Engineering Applications of Artificial Intelligence 171 (2026) 114321

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Real-time roadworks detection and high definition (HD) map updates for autonomous vehicles

Shaofan Sheng [a,*], Nicolette Formosa [b], Yuxiang Feng [a], Mohammed Quddus [a]

[a] Department of Civil and Environmental Engineering, Imperial College London, London, UK
[b] National Highways, Birmingham, UK

ARTICLE INFO

Keywords:
Roadworks
High definition (HD) maps
Visual language models (VLMs)
Autonomous vehicles

ABSTRACT

The increasing prevalence of roadworks poses significant challenges to maintaining accurate and up-to-date high-definition (HD) maps, crucial for autonomous vehicle (AV) safety and efficiency. Current methods for updating HD maps are expensive, time-consuming, and not responsive to real-time changes. This paper proposes a real-time, low-cost pipeline for updating HD maps with roadworks information using a monocular camera and a Contrastive Language–Image Pre-Training-based Vision Language Model (VLM), achieving robust few-shot sign recognition with minimal annotated data. The workflow is designed for low-cost, real-time deployment using only monocular camera input, and supports rapid, incremental HD map updates directly in OpenDRIVE format. Extensive experiments demonstrate that our system outperforms conventional baselines (e.g., fine-tuned You Only Look Once (YOLO) v11) not only in data-scarce settings but also across challenging environmental conditions. The recognition model is trained on a diverse dataset of 3752 real and virtual images, enhanced through data augmentation techniques. The model achieves a 97.12% recognition rate on a test dataset of 752 images and a root mean square error (RMSE) of less than 1.2 m for positional accuracy, processing single image inputs in 1.54 s. By leveraging the OpenDRIVE format, this approach ensures seamless data exchange between different HD map systems, facilitating real-time updates that accurately reflect current road conditions. The methodology demonstrates significant benefits in terms of responsiveness, cost and time efficiency, enhanced safety, and flexibility. Trials on the United Kingdom motorways validate the pipeline's effectiveness, offering a robust solution to dynamic road conditions and enabling safer, more efficient AV navigation.

1. Introduction

Autonomous vehicles (AVs) navigating seamlessly through dynamic urban environments and adapting effortlessly to dynamic road conditions rely heavily on the precision and reliability of high-definition (HD) maps. These maps provide the detailed contextual information necessary for safe and efficient navigation, containing intricate details about the road environment, including lane markings, traffic signals, and road geometries, all crucial for the vehicle's path planning and decision-making processes (Bao et al., 2023).

The use of the OpenDRIVE format, a widely accepted standard used for describing road networks, is essential for achieving this high level of precision and detail. OpenDRIVE is primarily an XML-based file format that provides a detailed representation of road geometry including lanes, lane markings and other road network attributes, ensuring compatibility and interoperability across different systems. However, maintaining up-to-date HD maps presents significant challenges. Traditional methods of HD map updating are often time-consuming and costly due to the reliance on manual data collection and processing. Frequent changes in road conditions, especially due to construction activities, necessitate regular updates to maintain accuracy. This constant need for updates, exacerbated by frequent road construction, demands continuous and labour-intensive map revisions and the cost of creating and updating HD maps can be as much as $1000/kilometre. In addition, delays in updating these maps can lead to discrepancies between the map data and the actual road conditions, potentially compromising the safety and efficiency of autonomous driving systems. Existing update pipelines are largely implementation-led (sensor fleets, offline refresh) and seldom articulate a method design that couples open-vocabulary recognition with a principled, single-camera geo-

* Corresponding author.
E-mail addresses: shaofan.sheng21@imperial.ac.uk (S. Sheng), nicolette.formosa@nationalhighways.co.uk (N. Formosa), y.feng19@imperial.ac.uk (Y. Feng), m.quddus@imperial.ac.uk (M. Quddus).

https://doi.org/10.1016/j.engappai.2026.114321
Received 17 April 2025; Received in revised form 18 November 2025; Accepted 23 February 2026
Available online 26 February 2026

projection and map-consistent serialisation. We target this gap by specifying the assumptions, invariants and failure modes of each stage.

Particularly challenging are scenarios such as roadworks, which alter the road layout and impact the routes that vehicles can travel through. Two common types of traffic management during roadworks - narrow lanes and lane closures - create substantial obstacles for AVs. Narrow lanes reduce the width of the travel path, creating a constrained environment for AVs, which rely heavily on precise lane-level data for navigation (Formosa et al., 2024; Singh et al., 2024). Lane closures force AVs to merge into adjacent lanes, requiring dynamic trajectory adjustments and real-time decision-making to ensure safety and increased traffic efficiency. These disruptions significantly impact traffic flow, leading to delays and increased traffic conflicts, particularly during peak hours or in high traffic density situations (Formosa et al., 2024).

HD maps contain elements such as road networks, static objects and semantic information that enable autonomous driving systems to achieve accurate navigation, sensing and route planning in complex environments (Mazhar et al., 2023). For AV controllers, they can achieve higher precision control and navigation with the precise environmental information provided by HD maps. For example, when encountering a traffic light, the behavioral decision controller can control the vehicle's deceleration and stopping by the location and status of the signal light from the HD map. Commonly used open-source HD map formats in autonomous driving research are OpenDRIVE and Lanelet2 (Poggenhans et al., 2018). They provide standardized data structures to represent road elements and lane attributes, which make HD maps highly compatible and easily accessible to different controllers. OpenDRIVE describes road geometry and lane topology in detail, which is suitable for path planning, while Lanelet2 has more semantic information for easy recognition of road rules.

Accurate detection of roadworks is essential for real-time HD maps updates. Current detection methods include on-board sensors, machine learning techniques and communication-based systems, and have some limitations. Traditional traffic objects detection methods tend to limit the type of detection and fix the labels for training leading to poor generalisation of the model (Zhang et al., 2025). Traditional methods of updating HD maps typically rely on large-scale road scanning by periodic data collection vehicles combined with manual labelling and calibration. These methods require large investments in labour and equipment and are costly (Wong et al., 2020). In addition to this, these methods mainly rely on offline processing, with a certain time delay between processing and uploading to the HD map platform, which may affect the correct perception and response of the real-time environment by AVs. For examples, camera-based systems and LiDAR while effective, require extensive labelled datasets and can misclassify objects (Formosa et al., 2024; O'Mahony et al., 2020). This can lead to false alarms and often lack contextual understanding of roadwork situations making it challenging for AVs to respond appropriately. Infrastructure-to-vehicle (I2V) systems, such as Smart Traffic Cones (STC), offer real-time information (Formosa et al., 2024), but face high costs and require extensive coverage and real-world testing.

To overcome these limitations, this paper proposes an innovative solution utilising Visual-Language Models (VLMs) such as CLIP (Contrastive Language–Image Pre-Training). CLIP leverages large pre-trained datasets, allowing it to generalise effectively with fewer labelled samples, significantly reducing the burden of data annotation (Gao et al., 2024). VLMs can utilise natural language-based labels, enhancing their adaptability to varied descriptions of roadworks (Zhu et al., 2025). This flexibility, combined with the ability to process real-time images from vehicle cameras, enables immediate and accurate updates to HD maps using the OpenDRIVE format. By integrating CLIP with OpenDRIVE, the map updating process becomes streamlined, improving the safety and efficiency of AVs. Additionally, incorporating virtual images during the training of the VLM augments the dataset, enhancing the model's ability to generalise and perform accurately under diverse conditions (Gaspar et al., 2025).

Our pipeline comprises three principled stages:

(i) **Open-vocabulary recognition with contextual prompts**, where compositional text (e.g., *roadworks sign with cones*) enlarges inter-class margins under co-occurrence;
(ii) **Single-camera geo-projection with minimal physical priors**, where calibrated intrinsics and a known marker height yield depth via the pinhole model, with first-order error propagation stated explicitly; and
(iii) **OpenDRIVE-schema-compliant insertion**, where detections are written as centreline-anchored objects $(s, t, z\text{Offset})$ under schema constraints to ensure auditability and safe roll-back.

Each stage specifies its assumptions (e.g., upright marker, stable extrinsics), expected gains, and dominant failure modes (confusion, occlusion, small-scale).

This research presents a novel solution for enhancing the precision and responsiveness of HD maps using Visual-Language Models like CLIP. By leveraging advanced simulation tools and integrating real-time image processing with the OpenDRIVE format, the proposed approach addresses the critical challenges of maintaining accurate and up-to-date navigational data for autonomous vehicles. The findings highlight the potential of this method to set a new standard for CAV navigation systems, improving safety, efficiency, and adaptability in rapidly changing road environments. We foreground the method design rather than implementation specifics and summarise our contributions together with the benefits they enable:

- Open-vocabulary roadworks recognition via descriptive prompts: Concise, compositional prompts (e.g., sign + cone context) improve separation and robustness across varied scenes and lighting.
- Single-camera geo-projection that is dependable in practice: Standard calibration and light-weight physical priors (approximate marker size) recover world positions without specialised sensors, supporting low-cost deployment.
- OpenDRIVE-consistent map updating: Detections are inserted as schema-valid, centreline-anchored entities, enabling minimal, auditable edits and safe roll-backs.
- Data efficiency through mixed real–synthetic training and prompt design: Targeted synthetic images plus descriptive prompts reduce annotation while maintaining performance under diverse conditions.
- Modular, plug-and-play pipeline: The components interface cleanly with existing HD-map stacks and can be adopted independently.

Each of these method-level choices is evaluated by focused studies (prompt ablations, localisation and map-validity checks, and efficiency), keeping the contribution at the level of method design rather than engineering throughput.

This paper is structured as follows: Literature Review section gives a detailed review of significance of HD map for AVs, challenges in maintaining HD maps and the applications of VLMs in addressing these challenges. Data section discusses the collection and composition of the data. Methodology section gives a detailed description of our roadworks recognition model, roadworks coordinate conversion from real-world to OpenDRIVE format and the HD map update process. Results section introduces the in-depth analysis of the performance of the roadworks recognition model and the HD map update accuracy. Limitation and Future Work section talks about potential problems of this paper and the future research directions. Finally, in Conclusion section, the main findings of this paper are synthesized.

2. Literature review

This literature review explores the critical role of HD maps in AVs, the challenges associated with maintaining these maps, and the application of VLMs, particularly the CLIP model, in overcoming these

challenges. HD maps are essential for the accurate positioning, robust perception, and effective route planning of AVs, but their maintenance is costly and infrequent. This review examines innovative solutions such as crowdsourced data and real-time sensor integration to overcome these obstacles. Additionally, it investigates the potential of VLMs to enhance roadworks detection and enable real-time HD map updates, highlighting their accuracy, flexibility, and responsiveness. By addressing these key areas, this review provides a comprehensive understanding of how advanced mapping technologies and artificial intelligence (AI) can enhance the safety and efficiency of autonomous driving.

2.1. Significance of HD map for AVs

HD maps have been instrumental in the development of AVs. Initially, digital maps provided a two-dimensional view with limited accuracy for traditional GPS navigation but insufficient for autonomous driving (Bao et al., 2023). By the turn of the century, enhanced digital maps emerged, incorporating precise roadway geometries such as curvature, gradient, and lane widths to an accuracy of around 50 cm (Bao et al., 2023). This advancement was fuelled by the growing need to support advanced driver assistance systems (ADAS) and autonomous driving. By the 2010s, the concept of HD maps dedicated to autonomous driving solidified, exemplified by projects such as the Mercedes-Benz Bertha Drive (Liu et al., 2020; Bao et al., 2023). These maps offer highly detailed three-dimensional lane-level displays with an accuracy of 10-20 cm, including key details such as lane markings, traffic signs, and road geometry.

HD maps are indispensable for AVs as they facilitate accurate vehicle positioning, robust sensing capabilities and efficient route planning. They provide detailed reference points that align vehicle location with the real world, critical in areas with low GPS coverage. Moreover, HD maps enhance vehicle perception by providing information about road features that onboard sensors might miss, such as accurate lane boundaries and traffic signs (Liu et al., 2020). This enriched perceptual capability facilitates better decision making and smoother navigation. In addition, HD maps are essential for performing complex driving manoeuvres and handling challenging scenarios, enabling AVs to anticipate and respond to upcoming road conditions (Bao et al., 2023). Advancements in data collection, processing and machine learning, continue to drive HD mapping technology forward, evolve, aiming to increase accuracy and real-time updating capabilities to ensure AVs can operate safely and efficiently in dynamic environments.

2.2. Challenges in maintaining HD maps

Maintaining HD maps remains a significant challenge. Traditional methods rely on specialised mapping vehicles equipped with high-precision sensors. However, these vehicles are costly and have low traversal frequency, resulting in rapidly outdated maps. To address this problem, researchers have developed crowdsourced data solutions that utilise anonymised data from fleets already on the road (Kim et al., 2021; Zhang et al., 2021). These innovative methods detect changes in real-time via camera and sensor data such as GNSS/IMU and update map features using techniques such as semantic segmentation and visual SLAM. By integrating this continuously collected data, HD maps can remain current, providing vital real-time information about lane markings, traffic signs and road geometry (Pannen et al., 2020). The partnership between Qualcomm and TomTom focuses on maintaining HD maps using crowdsourced data from vehicle sensors, facilitated through Qualcomm's Drive Data Platform. This platform generates map snippets based on camera input, which are then used by TomTom's HD Map to ensure real-time updates.

In recent years, end-to-end and vectorized HD map updating systems have also emerged, further advancing the field. For instance, Kim et al. (2024) proposes an approach that directly integrates CLIP-based semantic features into BEV (bird's-eye-view) representations, allowing for continuous, vectorized HD map construction without the need for traditional modular workflows. Similarly, Xia et al. (2025) presents an automated architecture capable of city-scale, lane-level map refreshing and topology updates. These systems demonstrate impressive automation and scalability, but they often require large-scale, continuous data collection and substantial computational resources, which may limit their practicality for incremental updates or deployment in resource-constrained environments.

While crowdsourcing offers a viable way of overcoming these difficulties by utilising the large amounts of data generated by daily vehicle operations, there are still issues with data quality and real-time performance. The primary issues include variability in data quality and reliability due to differences in sensor performance and environmental conditions. Additionally, concerns about privacy and data security arise from the extensive collection of vehicle data. Integrating diverse data sources and ensuring standardisation, such as adherence to the OpenDRIVE format, adds complexity. Lastly, achieving comprehensive coverage remains a challenge, especially in less traveled areas. Therefore, new approaches are needed to meet the real-time requirements of AVs for HD map updates while addressing both the economic and technical aspects to ensure the reliability and accuracy of navigational data. In this context, modular and flexible pipelines—such as the method proposed in this work—remain valuable, as they enable rapid deployment and incremental updates with lower data and hardware requirements, complementing the strengths of large-scale end-to-end systems.

2.3. Application of VLMs in addressing HD map challenges

The integration of visual data processing with natural language understanding has enabled models to interpret and analyse images in context with textual descriptions (Chowdhury and Soni, 2025), enhancing their ability to understand complex scenes and objects more closely aligned with human cognition. Among the existing VLMs, the CLIP model is particularly significant for its impact on the field of computer vision. CLIP has revolutionised the field by enabling more flexible and generalisable systems that perform well across a variety of tasks without the need for task-specific training data (Liu et al., 2024). This innovation has opened new possibilities for cross-modal learning and understanding, paving the way for more advanced and integrated AI systems.

For example, Lüddecke and Ecker (2021) proposed CLIPSeg for image segmentation, combining the capabilities of the CLIP model with a transformer-based decoder for dense prediction. This unified model handles various segmentation tasks, such as referring expression segmentation, zero-shot segmentation, and one-shot segmentation, using arbitrary text or image prompts.

Despite advancements in HD maps and their crucial role in CAV development, maintaining these maps remains challenging due to the high costs and inefficiencies of traditional update methods. Current solutions such as crowdsourcing and sensor integration face issues with data quality and real-time performance. Additionally, roadworks detection using on-board sensors and communication systems often lacks accuracy and contextual understanding, particularly under adverse conditions. The application of VLMs, specifically the CLIP model, offers potential improvements but has not been fully explored for roadworks detection and HD map updates. Addressing this gap, CLIP's robustness and ability to provide precise, real-time updates could significantly enhance the accurate update of HD maps, thereby improving CAV navigation and safety.

Using the CLIP model for roadworks detection presents several advantages over traditional methods. Firstly, CLIP's robustness allows it to maintain high detection accuracy across various environments and lighting conditions, where traditional methods often struggle. Secondly, CLIP requires smaller fine-tuning datasets because it has already been pre-trained on a vast number of image-text pairs, significantly reducing

the cost and time of data preparation. Thirdly, the flexibility of CLIP's text labels enables precise descriptions and classifications of roadworks. Unlike traditional rule-based or fixed-label systems, CLIP can distinguish between scenarios like road narrowing versus full closures and identify adjacent speed limit signs, providing more detailed and accurate roadwork information.

In terms of updating HD maps, the advantages of CLIP are even more pronounced. The model's ability to detect and describe roadwork changes in real-time allows for immediate updates to HD maps, ensuring they consistently reflect the latest road conditions. This capability not only improves the navigation accuracy and safety of autonomous vehicles but also mitigates the risks associated with delayed updates in traditional methods. Therefore, the application of CLIP in roadworks detection and HD map updates demonstrates significant benefits in terms of accuracy, flexibility, and real-time responsiveness, marking a crucial advancement for future intelligent transportation systems.

In summary, roadworks present unique challenges for AVs causing temporary changes in road layouts, leading to increased traffic conflicts, delays and potential safety risk. Current methods for detecting roadworks and updating HD maps are limited by environmental conditions, high costs, and a lack of contextual understanding. Effective navigation through these changing road conditions requires the use of up-to-date, accurate HD maps that reflect the current state of the road. Therefore, new approaches are necessary to detect roadworks more flexibly and accurately, providing comprehensive, real-time information. Despite advancements in mapping technologies and AI, integrating real-time roadwork data into HD maps remains a challenge. The integration of VLMs, such as the CLIP model, offers promising potential to bridge this gap. This paper aims to develop the models required to fully realise their capabilities in addressing the critical need for reliable and timely updates to HD maps to ensure that autonomous vehicles can navigate safely and efficiently in ever-changing environments.

3. Data

An instrumented vehicle was used to capture real-time traffic, location, and infrastructure data, simulating information an AV would receive on the road. The sensors include LiDAR, radar, camera sensors, a gyroscope, an accelerometer, and a MobilEye camera sensor. Fig. 1 shows the sensor configuration on the instrumented vehicle used for data collection. These sensors were selected to provide a comprehensive understanding of the vehicle's environment and to replicate the diverse range of data an AV relies on for navigation and decision-making.

In this paper, only the camera sensor was used to capture the driving scene and roadworks. The reason we chose to use only monocular cameras in this study is mainly due to the low cost and ease of deployment of monocular cameras, which is in line with our intention of pursuing low-cost real-time updates. The test car was equipped with a Grasshopper3® camera operating in the IR spectrum, recording at 4M pixels resolution and up to 90 fps. This high-resolution camera was chosen to ensure the clarity and detail of the images captured, which is critical for accurately identifying roadworks and other road features. The data collected was integrated and stored through a data integration design, ensuring consistency and accuracy. The study involved data from UK motorways, such as M1, M25, and M40. Nine trials were conducted, covering 1062 miles. This extensive coverage provided a diverse dataset for the analysis capturing a variety of road conditions and scenarios that an AV might encounter.

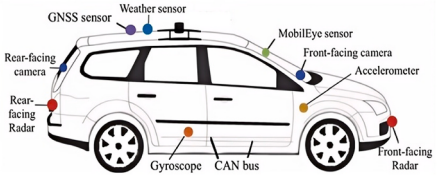

Fig. 1. Instrumented vehicle used for data collection.

A total of 1944 signs representing different kinds of roadworks were collected, including images by means of data augmentation. Specifically, methods such as rotation, flipping, and brightness adjustment were employed to enhance the dataset's diversity and robustness. These augmentation techniques simulate various real-world conditions and perspectives, thereby improving the model's ability to generalise. Additionally, images of speed limit signs, lane lines and speed limit markings which are directly related to roadworks were added to meet the model's contrastive needs. These images, encompassing a total of 1808, were sourced from Google Street View images (Google, 2015), covering diverse conditions, including different dates, weather conditions and driving situations. Virtual road marking images were also incorporated into the dataset, generated using advanced Text-to-Image models, such as DALL·E 3 from OpenAI (Betker et al., 2023). These virtual images capture complex textures and nuances, augment the dataset and introduce variations that may not be present in real-world data. In all, the 1944 roadworks sign images are real-world. Of the additional 1808 images used for contrastive training, 30% are virtual (text-to-image), and 70% are real (including Google Street View images).

Therefore, the dataset encompassed a total of 3752 images under diverse conditions. Information on data distribution and the number of images for each scenario in the dataset are shown in Figs. 2 and 3. These figures provide a clear breakdown of the training and testing datasets (80% for training and 20% for testing) and the conditions under which the images were captured.

Fig. 4 presents some examples from the dataset used, illustrating the variety and complexity of the images.

During the model training process, data augmentation techniques were employed to enhance the model's generalisation ability and robustness. These techniques included image resizing, random rotation, and colour dithering. Such operations increased the diversity of the training data and simulated various real-world variations, such as differing lighting conditions, angles, and sizes.

4. Methodology

This methodology outlines the development of a complete end-to-end system for real-time HD map update based on the CLIP model. This section describes the training process of roadworks signs recognition using the CLIP model, the using of RoadRunner as simulation, the conversion process of roadworks signs position in the image to real world coordinates and finally to the position in the map road network file, and the subsequent update process of these roadworks on HD map.

4.1. CLIP model

The CLIP model utilizes contrastive learning with dual encoders to map images and text into a shared high-dimensional space. This approach maximises similarity between semantically aligned image-text pairs while minimising it for mismatched pairs. The model is trained using the Information Noise Contrast Estimation (InfoNCE) loss function (Oord et al., 2019), expressed as:

$$L = -\log \frac{\exp\left(\text{sim}\left(u, v\right)/\tau\right)}{\sum_{j=1}^{N} \exp\left(\text{sim}\left(u, v_j\right)/\tau\right)} \tag{1}$$

where u and v are the feature representations of the image and the corresponding text, respectively, sim (u, v) is a cosine similarity measure between these two representations, τ is a temperature parameter controlling the smoothing of the distribution, N is the number of negative samples, and j represents the text representations in the negative samples.

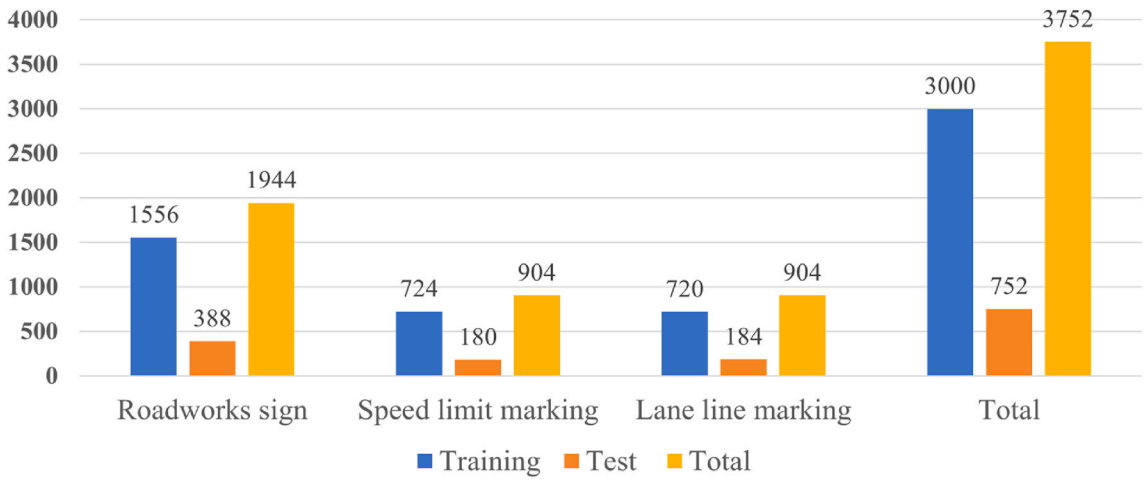

Fig. 2. Information on data distribution.

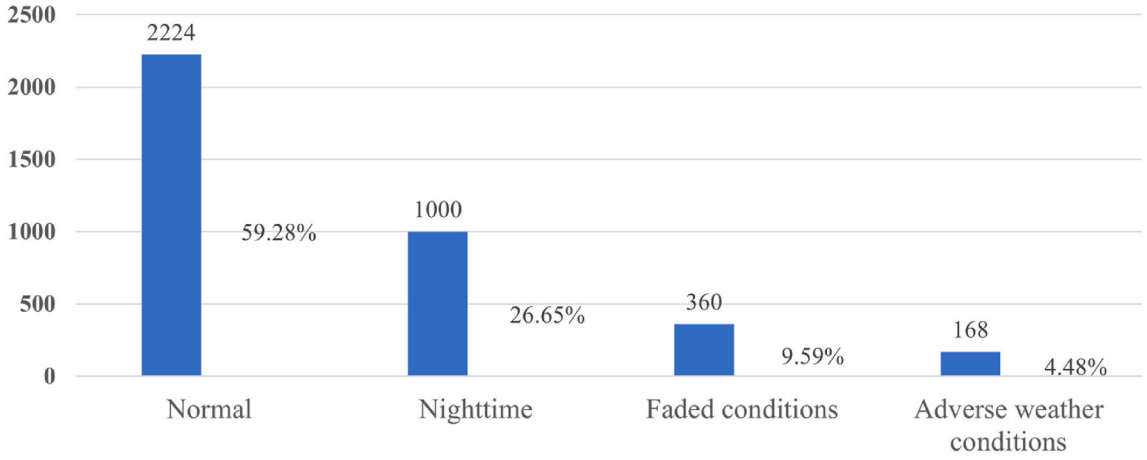

Fig. 3. Number of images for each scenario in the dataset.

Fig. 4. Sample images from the dataset.

This training methodology endows CLIP with powerful multimodal understanding, zero-shot learning, and cross-task generalization capabilities. The model's architecture enables it to form nuanced associations between visual and textual information, enhancing its performance across diverse tasks without task-specific fine-tuning. For instance, CLIP can accurately interpret images and generate detailed textual descriptions, identify objects within various contexts, and perform complex tasks such as image classification and object detection with high precision. These capabilities are particularly beneficial in scenarios where labelled data is scarce or unavailable, allowing the model to generalise from its extensive pre-trained knowledge. Additionally, CLIP's zero-shot learning ability enables it to handle tasks it has not been explicitly trained on by leveraging its broad understanding of visual and linguistic concepts. This flexibility makes CLIP a versatile tool in numerous applications, from autonomous driving to interactive AI systems, significantly reducing the need for extensive task-specific training and facilitating quicker deployment in real-world scenarios.

The proposed approach utilizes a pre-trained CLIP model, specifically the ViT-B/32 variant, which is built upon the Vision Transformer (ViT) architecture. This model incorporates 12 attention heads, enabling sophisticated image and text feature extraction. Inputs are resized with the shorter side to 256 and centre-cropped to 224 $\times$ 224, followed by CLIP normalisation. By leveraging these advanced capabilities, the ViT-B/32 CLIP model provides a strong foundation for the task of roadworks sign recognition, offering enhanced performance potential compared to traditional computer vision methods.

4.1.1. *Training Process*

The pre-trained CLIP model was used as backbone during the training process. Image-text pairs were input into the model, utilising

four different image data types: roadworks sign, roadworks and speed limit sign, speed limit marking and other markings. This diversity enabled the model to better learn the features of different types of markings through contrastive learning. The chosen loss function for fine-tuning is the cross-entropy loss function (Zhang and Sabuncu, 2018):

$$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log\left(\hat{y}_{ij}\right) \tag{2}$$

where N is number of samples, C is number of classifications, y_{ij} is an indicator that if sample i belongs to class j, then $y_{ij} = 1$, otherwise 0. $\hat{y}_{ij}$ is the probability that the model predicts that sample i belongs to category j.

The cross-entropy loss function was selected for its proven efficacy in classification tasks, guiding the model's outputs towards the target classification labels. For optimization, stochastic gradient descent (SGD) was employed with the following parameters: learning rate of 0.001, momentum of 0.9, and weight decay of 0.001. The model was trained for 50 epochs with a batch size of 8. The learning-rate is constant (no warm-up, no early stopping). Data augmentation is applied to training only: horizontal flip ($p = 0.5$), random rotation $\pm 15°$ ($p = 0.5$), and brightness scaling [0.8, 1.2] ($p = 0.5$). Unless stated otherwise, we fine-tune the image encoder while keeping the text encoder frozen and retain the CLIP image transforms. These training configurations were chosen via a Bayesian search to enhance the CLIP's model ability to accurately recognise roadworks signs by closely aligning its outputs with the intended classification labels. Fig. 5 introduces the fine-tune and inference process of the roadworks sign recognition model, illustrating the comprehensive roadworks recognition process. Two different text labels were used during fine-tune and inference. A detailed description of the signs' pattern, colour, and composition is provided when we fine-tune the model while in the prediction stage, we only use the names of signs as labels. For instance, a roadworks sign is labelled as “Roadworks sign” in prediction label and “Roadworks sign with cones” when fine-tuning. Why we do this is that the text description of the signs is more similar to the roadworks image because it contains more detailed information about the roadworks' composition and pattern. This detailed textual description helps the model better understand the relationship between the image and the text, thereby improving its performance. This demonstrates that providing more detailed textual descriptions can enhance the model's ability to learn and accurately identify roadworks. We report epoch-wise training/validation loss and validation F1. For the ViT-B/32 model family, we run six seeds and plot multi-seed mean $\pm 2\sigma$ training and validation loss over 50 epochs. We also show a representative single-run validation-F1 curve. Models are trained for 50 epochs, and results are reported from the final epoch. Fig. 6 summarises the diagnostics: multi-seed mean $\pm 2\sigma$ loss ($n = 6$) and a representative single-run validation-F1 curve.

4.2. RoadRunner simulation

Although we did collect data through instrumented vehicles in real environments, there are difficulties in obtaining high-precision geographic locations with accurate labeling of roadworks in real road environments. Since our goal is to update the HD maps of road sections containing roadworks, in the real world, due to the frequent changes in roadworks, it is difficult to find a road section that satisfies both the need to contain roadworks and the original HD maps, so the validity of our method was verified through the RoadRunner simulator (MathWorks, 2019). RoadRunner allows for the creation and editing of high-precision 3D road networks and driving scenarios, facilitating the accurate simulation of various roadwork situations. By using RoadRunner, developers can generate realistic driving scenarios that include narrow lanes and lane closures, allowing for thorough testing and validation of roadwork detection and HD map updating processes. These simulations help ensure that the HD maps are consistently updated with accurate roadwork information, improving the reliability and safety of AVs in real-world conditions. By using RoadRunner to simulate roadworks detection and update OpenDRIVE files in real-time further ensures the reliability of this approach.

4.3. Conversion of coordinates

After training the model, the next step involves converting the positions of the detected roadworks signs from image coordinates to real-world coordinates, which are essential for updating the HD map.

4.3.1. 2D coordinate acquisition for roadworks images

With the support of the existing HD map and roadworks recognition model, the exact positional coordinates of roadworks in the image can be obtained. This paper uses RoadRunner software to build virtual driving scenarios and generate maps containing roadworks, demonstrating the map update process by modifying its corresponding OpenDRIVE file. In the RoadRunner simulation environment, the bounding boxes for roadworks are automatically obtained by comparing the images before

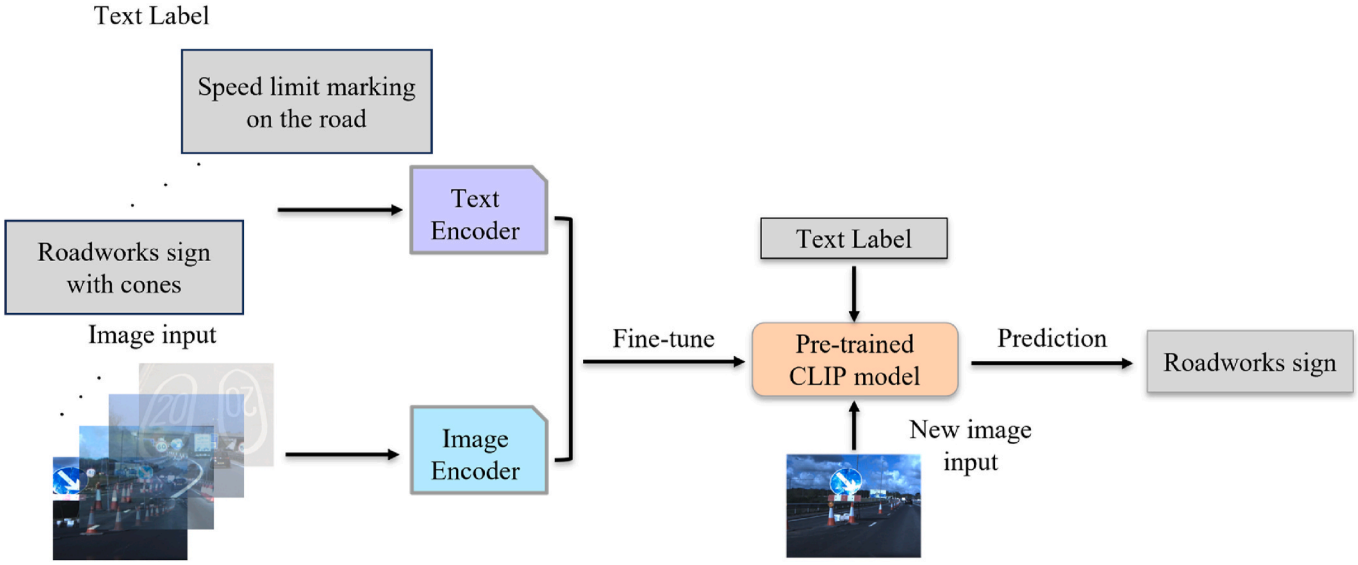

Fig. 5. Flowchart of the roadworks sign recognition model.

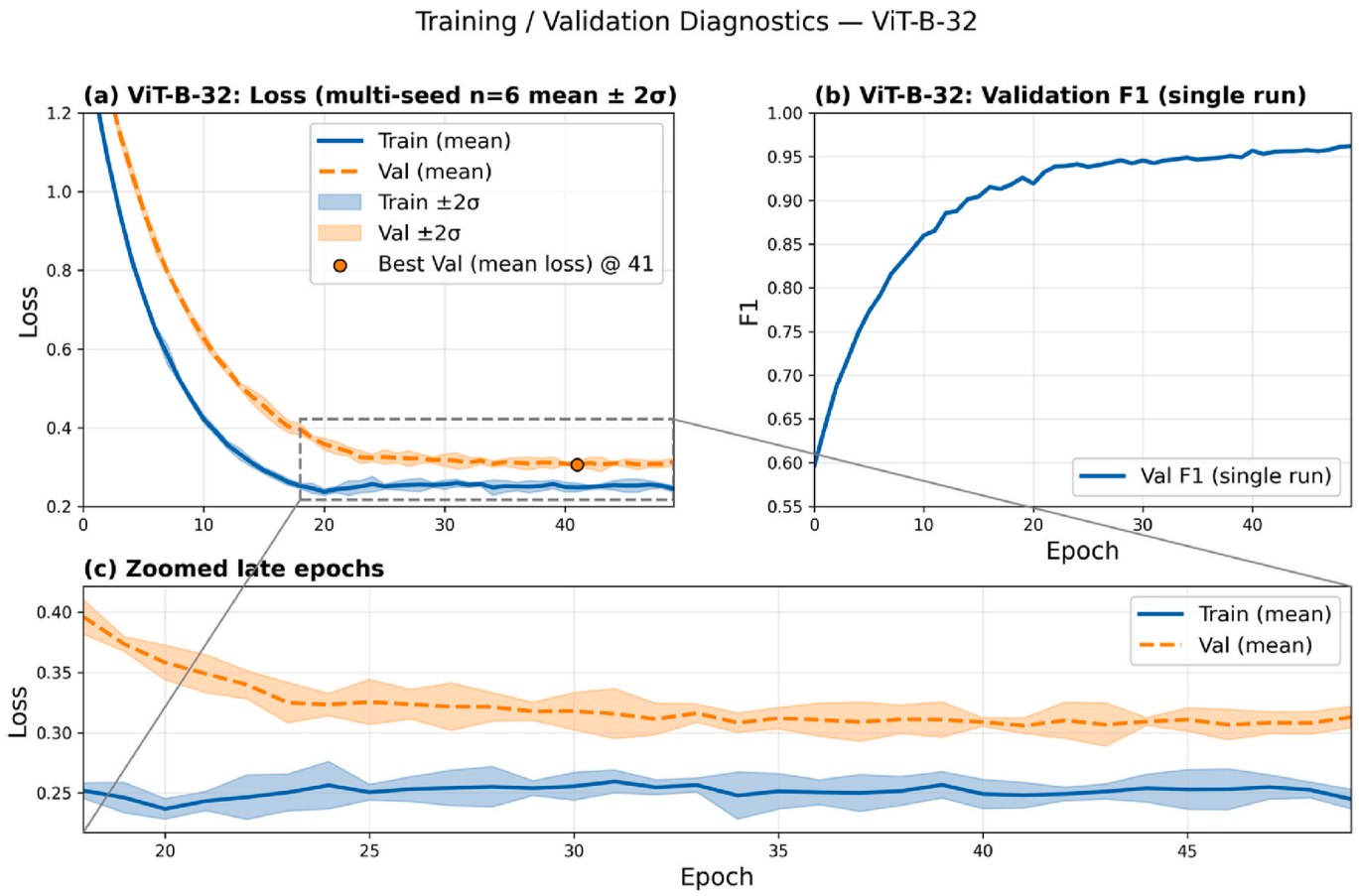

Fig. 6. Training/validation diagnostics (ViT-B/32). (a) Epoch-wise training/validation loss as multi-seed mean ±2σ (n = 6). (b) Validation-F1 for a representative single run. (c) Zoomed late epochs. Trained for 50 epochs; results reported from the final epoch.

Fig. 7. Driving scenarios containing roadworks generated by RoadRunner.

and after inserting the roadworks objects, ensuring precise localisation. The center of the bounding box is then used for subsequent coordinate conversion and map update steps. This automated process eliminates potential labeling errors that might otherwise arise from manual annotation or imperfect model detection, ensuring that coordinate extraction remains highly accurate in the simulation scenario. In real-world deployments, the detection bounding box would be produced by the recognition model, and the accuracy of the extracted coordinates would then depend on the detection performance.

RoadRunner is a specialised software developed by Mathworks for creating and editing high-precision 3D road networks and driving scenarios. It plays an important role in the development and testing of autonomous driving and advanced driver assistance systems (ADAS) (Frey et al., 2024). RoadRunner allows users to quickly build complex road environments and add detailed elements such as traffic signs, signals, and buildings. It supports the import of real-world map data or designing custom scenarios from scratch. In addition, RoadRunner generates industry-standard (e.g., OpenDRIVE) road network description files useable by various simulation platforms and autonomous driving algorithms (Frey et al., 2024). Fig. 7 shows the image of driving scenarios containing roadworks generated by RoadRunner.

When the RoadRunner-generated image containing roadworks is input into the roadworks recognition model, the 2D coordinates of roadworks can be easily obtained. Fig. 8 shows the 2D coordinates of roadworks.

4.3.2. Conversion of roadworks coordinate

To convert the 2D coordinates of roadworks' into useable location information for map updates, they need to be converted to real world coordinates. This process involves several steps:

1. **Camera Calibration:** The intrinsic and extrinsic parameters of the camera are used to establish the relationship between image pixels and real-world measurements. These parameters are typically obtained using a chessboard-based approach, where a known pattern is captured by the camera at different orientations to calculate the camera's intrinsic matrix and distortion coefficients. The intrinsic and extrinsic camera calibration parameters used in our experiments were derived from real-world measurements and standard calibration procedures (e.g., Zhang's chessboard calibration method), ensuring realistic camera settings. The internal reference matrix of the camera (*camera_matrix*), which describes the focal length and optical centre coordinates of the camera and facilitates the conversion from image coordinates to the camera frame, is of the form:

$$\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where f_x and f_y are focal lengths, c_x and c_y are optical center coordinates.

2. **Distortion coefficients**: The camera distortion coefficients describe the radial and tangential distortion of the lens, typically represented as:

$$[k_1, k_2, p_1, p_2, k_3] \tag{4}$$

where k_1 is the first-stage radial aberration coefficient, which mainly corrects radial aberrations close to the center of the light, k_2 is the second-stage radial aberration coefficient, which corrects radial aberrations further away from the center of the light that are not fully corrected by the first-stage coefficients, k_3 is the third-stage radial aberration coefficient, which corrects for higher-order radial aberrations, and is usually used less frequently for very fine correction and p_1 is the first tangential aberration coefficient, which describes aberrations caused by offsets of the sensor and the lens in one direction, p_2 is the second tangential aberration coefficient describing the aberration caused by the offset of the sensor and the lens in the other direction.

3. **External Reference Matrix:** The transformation from the world coordinate system to the camera coordinate system consists of a

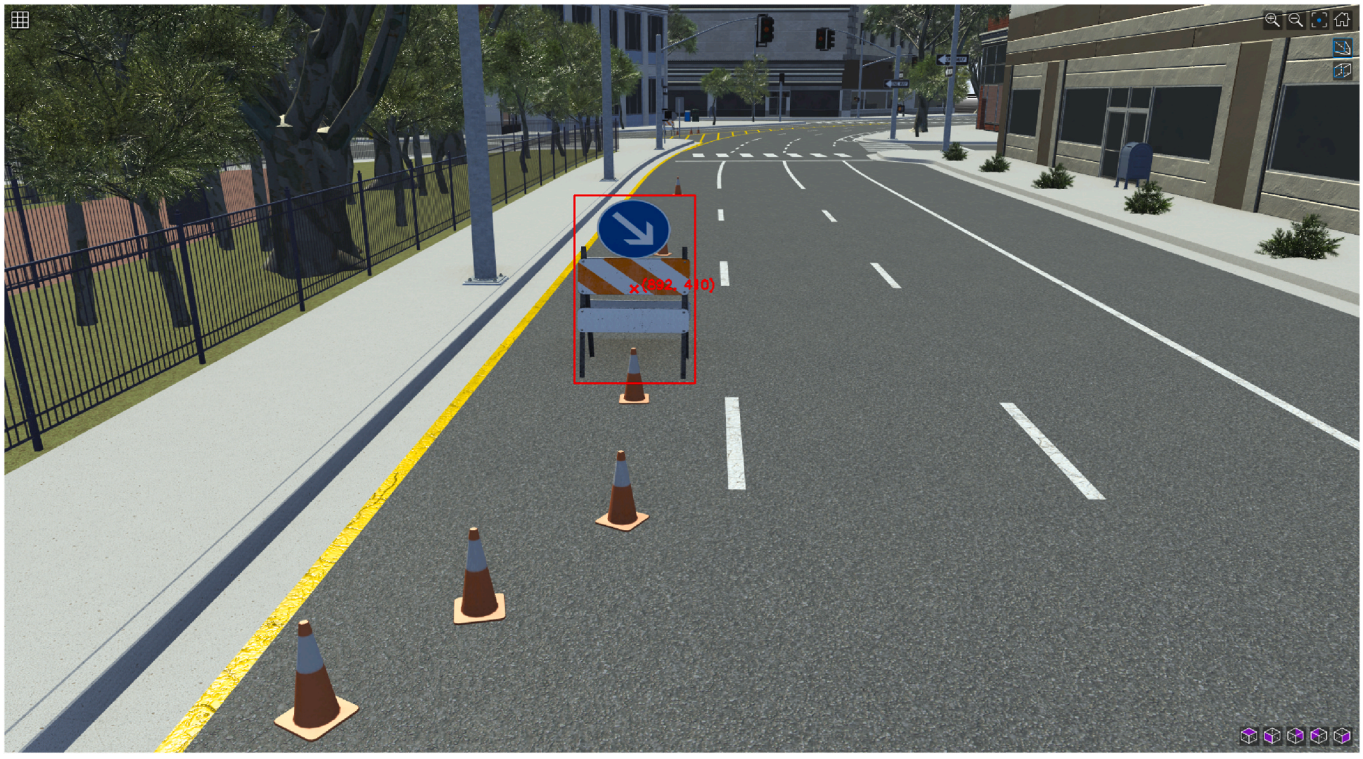

Fig. 8. 2D coordinates of roadworks.

rotation matrix R and a translation vector T:

$$R:\begin{bmatrix}\cos\theta_y\cos\theta_z & -\cos\theta_y\sin\theta_z & \sin\theta_y\\ \cos\theta_x\sin\theta_z+\sin\theta_x\sin\theta_y\cos\theta_z & \cos\theta_x\cos\theta_z-\sin\theta_x\sin\theta_y\sin\theta_z & -\sin\theta_x\cos\theta_y\\ \sin\theta_x\sin\theta_z-\cos\theta_x\sin\theta_y\cos\theta_z & \sin\theta_x\cos\theta_z+\cos\theta_x\sin\theta_y\sin\theta_z & \cos\theta_x\cos\theta_y\end{bmatrix}$$
$$\mathrm{T}:\left[t_x,t_y,t_z\right] \tag{5}$$

Where θ_x,θ_y andθ_z are the rotation angles around the X, Y, and Z axes, respectively. $t_x,\ t_y,t_z$ denote the translation of the camera in the X-axis, Y-axis and the Z-axis respectively.

4. **Coordinate Transformation**: Assuming (μ,υ) are the 2D coordinates of the roadworks sign in the image, the image coordinates are first converted to normalised image coordinates using the camera internal reference matrix:

$$\text{normalised\_coordinates}=\text{inverse}(\text{camera\_matrix})\times[\mu,\upsilon,1]^{\mathrm{T}} \tag{6}$$

The normalised image coordinates are then converted to world coordinates using the camera external reference matrix:

$$\text{camera\_point}=\text{normalized\_coordinates}\times\text{depth} \tag{7}$$

$$\text{world\_coordinates}=\begin{bmatrix}R & T\\ 0 & 1\end{bmatrix}^{-1}\times\begin{bmatrix}\text{camera\_point}\\ 1\end{bmatrix} \tag{8}$$

where depth is the distance of the object in camera's line of sight.

In our approach, the depth is estimated using monocular geometry and the known physical height of the roadworks marker. Given the camera's intrinsic calibration parameters (focal length, principal point) and the measured real-world height of the marker (e.g., a roadworks sign or cone), the distance from the camera to the object is calculated using the pinhole camera model as follows:

$$Z=\frac{f\cdot H_{obj}}{h_{img}} \tag{9}$$

where Z is the estimated depth from the camera to the marker, f is the focal length of the camera in pixels, H_{obj} is the real-world height of the roadworks marker, and h_{img} is the height of the detected marker in the image (in pixels).

The base of the marker in the image is identified, and its coordinates are projected into 3D space using the estimated depth together with the camera's intrinsic and extrinsic parameters. This approach enables approximate depth estimation from a single calibrated camera, provided the real-world object height is known.

5. **Error analysis**: We quantify how measurement and calibration uncertainties affect the monocular depth and the recovered 3D point used for s/t/z_Offset. First-order propagation gives $\Delta Z/\ Z\approx\sqrt{\left[(\Delta f/f)^2+(\Delta H/H)^2+(\Delta h/h)^2\right]}$. Using our calibration and a representative detection: $f\_x=2660.8\ \text{px}, f\_y=2680.18\text{px}, c\_x=1021.87\text{px}, c\_y=677.84\text{px}$; $H=1.50\text{m}$; $h=150\text{px}$; box centre $(u, v)=(814, 854)\text{px}$. Then $Z\approx(2680.18\times1.50)/150\approx26.8\text{m}$, and back-projected camera-frame coordinates are $X\approx((814-1021.87)\times26.8)/2660.82\approx-2.10\text{m}$ and $Y\approx((854-677.84)\times26.8)/2680.18\approx1.76\text{m}$. Taking $\Delta f=5$ px, $\Delta H=0.05$ m, $\Delta h=2$ px yields $\Delta Z/Z\approx3.6\%$ (so $\Delta Z\approx0.97\text{m}$ at 26.8 m). Pixel localisation noise propagates via $X=((u-c\_x)Z)/f\_x$ and $Y=((v-c\_y)Z)/f\_y$; with $\Delta u=\Delta v=1\text{px}$, the lateral terms contribute about $(Z/f\_x)\cdot\Delta u\approx0.010\text{m}$ each before combining with depth uncertainty, giving overall $\Delta X\approx7.6$ cm and $\Delta Y\approx6.5$ cm for this example (depth error dominates at longer ranges). These figures assume calibrated intrinsics, an approximately upright marker with known height, and stable extrinsics; violations (e.g., truncated boxes or pose drift) bias h or the transform and increase error.

4.3.3. OpenDRIVE integration

OpenDRIVE is a standardised file format for describing road networks, containing detailed road attribute information such as geometry, lane widths, pavement markings, traffic signs. To integrate roadworks coordinates into OpenDRIVE, the positional information is described using three parameters: (1) *s*: arc length coordinate on the road centreline, (2) *t*: lateral offset and (3) *z_Offset*: vertical offset. Assuming (x_{rw},y_{rw},z_{rw}) is the location of the roadworks sign the parameters are calculated as:

$$s_{rw}=\cos(hdg)\cdot(x_{rw}-x_0)+\sin(hdg)\cdot(y_{rw}-y_0) \tag{10}$$

$$t_{rw}=-\sin(hdg)\cdot(x_{rw}-x_0)+\cos(hdg)\cdot(y_{rw}-y_0) \tag{11}$$

$$z\_Offset_{rw}=z_{rw}-z_0 \tag{12}$$

These values are then used to update the HD map with the roadworks information in the OpenDRIVE file, ensuring that the map accurately reflects the current road conditions. Fig. 9 shows the complete process of coordinate conversion in the form of pseudocode. It should be noted that, when using the RoadRunner simulator, access to precise vehicle GPS coordinates and accurate camera mounting parameters is limited. For demonstration purposes, placeholder or estimated values were used for these parameters during simulation, which can introduce significant errors in some calculated values, such as z_Offset. As a result, certain z_Offset values in simulation (e.g., those shown in Fig. 13) may appear larger than what would be expected in real-world deployments. In practical applications, these calibration parameters would be obtained from actual vehicle sensors, ensuring that z_Offset and other map attributes remain within realistic and physically reasonable ranges.

This pseudocode outlines the steps required to transform 2D image coordinates into real-world coordinates and then into the format required for updating the HD map in an OpenDRIVE file.

4.3.4. Model training and evaluation

To ensure the CLIP model is effectively trained and evaluated for the task of roadworks sign recognition and HD map updates, the following steps are undertaken:

- Dataset Preparation: The dataset, as described earlier, is split into training and testing sets. The training set consists of 3000 images, while the testing set comprises 752 images, ensuring a robust evaluation. The training and testing sets were carefully curated to ensure they included a diverse range of scenarios, such as different lighting conditions, weather conditions, and various types of roadworks signs and markings. This diversity is essential for training a robust model that can generalise well to new, unseen data.

Algorithm 1 Pseudocode of coordinate conversion

Input known parameters:
camera_matrix = [[fx, 0, cx], [0, fy, cy], [0, 0, 1]]
dist_coeffs = [k1, k2, p1, p2, k3]
rotation_matrix (R) = [[cosθ_y cosθ_z ,-cosθ_y sinθ_z, sinθ_y], [cosθ_x sinθ_z+sinθ_x sinθ_y cosθ_z, cosθ_x cosθ_z-sinθ_x sinθ_y sinθ_z, -sinθ_x cosθ_y], [sinθ_x sinθ_z-cosθ_x sinθ_y cosθ_z, sinθ_x cosθ_z+cosθ_x sinθ_y sinθ_z, cosθ_x cosθ_y]]
translation_vector (T) = [tx, ty, tz]
roadworks_coords = (u, v)
starting point of the road section = (x0, y0, z0)
heading of the road section = hdg

% Convert image coordinates to normalized image coordinates
normalized_coordinates=inverse (camera_matrix) × [u, v, 1] ^ T

% Convert normalized image coordinates to world coordinates (x_rw, y_rw, z_rw)
camera_point = normalized_coordinates × depth
world_coordinates = $[[R, T], [0, 1]]^{-1}$ × [[camera_point], [1]]

% Convert world coordinates to (s_rw, t_rw, z_Offset_rw)
s_rw = cos (hdg) · (x_rw-x_0) + sin (hdg) · (y_rw-y_0)
t_rw = -sin (hdg) · (x_rw-x_0) + cos (hdg) · (y_rw-y_0)
z_Offset_rw = z_rw - z_0

END coordinate conversion

Fig. 9. Pseudocode of the coordinate conversion process.

- Training Process: The pre-trained CLIP model is fine-tuned using the prepared dataset, focusing on contrastive learning to map image-text pairs into a shared representation space. The fine-tuning process leverages the cross-entropy loss function for its proven efficacy in classification tasks, guiding the model's outputs towards the target classification labels. The optimization process employs stochastic gradient descent (SGD) with the following parameters: learning rate: 0.001, momentum: 0.9, weight decay: 0.001. The model is trained for 50 epochs with a batch size of 8. These training configurations were chosen via a Bayesian search to optimize the model's performance. The training process aims to enhance the CLIP model's ability to accurately recognise roadworks signs by closely aligning its outputs with the intended classification labels. Fig. 5 provides a visual representation of the entire training process, illustrating the comprehensive roadworks recognition workflow.
- Evaluation Metrics: The model's performance is evaluated using precision, recall, and F1-score, providing a comprehensive assessment of its accuracy and robustness. These metrics help quantify the model's ability to correctly identify roadworks signs (Precision), its ability to detect all relevant roadworks signs (Recall), and the balance between precision and recall (F1-score).
- Results Analysis: The results are analysed to identify areas of improvement and to fine-tune the model further. This analysis helps in understanding the model's strengths and weaknesses in real-world scenarios. This analysis helps in understanding the model's strengths and weaknesses in real-world scenarios, enabling iterative improvements to enhance its performance.

4.4. HD map update process

This section details the step-by-step process for updating HD maps using the CLIP model and OpenDRIVE file modifications. The goal is to ensure real-time updates of roadworks information on HD maps, thereby improving the navigation capabilities of autonomous vehicles.

The process begins with roadworks detection using the trained CLIP model. The model processes images captured by the vehicle's camera to identify and classify roadworks signs, assigning 2D coordinates in the image frame to each detected sign. These 2D coordinates are then transformed into real-world coordinates using the vehicle's GPS data and camera parameters. Camera calibration is performed to obtain the intrinsic and extrinsic parameters of the camera, including focal lengths, optical centre coordinates, and distortion coefficients. The transformation process involves converting image coordinates to normalised image coordinates, and subsequently to world coordinates using the camera's internal and external reference matrices.

Once the real-world coordinates of the roadworks signs are determined, the next step is to update the OpenDRIVE file, which contains the road network data. As shown in Fig. 10, using an XML parsing library in Python, the OpenDRIVE file is loaded and parsed to access specific elements required for modification. A new roadworks object is created with the obtained real-world coordinates $(s_{rw}, t_{rw}, zOffset_{rw})$ and this

Algorithm 2 Pseudocode of the OpenDRIVE file update

```
Input OpenDRIVE file:
xml_tree = load_xml("path_to_opendrive_file.xodr")
root = xml_tree.getroot ()
% Create new roadworks object
new_object = create_new_object (id = "new_id", name = "Roadworks", s = s_rw, t= t_rw, z_Offset = z_Offset_rw)
% Add new object to the road section
road = root.find (".//road")
objects_section = road.find ("objects")
objects_section.append (new_object)
% Save updated OpenDRIVE file
save_xml("path_to_opendrive_file_updated.xodr", xml_tree)
END OpenDRIVE file update
```

Fig. 10. Pseudocode of the OpenDRIVE file update process.

object is added to the appropriate section of the OpenDRIVE file. The roadworks object includes attributes such as position, size, and type of roadworks. After making these modifications, the updated OpenDRIVE file is saved to reflect the new roadworks information.

The pseudocode for the OpenDRIVE file update process is as follows:

The entire process from roadworks detection to HD map update is integrated into a seamless end-to-end update pipeline. The pipeline's input is the image data captured by the camera, and it directly outputs the updated OpenDRIVE file. This integration ensures that the conversion process of roadworks coordinates and the updating of the HD map are efficiently managed within a single workflow. The pipeline is optimised to enhance the speed and efficiency of the update process. By using only one monocular camera as the sensor, the update cost is significantly reduced. The system is designed to operate in real-time, ensuring that the HD map is continuously updated with the latest roadworks data.

The complete process, from capturing roadworks images to updating the HD map with the modified OpenDRIVE file is presented in Fig. 11. The system is designed to operate in real-time, ensuring that the HD map is continuously updated with the latest roadworks data.

By systematically detailing each step, this methodology ensures that the HD map update process is thorough, accurate, and efficient, thereby improving the overall functionality of AVs in dynamic road environments. By integrating advanced data processing techniques with cutting-edge AI models, this end-to-end system for real-time HD map updates aims to significantly improve the safety and efficiency of autonomous driving.

5. Results

This section presents the performance and accuracy of the roadworks sign recognition model and the HD map update process. It includes

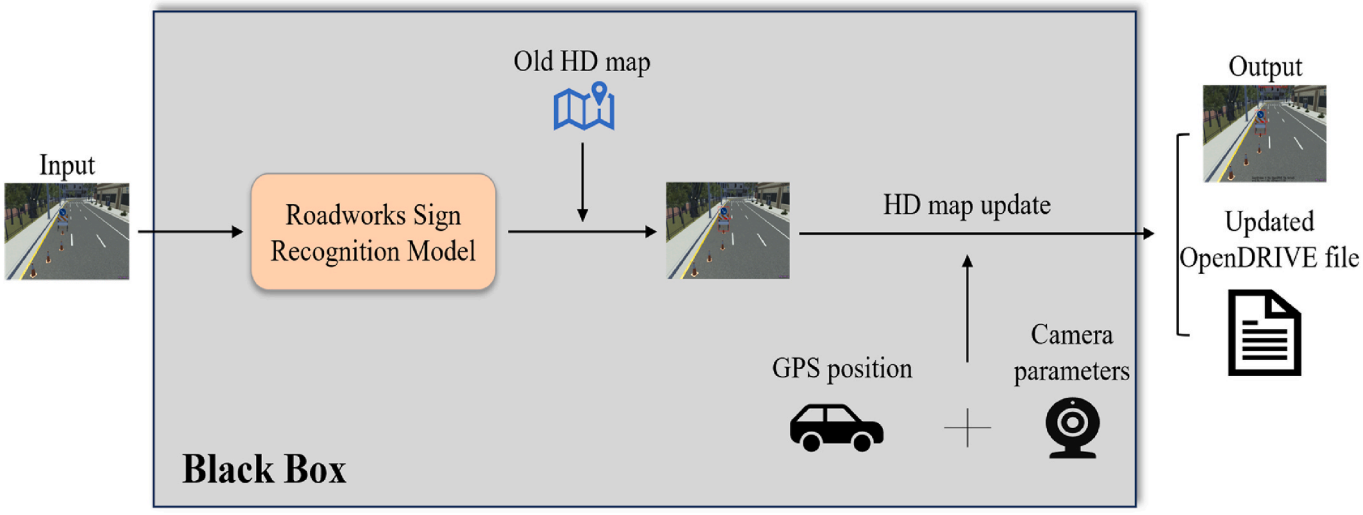

Fig. 11. Complete HD Map update process.

Fig. 12. Test results of roadworks sign recognition model.

Input	Model prediction with confidence level	Error tag	Note
	Speed limit marking 76.3	Confused	Confused with Speed limit marking
	Roadworks sign 23.1	Missed	Miss detected due to occlusion and low visibility
	Roadworks sign 45.7	Missed	Miss detected due to occlusion

Fig. 13. Qualitative examples of model outputs and common failure modes.

detailed evaluations and comparisons with state-of-the-art methods, as well as implementation details. Unless otherwise stated, results are real-world; we annotate environment and ground-truth sources: (real-world)/(simulation); GT: GPS for localisation metrics, human-labelled categories for recognition; RoadRunner is used only for simulation illustrations.

5.1. Roadworks sign recognition model performance

The roadworks sign recognition model was tested using both real-world and AI generated roadworks images. Fig. 12 shows the detection results, with captions indicating the most similar results as determined by the model from the provided labels. The test results demonstrate the model's capability to accurately recognise both real-world and AI-generated roadworks images. This capability is attributed to the mixed dataset used during training, which included virtual road marking images and emphasised their positive influence.

To evaluate the performance of the model, three metrics (Precision, Recall and F1-score) were calculated. These metrics are critical as they measure the model's accuracy in identifying roadworks signs (i.e. Precision), its ability to detect all relevant signs (i.e. Recall), and the balance between Precision and Recall (i.e. F1-score). Table 1 reports the evaluation metrics for roadworks detection, including comparisons with the latest YOLO model, OWL-ViT v2 (Minderer et al., 2022) and GLIP (Li

Table 1
Evaluation metrics for the roadworks sign recognition model (real-world).

	Precision	Recall	F1-score	AP
Roadworks sign **(Ours)**	97.10% ± 0.35	**92.51% ± 0.60**	94.68% ± 0.42	92.11% ± 0.42
Roadworks sign (fine-tuned YOLO v11)	93.33%	73.68%	82.30%	77.63%
Roadworks sign (fine-tuned OWL-ViT v2)	**98.32%**	92.06%	**95.12%**	**93.87%**
Roadworks sign (fine-tuned GLIP)	96.14%	90.81%	93.46%	90.35%
Speed limit marking	81.77% ± 1.10	93.26% ± 0.97	87.65% ± 0.88	81.57% ± 1.03
Lane line markings	98.81% ± 0.25	94.87% ± 0.43	96.79% ± 0.31	95.33% ± 0.35

GT: human labels.

Table 2
F1-score comparison under different conditions with 95% CI (real-world).

	Normal	Nighttime	Adverse weather	Faded condition
Roadworks sign **(Ours)**	**96.38% [95.84%, 96.79%]**	**94.46% [93.58%, 95.22%]**	**91.76% [88.95%, 93.26%]**	**96.32% [95.10%, 97.12%]**
Roadworks sign (fine-tuned YOLO v11)	85.81%	79.81%	80.66%	82.92%

GT: human labels.

et al., 2022). For our models, results are given as mean ± standard deviation across six seeds (n = 6), trained for 50 epochs and reported at the final epoch. Baselines are single-run results. Table 2 gives the F1-Score performance comparison under different conditions. For comparative evaluation, we include not only a fine-tuned YOLO v11 detector, but also recent open-vocabulary detectors, OWL-ViT v2 and GLIP. Detection accuracy and inference time per image are reported for all baselines to enable a comprehensive analysis of the trade-off between precision and processing speed in the context of HD map updates. In our experiments, the average inference time per image is 0.87 s for fine-tuned YOLO v11, 1.54 s for our proposed method, 2.93 s for fine-tuned OWL-ViT v2, and 3.15 s for fine-tuned GLIP. These results indicate that, while open-vocabulary models such as OWL-ViT v2 and GLIP achieve comparable or slightly higher detection accuracy, our method offers a more favorable balance between speed and precision, substantially reducing processing latency compared to state-of-the-art open-vocabulary detectors. This efficiency is especially important for real-time or near-real-time HD map updating applications. The model achieves a precision of 97.12% for roadworks sign recognition, indicating its effectiveness in correctly identifying roadworks signs without false positives. From the table, we achieved better results than YOLO v11 when fine-tuning using the same amount of data. For roadworks sign, they did not exist in the pre-trained YOLO model so that we need to re-training the model with larger dataset to get a better performance and manual labelling these images are time-costing and expensive. In addition, roadworks usually consist of signs and cones which increase the difficulty of training and there are few available datasets of roadworks we can use to train a YOLO model. However, CLIP, as an open vocabulary model, has rich natural language information, allowing the model to perform a certain degree of generalization even when it encounters signs that are not explicitly labelled in the training set which means we can use fewer datasets to fine-tune the model to get a better performance in this specific task. While YOLO excels in target detection, its main strength lies in dealing with the closed-set target detection problem. CLIP's capability in multimodal fusion allows us to more flexibly combine image and textual information to support finer-grained semantic distinctions and open-ended category extensions. This capability makes the subsequent transformation from the image plane to real-world more robust, providing the necessary semantic and geometric information to support real-time updating of high-precision maps.

However, some roadworks signs are misidentified as speed limit marking due to the simultaneous appearance of these signs in the training data. Also, the inclusion of multiple road marking types in other markings can lead to confusion and errors in the model's ability to differentiate between these markings and roadworks sign. Table 3 shows the comparison of our model with the state-of-the-arts. The effectiveness of the detection model has reached the level of existing state-of-the-arts, but the amount of data and computational resources needed is smaller, reflecting the superiority of the model for roadworks sign recognition.

Table 4 reports per-class detection performance (average precision, AP) together with 95% confidence intervals obtained by image-level bootstrap (B = 1000). AP is computed under the COCO convention (AP@ [0.50:0.95]); the CI bounds are the 2.5% and 97.5% percentiles of the bootstrap distribution and therefore quantify sampling uncertainty across images in the test set. For the roadworks sign class, we compare our method against a closed-vocabulary baseline (fine-tuned YOLOv11) and two open-vocabulary detectors (OWL-ViT v2 and GLIP); for Speed limit marking and Lane line markings the table reports our method only. All methods were evaluated under comparable settings (same input resolution, confidence/IoU thresholds and prompt lists for open-vocabulary models). In addition, we summarise the per-stage timing breakdown for our full pipeline and reports detection times for comparative detectors. Our pipeline's per-image end-to-end latency is 1.54s, composed of about 1.25s for roadworks detection (model inference including in-process pre/post steps), 0.03s for coordinate conversion (projection/back-projection and intrinsic/extrinsic transforms) and 0.26s for the OpenDRIVE update (local XML parse/serialise and disk write). By contrast, third-party detectors report detection-only times of 0.87s (YOLOv11, fine-tuned), 2.93s (OWL-ViT v2) and 3.15s (GLIP). YOLOv11 is optimised for real-time inference and thus has low latency, whereas OWL-ViT v2 and GLIP trade higher compute for stronger open-vocabulary/generalisation performance. Our pipeline provides a practical balance between accuracy and utility: it combines robust open-vocabulary detection with immediate, map-ready outputs, enabling accurate HD-map updates and scene understanding with a single-image end-to-end latency of 1.54s.

The comparison shows that the model developed in this paper achieves competitive performance with significantly less data and computational resources, highlighting its efficiency and practicality for real-world applications.

Fig. 13 presents three representative road scenes with the model's top 1 prediction and confidence level. **Top:** a roadworks layout with multiple cones and co-occurring speed-limit signs; the model occasionally confuses roadworks sign with speed-limit sign due to similar shapes/colours and spatial proximity—an appearance-level ambiguity rather than a calibration issue. **Middle:** a narrowed lane where the lead vehicle partially occludes the sign; truncation/occlusion reduces visible features and can cause misses or low confidence, especially when the visible height falls into the small-object regime. **Bottom:** an overpass scene where relevant signage appears small and distant near the vanishing point; limited pixel support leads to confusion with other markings. Together, these panels complement the quantitative results by illustrating when errors arise—(i) **confusion under co-occurrence**, (ii) **occlusion/truncation**, and (iii) **small-scale/long-range**—and motivate remedies such as integrating local context (cones/temporary layout

Table 3
Comparison with the State-of-the-art models (real-world).

	Precision	Recall	F1-Score
Road marking in VPGNet (Lee et al., 2017)	/	83.0%	/
Road marking in (Zhang et al., 2021)	/	90.2%	/
CeyMo (Jayasinghe et al., 2022)		/	90.6%
Road marking in (Gong et al., 2024)	86.31%	88.56%	85.20%
Traffic sign in (Zhang et al., 2025)	96.6%	94.6%	96.8%
Lane marking in **This work**	**98.84%**	**95.00%**	**96.90%**

GT: human labels.

Table 4
Per-class AP and 95% confidence intervals (real-world).

Class	Method	AP (mean)	95% CI (bootstrap)
Roadworks sign	**Ours**	**92.11**	**[91.46, 92.76]**
	YOLO v11	77.63	[75.98, 79.20]
	OWL-ViT v2	93.87	[93.18, 94.52]
	GLIP	90.35	[89.62, 91.03]
Speed limit marking	**Ours**	**81.57**	**[80.18, 82.96]**
Lane line marking	**Ours**	**95.33**	**[94.96, 95.72]**

Table 5
Ablation results in different virtual dataset proportion and text prompt.

Setting	Precision	Recall	F1-Score
Fine-tuned CLIP only with real-world dataset	88.03%	79.85%	83.74%
Fine-tuned CLIP with 25% virtual dataset	89.41%	82.05%	85.64%
Fine-tuned CLIP with 50% virtual dataset	92.87%	86.29%	89.54%
Fine-tuned CLIP with 75% virtual dataset	95.36%	90.08%	92.66%
Fine-tuned CLIP with 100% virtual dataset	97.12%	92.47%	94.73%
Text prompt: Roadworks sign	96.43%	88.72%	92.38%
Text prompt: Roadworks sign with cones	97.12%	92.47%	94.73%
Text prompt: Speed limit marking	80.24%	83.69%	81.94%
Text prompt: Speed limit marking on the road	81.77%	93.26%	87.65%
Text prompt: Lane line marking	96.01%	92.77%	94.46%
Text prompt: White Lane line marking	98.81%	94.87%	96.79%

cues), temporal evidence across frames, and high-resolution crops for small targets.

5.2. Ablation study

To validate the effectiveness and robustness of the proposed method, ablation experiments were performed to evaluate the impacts of synthetic data proportions, text-prompt selection, and hyper-parameter optimization, with detailed results summarised in Tables 4 and 5.

Table 5 presents the performance variations with different proportions of synthetic (virtual) data. Results indicate clear performance improvements as the proportion of synthetic data increases, with the optimal F1-score (94.73%) achieved at 100% synthetic data. Additionally, the influence of text prompt specificity was assessed, showing that a more descriptive prompt ("Roadworks sign with cones") outperforms a generic prompt ("Roadworks sign").

Table 6 illustrates the sensitivity of model performance to key hyper-parameters identified through Bayesian optimization. The learning rate demonstrated the strongest influence on performance, with notable accuracy degradation when substantially increased or decreased. Variations in momentum and weight decay produced only minor effects, whereas increasing the batch size significantly (from 8 to 32) resulted in performance deterioration, underscoring the suitability of smaller batch sizes for the dataset and task considered.

Overall, these ablation analyses confirm the benefits of synthetic data augmentation, appropriate prompt engineering, and carefully selected hyper-parameters in achieving robust and accurate roadworks detection for real-time HD map updates.

5.3. Map update accuracy

Our aim is to use the position marked by the roadworks sign as the starting point of roadworks, updating relevant information on the original HD map accordingly. To verify the accuracy of the map update, the real-world coordinates of the roadworks signs obtained during video collection were used as ground truth. These coordinates were derived by precisely recording the vehicle's GPS position at the timestamps when the vehicle passed the roadworks. Map matching against pre-existing HD maps further ensured high positional accuracy for these ground-truth measurements.

Subsequently, these independently measured real-world coordinates were compared with the coordinates computed by our roadworks sign recognition model, combined with the vehicle's GPS data and calibrated camera parameters. Positional accuracy error, quantified as Euclidean distance and Root Mean Squared Error (RMSE), was calculated to assess map update accuracy. Roadworks were encountered only on three of the nine motorways in our real-world trials; consequently, quantitative results for HD map update accuracy and localisation error are reported for these three roadworks instances. Three roadworks signs located between the A512 and the M1 motorway were specifically chosen for this accuracy evaluation. Fig. 14 illustrates the actual coordinates, computed coordinates, and accuracy errors for these signs.

In real-world motorway trials we encountered roadworks on only three of nine motorways; we therefore report these as a pilot validation. For the three instances, the planar localisation errors were 1.817 m, 1.136 m and 2.121 m, and the route-level RMSEs were 1.020 m, 0.629 m and 1.181 m (mean $\pm$ std: 1.69 $\pm$ 0.50 m and 0.94 $\pm$ 0.28 m, respectively). Under the calibrated monocular setup, the depth error derived from first-order propagation was approximately 0.97 m, 0.92 m and 1.01 m for the three cases (mean $\pm$ std: 0.97 $\pm$ 0.05 m). These pilot results are promising but limited in scope and will expand validation with more extensive, seasonally diverse motorway data.

To further verify the authenticity and validity of the map update pipeline, road network information containing roadworks segments was downloaded through OSM (Open Street Map). The.osm file was converted into an OpenDRIVE file which was then used to add and modify the roadworks information through SUMO (Simulation of Urban Mobility) (Eclipse SUMO, 2017).

Fig. 15 provides a step-by-step visualisation of the complete pipeline for roadworks detection and HD map update. Specifically, the four panels illustrate:

- **(a) Input image (no roadworks):** The original simulated road scene before any roadworks elements are present.
- **(b) Scene with roadworks inserted:** The same road segment after adding roadworks signs and cones in the simulation environment.
- **(c) Detection overlays and pixel geometry:** Output of the roadworks detector on the same frame, showing the bounding box, the pixel centre (u, v), the pixel height h used for monocular depth, and the estimated distance Z in metres.
- **(d) OpenDRIVE update:** Visualisation of the coordinate conversion and HD-map update, presenting the written OpenDRIVE fields $(s, t, zOffset)$ in metres and confirming that the roadworks object is inserted at the computed location.

This four-panel figure enables a clear understanding of the transformation from the initial input image, through detection and coordinate computation, to the final HD map update.

RoadRunner was used to create a driving scene with roadworks, exports its OpenDRIVE file (excluding roadworks information), input it into the pipeline, and finally output the visualisation. The top of the resulting map includes a label indicating the detection of roadworks sign by the recognition model, while the bottom shows the conversion and calculation of the roadworks signs from 2D pixel coordinates in the pipeline to the location information in the OpenDRIVE file. The final OpenDRIVE file generated includes the added information about the roadworks sign.

5.4. Implementation details

The entire HD map update pipeline, from the training of the roadworks sign recognition model to the inference, was performed in a high-performance computing environment equipped with AMD Ryzen 5 3600 6-Core Processor and an NVIDIA RTX 4070 graphics card with 32 GB RAM. In terms of inference performance, a single-image pass takes 1.54s end to end within the processing pipeline, comprising 1.25s for roadworks detection (model inference including in-process pre/post steps), less than 0.03s for coordinate conversion, and 0.26s for the OpenDRIVE update (XML parse/serialise). Unless otherwise stated, timings refer to on-device/on-server processing and exclude external I/O and any vehicle–cloud network transfer. The OpenDRIVE time includes local XML parsing/serialisation and local disk write on the benchmarking machine; camera capture and dataset reads are amortised and not counted unless indicated. The figure of 5.15 fps denotes steady-state throughput with a warmed model processing sequential images; it should not be interpreted as per-image latency. Our reported timings cover the processing segment (detection, coordinate conversion, and

Table 6
Ablation results in different hyper-parameters.

Config	Learning Rate	Momentum	Weight Decay	Batch Size	AP (%)
Baseline	0.001	0.9	0.001	8	92.14
LR-Down	0.0005	0.9	0.001	8	91.37
LR-Up	0.005	0.9	0.001	8	89.56
Mom-Up	0.001	0.95	0.001	8	92.08
Mom-Down	0.001	0.85	0.001	8	91.63
WD-Up	0.001	0.9	0.005	8	91.29
WD-Down	0.001	0.9	0.0005	8	91.85
BS-Up	0.001	0.9	0.001	16	91.44
BS-Up	0.001	0.9	0.001	32	90.07

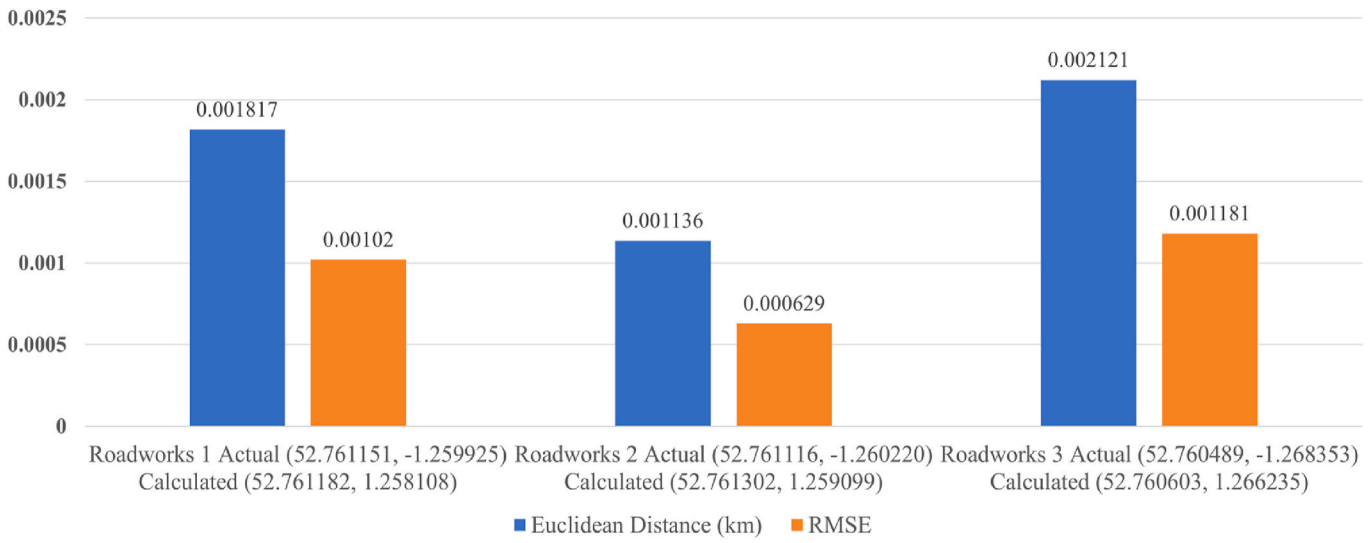

Fig. 14. Positioning accuracy error of roadworks sign (real-world).

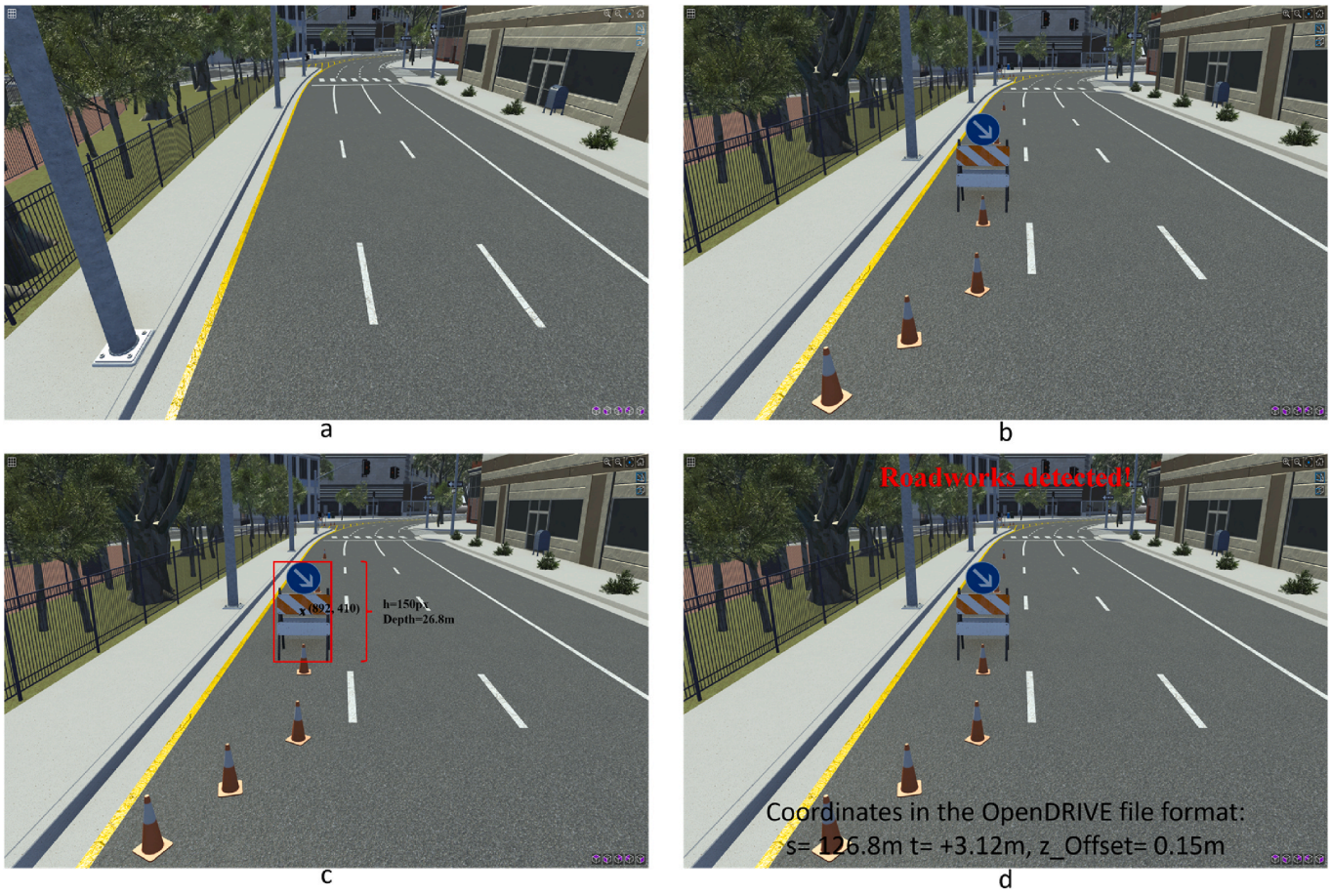

Fig. 15. End-to-End Roadworks Detection and Map Update pipeline (Simulation).

OpenDRIVE update). In a vehicle to cloud deployment, end-to-end delay additionally depends on capture/encode, uplink bandwidth, server-side queueing, processing placement (on-vehicle, edge, or central), and distribution/refresh. As these factors are deployment-specific, we report them qualitatively and present 1.54 s as the processing budget available to such a pipeline. The traditional HD map update relies on a long cycle and lagging information, which can't reflect the road changes in a timely manner, while the real-time update of HD maps can quickly respond to dynamic changes such as road construction, improve traffic fluency, safety, reduce congestion and bring significant economic benefits. For autonomous driving systems, timely reflection of roadworks, temporary closures or other changes in HD maps can provide more accurate information about the environment to help autonomous vehicles better plan their paths, choose lanes and perform maneuvers, reducing

decision-making errors caused by lagging map data. Since autonomous driving systems rely on HD maps as an important complement to environment awareness, rapidly updating map data can reduce the discrepancy between the map and the real-time scene, thus improving the robustness of the vehicle in complex or dynamic scenarios.

6. Discussion

Our results show that a lightweight, modular pipeline can deliver metre-level localisation for roadworks together with strong per-class detection while remaining practical for deployment. Under the calibrated setup, the per-image end-to-end latency is 1.54 s, composed of 1.25 s for detection including in-process pre- and post-processing, under 0.03 s for projection and coordinate conversion, and 0.26 s for OpenDRIVE updates. These timings refer to on-device or on-server processing and exclude external I/O; the reported 5.15 fps denotes warmed sequential throughput rather than single-pass latency. This balance of accuracy and immediacy maps well to incremental HD-map refresh, where rapid roll-out and interoperability with existing assets are priorities.

In the broader context, recent end-to-end and vectorized HD-map approaches infer bird's-eye-view structures and output polylines directly, demonstrating impressive city-scale automation when large multi-sensor datasets and substantial compute are available (Kim et al., 2021; Xia et al., 2025). By contrast, our design separates detection, monocular depth and projection, and OpenDRIVE serialisation. This modularity simplifies integration with legacy production stacks and OpenDRIVE-based tooling without retraining large models, and it enables distributed, small-batch updates when data and computing resources are constrained. Open-vocabulary detectors such as OWL-ViT v2 and GLIP provide stronger zero-shot generalisation at higher inference cost, whereas real-time families like YOLO tend to offer lower latency with closed vocabularies unless fine-tuned; our system occupies a pragmatic middle ground by coupling robust detection with immediate 3D and map-ready outputs (Minderer et al., 2022; Li et al., 2022).

There are important deployment implications. Using a single monocular camera reduces hardware cost and installation complexity, and representing updates in OpenDRIVE eases downstream consumption by mapping and simulation tools already using this exchange format. In practice, projection and serialisation are negligible compared with detection, so optimization should prioritise detector acceleration through mixed-precision or quantisation and compiled inference, together with asynchronous or batched map updates to amortise write costs. These choices align with the goal of rapid, field-deployable HD-map refresh rather than city-scale offline reconstruction (Kim et al., 2021; Chen et al., 2023).

Limitations are integrated with these findings. Real-world evidence is a pilot: during motorway runs, roadworks appeared on three of nine motorways, yielding three useable instances. We therefore provide per-instance planar error, route-level RMSE and depth-error magnitudes and temper claims accordingly. Monocular depth depends on known object height and stable calibration, so truncation, occlusion or large pose changes can bias depth and lateral offsets; future work will broaden trials across seasons and sites, and add multi-sensor fusion such as LiDAR-to-image projection to reduce long-range sensitivity while maintaining OpenDRIVE-first interoperability for adoption in brownfield deployments (Chen et al., 2023; Xia et al., 2025).

7. Conclusion

In this paper, a new, real-time, and cost-effective pipeline for updating HD maps with roadworks information, utilising a monocular camera and Visual-Language Models was proposed. This pipeline consists of three main components: a roadworks sign recognition model, a conversion process for roadworks sign coordinates, and the subsequent updating of the OpenDRIVE file. The roadworks sign recognition model demonstrates exceptional performance, achieving a 97.12% recognition rate for signs and a RMSE of less than 1.2m for sign position accuracy. Additionally, the pipeline processes a single image input in just 1.54 s. Compared to traditional HD map update methods, this pipeline offers significant advantages in terms of cost-efficiency, accuracy, and real-time performance. These attributes underscore its value and significance for the application of HD maps in autonomous vehicles, enhancing their capability to navigate and respond to dynamic road conditions effectively.

While the proposed pipeline for roadworks detection and HD map updating presents a significant advancement, it is at an initial stage and faces several limitations that require future improvements. One major limitation is the accuracy of depth information captured by monocular cameras, which can lead to errors in navigation data and camera parameters, impacting the precision of detected roadworks positions. Addressing this challenge is crucial for improving the overall reliability of the system. There is also a need to validate how the updated OpenDRIVE files affect the planning modules of AVs. Conducting simulations to ensure the reliability and effectiveness of the updates is essential. Furthermore, the current system lacks the capability to capture detailed and specific roadworks information, such as lane changes and diversions. Future efforts should focus on incorporating more detailed roadworks information into HD maps. This could involve integrating additional sensors or leveraging advanced image processing techniques to identify and classify more complex roadworks scenarios. By providing more comprehensive roadworks data, the system can enhance the decision-making and navigation for AVs. While the proposed pipeline presents a significant advancement in real-time HD map updating for autonomous vehicles, addressing its current limitations through further research and development will be essential.

In summary, while the proposed pipeline presents a transformative approach to real-time HD map updating for autonomous vehicles, further research and development will be essential to overcome its current limitations. The promising results and potential improvements make this pipeline a valuable foundation for future advancements in autonomous vehicle navigation technology.

CRediT authorship contribution statement

Shaofan Sheng: Writing – original draft, Methodology, Data curation, Conceptualization. **Nicolette Formosa:** Writing – review & editing, Supervision, Conceptualization. **Yuxiang Feng:** Writing – review & editing, Methodology, Formal analysis. **Mohammed Quddus:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shaofan Sheng reports writing assistance was provided by National Highways, Birmingham, UK. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Bao, Z., Hossain, S., Lang, H., Lin, X., 2023. A review of high-definition map creation methods for autonomous driving. Eng. Appl. Artif. Intell. 122, 106125.

Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., Manassra, W., 2023. Improving image generation with better captions. Computer Science 2 (3), 8. https://cdn.openai.com/papers/dall-e-3.pdf.

Chen, M., Liu, P., Zhao, H., 2023. LiDAR-camera fusion: dual transformer enhancement for 3D object detection. Eng. Appl. Artif. Intell. 120, 105815.
Chowdhury, S., Soni, B., 2025. ENVQA: improving visual question answering model by enriching the visual feature. Eng. Appl. Artif. Intell. 142, 109948.
Eclipse SUMO - Simulation of Urban MObility, 2017. Eclipse SUMO - Simulation of Urban MObility [online] Available at: https://eclipse.dev/sumo/.
Formosa, N., Quddus, M., Singh, M.K., Man, C.K., Morton, C., Masera, C.B., 2024. An experiment and simulation study on developing algorithms for cavs to navigate through roadworks. IEEE Trans. Intell. Transport. Syst. 25 (1), 120–132.
Frey, J., Patel, M., Atha, D., Nubert, J., Fan, D., Agha, A., Padgett, C., Spieler, P., Hutter, M., Khattak, S., 2024. Roadrunner-learning traversability estimation for autonomous off-road driving. IEEE Transactions on Field Robotics.
Gao, P., Geng, S., Zhang, R., Ma, T., Fang, R., Zhang, Y., Li, H., Qiao, Y., 2024. Clip-adapter: better vision-language models with feature adapters. Int. J. Comput. Vis. 132 (2), 581–595.
Gaspar, F., Carreira, D., Rodrigues, N., Miragaia, R., Ribeiro, J., Costa, P., Pereira, A., 2025. Synthetic image generation for effective deep learning model training for ceramic industry applications. Eng. Appl. Artif. Intell. 143, 110019.
Gong, Y., Zhang, X., Feng, J., He, X., Zhang, D., 2024. LiDAR-based HD map localization using semantic generalized ICP with road marking detection. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 3379–3386.
Google, 2015. Discover Street View and Contribute Your Own Imagery to Google Maps. https://www.google.com/streetview/.
Jayasinghe, O., Hemachandra, S., Anhettigama, D., Kariyawasam, S., Rodrigo, R., Jayasekara, P., 2022. Ceymo: see more on roads-a novel benchmark dataset for road marking detec-tion. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3104–3113.
Kim, K., Cho, S., Chung, W., 2021. HD map update for autonomous driving with crowdsourced data. IEEE Rob. Autom. Lett. 6 (2), 1895–1901.
Kim, N., Seong, H., Ji, D., Jang, S., 2024. Unveiling the hidden: online vectorized hd map construction with clip-level token interaction and propagation. Adv. Neural Inf. Process. Syst. 37, 111358–111381.
Lee, S., Kim, J., Shin Yoon, J., Shin, S., Bailo, O., Kim, N., Lee, T.H., Seok Hong, H., Han, S.H., So Kweon, I., 2017. Vpgnet: vanishing point guided network for Lane and road mark-ing detection and recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1947–1955.
Li, L.H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., Wang, L., Yuan, L., Zhang, L., Hwang, J.N., Chang, K.W., 2022. Grounded language-image pre-training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10965–10975.
Liu, D., Mao, Q., Gao, L., Wang, G., 2024. Leveraging contrastive language–image pre-training and bidirectional cross-attention for multimodal keyword spotting. Eng. Appl. Artif. Intell. 138, 109403.
Liu, R., Wang, J., Zhang, B., 2020. High definition map for automated driving: overview and analysis. J. Navig. 73 (2), 324–341.
Lüddecke, T., Ecker, A.S., 2021. Image segmentation using text and image prompts. 2022 IEEE. in CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7076–7086.
Mazhar, S., Atif, N., Bhuyan, M.K., Ahamed, S.R., 2023. Block attention network: a lightweight deep network for real-time semantic segmentation of road scenes in resource-constrained devices. Eng. Appl. Artif. Intell. 126, 107086.
Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z., Wang, X., 2022. Simple open-vocabulary object detection. In: European Conference on Computer Vision. Springer Nature Switzerland, Cham, pp. 728–755.
O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L., Riordan, D., Walsh, J., 2020. Deep learning vs. traditional computer vision. In: Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), 1 1. Springer International Publishing, pp. 128–144.
Oord, A.V.D., Li, Y., Vinyals, O., 2019. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748.
Pannen, D., Liebner, M., Hempel, W., Burgard, W., 2020. How to keep HD maps for automated driving up to date. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 2288–2294.
Poggenhans, F., Pauls, J.H., Janosovits, J., Orf, S., Naumann, M., Kuhnt, F., Mayr, M., 2018. Lanelet2: a high-definition map framework for the future of automated driving. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1672–1679.
Singh, M.K., Formosa, N., Man, C.K., Morton, C., Masera, C.B., Quddus, M., 2024. Assessing temporary traffic management measures on a motorway: lane closures vs narrow lanes for connected and autonomous vehicles in roadworks. IET Intell. Transp. Syst. 18 (7), 1210–1226.
Wong, K., Gu, Y., Kamijo, S., 2020. Mapping for autonomous driving: opportunities and challenges. IEEE Intelligent Transportation Systems Magazine 13 (1), 91–106.
Xia, D., Zhang, W., Liu, X., Zhang, W., Gong, C., Tan, X., Huang, J., Yang, M., Yang, D., 2025. LDMapNet-U: an end-to-end System for city-scale lane-level map updating. arXiv preprint arXiv:2501.02763.
Zhang, P., Zhang, M., Liu, J., 2021. Real-time HD map change detection for crowdsourcing update based on mid-to-high-end sensors. Sensors 21 (7), 2477.
Zhang, L., Yang, K., Han, Y., Li, J., Wei, W., Tan, H., Yu, P., Zhang, K., Yang, X., 2025. TSD-DETR: a lightweight real-time detection transformer of traffic sign detection for long-range perception of autonomous driving. Eng. Appl. Artif. Intell. 139, 109536.
Zhang, Z., Sabuncu, M., 2018. Generalized cross entropy loss for training deep neural networks with noisy labels. Adv. Neural Inf. Process. Syst. 31.
Zhu, W., Jiang, Z., He, Y., 2025. Geometry-sensitive semantic modeling in visual and visual-language domains for image captioning. Eng. Appl. Artif. Intell. 147, 110330.