of length estimation and detection confidence, especially at longer distances. However, its effect on width estimation was not uniformly positive, indicating that enhancement parameters may need to be adjusted depending on pothole size, lighting condition, and image contrast.

Case 3: Rainy Condition with Water-Filled Pothole

The original sample for Case 3 is shown in Figure 21. The target pothole has a ground-truth width of 0.690 m and a ground-truth length of 1.795 m.

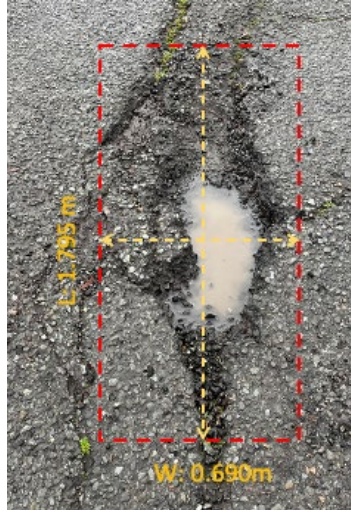


Figure 21. Original Image of Case 3 with Ground-Truth Width and Length

Figure 22 illustrates the pothole detection and size estimation results under rainy conditions with water-filled pothole at three different camera-to-pothole distances: 4.098 m, 4.975 m, and 5.950 m. For each distance, the left image shows the detection result without image enhancement, while the right image shows the corresponding result after applying image enhancement.

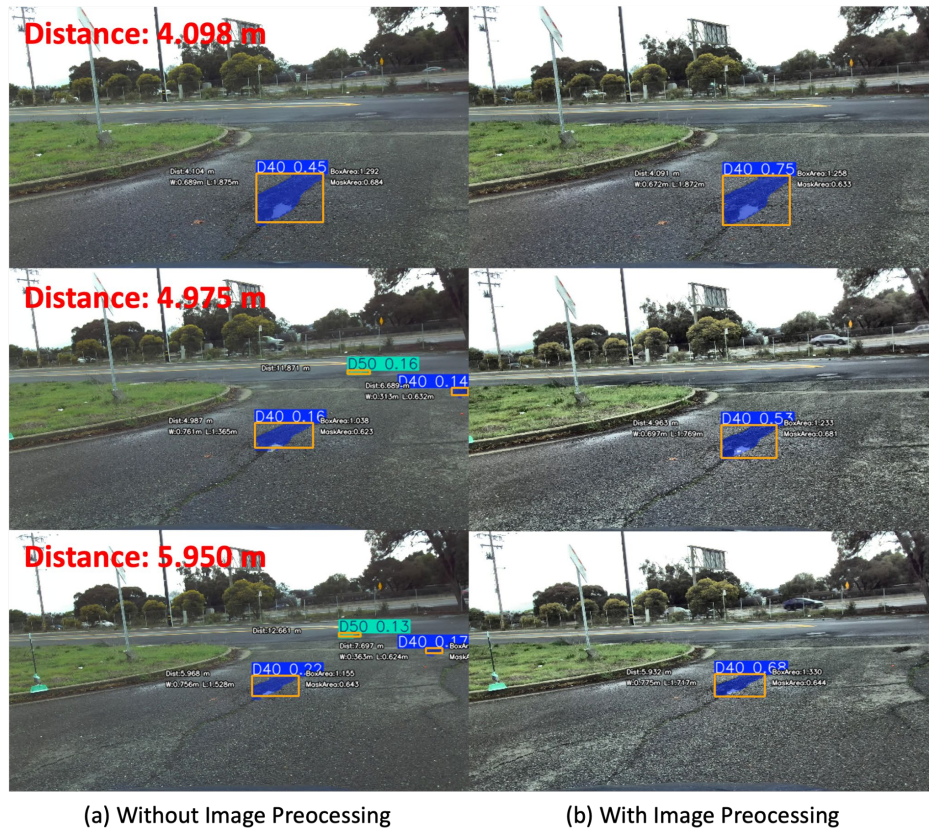


Figure 22. Illustration of the Rainy Condition Test with Water-Filled Potholes

Table 12 summarizes the size estimation results for Case 3. Without image enhancement, the width estimation errors were 0.145%, 10.290%, and 4.910% at the three measured distances, while the corresponding length errors were 4.457%, 23.955%, and 14.875%. The detection confidence scores were relatively low, particularly at longer distances, decreasing from 0.45 to 0.16 and 0.22.

Table 12. Estimate Results of Case 3

Measured Distance [m]	No Enhancement (Width, Length) [m]	No Enhancement error (%)	Confidence	With Enhancement (Width, Length) [m]	With Enhancement error (%)	Confidence
4.098	(0.689, 1.875)	(0.145, 4.457)	0.45	(0.672, 1.872)	(2.609, 4.290)	0.75
4.975	(0.761, 1.365)	(10.290, 23.955)	0.16	(0.697, 1.769)	(1.014, 1.448)	0.53
5.950	(0.756, 1.528)	(4.910, 14.875)	0.22	(0.775, 1.717)	(12.319, 4.345)	0.68

After applying image enhancement, the detection confidence improved at all three distances, increasing to 0.75, 0.53, and 0.68, respectively. The length estimation accuracy also improved substantially, especially at 4.975 m and 5.950 m. At 4.975 m, the length error decreased from 23.955% to 1.448%, and at 5.950 m, it decreased from 14.875% to 4.345%. At 4.098 m, the length error remained similar, changing slightly from 4.457% to 4.290%.

For width estimation, the effect of enhancement was mixed. The width error decreases significantly at 4.975 m, from 10.290% to 1.014%. However, it increased at 4.098 m, from 0.145% to 2.609%, and at 5.950 m, from 4.910% to 12.319%. Overall, image enhancement improved the stability of detection confidence and substantially reduced length estimation error, although its effect on width estimation was less consistent.

Case 4: Rainy Condition with Low Light

The original sample for Case 4 is shown in Figure 23. The target pothole has a ground-truth width of 1.128 m and a ground-truth length of 2.025 m.

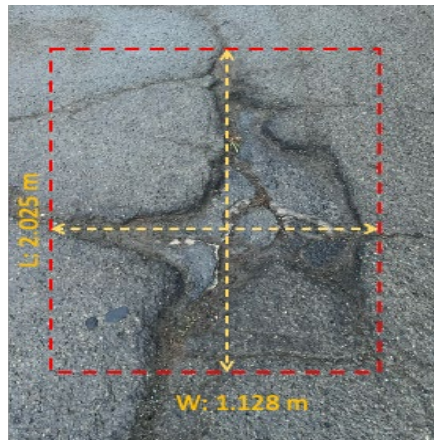


Figure 23. Original Image of Case 4 with Ground-Truth Width and Length

Figure 24 illustrates the pothole detection and size estimation results under rainy conditions with low light at three different camera-to-pothole distances: 4.740 m, 5.718 m, and 6.653 m. For each distance, the left image shows the detection result without image enhancement, while the right image shows the corresponding result after applying image enhancement.



Figure 24. Illustration of Rainy and Low-light Test

Table 13 summarizes the size estimation results for Case 4 under the rainy conditions with low light. Without image enhancement, the width estimation errors were 5.581%, 10.107%, and 29.521% at the three measured distances. The corresponding length errors were 2.617%, 14.519%, and 22.519%. As the distance increases, both the estimation error and detection instability become more pronounced. The confidence scores also decreased from 0.75 to 0.41 and 0.29, indicating that low-light rainy conditions made reliable detection more challenging, especially at longer distances.

Table 13. Estimate Results for Case 4

Measured Distance [m]	No Enhancement (Width, Length) [m]	No Enhancement error (%)	Confidence	With Enhancement (Width, Length) [m]	With Enhancement error (%)	Confidence
4.740	(1.194, 1.972)	(5.581, 2.617)	0.75	(1.163, 2.094)	(3.103, 3.407)	0.86

5.718	(1.014, 1.731)	(10.107, 14.519)	0.41	(1.070, 2.026)	(5.142, 0.049)	0.81
6.653	(0.795, 1.749)	(29.521, 22.519)	0.29	(0.915, 2.114)	(18.883, 4.395)	0.76

After applying image enhancement, the detection confidence improved at all three distances, increasing to 0.86, 0.81, and 0.76, respectively. The width estimation error was reduced at every measured distance. Specifically, it decreased from 5.581% to 3.103% at 4.740 m, from 10.107% to 5.142% at 5.718 m, and from 29.521% to 18.883% at 6.653 m.

For length estimation, the enhancement result was also generally improved. Although the length error slightly increased at 4.740 m, from 2.617% to 3.407%, it decreased sharply at the other two distances. At 5.718 m, the length error was reduced from 14.519% to 0.049%, and at 6.653 m, it decreased from 22.519% to 4.395%. These results indicate that image enhancement was particularly effective under rainy and low-light conditions, improving both detection confidence and size estimation accuracy.

5.3.4 Statistic Evaluation for Robustness

Figure 25 pooled the width and length estimation errors from all four test cases and compares the results under the no-enhancement and image-enhancement conditions. Since each case was tested at different measured distances, the x-axis uses standardized distance levels to enable consistent comparison across cases. Specifically, distance level 1 represents the near-distance condition, level 2 represents the mid-distance condition, and level 3 represents the far-distance condition. This allow the four cases to be compared on a common scale.

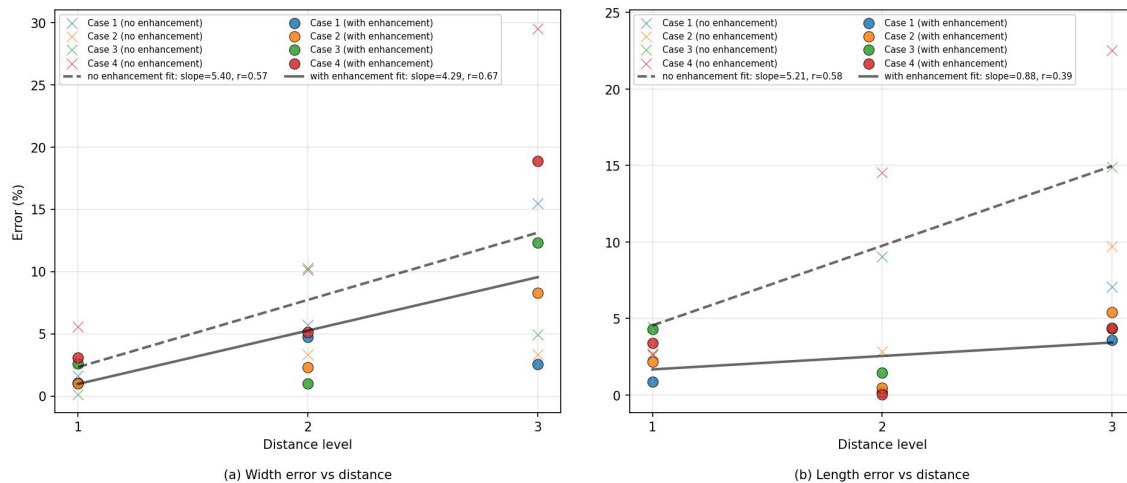


Figure 25. Pooled Width and Length Errors of Four Cases with Linear Fits

Figure 25(a) shows the pooled width estimation errors. Under both conditions, the width error generally increased as the distance increases. Without image enhancement, the error values were widely scattered and reach nearly 30% at the farthest distance level, with a Pearson correlation coefficient of $r = 0.568$. After applying image enhancement, most width error values were reduced to below 5%, and the fitted linear trend had a noticeably smaller slope. Although

the Pearson correlation slightly increases to $r = 0.671$, this does not indicate worse performance. Instead, the higher correlation suggests that the enhanced results are more tightly distributed around a clearer distance-dependent trend. In other words, image enhancement reduced the overall error magnitude, even though a mild relationship between distance and width error remains.

A few far-distance samples still show relatively large width errors after enhancement. This indicates that the enhancement process can broadly reduce width estimation error, but it cannot fully compensate for the loss of pixel resolution when the pothole occupied only a small region in the image at longer distances.

Figure 25(b) shows the pooled length estimation errors. The difference between the two conditions was more notable for length estimation. Without enhancement, the length error increased sharply with distance and exceeds 20% in the worst cases, with a Pearson correlation coefficient of $r = 0.579$. With enhancement, all twelve length error values were reduced to below 5.5%, the fitted slope became nearly flat, and the correlation decreased to $r = 0.392$. This suggests that the enhanced length estimation became much less sensitive to distance. The pooled mean length error also decreased substantially, from approximately 9.8% without enhancement to 2.5% with enhancement.

Overall, Figure 25 indicates that image enhancement consistently reduced both width and length estimation error and weakens their dependence on distance. The improvement was especially clear for length estimation. This suggests that the enhancement pipeline was more effective at recovering boundary information along the longer pothole axis than along the shorter axis. After enhancement, the remaining estimation error was mainly associated with width estimation at longer distances. For applications where pothole footprint or length-related severity metrics were important, the enhanced pipeline demonstrates reliable performance across the tested distance range.

Figure 26 plots detection confidence against the standardized distance level for all four cases, comparing the results with and without image enhancement.

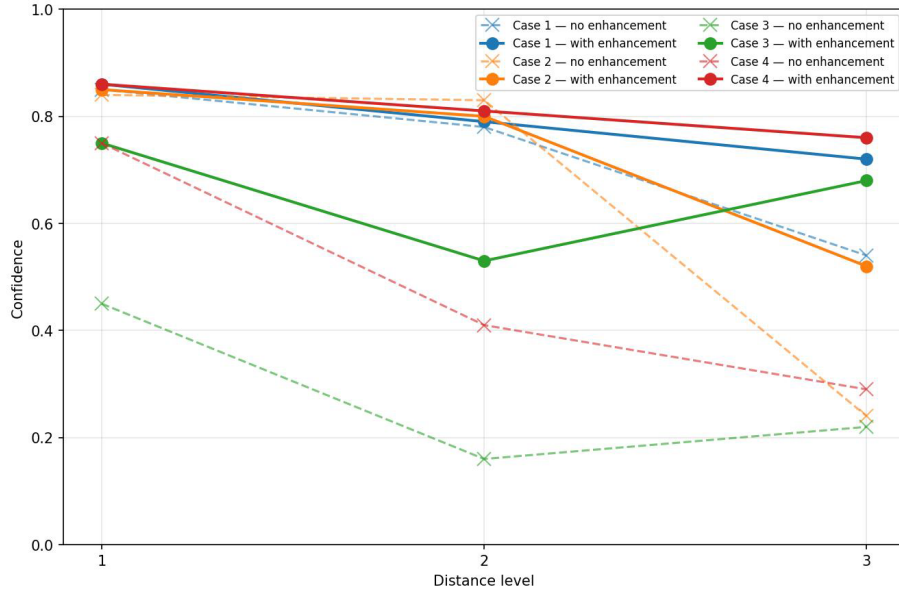


Figure 26. Detection Confidence of Four Cases

Without enhancement, the detection confidence decreased substantially as the distance level increases. The mean confidence across the four cases dropped from approximately 0.72 at the near-distance level to 0.55 at the mid-distance level and 0.32 at the far-distance level, corresponding to a total decrease of about 0.40. The confidence values were also widely distributed, ranging from 0.16 to 0.86, which indicated that the detector became less reliable at longer distances under different environmental conditions. The lowest confidence scores were concentrated in the rainy-day cases, where rain-induced texture noise and reduced image contrast made detection more difficult.

With image enhancement, the detection confidence was consistently higher at every distance level and in every case. The mean confidence values increased to approximately 0.83, 0.73, and 0.67 for the near-distance, mid-distance, and far-distance levels, respectively. As a result, the total confidence dropped across distance levels was reduced from about 0.40 to 0.16. The overall confidence range also became narrower, changing from 0.16–0.86 without enhancement to 0.52–0.86 with enhancement. This means that even the lowest post-enhancement confidence was comparable to the average confidence observed before enhancement.

The largest improvements occurred at the far-distance level under adverse conditions. In Case 3, the confidence increased from 0.22 to 0.68, and in Case 4, it increased from 0.29 to 0.76. These results indicate that image enhancement was especially effective when the original image quality was degraded by rain, low lighting, and long detection distance.

Overall, Figure 26 shows that image enhancement not only improved the average detection confidence but also stabilized confidence across distance levels. By increasing the lowest confidence values and reducing variability, the enhancement pipeline made confidence a more reliable indicator of detection quality. The improvement is most significant in the scenarios where the baseline model struggled the most, such as rainy and low-light conditions at longer distances.

6 Discussions and Conclusions

6.1 Findings

This project demonstrates the feasibility of leveraging automated vehicle sensing data and computer vision techniques to support roadway infrastructure condition monitoring. In this project, we developed a vision-based framework for detecting road surface damage and degraded pavement markings, with a particular focus on potholes, degraded pavement signs, and degraded lane markings.

A major outcome of the project is the construction of a large-scale annotated roadway damage dataset in the US. Through manual filtering, mask-based annotation, and an iterative semi-automated annotation pipeline, the project generated more than 20,000 annotated images for model training and validation. The use of mask-based polygon annotations allowed the model to capture irregular road damage boundaries more accurately than conventional bounding-box labels, which is especially important for pothole and pavement marking deterioration analysis.

The trained YOLOv12m-seg model achieved strong detection and segmentation accuracy across multiple roadway damage classes, including potholes, degraded pavement signs, degraded lane markings, and manholes. The overall model performance reached 0.849 mAP50 for bounding-box detection and 0.808 mAP50 for mask segmentation, confirming that YOLO-based segmentation is suitable for scalable roadway damage detection. The model also maintained reliable performance under challenging lighting and weather conditions, especially when image enhancement preprocessing was applied.

The trained model also showed promising robustness under different visual conditions, including normal lighting, low-light scenes, shadows, and rainy conditions. The prediction examples indicate that the model can identify roadway damage targets even in challenging environments. Additional preprocessing using image enhancement, particularly CLAHE-based contrast enhancement, further improved detection confidence and size estimation stability under low-light and rainy conditions.

Building upon the detection results, this project proposed a stereo vision-based pothole size estimation method. The method integrates instance segmentation, depth sensing, camera geometry, and calibration-based correction to estimate pothole width, length, and area. Experimental results showed that raw perspective-based estimation can produce large errors, particularly for pothole length estimation. After applying a distance-dependent correction factor, the estimation accuracy improved substantially. In representative test cases, the corrected width and length errors were generally reduced to below 5–10%, and many cases achieved errors within approximately 5% compared with ground-truth measurements.

The robustness tests further demonstrated that image enhancement can significantly improve the reliability of pothole size estimation in adverse conditions. Without enhancement, confidence and size estimation accuracy degraded notably as measurement distance increased, especially in rainy or low-light conditions. With enhancement, detection confidence remained more stable, and length estimation errors were consistently reduced. These findings suggest that the proposed enhancement and calibration pipeline can support field deployment.

The framework can be integrated into AV company data pipelines to process newly collected road condition videos, automatically detect pavement defects, estimate pothole dimensions, and generate roadway damage assessment outputs for Caltrans maintenance planning.

6.2 Areas of improvement

Although this project demonstrated strong potential, several limitations remain. The current dataset is large but still limited in geographic, environmental diversity, and long tail scenarios. While the dataset includes California roadway data, AV demonstration data, and open-source road damage images, additional samples from different regions, roadway types, and lighting conditions would further improve generalization.

The pothole size estimation experiments were conducted under controlled field conditions. Although the results are promising, further testing is needed in real-world traffic environments where vehicle speed, vibration, camera angle, occlusion, and motion blur may affect detection and depth estimation.

The size estimation method relied on a correction-factor-based calibration approach, which was derived from empirical fitting. This method was sensitive to the camera installation configuration, including camera height, tilt angle, and mounting position. As a result, when the camera setup changes, value of the correction factors may no longer be valid and require recalibration, which limits the generalizability of the approach across different platforms and deployment.

The current detection framework is built upon the YOLO-based segmentation model, which, although effective, represents only one class of state-of-the-art methods. This study did not explore whether alternative semantic segmentation architectures (e.g., transformer-based or multi-scale segmentation models) could achieve higher detection and segmentation accuracy. Therefore, there remains potential for further performance improvement through model architecture exploration and optimization.

6.3 Future Research Directions

We will continue expanding AV data-sharing partnerships, since AV fleets provide repeated roadway coverage and can serve as a valuable supplemental data source for infrastructure monitoring.

The trained model should be tested on broader roadway environments, including freeways, arterials, rural highways, and urban streets. This will help evaluate generalization and identify where additional training data are needed.

Future development should combine detection outputs with GPS trajectories and map-matching tools so that each detected defect can be automatically geolocated and summarized in a maintenance report.

Future model versions should include additional roadway infrastructure issues, such as rutting, patching, missing signs, damaged and obstructed signs. This would make the system more comprehensive for transportation department maintenance planning.

Future work should integrate a Large Language Model (LLM) based decision-support module into the proposed detection and size estimation framework. This module can analyze detected damage type, estimated dimensions, location, confidence score, and roadway context to classify damage severity and generate preliminary maintenance recommendations. By translating computer vision outputs into actionable maintenance plans, the LLM-enhanced system can further support Caltrans in prioritizing roadway maintenance activities and improving regular infrastructure management.

References

1. Lu, X., Madadi, B., Farah, H., Snelder, M., Annema, J. A., & Arem, B. V. (2019). Scenario-Based Infrastructure Requirements for Automated Driving. CICTP 2019, 5684–5695. <https://doi.org/10.1061/9780784482292.489>
2. Wang, P., McKeever, B., & Chan, C.-Y. (2022). Automated Vehicles Industry Survey of Transportation Infrastructure Needs. *Transportation Research Record: Journal of the Transportation Research Board*, 2676(7), 554–569. <https://doi.org/10.1177/03611981221080135>
3. Farah, H., Erkens, S. M. J. G., Alkim, T., & Van Arem, B. (2018). Infrastructure for Automated and Connected Driving: State of the Art and Future Research Directions. In G. Meyer & S. Beiker (Eds.), *Road Vehicle Automation 4* (pp. 187–197). Springer International Publishing. https://doi.org/10.1007/978-3-319-60934-8_16
4. Kim, J., & Lee, M. (2014). Robust Lane Detection Based On Convolutional Neural Network and Random Sample Consensus. In C. K. Loo, K. S. Yap, K. W. Wong, A. Teoh, & K. Huang (Eds.), *Neural Information Processing* (pp. 454–461). Springer International Publishing. https://doi.org/10.1007/978-3-319-12637-1_57
5. Li, J., Mei, X., Prokhorov, D., & Tao, D. (2017). Deep Neural Network for Structural Prediction and Lane Detection in Traffic Scene. *IEEE Transactions on Neural Networks and Learning Systems*, 28(3), 690–703. <https://doi.org/10.1109/TNNLS.2016.2522428>
6. Kawano, M., Mikami, K., Yokoyama, S., Yonezawa, T., & Nakazawa, J. (2017). Road marking blur detection with drive recorder. 2017 IEEE International Conference on Big Data (Big Data), 4092–4097. <https://doi.org/10.1109/BigData.2017.8258427>
7. Wei, C., Li, S., Wu, K., Zhang, Z., & Wang, Y. (2021). Damage inspection for road markings based on images with hierarchical semantic segmentation strategy and dynamic homography estimation. *Automation in Construction*, 131, 103876. <https://doi.org/10.1016/j.autcon.2021.103876>
8. Wang, Z., Ikeura, R., Hayakawa, S., & Zhang, Z. (2024). Road Marking Damage Degree Detection Based on Adaptive Feature Fusion. 2024 IEEE International Conference on Mechatronics and Automation (ICMA), 1819–1823. <https://doi.org/10.1109/ICMA61710.2024.10633191>
9. Zhang, L., Yang, F., Daniel Zhang, Y., & Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. 2016 IEEE International Conference on Image Processing (ICIP), 3708–3712. <https://doi.org/10.1109/ICIP.2016.7533052>

10. Majidifard, H., Adu-Gyamfi, Y., & Buttlar, W. G. (2020). Deep machine learning approach to develop a new asphalt pavement condition index. *Construction and Building Materials*, 247, 118513. <https://doi.org/10.1016/j.conbuildmat.2020.118513>
11. Matouq, Y., Manasreh, D., & Nazzal, M. D. (2024). AI-Driven Approach for Automated Real-Time Pothole Detection, Localization, and Area Estimation. *Transportation Research Record: Journal of the Transportation Research Board*, 03611981241246993. <https://doi.org/10.1177/03611981241246993>
12. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection (arXiv:1506.02640). *arXiv*. <https://doi.org/10.48550/arXiv.1506.02640>
13. Ettinger, S., Cheng, S., Caine, B., Liu, C., Zhao, H., Pradhan, S., Chai, Y., Sapp, B., Qi, C., Zhou, Y., Yang, Z., Chouard, A., Sun, P., Ngiam, J., Vasudevan, V., McCauley, A., Shlens, J., & Anguelov, D. (2021). Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9690–9699. <https://doi.org/10.1109/ICCV48922.2021.00957>
14. Wilson, B., Qi, W., Agarwal, T., Lambert, J., Singh, J., Khandelwal, S., Pan, B., Kumar, R., Hartnett, A., Pontes, J. K., Ramanan, D., Carr, P., & Hays, J. (2023). Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting (arXiv:2301.00493). *arXiv*. <https://doi.org/10.48550/arXiv.2301.00493>
15. Liao, Y., Xie, J., & Geiger, A. (2022). KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D (arXiv:2109.13410). *arXiv*. <https://doi.org/10.48550/arXiv.2109.13410>
16. Iparraguirre, O., Iturbe-Olleta, N., Brazalez, A., & Borro, D. (2022). Road Marking Damage Detection Based on Deep Learning for Infrastructure Evaluation in Emerging Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(11), 22378–22385. <https://doi.org/10.1109/TITS.2022.3192916>
17. Khanam, R., & Hussain, M. (2024). *YOLOv11: An Overview of the Key Architectural Enhancements*. <http://arxiv.org/abs/2410.17725>
18. Maeda, H., Sekimoto, Y., Seto, T., Kashiya, T., & Omata, H. (2018). Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. *Computer-Aided Civil and Infrastructure Engineering*, 33(12), 1127–1141. <https://doi.org/10.1111/mice.12387>
19. Reza, A. M. (2004). Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. In *Journal of VLSI Signal Processing* (Vol. 38).

20. Sapkota, R., Cheppally, R. H., Sharda, A., & Karkee, M. (2026). *YOLO26: Key Architectural Enhancements and Performance Benchmarking for Real-Time Object Detection*.
<http://arxiv.org/abs/2509.25164>
21. Smolyanskiy, N., Kamenev, A., & Nvidia, S. B. (2018). *On the Importance of Stereo for Accurate Depth Estimation: An Efficient Semi-Supervised Deep Neural Network Approach*.
<https://youtu.be/0FPQdVOYoAU>.
22. Sturm, P. (2005). *Multi-View Geometry for General Camera Models*.
23. Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-Centric Real-Time Object Detectors. <http://arxiv.org/abs/2502.12524>