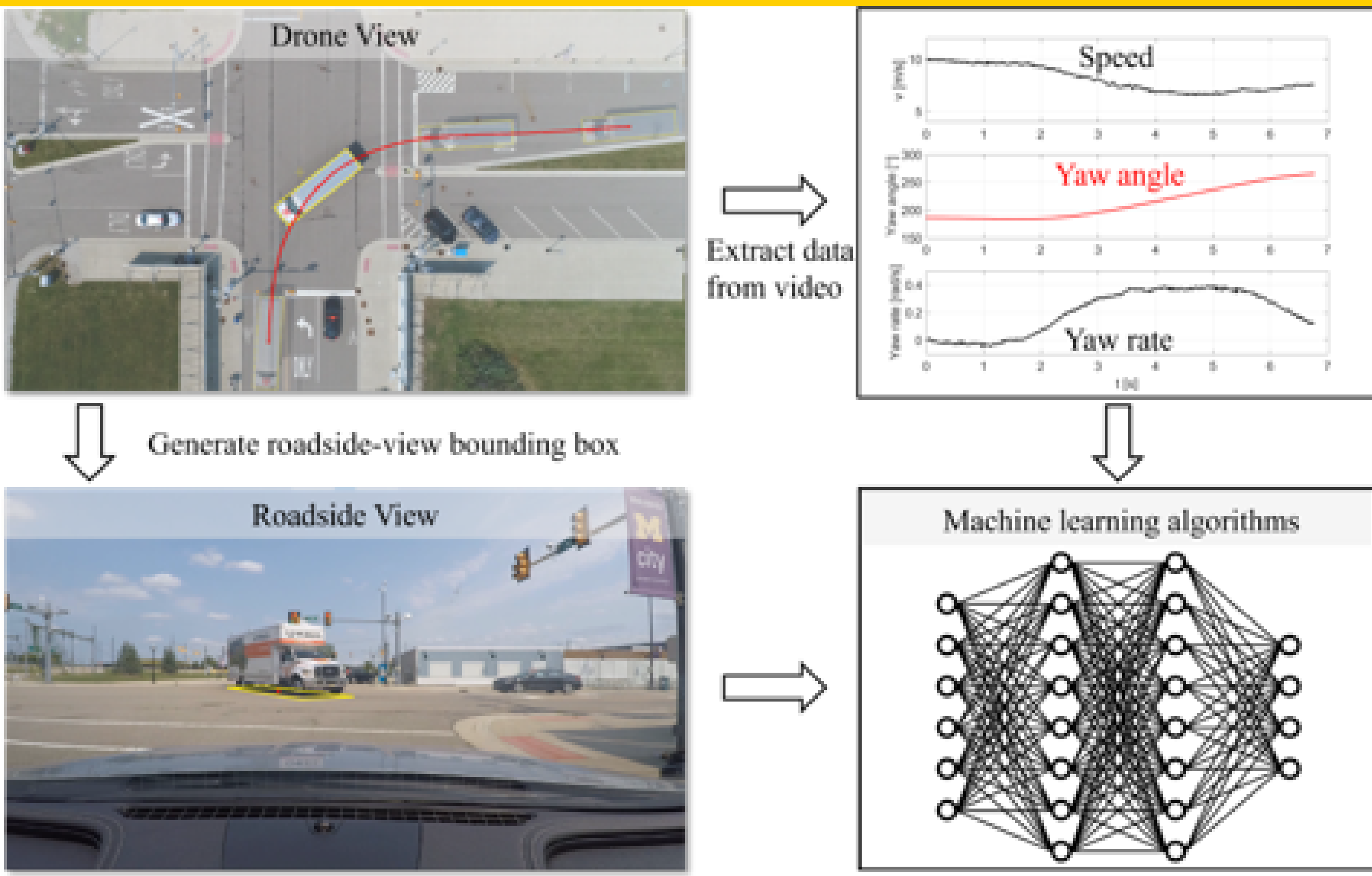
CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Final Report #89
September 2026



Generating High-Accuracy Transportation Datasets with Unmanned Aerial Vehicles

Ákos T. Köpeczi-Bócz
Tian Mi
Dénes Takács
Gábor Orosz

UNIVERSITY OF MICHIGAN

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Report No. 89
Project Start Date: June 1, 2023
Project End Date: August 31, 2026
September 2026

Generating high-accuracy transportation datasets with unmanned aerial vehicles

by
Ákos T. Köpeczi-Bócz,
Tian Mi, Dénes Takács,
and Gábor Orosz
University of Michigan,
Ann Arbor

DISCLAIMER

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant No. 69A3552348305 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Suggested APA Format Citation: Köpeczi-Bócz, Á.T., Mi, T., Takács, D., Orosz, G. (2026). Report 89. Generating high-accuracy transportation datasets with unmanned aerial vehicles. Center for Connected and Automated Transportation. Final Report.
DOI: https://doi.org/10.7302/dspace/30322

Contacts

For more information:

Gabor Orosz
1231 Beal Ave
Ann Arbor, MI 48109
Phone: 734 763 0567
Email: orosz@umich.edu

CCAT
University of Michigan Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48109
umtri-ccat@umich.edu
(734) 763-2498

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
CCAT Report 89	Leave blank – not used	Leave blank - not used

4. Title and Subtitle
Generating high-accuracy transportation datasets with unmanned aerial vehicles
DOI: https://doi.org/10.7302/dspace/30322

5. Report Date
September 2, 2026

6. Performing Organization Code
Enter any/all unique numbers assigned to the performing organization, if applicable.

7. Author(s)
Ákos T. Köpeczi-Bócz: 0009-0009-7171-7037
Tian Mi: 0000-0002-3780-3216
Dénes Takács: 0000-0003-1226-8613
Gábor Orosz: 0000-0002-9000-3736

8. Performing Organization Report No.
Enter any/all unique alphanumeric report numbers assigned by the performing organization, if applicable.

9. Performing Organization Name and Address
Center for Connected and Automated Transportation
University of Michigan Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48109

University of Michigan Mechanical Engineering Department
1231 Beal Ave
Ann Arbor, MI 48109

10. Work Unit No.

11. Contract or Grant No.
Contract No. 69A3552348305

12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
1200 New Jersey Avenue, SE
Washington, DC 20590

13. Type of Report and Period Covered
Final Report (June 2023 – August 2026)

14. Sponsoring Agency Code
OST-R

15. Supplementary Notes
Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.

16. Abstract
A new trend in the area of connected and automated vehicles (CAVs) is infrastructure-based sensing and tracking. By installing roadside cameras at the infrastructure, one can monitor the overall traffic situation and, after processing the collected data via edge computing, the information can be communicated to CAVs via V2X connectivity. This has a huge potential to improve traffic safety and efficiency. We identify two main challenges to deploy such systems: (i) to obtain high precision data and (ii) to label the obtained data. It is necessary to have sufficient high-precision labelled datasets for training the underlying machine learning (ML) algorithms to detect, identify, localize and track the road participants. In other words, one needs to know the "ground truth" (with high precision) for large variety of different scenarios. As of now, this may be done by hand-labeling images, which is immensely labor intensive, or by using probe vehicles equipped with high precision GPS, which can only provide data about a few specific vehicles.

17. **Key Words**			18. **Distribution Statement**
Vehicle localization, infrastructure-based perception,monocular 3D detection, unmanned aerial vehicle, Imageprocessing, Connected and Autonomous Vehicles			No restrictions.
19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	12	Leave blank – not used

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Final Project Report

Table of Contents

List of Figures

List of Tables

1. Introduction

Intersections are among the most hazardous locations in the road network, accounting for roughly one-quarter of all traffic fatalities and about half of all traffic injuries in the United States. Improving situational awareness at these locations is a key enabler for both traditional traffic safety programs and emerging intelligent transportation systems (ITS) that rely on vehicle-to-infrastructure (V2I) communication and connected and automated vehicles (CAVs). This project addressed a central bottleneck in deploying infrastructure-based perception at scale: existing roadside-sensing approaches typically require expensive multi-sensor rigs (e.g., LiDAR) or complex multi-stage image-processing pipelines that first detect vehicles in the camera's image plane and only then estimate their real-world position and orientation through separate geometric post-processing.

The project's overarching objective was to develop and experimentally validate a low-cost, camera-only methodology that can recover full vehicle kinematic states — position, dimensions, orientation (yaw angle), velocity, and yaw rate — directly from a single, fixed roadside camera, without requiring LiDAR, multiple cameras, or onboard vehicle instrumentation. To achieve this, the project combined three complementary lines of work:

- Developing a high-precision method to extract "true," oriented bounding-box kinematic data for road vehicles from unmanned aerial vehicle (UAV) video, to be used as ground truth for machine learning.
- Designing and training a modified one-stage AI detector that maps roadside-camera images directly to vehicle states expressed in a ground-fixed coordinate frame, removing the need for a separate geometric post-processing stage.
- Validating the complete pipeline experimentally at the Mcity Test Facility (University of Michigan, Ann Arbor) using synchronized drone and roadside-camera measurements across multiple vehicles and driving scenarios.

This research produced two peer-reviewed/submitted journal papers, one conference paper presented at an international symposium, and two U.S. patent applications protecting the underlying methodologies. The remainder of this report presents the findings from this work, our recommendations for future research and technology transfer, and the formal documentation of outputs, outcomes, and impacts required for final reporting.

2. Findings

2.1 UAV-based ground-truth extraction methodology

The project first established a reliable way to generate high-fidelity, oriented bounding-box ground truth data for vehicles using drone video, since no adequate labeled dataset previously existed for training roadside camera-based perception models with full kinematic states (position, dimensions, and orientation). A DJI Phantom 4 Pro drone equipped with a 3-axis gimbal camera was flown at approximately 75 m (250 ft) altitude above test intersections at the Mcity Test Facility, University of Michigan, recording at 60 frames per second in 4K resolution.

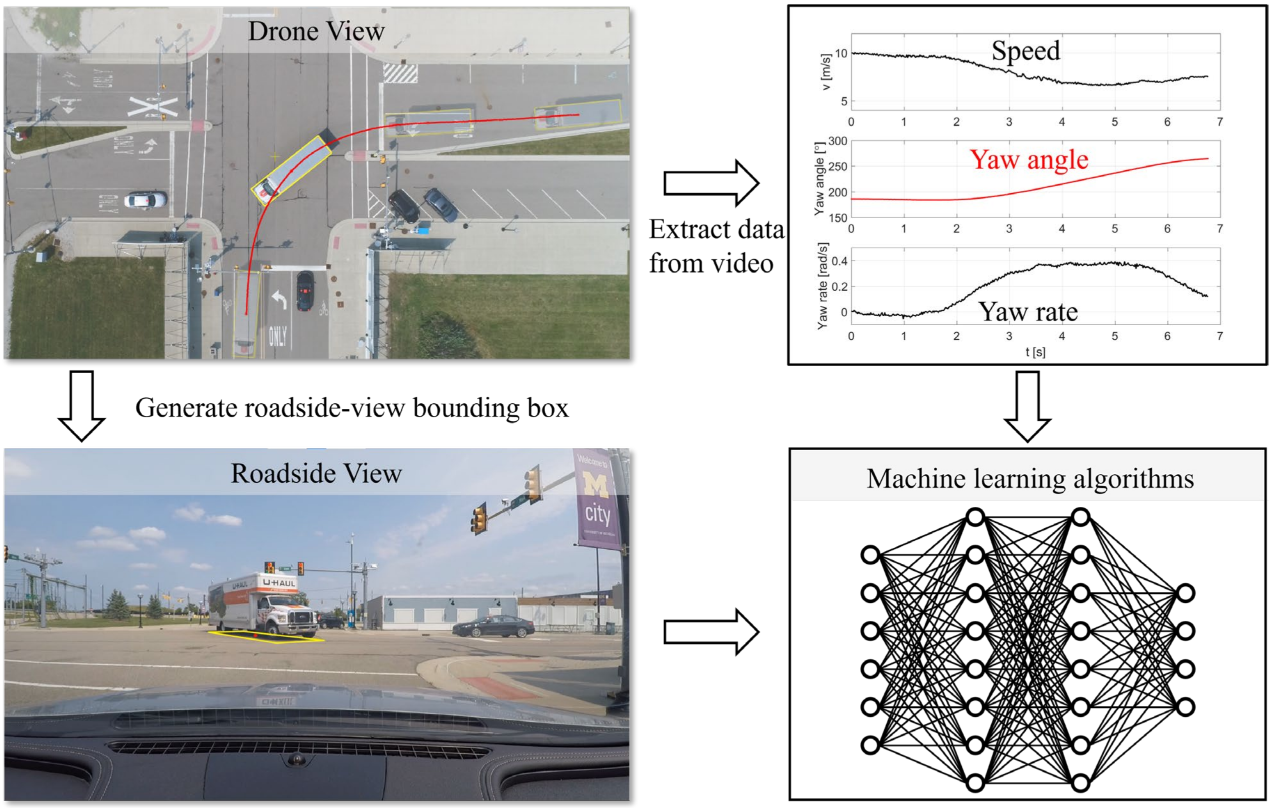


Figure 1: UAV-based data collection framework for generating datasets for machine-learning algorithms.

Key technical findings from this line of work include:

- Image stabilization combined with template-matching/correlation-based vehicle detection (using rotated-image cross-correlation) recovers vehicle position and yaw angle from drone footage with sub-pixel precision; one pixel of position error corresponds to roughly 0.0275 m on the ground.
- Direct numerical differentiation of raw position data to obtain velocity is highly noise-sensitive (a single pixel of jitter can translate into an artificial 1.65 m/s speed error at 60 fps); applying a moving-average or Savitzky–Golay filter (order 2, ~0.68 s window) is necessary to obtain physically consistent velocity and yaw-rate estimates.
- Comparison against onboard RTK-grade GPS confirmed that the UAV-derived trajectories are smoother and richer in kinematic detail (providing heading angle, slip angle, and yaw rate in addition to position and speed) than GPS alone, and do not require any additional instrumentation on the tracked vehicles.

- A kinematic bicycle-model analysis showed that the lateral velocity (slip angle) is close to zero at the vehicle's rear axle, confirming that the extracted yaw angle is a physically valid proxy for heading at that reference point — an important validation step for downstream motion-prediction and control applications.
- A homography transformation was derived to project the drone-view (top-view) oriented bounding boxes into the roadside camera's image plane, enabling one drone flight to automatically generate labeled training data for an arbitrary number of ground-based roadside cameras observing the same intersection.

2.2 One-stage monocular roadside vehicle-localization network ("YOLOgraphy")

Building on the UAV-generated ground truth, the project developed a novel one-stage deep-learning framework — internally named "YOLOgraphy" — that maps a single roadside-camera image directly to each vehicle's ground-plane position, dimensions, and yaw angle, without an intermediate image-plane detection or geometric post-processing step.

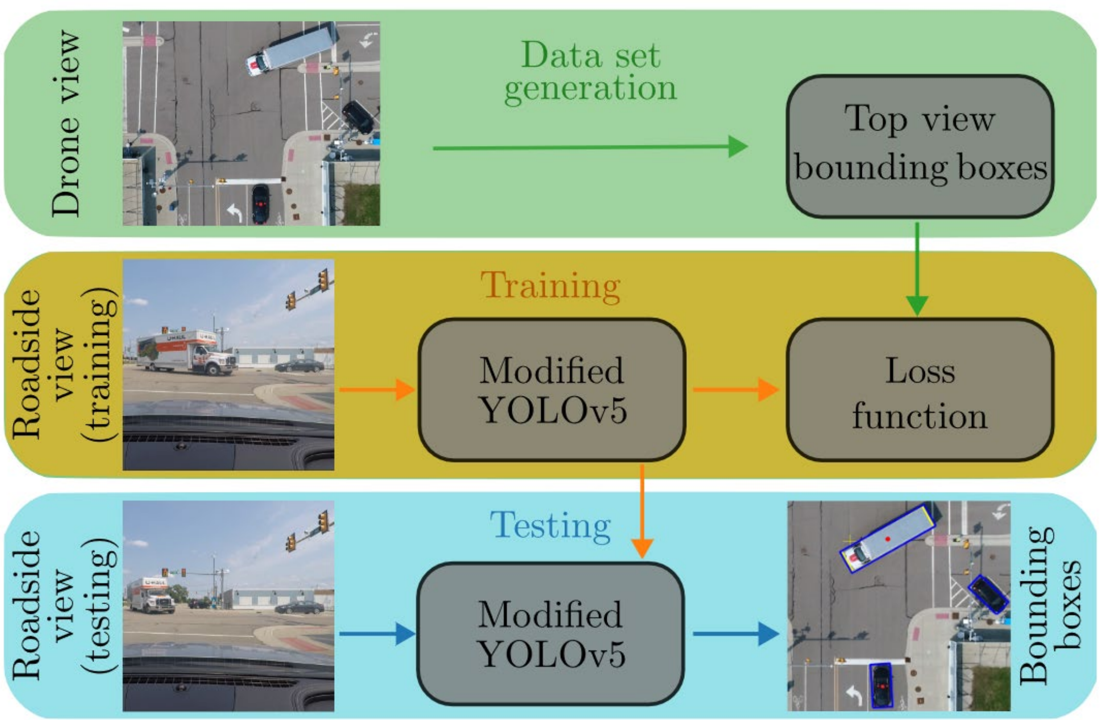


Figure 2: Training framework of a one-stage vehicle detector for monocular 3D localiaztion.

Principal findings include:

- Starting from the pretrained YOLOv5 object detector, only the final detection head (Layer 25 in the network's 25-layer structure) needed to be redesigned; the convolutional backbone and Path Aggregation Network neck were reused largely unchanged, preserving transferable visual features learned from large-scale pretraining.
- The detection head was reduced to a single object class and re-parameterized to output position, width, length, and yaw angle directly in the ground-fixed coordinate frame, replacing the conventional image-plane bounding-box parameterization.
- Systematic experiments freezing different numbers of pretrained backbone layers showed that freezing layers 1–2 gave the best trade-off, yielding a 5% reduction in validation box loss relative to a fully unfrozen network (1.8855 vs. 1.9788), while freezing beyond layer 10 significantly degraded training performance because too many task-relevant parameters were excluded from optimization.
- The trained network was evaluated at three levels of generalization: (1) on training/validation data, (2) on unseen runs of familiar scenarios ("in-context" generalization), and (3) on a completely unseen maneuver scenario ("out-of-context" extrapolation). The model accurately recovered vehicle trajectories and yaw angles for in-context test data, confirming the viability of learning the image-to-ground-plane mapping implicitly through network weights, while performance degraded predictably when extrapolating to unrepresented trajectories, revealing a clear boundary condition for safe deployment.

2.3 Experimental validation and dataset

The complete pipeline was validated using data from four designed traffic scenarios (straight crossings, left/right turns, car-following, and vehicle-interaction maneuvers), each repeated five times, for a total of 20 synchronized drone/roadside-camera measurement runs involving five vehicles. This dataset construction and experimental protocol underpin both journal papers and the conference paper, and demonstrate a scalable, repeatable methodology that other researchers or agencies could replicate at additional intersections using only a drone and a single fixed camera — substantially lower-cost instrumentation than LiDAR-based alternatives.

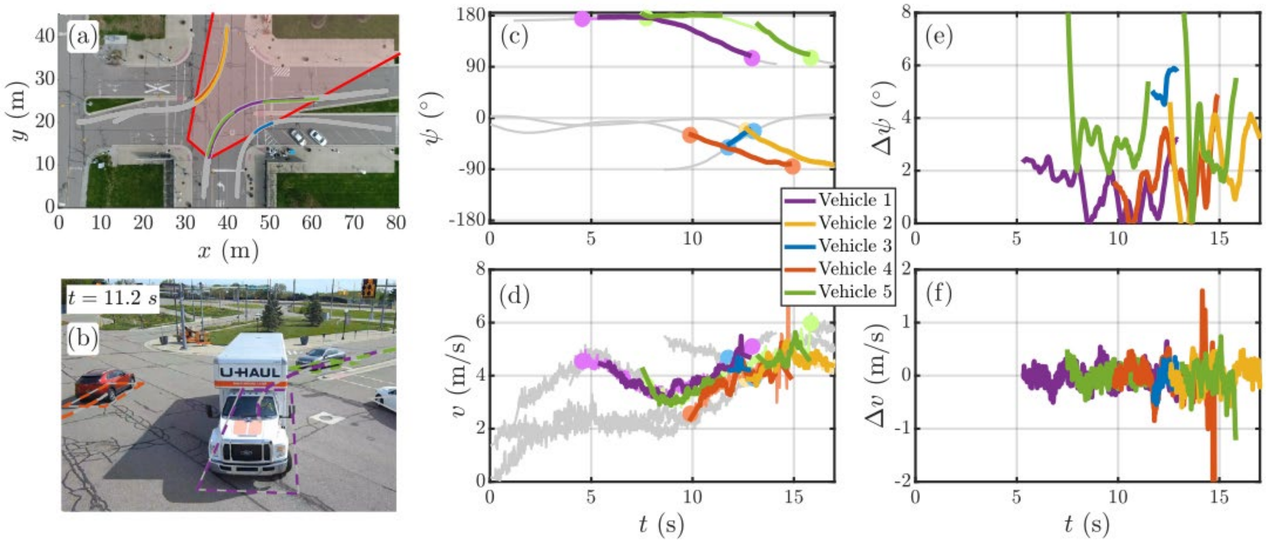


Figure 3: Data-set and detection results for an experiment in the dataset.

3. Recommendations

Based on the findings summarized above, the project team offers the following recommendations for future research, deployment, and technology transfer:

- **Expand training data to a real traffic intersection.** All data collected under this project were captured at the closed, controlled Mcity Test Facility with a limited set of designed maneuvers and vehicles; we recommend extending data collection to a real, open-traffic intersection to capture the greater diversity of vehicle types, driving behaviors, traffic density, and lighting/weather conditions found in everyday traffic.
- **Investigate multi-camera fusion.** A single roadside camera can be occluded by large vehicles at close range; incorporating input from multiple roadside cameras observing the same intersection from different angles is recommended to increase robustness, particularly at high-traffic-density intersections.
- **Extend to kinematic filtering and prediction.** We recommend integrating vehicle kinematic models (e.g., the bicycle model used for ground-truth validation) as a post-processing or Kalman-filtering layer on top of the network's frame-by-frame outputs, to smooth trajectory estimates and support short-horizon motion prediction for CAV planning and control.
- **Broaden scenario coverage before field deployment.** Given the observed degradation in "out-of-context" extrapolation, we recommend deliberately including edge-case maneuvers (e.g., unusual

turning radii, emergency stops, and mixed vehicle classes) in future training datasets prior to any operational deployment at live intersections.

- **Pursue technology transfer through the filed patents.** The two patent applications filed under this project (see Section 4.2) represent tangible commercialization pathways; we recommend continued engagement with the University of Michigan's and BME's technology transfer offices to explore licensing to intersection-management, traffic-signal, or CAV-infrastructure vendors.
- **Adapt the approach to other one-stage detectors.** Because the core innovation (ground-plane output parameterization decoupled from the input image plane) is not specific to YOLOv5, we recommend evaluating the methodology on newer one-stage architectures (e.g., later YOLO versions with native oriented-bounding-box support) to assess potential accuracy and inference-speed gains.

4. Final Documentation of Outputs, Outcomes, and Impacts

4.1 Synopsis of Performance Indicators

The project met or exceeded its planned technical milestones. Over the reporting period, the team produced two archival journal publications, one peer-reviewed international conference paper with in-person presentation, and two U.S. patent applications protecting the ground-truth-extraction methodology and the monocular roadside vehicle-localization framework, respectively. The work also delivered a validated, reproducible 20-run experimental dataset collected at a dedicated connected/automated-vehicle test facility (Mcity), directly supporting the Part I/II performance indicators for research productivity, knowledge creation, and technology-transfer activity.

4.2 Output(s) Description

Table 1. Overview of project outputs

Output type	Title	Authors	Venue / Status	Date
Journal paper	Capturing the True Bounding Boxes: Vehicle Kinematic Data Extraction Using Unmanned Aerial Vehicles	Tian Mi, Dénes Takács, Henry Liu, Gábor Orosz	Journal of Intelligent Transportation Systems, vol. 29, no. 5, pp. 566–578	Published online 18 Apr 2024; DOI 10.1080/15472450.2024.2341395
Journal paper	Infrastructure-based Monocular 3D Vehicle Localization Framework with Experimental Validation	Ákos T. Köpeczi-Bócz, Tian Mi, Gábor Orosz, Dénes Takács	IEEE Transaction on Intelligent	Under review

Conference presentation & paper	YOLOgraphy: Image Processing Based Vehicle Position Recognition	Ákos T. Köpeczi-Bócz, Tian Mi, Gábor Orosz, Dénes Takács	AVEC (International Symposium on Advanced Vehicle Control), 2024	2024
Patent application	Vehicle Kinematic Data Extraction Using Unmanned Aerial Vehicles (algorithm for extracting kinematic variables — position, speed, yaw angle, yaw rate — and generating labeled ground-truth data sets, including oriented bounding boxes converted between drone and roadside camera views via homography)	Gabor Orosz, Tian Mi, Denes Takacs, Henry Liu	U.S. Patent Application Publication No. US2024/0386580 A1	Filed 15 May 2024; published 21 Nov 2024
Patent application	Infrastructure-based monocular vehicle localization / detection method (one-stage ground-plane vehicle-state estimation framework)	Project team (BME / University of Michigan)	U.S. Patent Application Publication No. US2026/0065685	Filed/published 2025–2026 (details per applicant's provided reference)

Full citations for the two journal papers:

- Mi, T., Takács, D., Liu, H., & Orosz, G. (2024). Capturing the true bounding boxes: vehicle kinematic data extraction using unmanned aerial vehicles. *Journal of Intelligent Transportation Systems*, 29(5), 566–578.
- Köpeczi-Bócz, Á. T., Mi, T., Orosz, G., & Takács, D. Infrastructure-based Monocular 3D Vehicle Localization Framework with Experimental Validation. *Under review in IEEE Transaction on Intelligent Transportation Systems*

Full citation for the conference paper:

- Köpeczi-Bócz, Á. T., Mi, T., Orosz, G., & Takács, D. (2024). YOLOgraphy: Image processing based vehicle position recognition. In *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC 2024)*.

Patents/inventions (technology transfer):

- U.S. Patent Application Publication No. US2024/0386580, "Vehicle Kinematic Data Extraction Using Unmanned Aerial Vehicles," inventors Gabor Orosz, Tian Mi, Denes Takacs, and Henry Liu; filed May 15, 2024; published November 21, 2024. The application discloses an algorithm to extract ground-vehicle kinematic variables (position, speed, yaw angle, yaw rate) from UAV footage and to generate labeled machine-learning data sets, including oriented top-view bounding boxes converted to a fixed roadside-camera perspective via homography, for use in intelligent-transportation planning and CAV control.
- U.S. Patent Application Publication No. US2026/0065685, "Vehicle Position Recognition Using Modified Neural Network" inventors Gabor Orosz, Tian Mi, Denes Takacs and Akos T. Kopeczi-Bocz; Covering the infrastructure-based, one-stage monocular vehicle-localization framework developed under this project, which maps roadside-camera images directly to ground-plane vehicle position, dimensions, and yaw angle using a modified one-stage object-detection network.

New methodologies, technologies, or techniques produced by this project:

- A synchronized UAV–roadside-camera data-collection and homography-based label-generation pipeline that removes the need for costly LiDAR instrumentation when building training datasets for infrastructure-based perception.
- A modified YOLOv5 detection-head architecture and custom loss function ("YOLOgraphy") that jointly predicts vehicle ground-plane position, size, and orientation in a single network pass, eliminating the separate geometric post-processing stage used in prior two-stage approaches.
- A validated experimental protocol (four scenarios × five repetitions, five vehicles) at the Mcity Test Facility that can serve as a template for future infrastructure-perception data collection efforts.

4.3 Outcome(s) and Impact(s) Description

Outcomes. The project's outputs have already produced several near-term outcomes consistent with UTC reporting objectives:

- *Increases in the body of knowledge*: two archival journal publications and one international conference paper document a new class of one-stage, ground-plane-parameterized monocular perception methods, expanding the published literature on infrastructure-based vehicle localization beyond other two-stage, LiDAR-dependent approaches.

- *Improved processes and techniques*: the demonstrated UAV-based ground-truth generation pipeline offers transportation agencies and researchers a lower-cost alternative to LiDAR-based label generation for training roadside perception systems.
- *Adoption of new technologies*: the filing of two U.S. patent applications formalizes a technology-transfer pathway, positioning the underlying algorithms for potential licensing by intersection-management or CAV-infrastructure vendors.
- *Enlargement of the trained professional pool*: the project supported graduate/doctoral-level research training, directly contributing to the pipeline of transportation and computer-vision researchers.

Impacts. Looking forward, successful adoption of this monocular, camera-only infrastructure-perception approach has the potential to:

- Improve intersection safety and situational awareness by making broad-coverage vehicle tracking economically feasible at a much larger number of intersections than LiDAR-based systems currently allow, directly addressing the disproportionate share of fatalities and injuries occurring at intersections.
- Lower the capital cost of deploying V2I-capable infrastructure, since the method requires only a standard monocular camera at inference time (no LiDAR, no onboard vehicle instrumentation), which can reduce both capital and operating costs for public agencies.
- Provide higher-fidelity kinematic data (yaw angle, yaw rate, lateral/longitudinal velocity) than conventional single-point GPS-based traffic sensors, supporting more accurate motion prediction and control algorithms for connected and automated vehicles, with downstream safety and traffic-efficiency benefits.

4.4 Challenges and Lessons Learned

- **Balancing pretrained-feature reuse against task adaptation.** Determining how many layers to freeze required systematic experimentation; freezing too few layers caused overfitting/slower adaptation of the new ground-plane output space, while freezing too many layers excluded parameters critical to the new task, degrading accuracy. The optimal point was not obvious a priori and required an empirical sweep.
- **Field-of-view and occlusion mismatches between sensing platforms.** The drone captures the full intersection while the roadside camera sees only a subset; the team had to develop a spatial-filtering step to ensure training labels correspond only to vehicles genuinely visible to the roadside camera.

- **Labeling challenges.** The drone video evaluation still requires significant human resources, that require the further development of the data acquisition system. We have promising results in this direction with the help of oriented bounding box detection with machine learning.
- **Limits of extrapolation.** The out-of-context test scenario clearly showed that the network's accuracy degrades for vehicle trajectories and orientations not represented during training — an important reminder that data diversity, not just data volume, governs real-world robustness for this class of model.
- **Coordinating multi-institution, multi-sensor field experiments.** Synchronizing drone flights and fixed roadside-camera recordings across 20 measurement runs, and aligning two research groups (BME and University of Michigan) working across two continents, required careful advance protocol design (scenario scripting, repeated runs, consistent camera mounting) to yield a consistent, high-quality dataset usable across both papers and the patent applications.