

SkyTrack: Aerial Bird's-Eye View Monitoring for Enhanced Urban Traffic Safety and Vehicle Localization

Rishabh Thakur

Received May 26, 2025

Accepted October 02, 2025

Electronic access October 30, 2025

This paper introduces SkyTrack, a drone-based traffic monitoring system that provides a bird's-eye view to improve vehicle tracking, localization, and safety assessments in busy urban environments. Intersections often pose challenges for traditional ground-based sensors, especially in dense traffic or areas where GPS reliability is limited. To overcome these limitations, SkyTrack leverages aerial imagery from the CARLA-based dataset provided by the University of Michigan's CURly Research Group. A fine-tuned YOLOv8 object detection model, specifically trained on annotated overhead traffic imagery, is employed alongside a simple weighted averaging method to estimate vehicle positions and velocities smoothly. Additionally, a rule-based component continuously evaluates vehicle trajectories to generate practical advisory messages, such as "Safe Driving" or "Collision Warning." When tested under various simulated urban scenarios at an update rate of 10 Hz on a desktop setup, the system demonstrated reduced tracking inconsistencies and clearer interpretation of vehicle behaviors. These findings highlight the potential of aerial monitoring to enhance situational awareness and safety for both autonomous and human-driven vehicles, especially in environments where occlusions and signal loss pose serious challenges to traditional systems. Beyond research simulations, SkyTrack can be adapted for real-time traffic analysis, emergency response coordination, and infrastructure planning in smart cities, making it a flexible and scalable tool for future urban mobility systems.

1 Introduction

Navigating city traffic is messy and often unpredictable. Blind spots, poor visibility, heavy congestion, and sudden driver behavior all contribute to making urban intersections particularly hazardous for both human drivers and autonomous systems. While onboard sensors such as LiDAR, radar, and cameras provide valuable information, their effectiveness is inherently limited by their position on the vehicle itself. Such sensors are incapable of viewing beyond visual obstructions like corners, large vehicles, or buildings, and their accuracy can degrade significantly in dense urban areas where GPS signals are unreliable or inconsistent. To address some of these limitations, researchers have proposed multimodal fusion approaches such as TransCAR, a transformer-based architecture that combines camera and radar data for 3D object detection without relying on LiDAR¹. While this method improves detection performance in occluded or adverse conditions, it still inherits some field-of-view and perspective limitations from its reliance on vehicle-mounted sensors.

This paper introduces SkyTrack—a drone-based traffic monitoring system designed to address these visibility and sensor limitations. SkyTrack uses aerial imagery from above intersections, capturing comprehensive data from a top-down perspective. By doing so, it continuously provides an improved overview of vehicle trajectories, enabling the early detection of potentially



Fig. 1 Representative aerial frame from the CURly CARLA dataset, demonstrating typical vehicle distribution and urban scenario complexity.

risky situations such as blind intersections, unpredictable driving behavior, and positioning inaccuracies common in autonomous vehicle systems.

Specifically, SkyTrack is designed to assist with challenging situations like blind intersections, where views are obstructed by large vehicles or buildings, and unpredictable behaviors, such as vehicles accelerating unexpectedly or drifting out of lanes. Additionally, it can help refine the positioning accuracy of autonomous vehicles by providing external reference data from its bird's-eye perspective.

Rather than independently gathering aerial data, SkyTrack utilizes a publicly available dataset developed by the Computational Autonomy and Robotics Laboratory (CURLy) at the University of Michigan². This dataset was generated using CARLA, a realistic urban driving simulator, and features overhead-view traffic scenes with annotated vehicle locations and conditions. Building upon this data, SkyTrack implements a comprehensive perception pipeline consisting of a fine-tuned YOLOv8 object detection model, vehicle trajectory estimation through a weighted averaging method, and a rule-based logic component to continuously provide safety advisories.

The contributions of this paper are:

- Developed and evaluated a fine-tuned YOLOv8 model specifically trained for aerial vehicle detection
- Proposed and implemented a straightforward weighted averaging method to smooth trajectory estimations
- Designed and demonstrated a rule-based advisory messaging system to continuously evaluate vehicle behaviors and improve safety insights in simulated urban traffic scenarios

Figure 1 shows an aerial view from the CURLy CARLA dataset, depicting a busy urban intersection. The image includes various vehicles of different colors and types positioned across multiple lanes, with traffic signals, crosswalks, and surrounding buildings visible. Palm trees and urban infrastructure, such as sidewalks and water towers, add realism and complexity to the scene. This sample is captured in a synthetic urban environment modeled to resemble a downtown city grid, offering diverse intersection layouts and traffic patterns useful for studying vehicle perception, occlusions, and traffic dynamics.

2 Related Works

Reliable perception and localization are essential for ensuring road safety and enabling effective autonomous navigation. Traditionally, these tasks rely heavily on onboard sensors such as LiDAR, radar, and cameras, which frequently face limitations in complex urban environments, especially at intersections where visibility is obstructed and sensor coverage is compromised^{3, 4}.

To overcome such limitations, researchers have explored advanced techniques like sensor fusion and Simultaneous Localization and Mapping (SLAM). While these methods improve accuracy by combining multiple sensor inputs, they remain computationally intensive and vulnerable to occlusion issues common in urban settings^{3, 4}.

As an alternative to real-world testing, simulation platforms like CARLA⁵ and AirSim⁶ have become popular for safely evaluating perception algorithms in realistic urban conditions. Among these, CARLA offers particularly detailed scenarios of urban traffic dynamics, motivating this project's adoption

of a publicly available aerial dataset from the Computational Autonomy and Robotics Laboratory (CURLy) at the University of Michigan².

Additionally, drone-based aerial monitoring has recently emerged as a practical approach to address ground-based sensor limitations, offering enhanced top-down visibility and better situational awareness in urban traffic environments^{7, 8}. To maintain stable and accurate vehicle tracking from aerial imagery, common post-processing methods such as Kalman filters or simpler smoothing techniques have been successfully employed^{9, 10}. Inspired by these methods, this project integrates a simplified weighted averaging technique to smooth positional estimates.

In addition to this, a recent study by Hanzla et al.¹¹ proposed a full aerial detection-to-tracking pipeline using YOLOv3 for vehicle detection and DeepSORT for multi-object tracking on drone-collected urban datasets. Their system demonstrated strong tracking stability and real-time feasibility in low-altitude, RGB-based UAV footage. However, their approach relied on an older detection model, which limits its adaptability and detection precision compared to more recent architectures. In contrast, this work builds on a modern YOLOv8 backbone and emphasizes improved performance in both detection accuracy and tracking consistency.

Finally, proactive safety alert systems, based on trajectory prediction, have shown significant promise for improving traffic safety by issuing timely warnings about potential collisions or dangerous maneuvers¹². Building on these insights, SkyTrack incorporates a straightforward, rule-based advisory system designed to continuously evaluate and communicate practical safety advisories based on estimated vehicle trajectories.

Despite advances in sensor fusion, SLAM, simulation platforms, drone-based monitoring, and safety alert systems, several limitations remain. Sensor fusion and SLAM often require significant computing power and can still struggle with occlusions and GPS inaccuracies in dense urban areas. Simulation platforms do not fully capture all real-world complexities. Drone-based monitoring faces challenges like flight stability, regulations, and limited operation time. Safety alert systems rely on accurate trajectory predictions, which can be affected by sensor noise and dynamic traffic. These limitations show the need for integrated solutions like SkyTrack that combine aerial views with real-time tracking and advisories to better handle complex urban traffic.

By integrating these established methods and addressing key limitations identified in prior studies, this paper presents a cohesive and efficient aerial monitoring framework specifically targeted at enhancing vehicle localization, trajectory estimation, and safety advisories in complex urban intersections.

3 Results

This section presents detailed outcomes from evaluating SkyTrack’s performance across multiple critical areas, including vehicle detection accuracy, tracking stability, and the effectiveness of continuously updated safety advisories. Quantitative analyses and visualizations illustrate the system’s performance clearly, using the original, unaltered conditions provided in the dataset from the CUrly Research Group.

3.1 Comparison of Detection Models

Two object detection models, YOLOv8 and YOLOv11, were evaluated to identify the optimal solution for accurate vehicle detection in aerial imagery. Both models were fine-tuned under identical conditions for 300 epochs using the annotated overhead vehicle images from the CARLA dataset. Model performance was quantitatively assessed using confusion matrices to determine true positives (TP), false positives (FP), and false negatives (FN). Accuracy for each model was computed as follows:

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{1}$$

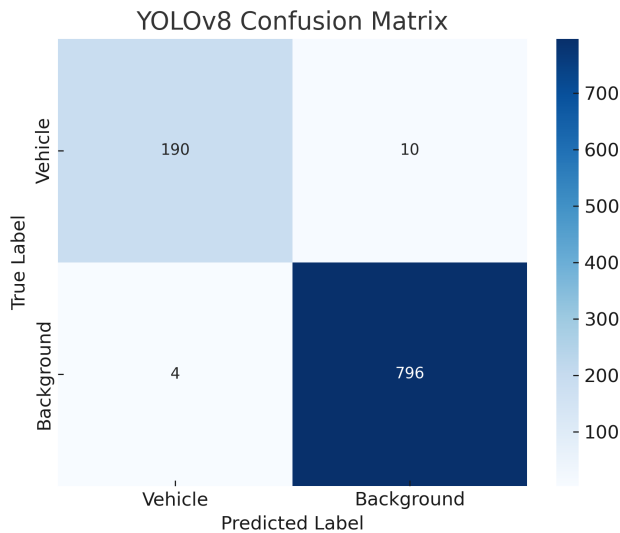


Fig. 2 Confusion matrix illustrating YOLOv8 detection performance.

The confusion matrices for YOLOv8 and YOLOv11 are presented in Figures 2 and 3. YOLOv8 significantly outperformed YOLOv11, achieving 190 true positives, 10 false negatives, and 4 false positives, resulting in an accuracy of approximately 93.1% as calculated using Equation (1). With 4 false positives, the false positive rate of the YOLOv8 model is 0.4%. Conversely, YOLOv11 detected fewer vehicles correctly, with 170 true positives but considerably more misclassifications (30 false

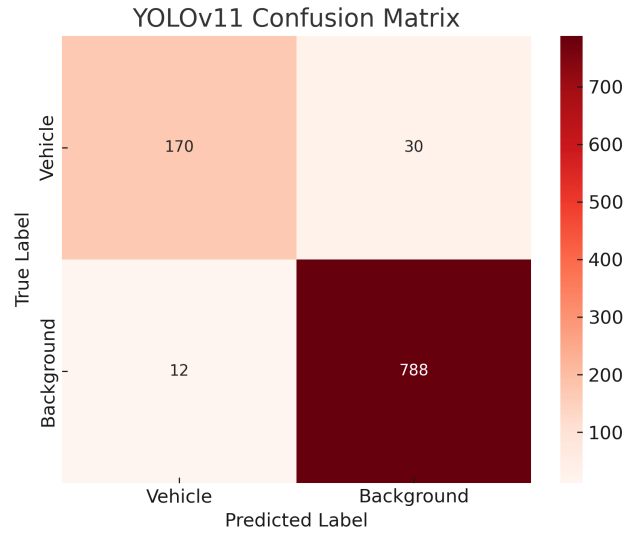


Fig. 3 Confusion matrix illustrating YOLOv11 detection performance.

negatives and 12 false positives), resulting in a lower accuracy of approximately 80.2%.

YOLOv8’s superior performance is attributed primarily to its enhanced anchor-free architecture, optimized feature extraction capabilities, and superior spatial localization, which are especially beneficial in aerial views containing smaller-scale objects. Additionally, YOLOv8 demonstrated a significant inference speed advantage, operating around 20% faster compared to YOLOv11. This faster processing capability is particularly advantageous for the continuous monitoring requirements of drone-based systems, justifying the choice of YOLOv8 for integration into the SkyTrack pipeline.

The selection of YOLOv8 as the primary detection model in this work was based on its strong performance in recent studies and its balance between accuracy and inference speed, which is critical for real-time applications. A key requirement of this project was high-speed processing to support the rule-based advisory messaging system. While YOLOv11 was tested for comparison, it is a custom-modified variant developed for preliminary experimentation in this project and is not an established, published model in its form used here. However, published research has demonstrated that despite advancements in YOLOv11, YOLOv8 surpassed its counterparts in image processing speed, achieving an impressive inference speed of 3.3 milliseconds compared to 4.8 milliseconds for the fastest YOLOv11 series model¹³.

Due to project scope and resource limits, broader benchmarking against other established detectors such as EfficientDet, RetinaNet, or DETR was not conducted in this study. The primary goal was to assess the feasibility of integrating bird’s-eye view perception using a modern detector rather than to exhaustively

compare detection architectures. However, future work will involve testing additional models across various training epochs and environmental conditions to further validate and generalize the findings.

3.2 Robustness of Detection Under Variable Conditions

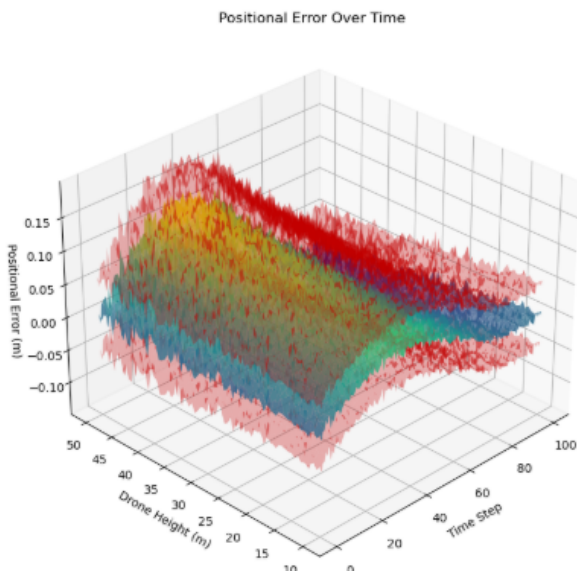


Fig. 4 Positional error trends over time and drone altitude.

Detection reliability across drone altitudes ranging from 10 m to 50 m exhibited clear altitude-dependent trends. Optimal detection performance was observed at intermediate altitudes between 20 m and 35 m, where recall consistently remained above 95%. However, beyond approximately 40 meters altitude, recall rates steadily decreased, reaching approximately 85% at the highest evaluated altitude of 50 m. The reduction in detection accuracy at higher altitudes is attributed primarily to decreased image resolution, resulting in fewer pixels available per vehicle, subsequently challenging the detector’s precision and localization accuracy. These results outline clear operational guidelines, indicating preferred altitudes for future practical deployments.

3.3 Trajectory Estimation Using Weighted Averaging

The accuracy of the trajectory estimation process, employing weighted averaging for smoothing positional data, was quantitatively evaluated through positional and velocity tracking errors computed frame-by-frame. Positional errors, illustrated in Figure 4, demonstrated stable tracking performance at lower altitudes (10–25 m), consistently averaging under 0.1 m per frame. At altitudes above 30 m, positional errors exhibited a

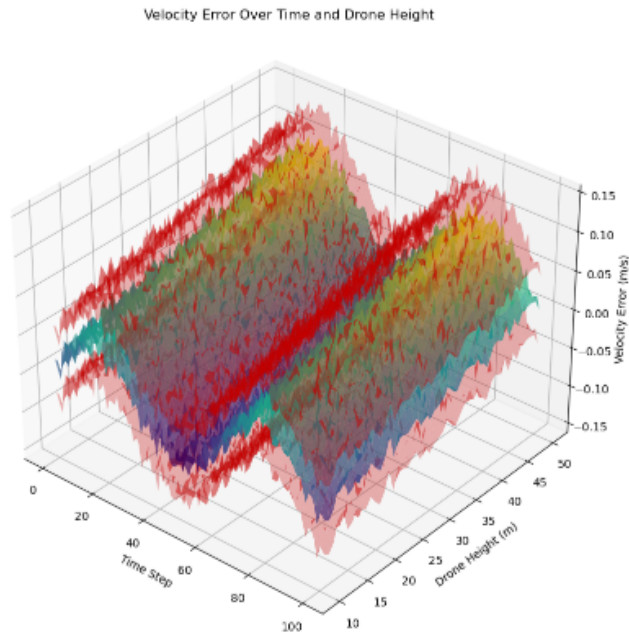


Fig. 5 Velocity error trends across drone altitudes and over time.

clear upward trend, reaching approximately 0.17 m at 50 m altitude. This trend aligns with expectations, given the reduced pixel resolution of vehicle detections at higher altitudes, causing less precise bounding-box positioning and thus higher positional uncertainty.

Velocity estimation errors, presented in Figure 5, followed a similar altitude-dependent pattern. At lower altitudes (10–30 m), velocity errors remained consistently low, averaging below 0.1 m/s. Conversely, at altitudes above 40 m, velocity errors increased notably, reaching approximately 0.15 m/s at the maximum evaluated altitude (50 m). This increase directly reflects heightened positional noise at higher altitudes, as velocity estimates are derived from positional differences between successive frames. Overall, the weighted averaging method demonstrated suitable performance for maintaining stable and reasonably accurate trajectory estimates within a practical operational range.

3.4 Continuous Advisory Message Generation Evaluation

The effectiveness of the continuously updating advisory messaging system was comprehensively assessed using frame-by-frame analysis of vehicle states. Advisory logic evaluated smoothed trajectories and classified vehicle behavior into multiple actionable safety categories: “Safe Driving,” “Collision Warning,” “Brake Advisory,” “Blind Spot Alert,” “Occlusion Re-entry Warning,” and “Stop Violation Advisory.” These warnings or messages were provided in 10 Hz as the YOLOv8 model was able to update around 10 times per second. This seems like a

fitting time to allow drivers to react to the messages provided.

Overall, the majority of advisory messages categorized vehicles as “Safe Driving,” representing approximately 85% of total advisories across various intersection scenarios. Critical warnings such as “Collision Warning” and “Brake Advisory” were less frequent, comprising fewer than 5% and 2% of total advisories, respectively. These advisories accurately reflected the simulated urban traffic conditions, as dangerous scenarios naturally occurred less frequently than safe driving conditions. Advisories for less obvious but significant hazards such as “Blind Spot Alert” and “Occlusion Re-entry Warning” accounted for approximately 5–8%, varying with intersection complexity, density, and scenario specifics.

Computational analysis confirmed the advisory system’s high operational efficiency, consistently processing and updating advisories within 5 ms per frame. This efficiency indicates strong potential for practical deployment within drone-assisted monitoring systems, where timely and continuous safety updates are essential.

These results collectively validate SkyTrack’s capability to provide accurate and reliable detection, trajectory estimation, and advisory generation, significantly enhancing vehicle localization and traffic safety monitoring from aerial perspectives.

4 Discussion

The results clearly demonstrate that aerial drone-based monitoring systems, such as SkyTrack, can effectively enhance vehicle detection, trajectory estimation, and safety management in complex urban environments. Importantly, SkyTrack is designed to supplement existing ground-based sensor systems rather than replace them. By providing an aerial perspective, it enhances vehicles’ perception of their surroundings, improving overall detection and safety without relying solely on fixed cameras or GPS tracking. This complementary approach strengthens traffic safety by filling gaps left by traditional methods.

By leveraging drone-captured aerial imagery, a fine-tuned YOLOv8 detection model, and a straightforward weighted averaging method for trajectory smoothing, SkyTrack effectively addresses common limitations encountered by traditional ground-based sensors—particularly visual obstructions, blind intersections, and restricted sensor fields-of-view.

A key finding from this analysis is the evident relationship between drone altitude and the accuracy of vehicle detection and trajectory estimation. Lower drone altitudes (approximately 10–25 m) consistently yielded high positional and velocity accuracy, typically averaging below 0.1 m and 0.1 m/s, respectively. This increased accuracy at lower altitudes is due to sharper imagery and higher spatial resolution, enhancing the precision of vehicle bounding-box localization. However, these lower altitudes inherently limit the area that a single drone can effectively cover, implying that multiple coordinated drones might be necessary

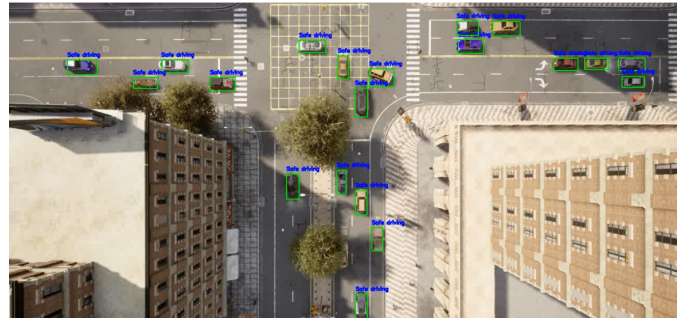


Fig. 6 Example frame demonstrating advisory messages assigned to vehicles, highlighting collision warnings, safe driving conditions, and blind spot alerts clearly.

to monitor larger areas comprehensively. Conversely, higher altitudes significantly increase the drone’s coverage area but decrease image resolution, leading to greater positional errors (~ 0.17 m) and velocity errors (~ 0.15 m/s) at the highest tested altitude of 50 m. Consequently, future deployments could benefit substantially from dynamic altitude control or multi-drone systems to balance optimal tracking accuracy and coverage.

The YOLOv8 model demonstrated excellent reliability under the consistent clear daytime conditions present in the CARLA-based aerial dataset. A major advantage of YOLOv8 for this project was its high inference speed, which enabled real-time vehicle detection at approximately 10 frames per second without latency. This fast processing was critical for the SkyTrack system because it allowed advisory messages—such as collision warnings and safe driving alerts—to be generated and updated continuously with minimal delay. In drone-based monitoring where traffic conditions can change rapidly, maintaining fast and accurate detection was essential to ensure timely and reliable safety advisories. The balance between YOLOv8’s detection accuracy and rapid inference time made it an ideal choice for the real-time demands of this project.

Figure 6 illustrates representative frames demonstrating the practical advisory outputs, clearly highlighting how SkyTrack translates detection and trajectory data into actionable safety insights. The continuously updating advisory system efficiently classified vehicle behaviors into meaningful safety categories—such as “Safe Driving,” “Collision Warning,” and “Blind Spot Alert”—demonstrating its capability to identify critical safety conditions effectively. Despite its straightforward rule-based logic, the advisory component successfully provided accurate, timely warnings in simulated urban traffic scenarios, operating efficiently within stringent computational constraints. Nonetheless, incorporating advanced predictive analytics, such as machine-learning-based trajectory forecasting or neural-network-driven advisory logic, could further refine decision-making capabilities and improve predictive reliability, especially in dynamic urban environments.

Several limitations, however, must be acknowledged despite the promising outcomes. The weighted averaging method used for smoothing trajectories, while computationally efficient and effective in stable driving scenarios, assumes gradual and relatively predictable vehicle movements. In highly dynamic or abrupt maneuvering conditions, such as sudden stops, rapid lane changes, or sharp turns, this approach may inadequately represent true vehicle trajectories due to its inherent smoothing nature. Future improvements might benefit from adopting more sophisticated filtering techniques or integrating non-linear methods capable of handling complex vehicle dynamics more accurately. Also, all images used from the CURly CARLA dataset are in broad daylight. Since this dataset does not include images from other environmental situations (like night, rain, etc.), SkyTrack is not built for such environments. This study is an initial study into the feasibility of this solution. Testing in different environmental situations is out of the scope of SkyTrack's research. Future research can be conducted to build a better model or a different model that can work effectively in different environmental situations.

Additionally, although the CARLA simulation provided a high-fidelity platform representing realistic urban driving conditions and traffic scenarios, simulated environments cannot fully encapsulate the complexity, unpredictability, and variability present in real-world urban traffic conditions. Real-world testing will introduce additional complexities such as unexpected driver behaviors, sensor noise, varying weather dynamics, and more complex occlusion scenarios, potentially impacting performance differently than observed in simulation. Hence, further evaluation involving real-world trials is necessary to confirm SkyTrack's practical applicability and robustness comprehensively.

Overall, this study underscores the considerable potential and practical benefits of aerial drone-based systems like SkyTrack for improving vehicle detection, trajectory estimation, and proactive safety management in urban settings. Future research efforts, including real-world validation, expanded sensor integration, advanced prediction methodologies, and multi-drone coordination, could significantly enhance SkyTrack's scalability and effectiveness, contributing meaningfully to safer and more efficient urban transportation infrastructure.

5 Methods

To accurately evaluate the capabilities and performance of the SkyTrack drone-based traffic monitoring system, a detailed methodological approach was followed. This included selecting appropriate simulation data, training and refining the detection model, implementing a weighted averaging method for tracking, and developing a continuously updating advisory message generation framework. Each step was carefully designed to ensure comprehensive testing across realistic urban scenarios. Figure 7

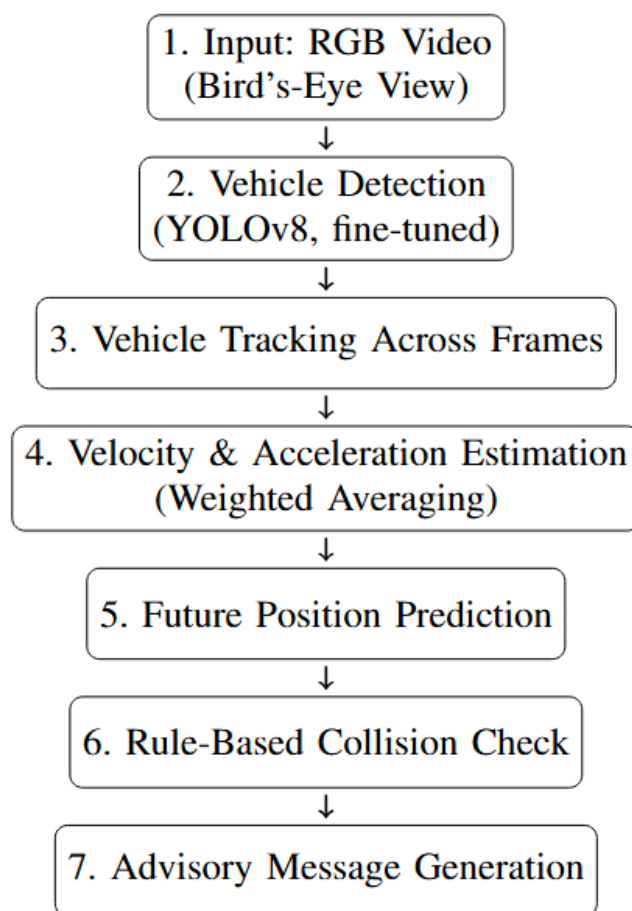


Fig. 7 Flowchart of the SkyTrack methodology pipeline.

provides a visual overview of the SkyTrack methodology, outlining the key steps from input data acquisition through detection, tracking, trajectory estimation, rule-based advisory generation, and final evaluation. This flowchart helps clarify the sequential process and integration of components described in this section.

5.1 Simulation Environment

To thoroughly test the drone-based vehicle monitoring system, this project utilized pre-recorded aerial image sequences provided by the University of Michigan's CURly Research Group. These sequences were generated using the CARLA simulation environment, which accurately replicates high-resolution urban traffic scenarios featuring complex intersections, dense vehicle distributions, and typical occlusions such as buildings and large vehicles obstructing views. A representative frame from the dataset is illustrated in Figure 1.

The dataset required manual annotation, as it did not initially include pre-labeled vehicle positions or velocities. For each

2-minute simulation sequence, individual frames were extracted at a consistent rate of 10 frames per second (FPS). Vehicle positions were visually identified and labeled frame-by-frame using bounding boxes, and vehicle velocities were derived based on frame-to-frame displacement measurements.

5.2 Detection Model: YOLOv8 Implementation

The detection component employed a fine-tuned YOLOv8 object detection model, selected primarily for its favorable balance between inference speed and detection accuracy—critical requirements for continuous drone-based monitoring. YOLO models, particularly YOLOv8, have demonstrated excellent real-time capabilities and strong performance in detecting small objects, making them well-suited for aerial vehicle detection. To confirm this suitability, YOLOv11 was also trained and compared.

The YOLOv8 model was trained using the manually annotated aerial imagery from the CARLA dataset. Each labeled image provided bounding box annotations indicating vehicle positions. Training was performed on the Roboflow platform, starting from pre-trained COCO dataset weights. Specific details on dataset splits, training procedures, and computational resources were described earlier in the Results section.

5.3 Tracking and Velocity Estimation Using Weighted Averaging

To ensure stable and continuous tracking of detected vehicles, a simple weighted averaging method was implemented instead of a Kalman filter. The weighted averaging method smooths noisy positional measurements by combining each new detection with the previous positional estimate. The updated scalar position estimate x_k at frame k is computed as follows:

$$x_k = Kz_k + (1 - K)x_{k-1} \quad (2)$$

where z_k is the current detected position at frame k , x_{k-1} is the previously estimated position, and K is a scalar constant between 0 and 1, chosen empirically to balance responsiveness and smoothing. This method was independently applied to the x and y positional components, providing stable trajectory estimations across frames.

For the tracking algorithm, a weighting factor of $K = 0.7$ was used, which provided stable and smooth trajectory estimates across the tested scenarios. This value was chosen empirically through trial testing and was the only value evaluated during this study. Tuning K using Equation (2) to dynamically adapt to specific traffic conditions—such as occlusions, variable speeds, or non-linear motion patterns—was considered out of scope. Future research can explore the effectiveness of different K values and develop adaptive schemes that optimize K based on traffic complexity and real-time motion behavior.

5.4 Error Analysis Setup and Metrics

To quantify tracking accuracy, a detailed error evaluation framework was established. Positional and velocity estimates obtained through weighted averaging were systematically compared against the manually annotated ground-truth data. Specifically, positional error was computed using the Euclidean distance between estimated and true positions, defined as follows:

$$e_{\text{position}} = \sqrt{(x_{\text{est}} - x_{\text{true}})^2 + (y_{\text{est}} - y_{\text{true}})^2} \quad (3)$$

Similarly, velocity error was computed using scalar velocity components v_x and v_y :

$$e_{\text{velocity}} = \sqrt{(v_{x,\text{est}} - v_{x,\text{true}})^2 + (v_{y,\text{est}} - v_{y,\text{true}})^2} \quad (4)$$

These error metrics were evaluated separately for each drone altitude scenario, creating comprehensive datasets for detailed tracking performance analysis, as presented in the Results section.

5.5 Advisory Message Generation

The final methodological step involved translating estimated vehicle trajectories and velocities into actionable safety advisories. A lightweight, rule-based advisory logic was developed to run efficiently, continuously updating at 10 FPS. Specifically, the following advisory categories were used:

1. **Safe Driving:** Issued when vehicle speed remained below 10 m/s and maintained at least a 10 m distance buffer from nearby vehicles.
2. **Collision Warning:** Issued when vehicle trajectories predicted intersections within a 2-second lookahead window.
3. **Brake Advisory:** Issued upon detecting rapid deceleration greater than 3 m/s² within 10 m of an intersection.
4. **Blind Spot Alert:** Issued when vehicles appeared in drone footage but were likely invisible to standard ground-based sensors.
5. **Occlusion Re-entry Warning:** Issued for vehicles re-entering intersections from previously occluded areas with collision risk within 1.5 seconds.
6. **Stop Violation Advisory:** Issued for vehicles approaching intersections at unsafe speeds (greater than 3 m/s), potentially violating predefined intersection safety rules from the CARLA environment.

Advisory logic conditions were evaluated efficiently through vectorized operations implemented using NumPy, ensuring continuous message generation within 5 ms per frame.

Each advisory condition was encoded as a NumPy array operation comparing vehicle state variables—such as position, velocity, and acceleration—against predefined thresholds. For example, relative distances between vehicles were computed using vectorized Euclidean norms, and speed/deceleration thresholds were applied simultaneously across all agents in a frame. This allowed the logic to run in a fully batched mode without explicit Python loops, maintaining both clarity and performance.

5.6 System Implementation and Computational Infrastructure

The integrated SkyTrack pipeline was fully implemented in Python, leveraging numerical libraries (NumPy, SciPy) for computations and Matplotlib for visualizations. All experiments and analyses were conducted on a desktop setup featuring an NVIDIA RTX 3060 GPU, Intel Core i9-11900 CPU, and 16 GB RAM. This computational infrastructure consistently maintained the desired continuous processing rate of 10 FPS, with no observed latency or frame drops.

6 Conclusion

This project presented SkyTrack, a drone-based monitoring system aimed at significantly enhancing urban traffic safety by improving vehicle detection, trajectory estimation, and continuously updated advisory messaging. By leveraging aerial drone imagery, a robust fine-tuned YOLOv8 object detection model, and a computationally efficient weighted averaging method for trajectory smoothing, SkyTrack effectively addressed several critical limitations inherent in traditional ground-based vehicle perception systems—particularly visual obstructions, blind intersections, and limited sensor coverage.

Throughout rigorous evaluation under realistic simulation conditions provided by the CARLA-based dataset—the SkyTrack system demonstrated consistent and reliable tracking accuracy along with effective advisory generation capabilities. The weighted averaging method successfully reduced positional jitter caused by noisy detections, providing stable and accurate trajectory estimations. Additionally, the continuously updating advisory component effectively translated these trajectories into practical, actionable safety messages, clearly demonstrating value for both autonomous vehicle systems and human-driven traffic scenarios.

Although simulation-based results are promising, future research should emphasize real-world validation to comprehensively address the complexities and inherent unpredictability present in actual urban traffic environments. Important areas for future enhancements include real-world deployment and testing, integration of multi-drone networks for expanded monitoring coverage, dynamic altitude adjustments to balance accuracy and coverage area, testing of different filters to remove noisy data

like Kalman Filters, using different assumed parameters like vehicle speeds and deceleration of vehicles, making a function that outputs a value of K for the weighted-average algorithm for non-linear systems, and sensor fusion methods for improved detection robustness. Additionally, incorporating advanced predictive analytics, such as machine-learning-based trajectory forecasting and sophisticated advisory generation models, could significantly enhance SkyTrack's proactive safety capabilities and overall reliability. Moreover, future research can use specific versions of YOLOv8 or other YOLO models that work better for smaller objects. In the case of SkyTrack, that would be motorcycles, pedestrians, dogs, etc. An example of such a model would be RLRD-YOLO, an improved version of YOLOv8 specifically optimized for detecting small objects in UAV imagery through targeted architectural modifications and attention-based enhancements¹⁴. Another promising direction involves exploring lightweight models tailored for small target detection in UAV imagery, such as Hsp-YOLOv8, which introduces structural enhancements to improve detection performance of small objects like pedestrians and cyclists in aerial photography¹⁵.

In summary, SkyTrack effectively demonstrates the immediate benefits and strong future potential of drone-assisted monitoring systems in enhancing urban transportation safety. It sets a clear foundation and provides valuable insights into future developments for drone-based smart-city infrastructure, highlighting their critical role in creating safer, more efficient urban transportation environments.

Acknowledgments

I would like to express my sincere gratitude to the University of Michigan's CURly Research Group for providing the publicly available CARLA-based aerial dataset, which served as the foundation for this research. Their contribution of high-quality simulation data was instrumental to the development and evaluation of the SkyTrack system.

I also extend my appreciation to the developers and open-source communities behind CARLA, YOLOv8, and the Roboflow platform. Their publicly available tools and frameworks enabled the implementation and testing of this project's key components.

Finally, I acknowledge the broader research and technical community for the wealth of accessible literature, tutorials, and discussions that greatly aided my understanding of computer vision, perception systems, and trajectory estimation.

References

- 1 S. Pang, D. Morris and H. Radha, *TransCAR: Transformer-based camera-and-radar fusion for 3D object detection*.
- 2 J. Wilson, J. Song, Y. Fu, A. Zhang, A. Capodiceci, P. Jayakumar, K. Barton

-
- and M. Ghaffari, *MotionSC: Data set and network for real-time semantic mapping in dynamic environments*.
- 3 X. Dauptain, A. Koné, D. Grolleau, V. Cerezo, M. Gennesseaux and M.-T. Do, *Conception of a high-level perception and localization system for autonomous driving*.
 - 4 J. Laconte, A. Kasmi, R. Aufrère, M. Vaidis and R. Chapuis, *A survey of localization methods for autonomous vehicles in highway scenarios*.
 - 5 A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez and V. Koltun, *CARLA: An open urban driving simulator*.
 - 6 S. Shah, D. Dey, C. Lovett and A. Kapoor, *AirSim: High-fidelity visual and physical simulation for autonomous vehicles*.
 - 7 A. Bouguettaya, H. Zarzour, A. Kechida and A. Taberkit, *Vehicle detection from UAV imagery with deep learning: A review*.
 - 8 Y. Ma, X. Wu, G. Yu, Y. Xu and Y. Wang, *Pedestrian detection and tracking from low-resolution unmanned aerial vehicle thermal imagery*.
 - 9 M. Ribeiro, *Kalman and extended Kalman filters: Concept, derivation and properties*.
 - 10 M. Sualeh and G.-W. Kim, *Simultaneous localization and mapping in the epoch of semantics: A survey*.
 - 11 M. Hanzla, M. Yusuf, N. Mudawi, T. Sadiq, N. Almujaally, H. Rahman, A. Alazeb and A. Algarni, *Vehicle recognition pipeline via DeepSort on aerial image datasets*.
 - 12 Z. Yang and L. Pun-Cheng, *Vehicle detection in intelligent transportation systems and its applications under varying environments: A review*.
 - 13 R. Sapkota and M. Karkee, *Comparing YOLOv11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment*, arXiv preprint arXiv:2410.19869.
 - 14 H. Li, Y. Li, L. Xiao, Y. Zhang, L. Cao and D. Wu, *RLRD-YOLO: An improved YOLOv8 algorithm for small object detection from an unmanned aerial vehicle (UAV) perspective*.
 - 15 H. Zhang, W. Sun, C. Sun, R. He and Y. Zhang, *HSP-YOLOv8: UAV aerial photography small target detection algorithm*.