# Enhancing Micromobility Safety and Infrastructure Monitoring through Computer Vision and Real-Time Data Integration

Hugo Martínez[1], Josep Ramon Morros[1], Andreas Nikiforiadis[2,3], Josep Maria Salanova Grau[2], Annie Kortsari[2] and Evangelia Manola[2]

[1] Universitat Politècnica de Catalunya, Barcelona 08034, Spain
[2] Hellenic Institute of Transport, Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece
[3] School of Civil Engineering, Democritus University of Thrace, 67100 Xanthi, Greece

hugo.martinez@upc.edu

**Abstract.** Micromobility has emerged as a key solution to urban mobility challenges, offering low-carbon, space-efficient, and health-promoting alternatives to traditional transport. However, safety and infrastructure quality are still significant barriers to its widespread adoption. This paper presents the development of an integrated system that combines computer vision, GNSS geolocation, and real-time data processing to enhance micromobility safety and infrastructure monitoring. The system is composed of three specialized computer vision models: a) classification of the type of lane on which the vehicle circulates (bike lane, sidewalk, etc.), b) estimation of pedestrian density along the vehicle's path, and c) detection of pavement and infrastructure defects, including potholes and cracks. Video feeds captured from micromobility vehicles are analyzed to identify safety hazards and infrastructure issues. Detected defects are geolocated and mapped, providing a visual interface for operators and policymakers. The system also includes a dashboard for real-time monitoring, allowing authorities to make evidence-based decisions, dynamically manage safety concerns, and optimize public space usage. By integrating data on vehicle-pedestrian and vehicle-vehicle interactions, infrastructure conditions, and traffic density, the platform supports holistic governance of shared and active mobility. This approach facilitates regulation enforcement, user awareness and promotes proactive infrastructure maintenance. Overall, the proposed solution shows how advanced sensing technologies and data-driven tools can foster safer, more efficient, and sustainable urban mobility systems, while providing tools for both operators and public authorities.

**Keywords:** Micromobility, computer vision, infrastructure monitoring, urban mobility safety

## 1 Introduction

Cities are increasingly turning to micromobility solutions such as bicycles, e-bikes, and e-scooters to reduce congestion, cut transport emissions, and promote active travel. These lightweight vehicles use significantly less road space than private cars and can substantially reduce greenhouse gas emissions, making them an important component of sustainable urban mobility strategies [1]. By integrating micromobility services with public transport, several studies report potential reductions in car use and improved access to sustainable modes, especially for first- and last‑mile connections [2,3].

However, safety concerns are still a barrier to the wider adoption and social acceptance of micromobility. Technical reports and studies [4,5] highlight high injury rates for e-scooter riders compared with other transport modes and emphasize the frequency of single‑vehicle crashes, injuries, and falls. Infrastructure-related factors play a central role: surface defects such as potholes, changes in surface level, and curb impacts are major contributors to loss of control and falls among users. The lack of continuous, well-designed cycling infrastructure, ambiguous or degraded markings, and mixed traffic conditions further increase interactions with pedestrians and motor vehicles, amplifying perceived and actual risk [6].

Therefore, advanced monitoring of micromobility infrastructure is essential to support a safer design, maintenance, and context-aware regulation. Yet current monitoring practices are fragmented and often reactive. Municipalities typically rely on periodic inspections, citizen complaints, or crash statistics, which provide limited spatial and temporal resolution and may under-represent minor incidents and near-misses linked to infrastructure quality [3,6]. Even where crowdsourcing and open data initiatives are in place, they depend on voluntary reporting and rarely capture detailed information about lane types, surface conditions, or local pedestrian densities along specific routes. Then, many safety-critical issues affecting micromobility users are unnoticed in day-to-day governance and planning processes.

Advances in computer vision and low-cost hardware create new opportunities to monitor infrastructure and user environments directly from vehicles in motion. Vision-based methods have been successfully applied to detect potholes, cracks, and general pavement distress, as well as to identify traffic signs and lane markings in automotive and intelligent transport system applications [7]. Recent work [8,9] demonstrates that deep learning models, such as YOLO [10], can localize road defects, recognize lane markings in real time and enable pedestrian detection for crowd density estimation, all of which can be used to better understand safety risks.

The goal is to adapt these models and architectures to work on micromobility infrastructure to obtain highly accurate and relevant data. This paper presents an integrated framework, containing a user application and a data based platform, that utilizes egocentric images and GNSS geolocation for infrastructure monitoring and enhancing micromobility safety. Video streams captured from micromobility vehicles (either users or municipality workers) are processed by three specialized computer vision modules: (i) lane type classification and localization, which identifies whether the vehicle is traveling on a bike lane; (ii) pedestrian density estimation along the vehicle's trajectory, capturing the intensity of interactions between micromobility users and pedestrians; and (iii) pavement and infrastructure defect detection, focusing on hazards such as potholes and cracks. Detected events are geolocated and aggregated into a database, generating dynamic infrastructure maps and interaction heatmaps for operators and authorities. Defect detections make maintenance more efficient and allow prioritizing and verifying repairs, while pedestrian density profiles show conflict areas to improve bike lane network planning and design. All these procedures are executed in a user application operating directly on the vehicle hardware.

Beyond data acquisition and perception, the effective use of these observations requires an integrated environment capable of transforming raw detections into useful knowledge for cities (see Figure 1). To this end, the proposed framework is complemented by a Platform and Decision-Support layer that consolidates computer vision outputs with crowdsourced user reports into a unified analytical and governance tool. This platform enables systematic statistical analysis of safety-related events, supports the data-driven establishment and adaptation of circulation restrictions for micromobility users (which are handled by the user application mentioned previously), and provides a structured assessment of infrastructure quality. By closing the loop between monitoring, analysis, and policy intervention, the platform aims to support more proactive, evidence-based management of active and micromobility systems.

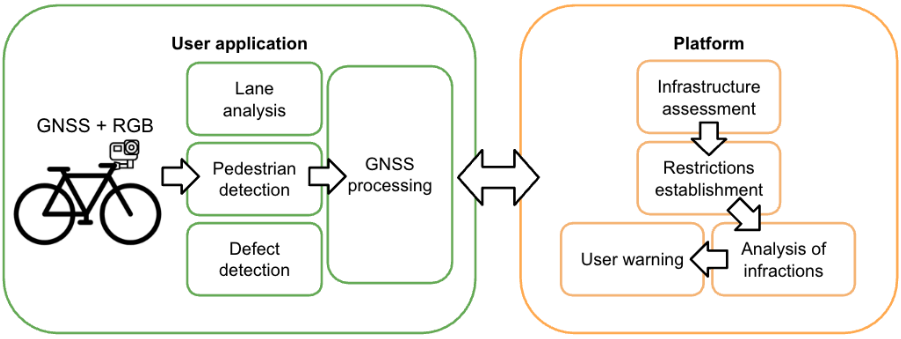


**Fig. 1.** Overview of the system.

## 2 Computer Vision Framework

The proposed framework is structured around three computer vision modules that tackle: lane type classification, pedestrian density estimation, and infrastructure defect detection. Each module has to extract relevant information from an egocentric viewpoint and to operate under realistic constraints such as limited compute, variable lighting, and diverse urban layouts.

The goal is to adapt recent advances in deep learning techniques in scene understanding, object detection, and road surface analysis, to the specific characteristics of micromobility video data: narrow field of view, unstable sequences and complex urban environments.

### 2.1 Lane Analysis

The lane analysis module has two submodules, see Figure 2. First, an image classifier determines the type of infrastructure the micromobility user is going through (bike lane, road, or sidewalk). When the image is classified as a bike lane it activates the lane detection module, which detects and locates the bike lane, information that is combined with the output of the other modules (pedestrian/vehicle detection and pavement defect detection). This allows us to be certain that the output from the overall pipeline belongs to micromobility infrastructure.

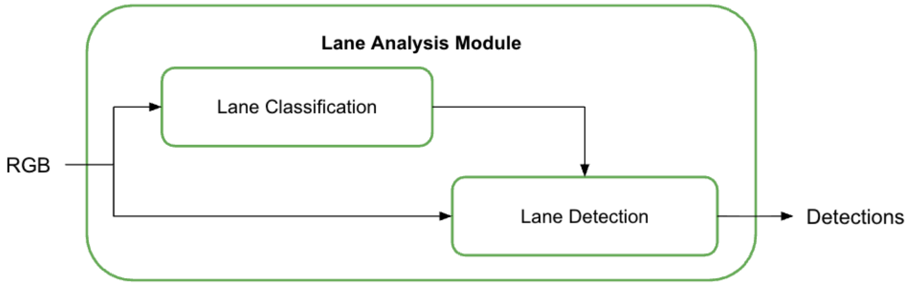


**Fig. 2.** Lane Analysis module architecture.

Following previous work from the RideSafe project [11], lane classification is performed using an image-based scene classification model with RGB images captured from the vehicle as input. The used model is ShuffleNet v2 1x (1.3M parameters), which is pretrained on ImageNet, and fine-tuned on manually annotated images captured from the vehicle. Its output acts as a base information for subsequent modules, in particular using the lane detection module only when the image is classified as bike lane.

Following classification, the lane detection aims to locate the bike lane geometry. Early lane detection relied on classical image processing techniques such as Canny edge detection [12] and the Hough Transform [13] to recover lane markings as straight or curved lines from binary edge maps. While simple and computationally efficient, these methods depend on clear markings and carefully tuned thresholds, and they degrade with shadows, faded paint, and complex urban geometries. This can particularly affect when dealing with micromobility settings, where lanes are usually narrow, partially occluded, or embedded in shared spaces without clear painted boundaries.

Recent work focuses on deep-learning-based approaches, which can learn richer visual representations and are more robust to appearance changes [14]. Two main methods are relevant for micromobility lane detection: segmentation-based methods and anchor-based methods [15].

Segmentation-based approaches treat lane detection as a pixel-wise prediction problem, producing a mask where each pixel is assigned to a specific lane instance. Architectures such as LaneNet [16], SCNN [17], and RESA [18] use encoder-decoder backbones with dedicated heads to generate binary lane masks and per-pixel embeddings that can be clustered into individual lanes. These models can represent multiple lanes, capture curved geometries, and, when extended, jointly segment surrounding structures such as sidewalks or curbs, which is advantageous for distinguishing bike lanes from road lanes or pedestrian areas in dense urban scenes [17]. However, their outputs are dense and may require intense post-processing which can be computationally demanding for edge applications.

Anchor-based approaches, e.g. LaneAtt [20], UFLD [21], and UFLDv2 [22], represent lanes through a set of predefined anchors and regress lane coordinates relative to these reference points. The network outputs a small number of control points of each lane, along with probabilities of existence, providing a structured representation, better for real-time performance. Hybrid anchor setups help reduce location errors by combining row- and column-based anchors, improving robustness under perspective distortion and partial occlusions [22]. These properties make anchor-based methods interesting for deployment on low-power micromobility devices [23].

In our case, an anchor-based architecture is selected because these models directly output lane point coordinates. Additionally, this choice is motivated by the availability of open-source models and their pretrained weights, concretely, UFLDv2 is used as the lane detection model.

### 2.2 Pedestrian Density Estimation

The goal of the pedestrian density estimation module is to count the number of pedestrians in the visible field of view of a micromobility vehicle, helping to obtain information on the interaction between these two groups. This can help identify locations of high pedestrian and micromobility vehicles interaction that may require infrastructure interventions or operational measures, since high densities increase the likelihood of conflicts, near accident misses, and low-speed collisions. In addition, it reveals patterns such as peak interaction times and locations.

This task requires continuous pedestrian detection from moving vehicle images that may contain partial occlusions as the vehicle navigates intersections, turns, and dense urban scenes which can be a challenge. Recent methods for person detection divide into detection-based and density-map approaches.

Detection-based approaches apply object detectors (YOLO [11,25], Faster R-CNN [24]) to count persons within a region of interest. These methods have great performance when pedestrians are well-separated but may miss detections in highly dense crowds due to heavy occlusion. In our scenario, a single-stage detector like YOLO offers real-time performance and can be fine-tuned for pedestrian detection in urban scenarios.

Density-map approaches directly regress a density map where each pixel contributes to the total count via Gaussian kernels centered at annotated head positions [26]. CSRNet [27], MCNN [28], and transformer-based variants like ADCrowdNet [29] handle scale variation and occlusion better by learning multi-scale features and global context. Density maps handle overlapping pedestrians, making them robust to occlusion. However, they require dense annotations and higher compute capacity.

In this case, we select a pretrained object detector (YOLOv26) for pedestrian detection (density estimation) due to its real-time performance if deployed on edge devices, the availability of open-source code and pretrained models, and straightforward fine-tuning process. In addition, it can be extended to detect multiple classes of objects detection (pedestrians, vehicles, obstacles, etc.).

### 2.3 Infrastructure Defect Detection

The goal of the infrastructure defect detection module is to detect surface irregularities (potholes and cracks) and obstacles in the way of users which can be safety risks. Mapping these can make micromobility infrastructure maintenance more efficient, geolocated defects/obstacles enabling making more effective interventions, and can help to obtain the correlation with incident patterns and infrastructure usage.

Different sensors are typically employed, laser scanning, depth sensing, infrared thermography, and ground-penetrating radar which provide high-accuracy pavement assessment. [30] However, this work focuses on RGB camera data for its simplicity, low cost, and availability, also on consumer devices.

Typical defects in lanes can be categorized into a few, predetermined classes: potholes, cracks, etc. This allows using object detection technology to detect and localize these defects. Multi-class detectors like YOLO [25] have real-time performance suitable for edge deployment, and can be easily fine-tuned. They produce bounding boxes that locate and classify objects (defects) in a scene. Semantic segmentation models provide pixel masks of the objects of interest using encoder-decoder architectures such as U-Net [31] and there are specialized variants (DeepCrack [32]) that have great performance detecting cracks.

In addition to lane defects, objects in the lane can pose hazards to the drivers. These objects can be generic, so their classes form an open set, which makes it impossible to train an object detector. In this case, we propose to detect these objects by looking at their volumetric properties: as the lanes are supposed to be flat, objects that protrude from the road are considered dangerous. While it is difficult to estimate the volumetric properties directly from RGB images, Monocular Depth Estimation (MDE) models can predict depth for each pixel using RGB images. These models can be pretrained on diverse scenes and generalize well to road images without the need for actual depth sensors. Recent models include MiDaS [34] or DepthAnything [35] and their pretrained versions are fully available.

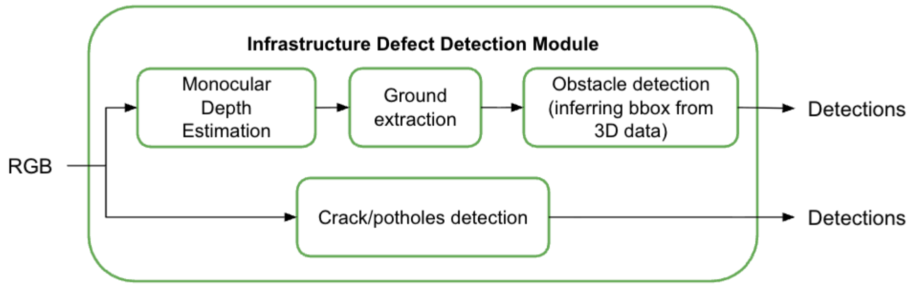


**Fig. 3.** Defect detection module architecture.

The architecture of this module combines MDE and object detection in a complementary way, both using RGB images as input, see Figure 3. First, a pretrained MDE model (DepthAnything) is used to obtain the point cloud of the scene (3D mapping), then it is possible to extract the ground plane and identify objects protruding from it, which are potential obstacles when located in the bike lane. For defects (cracks and potholes), the core component is a pretrained and fine-tuned object detector (YOLOv26) trained on an annotated defect dataset. If the MDE has enough accuracy, a potential solution is to leverage the depth information to detect objects below the ground plane (potholes).

Overall, the three modules have to take into account the low-angle perspective from cameras in micromobility vehicles, edge inference constraints (latency, memory), battery life limitations favoring compact architectures, and the vehicle motion noise, to enable robust, lane detection, density monitoring and defect detection under realistic conditions.

## 3 Platform and Decision-Support Layer

Figure 4 provides an overview of the Platform and Decision-Support layer, illustrating the flow of information from heterogeneous data sources to a set of analytical and policy-oriented outputs. On the input side, the platform ingests (i) crowdsourced reports submitted through a mobile application by active and micromobility users, and (ii) automated observations generated with the computer vision models. These inputs are harmonized within a common data management environment and processed through a series of analytical modules. The resulting outputs support city authorities and stakeholders by enabling statistical analysis of infractions, the establishment and continuous update of circulation restrictions for active and micromobility users, and a systematic assessment of the quality of active and micromobility infrastructure.

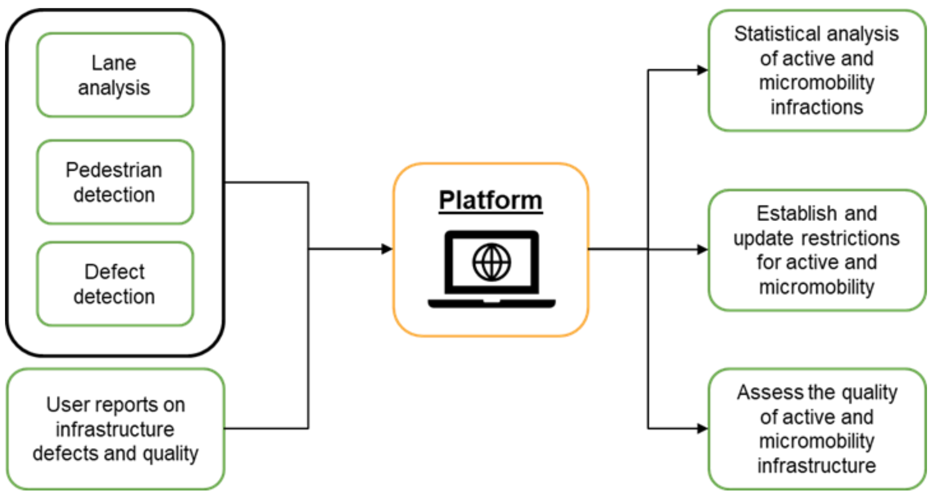


**Fig. 4.** Platform's main inputs and outputs.

### 3.1 Data Integration and Statistical Analysis

A core function of the platform is the transformation of raw, heterogeneous inputs into actionable statistical knowledge. Crowdsourced data from the mobile application include infractions, perceived safety issues, and infrastructure defects, typically enriched with spatial and temporal metadata. In parallel, the computer vision models described in the previous section contribute objective observations, such as road and lane classification, detection of pedestrians in front of bicycles or e-scooters, and automated identification of surface defects or obstacles. The platform integrates these data streams by aligning them spatially (e.g., at street segment) and temporally, allowing for joint analysis and cross-validation between subjective user reports and automated detections.

The statistical analysis component is designed to provide both a real-time and a longitudinal view of active and micromobility-related phenomena in the urban environment. Rather than focusing on a single indicator, the platform supports a multi-dimensional analysis that captures the frequency, location, and evolution of accidents and infractions. Typical indicators include the number of reported or detected incidents, their spatial distribution across the monitored network, and their temporal variation at different resolutions (hourly, daily, monthly, or yearly). By aggregating data over time, the platform enables the identification of recurring patterns, emerging hotspots, and seasonal trends that would not be visible through isolated observations.

Spatial analyses rely on map-based representations, such as density or risk-level aggregations at street or area level, while temporal analyses highlight changes in incident frequency over time. The combination of spatial and temporal perspectives supports evidence-based decision-making by enabling authorities to relate observed safety outcomes to specific locations, time periods, or policy interventions.

Importantly, the statistical analysis layer is not conceived as a static reporting tool but as an adaptive component that evolves as new data are collected. As the volume and diversity of crowdsourced and computer vision data increase, the reliability of the indicators improves, allowing the platform to progressively shift from descriptive statistics toward more advanced analytics, such as trend analysis or before–after comparisons of interventions. In this sense, the statistical analysis serves as a foundational layer that feeds directly into the decision-support functionalities described in the following subsections.

### 3.2 Establishment and Update of Restrictions for Active and Micromobility Users

Building on the knowledge extracted from the statistical analysis and perception modules, the platform supports the establishment and dynamic update of circulation restrictions and warnings for active and micromobility users. The underlying logic is that restrictions should not be defined in isolation, but rather derived from an integrated understanding of user behavior, safety outcomes, and infrastructure characteristics. To this end, the platform enables authorities to translate analytical insights, such as high-risk locations, frequent infractions, into targeted regulatory measures.

Restrictions are represented within the platform using geospatial constructs, typically geofenced areas defined over the urban network. These areas can correspond to entire neighborhoods, specific street segments, or localized zones around sensitive locations. Within each geofenced area, authorities can associate a set of rules or warnings that apply to specific categories of users, such as cyclists or e-scooter riders. The platform distinguishes between location-based restrictions, which apply uniformly within an area, and infrastructure-based restrictions, which are conditional on the type or quality of infrastructure present (for example, restrictions that apply only on sidewalks, shared paths, or narrow lanes). The restriction data is delivered to the user application, which leverages GNSS location information to assess compliance and generate warnings in case of violations.

A key aspect of this component is its iterative and data-driven nature. Restrictions can be created, modified, or deactivated as new evidence becomes available. For instance, a concentration of speeding infractions detected through user reports and corroborated by computer vision-based speed estimates may motivate the introduction of a speed limitation or warning in a specific corridor. Conversely, improvements in infrastructure quality or changes in user behavior over time may justify the relaxation or removal of previously imposed restrictions. In cases where multiple restrictions overlap spatially, prioritization rules ensure that the most relevant or stringent measure is applied, reflecting the hierarchical logic defined by the authorities.

## 3.3 Assessment of Active and Micromobility Infrastructure Quality

The third major output of the Platform and Decision-Support layer is the assessment of the quality of active and micromobility infrastructure. This assessment relies explicitly on the combination of crowdsourced inputs and computer vision outputs, recognizing that infrastructure quality is a multi-faceted concept that cannot be fully captured by a single data source. User reports provide valuable insights into perceived comfort, safety, and usability, while automated vision-based detections offer objective and scalable measurements of physical conditions, such as surface defects, lane continuity, or the presence of obstacles.

The platform aggregates these complementary inputs to compute a set of widely accepted key performance indicators (KPIs) and metrics related to active and micromobility infrastructure. By normalizing and aggregating indicators at appropriate spatial units (e.g., street segments or zones), the platform enables comparative assessments across the network and over time.

Crucially, the methodology emphasizes consistency and interpretability rather than exhaustive coverage. The selected KPIs are aligned with established practices in transport planning and urban mobility assessment, ensuring that results can be understood and used by practitioners. Over time, as more data is collected and validated, the infrastructure quality assessment can support prioritization of maintenance actions, identification of critical bottlenecks, and evaluation of the impact of infrastructure upgrades.

## 4 Conclusion

The Platform and Decision-Support layer, presented in this paper, demonstrates how heterogeneous data sources can be transformed into actionable knowledge for active and micromobility planning and management. By combining crowdsourced inputs from users with automated observations from computer vision models, the platform enables a unified view of safety-related events, user behavior, and infrastructure conditions. Its analytical capabilities support evidence-based statistical assessments, while the tight coupling between analysis and policy mechanisms allows circulation restrictions and warnings to be dynamically established and updated based on observed conditions. At the same time, the integration of subjective user feedback with objective infrastructure monitoring enables a robust and scalable assessment of active and micromobility infrastructure quality. Overall, the proposed platform highlights the value of data-driven decision support for cities, offering a systematic approach to improving safety, regulatory effectiveness, and infrastructure quality for active and micromobility users, while remaining adaptable to different urban contexts.

**Acknowledgement:** This research received funding by the European Union through the Horizon Europe research and innovation programme under grant agreement No 101203240 and the Spanish PID2024-161868OB-I00 [C3DRUM] research project funded by MCIN/AEI/10.13039/501100011033 and FEDER, EU.

## References


1. D Yuriy Vovk, Oleg Tson, Marcin Kicinski, Iryna Vovk, Volodymyr Dzyura, Yaroslav Vovk. Microtransportation: A sustainable solution for urban mobility challenges. Journal of Chinese Architecture and Urbanism 7(4), 6561 (2025).
2. Szakonyi P.: Mapping the Potential of e-Micromobility Vehicles for Sustainable Urban Development – the Case of Gyor, Hungary, Chemical Engineering Transactions, 114, 703-708 (2024).
3. Dias, Gabriel José Cabral, Paulo Ribeiro, and Elisabete Arsenio.: Micromobility: a systematic literature review on the measurement of its environmental, social, and economic impacts on urban sustainability. (2021).
4. Turoń, Katarzyna, and Piotr Czech.: The concept of rules and recommendations for riding shared and private e-scooters in the road network in the light of global problems. Scientific and Technical Conference Transport Systems Theory and Practice. Cham: Springer International Publishing, (2019).
5. Rix, Kevin, et al. "Injury rates per mile of travel for electric scooters versus motor vehicles." The American Journal of Emergency Medicine 40 (2021): 166-168.
6. Yannis, George, P. Crist, and V. Petraki.: Safer micromobility: technical background report. International Transport Forum. (2024).
7. Rathee, Munish, Boris Bačić, and Maryam Doborjeh.: Automated road defect and anomaly detection for traffic safety: A systematic review. Sensors 23.12 (2023): 5656.
8. Li, Tianwen, and Gongquan Li. "Road defect identification and location method based on an improved ML-YOLO algorithm." Sensors 24.21 (2024): 6783.
9. Dewi, Christine, et al. "Recognizing road surface traffic signs based on YOLO models considering image flips." Big data and cognitive computing 7.1 (2023): 54.
10. Redmon J, Divvala S, Girshick R, Farhadi A.: You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). p. 779–88 (2016).
11. Morros, Josep Ramon, et al.: Real-time lane classification and accident detection for safer micromobility. 11th International Congress on Transportation Research. Heraklion, Crete. 2023.
12. Canny, John.: A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence 6 (2009): 679-698.
13. Hough, Paul VC. "Method and means for recognizing complex patterns." U.S. Patent No. 3,069,654. 18 Dec. (1962).
14. Tang, Jigang, Songbin Li, and Peng Liu.: A review of lane detection methods based on deep learning. Pattern Recognition 111 (2021): 107623.
15. Qin, Zequn, Huanyu Wang, and Xi Li.: Ultra fast structure-aware deep lane detection. European conference on computer vision. Cham: Springer International Publishing, (2020).
16. Neven, Davy, et al.: Towards end-to-end lane detection: an instance segmentation approach. 2018 IEEE intelligent vehicles symposium (IV). IEEE, (2018).
17. Pan, Xingang, et al.: Spatial as deep: Spatial cnn for traffic scene understanding. Proceedings of the AAAI conference on artificial intelligence. Vol. 32. No. 1. (2018).
18. Zheng, Tu, et al.: Resa: Recurrent feature-shift aggregator for lane detection. Proceedings of the AAAI conference on artificial intelligence. Vol. 35. No. 4. (2021).
19. CHEN, Zheng-bin, and Dong-yi YE.: Lightweight Lane Detection Algorithm with Semantic Segmentation. Journal of Chinese Computer Systems 42.9 (2021): 1877-1883.
20. Tabelini, Lucas, et al.: Keep your eyes on the lane: Real-time attention-guided lane detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. (2021).

21. Qin, Zequn, Huanyu Wang, and Xi Li.: Ultra fast structure-aware deep lane detection." European conference on computer vision. Cham: Springer International Publishing, 2020.
22. Qin, Zequn, Pengyi Zhang, and Xi Li.: Ultra fast deep lane detection with hybrid anchor driven ordinal classification.: IEEE transactions on pattern analysis and machine intelligence 46.5 (2022): 2555-2568.
23. Yao, Zhiting, and Xiyuan Chen.: Efficient lane detection technique based on lightweight attention deep neural network. Journal of Advanced Transportation 2022.1 (2022): 5134437.
24. Ren, Shaoqing, et al.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1137-1149.
25. Sapkota, Ranjan, and Manoj Karkee.: Ultralytics YOLO evolution: An overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 object detectors for computer vision and pattern recognition. arXiv preprint arXiv:2510.09653 (2025).
26. Sindagi, Vishwanath A., Rajeev Yasarla, and Vishal M. Patel.: Pushing the frontiers of unconstrained crowd counting: New dataset and benchmark method. Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019).
27. Li, Yuhong, Xiaofan Zhang, and Deming Chen.: Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2018).
28. Zhang, Yingying, et al.: Single-image crowd counting via multi-column convolutional neural network. Proceedings of the IEEE conference on computer vision and pattern recognition. (2016).
29. Liu, Ning, et al.: Adcrowdnet: An attention-injective deformable convolutional network for crowd understanding. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. (2019).
30. Schnebele, E., et al. : Review of remote sensing methodologies for pavement management and assessment. European Transport Research Review 7.2 (2015).
31. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international publishing, (2015).
32. Zou, Qin, et al.: Deepcrack: Learning hierarchical convolutional features for crack detection. IEEE transactions on image processing 28.3 (2018): 1498-1512.
33. Li, Peigen, et al.: CNN-based pavement defects detection using grey and depth images. Automation in Construction 158 (2024): 105192.
34. Ranftl, René, Alexey Bochkovskiy, and Vladlen Koltun.: Vision transformers for dense prediction. Proceedings of the IEEE/CVF international conference on computer vision. (2021).
35. Yang, Lihe, et al.: Depth anything: Unleashing the power of large-scale unlabeled data. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. (2024).