An End-to-End Multimodal System for Large-Scale Road Infrastructure Monitoring Using Computer Vision and Geospatial Integration

Paulina Irene Velasquez Ferrufino[1], Matheus Gois Paiva[2], Cornélio Albuquerque de Sousa[1], José Florêncio de Queiroz Neto[1], Wellington Clay Porcino[3], José Antônio Fernandes de Macedo[1]

[1] Department of Computer Science – Federal University of Ceará (UFC)
Fortaleza, CE – Brazil

[2]University of Fortaleza (UNIFOR)
Fortaleza, CE – Brazil

[3]National Department of Public Security
Federal District – Brazil

paulinavelasquez@alu.ufc.br, mgoispaiva@edu.unifor.br

corneliosousa@alu.ufc.br, florencio@insightlab.ufc.br

wellington.porcino@mj.gov.br, jose.macedo@dc.ufc.br

Abstract. *We present a multimodal road-monitoring system for Brazil using onboard videos, telemetry, computer vision, and geospatial references. It aligns frames with Global Navigation Satellite System (GNSS) telemetry, filters them using motion evidence, applies models to pavement defects, horizontal markings, traffic signs, and roadside vegetation, and aggregates auditable evidence by segment. Trajectories undergo map matching to OpenStreetMap (OSM) with GraphHopper through mmlib, then are linked to official federal and state road references. In operation, 23.6 million telemetry points supported 6.26 million frame-level inferences, ranging from crack detection to traffic-sign type detection at 97.57% mAP@50.*

1. Introduction

Manual road inspection remains constrained by cost, periodicity, and subjective assessment, especially when the monitored network spans long corridors and heterogeneous operating conditions [Ma and Chen 2025]. Computer vision has become a practical alternative for pavement-damage detection because vehicle-mounted imagery can be transformed into structured maintenance evidence [Dong et al. 2025]. Mobile mapping systems combine Global Navigation Satellite System (GNSS), cameras, Light Detection and Ranging (LiDAR), and inertial sensors to expand continuous infrastructure observation in the field [Bruno et al. 2025]. Recent studies also show that traffic signs, horizontal markings, and other road assets require robust perception under scale variation, changing illumination, occlusion, and wear [Kulawik and Kuczyński 2025].

Despite this progress, a gap remains between accurate visual models and deployable systems that operate on long videos, imperfect telemetry, missing frames, redundant observations, and heterogeneous official databases [Zhumadillayeva et al. 2025].

Road-asset monitoring studies emphasize that useful inventories must be georeferenced, updateable, and compatible with Geographic Information System (GIS) workflows [Katsamenis et al. 2025]. Map matching is therefore a core requirement: noisy GNSS observations must be associated with the road network before model outputs can become indicators per road segment [Stipancic et al. 2025]. In Brazil, federal road analysis relies on the National Road System (Sistema Nacional de Viação, SNV), maintained by the Brazilian Department of Transport Infrastructure (Departamento Nacional de Infraestrutura de Transportes, DNIT) [DNIT 2026].

This work addresses that operational gap by treating transportation company bus-operation data as a real-world opportunity to develop and evaluate large-scale road-monitoring systems. Onboard videos recorded between June 28 and July 14, 2025, together with large-scale telemetry, provided the operational basis for training and evaluating the visual modules. The system reduced 584.6 million raw telemetry records to 23.6 million points for map matching, covered 512 trips and 259 bus lines, and supported 6.26 million frame-level model applications. In the most recent pavement-defect run, the updated model analyzed 877,070 frames and produced 37,203 predicted defect events, later consolidated and aggregated by road segment. These values emphasize that the novelty is not a detector in isolation, but the transformation of moving-vehicle observations into auditable, segment-level infrastructure indicators.

The contributions are fourfold. First, we present an end-to-end architecture that joins onboard acquisition, image–telemetry synchronization, multimodal inference, temporal persistence, map matching, and validation. Second, we formalize data-curation controls for stationary buses, rain, visually irrelevant contexts, and candidate-segment prioritization. Third, we integrate GNSS trajectories with OpenStreetMap (OSM), GraphHopper, mmlib as a map-matching interface, SNV, the state road reference associated with the Contribution for Intervention in the Economic Domain (in Portuguese Contribuição de Intervenção no Domínio Econômico, CIDE), and kilometer-level road attribution. Fourth, we show how heterogeneous visual modules can feed a common segment-level indicator layer while preserving image evidence for human review.

2. Related Work

Vision-based road-infrastructure monitoring has evolved from isolated field surveys toward onboard perception, where cameras and geospatial sensors record infrastructure condition during regular vehicle operation [Bruno et al. 2025]. In pavement assessment, You Only Look Once (YOLO) models remain widely used because they balance inference speed and spatial localization [Ma and Chen 2025]. Fine cracks remain difficult to detect because contrast, shadows, perspective, and crack morphology reduce visual separability in road imagery [Dong et al. 2025]. For deformation-related distresses, Sheikholeslami et al. integrate a Physics-Informed Neural Network (PINN) penalty into the YOLO loss to improve bounding-box reliability and reduce false positives for rutting and corrugation [Sheikholeslami et al. 2026]. These studies support model development, but operational deployment also requires synchronization, temporal consolidation, and spatial aggregation beyond benchmark images.

Road monitoring is not limited to pavement damage, since maintenance decisions also depend on signs, horizontal markings, roadside vegetation, and other assets observed

from moving vehicles. Traffic-sign detection is challenged by small objects, variable distance, occlusion, and dynamic backgrounds [Zhang et al. 2025a]. Traffic-sign recognition in Intelligent Transportation Systems (ITS) also requires environmental robustness and low-latency inference [Wang et al. 2025]. Horizontal road markings are less explored than vertical signs and lane markings, although their condition directly affects navigation, safety, and maintenance planning [Kulawik and Kuczyński 2025]. Recent road-infrastructure datasets and benchmarks have highlighted that infrastructure elements such as traffic signs, guardrails, delineators, and other roadside assets remain underrepresented in conventional autonomous-driving datasets [Katsamenis et al. 2025].

To make visual observations actionable, vehicle GNSS traces must be linked to road geometry through map matching, enabling analysis by road, kilometer, and operational segment [Stipancic et al. 2025]. Hidden Markov Model (HMM)-based methods model observation uncertainty and plausible road transitions [Zhang et al. 2025b], while diffusion approaches address noisy and sparse trajectories [Han et al. 2026]. For deployment, GraphHopper supports routing and OSM-based map matching [GraphHopper Project 2026], while mmlib is a public Python interoperability framework used here to submit trajectories to GraphHopper and recover matched routes [Dias de Souza 2026]. In Brazil, SNV provides official federal road and kilometer references [DNIT 2026], while DNIT structures maintenance assessment through the Maintenance Condition Index (in Portuguese Índice de Condição da Manutenção, ICM), composed of the Pavement Index (in Portuguese Índice de Pavimentação, IP) and the Conservation Index (in Portuguese Índice de Conservação, IC), with video records, computer vision, and manual post-processing [DNIT 2024]. RoadAtlas integrates georeferenced imagery, defect detection, asset management, and visualization [Chen et al. 2021]. Pavimentados processes georeferenced imagery for pavement-condition assessment [IDB 2025]. Unlike these alternatives, the proposed system combines four visual tasks, bus telemetry, temporal consolidation, map matching, and official Brazilian road references into auditable segment-level indicators; different datasets and taxonomies preclude direct metric comparison.

3. System Contribution and Design Principles

The central contribution is to move the analytical unit from an isolated frame to a localized road event or segment state. In real operation, a visual prediction is only actionable after it becomes a spatial decision: a defect count, a conservation state, a sign inventory entry, or a vegetation-condition indicator associated with a defined road section. In this paper, a model-selection problem refers mainly to choosing or improving a detector or classifier, whereas a system-level problem involves coordinating acquisition, synchronization, curation, temporal validation, geospatial attribution, and audit mechanisms so that predictions become road-management evidence.

Three design principles support this formulation. The first separates visual quality from spatial quality: a confident detection may still require review if telemetry, map matching, or road identity is uncertain. The second principle standardizes how different visual outputs are converted into spatial evidence: detections, scene-level classifications, and traffic-sign results all feed the same geospatial aggregation layer, allowing new modules to be added without redesigning the segment-level logic. The third treats curation and label formulation as part of the method: vehicle motion, rain, pavement type, official

maintenance indicators, and labeling history guide the selection of candidate segments and the frames associated with training and inference. Table 1 summarizes how these principles address operational constraints in large-scale road monitoring.

Table 1. Innovation mechanisms embedded in the proposed road-monitoring system. The table links operational constraints to architectural decisions and expected impact.

Innovation axis	Operational challenge	System mechanism	Impact on road management
Evidence-centered monitoring	Model outputs are weak if detached from images, time, and location.	Each prediction is linked to visual evidence, telemetry, vehicle context, and a road segment.	Enables auditing, review, and institutional trust in automated indicators.
Temporal reliability	Video streams may include clock offsets, missing frames, and repeated observations of the same object.	Frame-time alignment, confidence gates, temporal tracking, and deduplication.	Reduces duplicated events and separates visual uncertainty from temporal uncertainty.
Multitask extensibility	Pavement defects, markings, signs, and vegetation require different visual formulations.	Detection, classification, and multi-stage pipelines feed the same spatial aggregation layer.	Allows new visual modules without redesigning the geospatial core.
Official-network integration	Raw GNSS trajectories do not directly correspond to official road segments.	Map matching against federal and state road references with kilometer-level attribution.	Produces indicators compatible with road-management language and official inventories.
Human-in-the-loop correction	Coincident roads, unmapped stretches, and ambiguous evidence cannot be fully solved by automation.	Manual validation, road-identity correction, and recoverable visual records.	Supports operational deployment under imperfect data and evolving road databases.

4. Proposed System

4.1. Operational Setting and Data Curation

The system considers transportation company buses equipped with frontal cameras and GNSS telemetry, operating over long and visually heterogeneous routes. Since the videos were collected during regular service rather than under controlled imaging conditions for AI training, the data contain low-resolution frames, motion blur, occlusions, lighting variation, stops, rain, and scenes outside the road corridor. The methodological challenge is therefore to convert operational video and telemetry streams into decisions associated with time, position, road identity, and segment, despite missing frames, temporal offsets, rare classes, high-speed motion, pavement-type variability, and stretches without a clear official-network match.

Data curation is treated as a method step, not merely preprocessing. From 584.6 million records in two telemetry sources, entries lacking vehicle, timestamp, or valid geometry were removed, identifiers normalized, and duplicates suppressed. Points were restricted to executed trip windows; the two densest operations per line yielded 23.6 million points across 512 trips and 259 lines from April 1–26, 2026. The densest trajectory per line was then map-matched to reduce redundancy while preserving network coverage.

Frames with low road-monitoring value are removed before labeling or inference. Stationary-bus imagery is identified primarily from telemetry, using speeds below 0.1

km/h, and complemented by dense optical-flow analysis when visual motion evidence is needed. This strategy follows Farneback's dense optical-flow formulation, which estimates pixel-level motion between consecutive frames through local polynomial expansion [Farnebäck 2003]. It reduces stationary scenes from terminals, garages, and prolonged stops, which otherwise overrepresent contexts unrelated to the monitored road corridor.

Rain, garage, construction, and non-representative pavement intervals are manually registered by vehicle, date, frame range, and exclusion reason, then removed from labeling and training. DNIT severity indicators, together with IP, IC, and ICM, guide where human labeling effort is more likely to find relevant defects, but are never used as image labels. This distinction is important because annotation defines what each model is expected to learn; for example, block-paved or cobblestone surfaces may introduce joints and discontinuities that resemble cracks, increasing class-boundary ambiguity.

4.2. End-to-End Architecture

The architecture is organized into operational layers. Acquisition receives videos and telemetry; synchronization generates image–position observations; multimodal inference applies detection, classification, and staged pipelines; temporal validation reduces duplicates; geospatial integration converts events into road-segment indicators; and validation interfaces expose maps, images, metrics, and correction mechanisms. Figure 1 presents this end-to-end workflow.

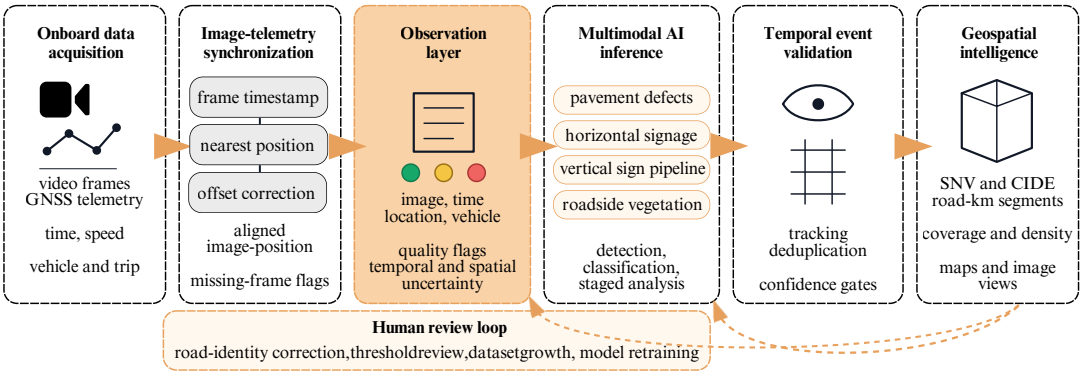


Figure 1. End-to-end road-monitoring workflow linking onboard acquisition, synchronization, multimodal inference, temporal validation, geospatial attribution, and human review.

The design decouples visual and geospatial layers: telemetry normalization, motion filtering, inference, temporal consolidation, map matching, and aggregation are automated, while humans validate temporal offsets, register excluded visual intervals, review errors, resolve ambiguous road assignments, and audit evidence. Models can thus evolve without changing attribution logic.

4.3. Image–Telemetry Synchronization

Let f_i be the frame number, r the frame rate, t_0 the estimated video-start time, and Δ the temporal offset between camera and telemetry clocks. The timestamp associated with frame i is

$$\hat{t}_i = t_0 + \frac{f_i}{r} + \Delta, \tag{1}$$

where $\hat{t}_i$ is the estimated timestamp of frame i, f_i/r converts frame index into elapsed video time, and Δ corrects sensor misalignment. Once $\hat{t}_i$ is available, the nearest telemetry observation is selected as

$$p_i = \arg\min_{p_j \in P} |\tau_j - \hat{t}_i|, \tag{2}$$

where P is the set of GNSS observations for the trip, $p_j = (lat_j, lon_j, v_j, \tau_j)$ contains latitude, longitude, speed, and timestamp, and $|\tau_j - \hat{t}_i|$ measures the temporal distance between telemetry and frame. The synchronization stage also records missing images, absent frames, and camera- or vehicle-specific offsets, so ambiguous observations can be discarded or flagged.

4.4. Multimodal Inference and Temporal Persistence

The system integrates independently trained, task-specific YOLO detection and classification modules, using temporal splits and task-specific hyperparameter selection. Pavement defects and sign types are localized for counting, deduplication, inventory, and segment density; horizontal markings and vegetation represent scene condition, while sign quality is classified on detected crops. Traffic signs follow a staged pipeline: a single-class detector first retrieves candidate images with signs from large video collections, reducing the labeling burden; the operational outputs are then produced by a sign-type detector and a quality classifier applied to sign crops.

Temporal persistence converts frame-level predictions into consolidated events. For pavement defects, tracking uses the expected vertical motion of road-surface objects as they approach the camera. For objects without this motion pattern, such as signs and lateral assets, association uses centroid proximity and persistence in consecutive frames. The event-validation rule is

$$1(o) = \{\, 1 \,, n_o \geq m \;\wedge\; \bar{c}_o \geq \theta_{c(o)}, 0, otherwise. \tag{3}$$

where o is a candidate event, n_o is the number of frames in which it remains associated, m is the minimum required persistence, $\bar{c}_o$ is the average track confidence, and $\theta_{c(o)}$ is the threshold for class $c(o)$. This rule reduces isolated false positives and transforms frame-by-frame outputs into event records.

For pavement defects, operational inference uses class-specific confidence thresholds defined through sensitivity analysis on the inference set: 0.60 for potholes, 0.70 for patches, and 0.30 for cracks. A low global threshold increased patch noise, especially on shadows, pavement stains, and surface irregularities, whereas a high global threshold suppressed cracks, which often appeared with lower confidence due to their subtle morphology, texture dependence, and pavement-type variability. Class-specific thresholds therefore preserve crack recall while requiring stronger evidence for classes more prone to false positives. High vehicle speed may also produce blur, streak-like marks, or reduced temporal continuity between consecutive frames, reinforcing the need for temporal persistence and class-specific decision rules.

These class-specific thresholds are propagated to the operational audit layer, where reviewers can inspect the original evidence and adjust the decision filters used for pavement-defect visualization. The normalized output of each visual module contains

time, location, class, confidence, and links to visual evidence. This common event representation allows detection, classification, and staged analysis to feed the same geospatial layer.

5. Geospatial Integration and Segment-Level Indicators

Geospatial integration has two phases. First, raw GNSS trajectories are matched to OSM using GraphHopper [GraphHopper Project 2026], with mmlib providing the interface used to submit trajectories and recover matched route geometries [Dias de Souza 2026]. Second, the matched trajectory is attributed to official references, including SNV and CIDE, to generate road identity, kilometer ranges, coverage, and segment-level indicators. Figure 2 summarizes this workflow, highlighting the separation between OSM map matching, official attribution, and operational refinement of ambiguous road identity.

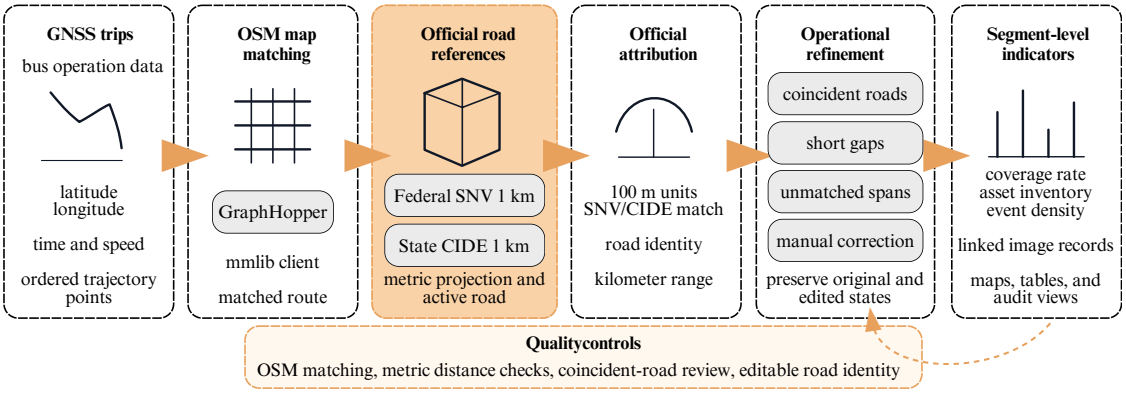


Figure 2. Geospatial integration workflow. GNSS trajectories are matched to OSM, attributed to federal and state road references, refined under coincident-road and short-gap cases, and converted into segment-level indicators linked to image records.

The spatial attribution of each trip unit is

$$s_i^* = \arg \min_{s \in S,\, d(g_i,s) \leq \rho} d(g_i, s), \tag{4}$$

where g_i is a metric road segment of the trip, S is the set of candidate official-road segments, $d(g_i, s)$ is the distance between both geometries, ρ is the maximum search radius, and s_i^* is the selected road segment. The matched trajectory is transformed to a metric coordinate system, subdivided into 100 m units, and attributed using empirically configured, non-normative tolerances: up to 100 m for SNV and 30 m for CIDE. These values accommodate observed displacement between GNSS trajectories and reference geometries; controlled sensitivity analysis remains future work.

The attribution is refined beyond nearest-neighbor assignment. When SNV reports coincident roads, identifiers associated with the same physical axis are preserved, while the main road is used only to group contiguous stretches. Short gaps between two stretches assigned to the same road may inherit that identity, reducing artificial fragmentation caused by local ambiguity. Original and corrected states are preserved so automated decisions and manual edits remain auditable.

Once segments are attributed, visual indicators are aggregated by count or density as

$$D_{k,c} = \frac{1}{L_k} \sum_{o \in O_c} 1(o \in s_k), \tag{5}$$

where $D_{k,c}$ is the density of class c in segment k, L_k is the segment length in kilometers, O_c is the set of validated events of class c, and $1(o \in s_k)$ indicates whether event o belongs to segment s_k. This aggregation transforms visual observations into indicators compatible with official maintenance surveys. Consequently, comparison with DNIT data occurs at the road-segment level, not at the isolated-image level.

6. Experimental Results and Operational Validation

Table 2 summarizes the four visual tasks, their data basis, performance, and operational outputs. Detection performance is reported using precision and mAP, while classification accuracy is retained for cross-cycle comparability and interpreted within each task. Because the tasks use different formulations and evaluation protocols, their metrics should not be interpreted as a direct ranking across modules. Pavement defects require localized event detection, markings and vegetation represent scene-level condition, and traffic-sign monitoring combines type detection with crop-level quality classification.

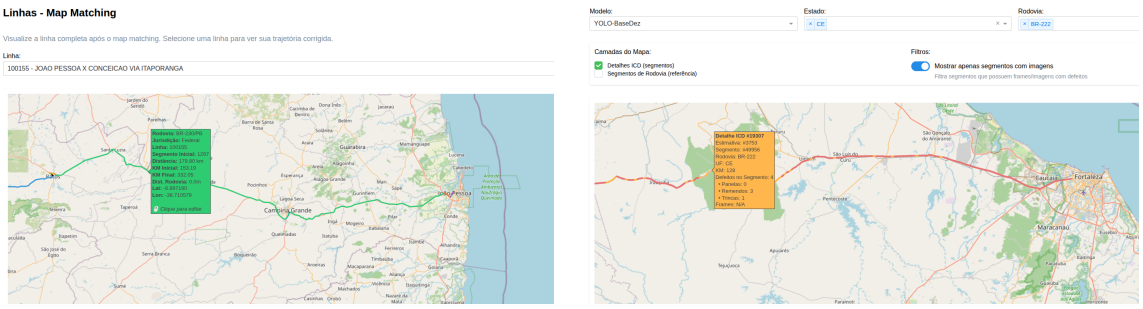
Table 2. Visual modules integrated into the operational monitoring system. Metrics summarize the experimental performance of each visual task and its operational role in segment-level road management.

Visual task	Learning formulation	Evaluation base	Result	Operational output
Pavement defects	Object detection: crack, patch, pothole	25,795 reviewed images; 4,212 annotated images; 11,072 boxes	Precision = 64.05%; $mAP_{@50--95}$ = 38.00%	Defect inference, counting, deduplication, and density per road segment.
Horizontal road markings	Scene classification: good, regular, poor, missing	26,951 labeled images	Accuracy = 85.96%	Segment-level conservation indicator for road markings.
Traffic signs	Candidate retrieval + type detection + quality classification	75,923 reviewed images; 27,036 annotated images; 32,163 boxes; 31,175 quality crop labels	Candidate retrieval: $mAP_{@50}$ = 99.38%; type detection: $mAP_{@50}$ = 97.57%; quality accuracy = 97.54%	Assisted sign inventory by type and quality, with recoverable image evidence.
Roadside vegetation	Scene classification: good, poor, not applicable; extension to regular	26,954 labeled images under class-structure review	Accuracy = 75.23% in the optimized model	Vegetation-condition indicator and potential obstruction assessment.

For the final pavement-defect cycle, 25,795 images were reviewed, yielding 4,212 annotated images and 11,072 boxes; horizontal-marking and optimized vegetation cycles used 26,951 and 26,954 image-level labels. For traffic signs, the 99.38% mAP@50 candidate detector served only to reduce manual review, while operational outputs came from type detection (27,036 images; 32,163 boxes) and quality classification (31,175 crops). Sign-quality data comprised 30,099 good and 1,076 bad crops; the good class reached 99.74% precision and 97.70% recall, whereas the bad class reached 59.77% precision

and 93.18% recall, reflecting high sensitivity to degraded signs but more false positives. Across tasks, 6.26 million frame-level model applications were recorded, potentially reusing frames across modules. The latest pavement run processed 877,070 frames from videos recorded by one bus between June 28 and July 14, 2025, yielding 37,203 predicted events: 2,143 potholes, 25,022 patches, and 10,038 cracks. After temporal consolidation, these events were aggregated by segment and should be interpreted as system outputs rather than verified field prevalence.

Contiguous frames remain within one train, validation, or test split to limit leakage. Each model version preserves data origin, split, configuration, metrics, binary reference, and annotation links; recurrent errors are reviewed with neighboring frames and accepted corrections form new labeling cycles. Motion filtering removes stationary scenes, exclusions remove rain and unsuitable contexts, and DNIT-based prioritization focuses review without replacing human labels. Validation combines held-out temporal performance with operational traceability from image evidence through telemetry and map matching to official segments. Figure 3 shows matched routes and segment-level aggregation by road, state, kilometer range, and predicted counts. Current operational validation establishes traceability rather than event-level accuracy; estimating operational precision and recall requires manual annotation of sampled clips.



(a) Line-level map matching against official road references.

(b) Segment-level pavement-defect aggregation with map tooltip.

Figure 3. Spatial validation interfaces linking map matching, official road references, and segment-level pavement-defect estimates.

The segment map shows how visual detections become spatial readings of a road stretch, but operational use still requires inspection of the evidence behind each count. Figure 4 shows the audit interface for a selected segment, including the original frame, bounding box, confidence score, class-specific confidence filters, defect-type toggles, and segment metadata. This view closes the audit loop: the user can start from an official segment, inspect estimated defects, and recover the visual observations that produced the aggregate.

7. Discussion

The results show that large-scale road monitoring is a system-level problem rather than only a model-selection problem. Performance depends not only on the selected visual models, but also on acquisition conditions, synchronization, temporal consolidation, and spatial attribution. The visual modules reached different levels of maturity because each task imposes distinct constraints: traffic-sign detection and quality assessment benefit

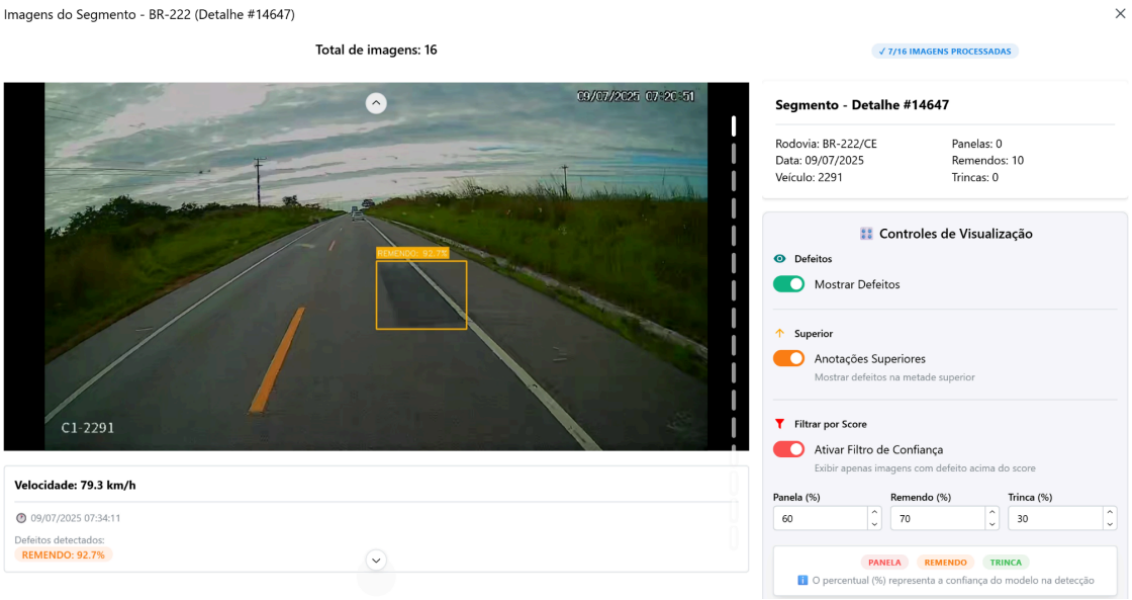


Figure 4. Visual audit interface for a selected road segment. The evidence view exposes the original frame, detected defect, model confidence, segment metadata, defect counts, visualization toggles, and class-specific confidence thresholds used to filter pothole, patch, and crack detections.

from compact targets and focused crops, whereas pavement-defect detection is more affected by texture variation, illumination, pavement type, motion artifacts, and subtle class boundaries. This reinforces the main premise of the work: operational value emerges when heterogeneous predictions become localized, auditable, and segment-level road evidence.

The transportation company case study illustrates this challenge under real deployment conditions. Regular bus operation expands coverage and reduces collection cost, but it also introduces repeated observations, stationary scenes, missing frames, telemetry irregularities, temporal offsets, and variable image quality. Therefore, data curation, task-specific model formulation, temporal persistence, and geospatial attribution must operate together to convert frame-level outputs into usable infrastructure indicators. Map matching and attribution to official references allow model outputs to be compared with DNIT information at the segment level rather than at the isolated-image level, which is essential when coincident roads, short gaps, incomplete geometries, and manual corrections must remain traceable.

The reported scale demonstrates feasibility within the current corpus, not throughput beyond it. Generalization remains limited to one operator and selected Brazilian corridors, while rain exclusion restricts adverse-weather claims. Synchronization depends on camera offsets and telemetry quality. Map matching may fail in urban corridors, ramps, parallel lanes, or incomplete road databases. Rare classes, such as low-quality traffic signs, remain imbalanced. Roadside-vegetation classification depends on indirect visual cues of height, distance, and obstruction. Older low-resolution frames limit the legibility of small assets. These limitations are mitigated through curation filters, spatial tolerances, versioned states, and incremental expansion of training data, but they remain important targets for future work.

8. Conclusion

This paper presented an end-to-end multimodal road-monitoring system integrating onboard video, telemetry, visual models, temporal reasoning, and official road references. It reduced 584.6 million telemetry records to a 23.6-million-point candidate pool and supported 6.26 million frame-level model applications. The latest pavement run converted 877,070 frames into 37,203 predicted events for segment aggregation. The tasks showed complementary maturity: traffic-sign type detection reached 97.57% mAP@50 and quality classification 97.54% accuracy, horizontal markings 85.96% accuracy, roadside vegetation 84.29% accuracy, and pavement defects 38.00% mAP@50–95. The contribution is to show how computer vision, telemetry, official road references, and validation can produce actionable road-management indicators at operational scale.

Future work will estimate operational precision and recall on manually annotated clips, evaluate higher-frequency telemetry, enrich minority classes, analyze spatial-tolerance sensitivity, and transfer the system to other regions. It will also strengthen synchronization with updated DNIT/SNV and CIDE references, move toward near-real-time processing, generate executive segment-level reports, and longitudinal analysis to track infrastructure-condition changes over time and support faster operational decision-making. GraphHopper and mmlib are public; access to the complete prototype, versioned experimental artifacts, and field imagery remains governed by the applicable operational data-use agreement.

Declaration

For open access purposes, the authors have applied a Creative Commons Attribution (CC BY) licence to any accepted version of this manuscript.

9. References

References

Bruno, S., De Amicis, A., Loprencipe, G., Orlandini, S., Rotondi, G., and Vita, L. (2025). Advancing road infrastructure: Cross-slope analysis of road sections using a mobile mapping system (mms). *Transportation Research Procedia*, 90:456–463.

Chen, Z., Zhang, Y., Luo, Y., Wang, Z., Zhong, J., and Southon, A. (2021). Roadatlas: Intelligent platform for automated road defect detection and asset management. In *Proceedings of the 3rd ACM International Conference on Multimedia in Asia*, pages 1–3.

Dias de Souza, J. V. (2026). mmlib: Interoperability framework for map matching. `https://github.com/JoseEdSouza/mmlib`. Accessed 13 May 2026.

DNIT (2024). Índice de condição da manutenção. `https://www.gov.br/dnit/pt-br/assuntos/infraestrutura-rodoviaria/icm`. Page last updated on 9 October 2024.

DNIT (2026). Plano nacional de viação e sistema nacional de viação. `https://www.gov.br/dnit/pt-br/servicos/plano-nacional-de-viacao-e-sistema-nacional-de-viacao`. Page last updated on 26 February 2026.

Dong, S., Wang, Y., Cao, J., Ma, J., Chen, Y., and Kang, X. (2025). Advanced lightweight deep learning vision framework for efficient pavement damage identification. *Scientific Reports*, 15(1):12966.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Scandinavian Conference on Image Analysis*, pages 363–370. Springer.

GraphHopper Project (2026). Graphhopper routing engine. `https://github.com/graphhopper/graphhopper`. Accessed 13 May 2026.

Han, C., Yang, S. B., and Hu, J. (2026). Diffmm: Efficient method for accurate noisy and sparse trajectory map matching via one step diffusion. *arXiv preprint arXiv:2601.08482*.

IDB (2025). Pavimentados: Library for detecting cracks and failures on roads. `https://github.com/EL-BID/pavimentados`. Inter-American Development Bank. Open-source software repository. Accessed 13 May 2026.

Katsamenis, I., Bakalos, N., Lappas, A., Protopapadakis, E., Martín-Portugués Montoliu, C., Doulamis, A., Doulamis, N., Rallis, I., and Kalogeras, D. (2025). Dorie: Dataset of road infrastructure elements—a benchmark of yolo architectures for real-time patrol vehicle monitoring. *Sensors*, 25(21):6653.

Kulawik, J. and Kuczyński, Ł. (2025). Ai-based detection and classification of horizontal road markings in digital images dedicated to driver assistance systems. *Applied Sciences*, 15(22):12189.

Ma, L. and Chen, M. (2025). Road damage detection based on improved yolo algorithm. *Scientific Reports*, 15(1):28506.

Sheikholeslami, D., Golroo, A., Khodaii, A., Daneshvar, A., and Moghadas Nejad, F. (2026). A physic-guided yolo framework for pavement deformation distress detection. *Scientific Reports*.

Stipancic, J., Saunier, N., Navidi, N., Racine, E. B., Miranda-Moreno, L., and Labbe, A. (2025). Evaluating map matching algorithms for smartphone gnss data: Matching vehicle trajectories to an urban road network. *Transportation Research Procedia*, 82:303–322.

Wang, C., Zheng, B., and Li, C. (2025). Efficient traffic sign recognition using yolo for intelligent transport systems. *Scientific reports*, 15(1):13657.

Zhang, H., Liang, M., and Wang, Y. (2025a). Yolo-bs: a traffic sign detection algorithm based on yolov8. *Scientific Reports*, 15(1):7558.

Zhang, Y., Yan, H., and Lu, X. (2025b). An enhanced hmm map matching algorithm incorporating personal road selection preferences. *Scientific Reports*, 15(1):35962.

Zhumadillayeva, A., Ahanger, T. A., and Matkarimov, B. (2025). An intelligent yolo and cnn-bigru framework for road infrastructure based anomaly assessment. *Scientific Reports*, 15(1):41193.