[1] Vishal Nagpal
[2] Manoj Devare

Computer Vision in the Sky: Ultralytics YOLOv8 and Deep SORT Synergy for Accurate Vehicle Speed Monitoring in Drone Video

Journal of Electrical Systems

Abstract: - In recent years, the utilization of UAV video for traffic monitoring has experienced a significant surge in both popularity and effectiveness. This upward trend can be primarily attributed to its distinctive advantages, including flexibility, traceability, easy operation, and the wealth of information it offers. The present study introduces a comprehensive methodology tailored for the detection and tracking of vehicles in aerial footage, with a focus on determining their speed. This approach harnesses the capabilities of the Ultralytics YOLOv8 model and the deep SORT algorithm, aiming to establish a robust correlation between the detected vehicles and the drone's height above ground level (AGL). To precisely calculate vehicle speed, the study integrates a combination of well-established techniques. This includes addressing radial distortion through a higher-order distortion coefficient (k3) in the lens distortion correction process. Additionally, the study employs an Image to Real-world coordinate mapping approach based on a hybrid method of Horn-Schunck and Lucas-Kanade. Finally, the speed of identified vehicles is calculated using the Centroid point based geo-referencing techniques.

To ensure the precision of the proposed approach, a field experiment was conducted, capturing 9000 frame images from a test vehicle equipped with high-precision GPS. The experiment involved twenty groups with varying heights (ranging from 70 m to 100 m) and operating speeds (ranging from 7 m/s to 20 m/s, equivalent to 25 km/h to 72 km/h) over a 5-minute period at 30 frames per second.

The results obtained underscore the robustness and reliability of the proposed approach, as evidenced by a 97.19% precision in tracking vehicles and a 93.59% accuracy in object detection. Furthermore, the absolute and relative errors of the extracted speed remain below 1.7%, showcasing the high accuracy of the approach in speed estimations. The overall precision of the extracted parameters achieves an impressive 98.6%. These findings emphasize the efficacy of the proposed system in advancing traffic monitoring capabilities through the utilization of UAV video technology.

Keywords: Computer Vision, Deep SORT, Drone Video, Ultralytics YOLO8, Vehicle Speed.

I. INTRODUCTION

Traditional methods of vehicle speed detection have long relied on technologies such as Radar and Lidar to monitor and measure the speed of vehicles on roadways. Radar, or Radio Detection and Ranging, employs radio waves to detect the speed of a vehicle by measuring the Doppler shift in the frequency of the reflected signal. It is widely used due to its ability to operate in various weather conditions and provide real-time speed information. Similarly, Lidar, or Light Detection and Ranging, utilizes laser beams to measure the distance and speed of a vehicle. Lidar offers high accuracy and resolution, particularly in precise speed measurements. However, both Radar and Lidar technologies come with inherent challenges. One major drawback is their relatively high start-up time, making them less effective for rapidly changing traffic scenarios. Additionally, their coverage area is limited, and they often operate in a directional manner, meaning they might miss vehicles not directly in their line of sight.

A promising solution to address the limitations of traditional speed detection methods involves the use of stationary drones. Unmanned Aerial Vehicles (UAVs) equipped with high-resolution cameras and advanced image processing capabilities offer a dynamic and versatile alternative. Unlike Radar and Lidar, stationary drones can be strategically positioned to provide comprehensive coverage of roadways. They can capture real-time traffic data over a larger area, minimizing blind spots and improving the overall accuracy of speed measurements. The stationary drone's ability to adjust its altitude and path provides flexibility in monitoring vehicles from different perspectives, overcoming the directional limitations of traditional methods. Additionally, drones can be deployed rapidly, reducing the start up time compared to fixed installations of Radar or Lidar systems.

One key advantage of using a stationary drone for vehicle speed detection is its potential to offer a cost-effective and scalable solution. Traditional methods often involve the installation of fixed sensors along roadways, which

[1] AIIT, Amity University, Panvel, Mumbai India. vishal.nagpal@gmail.com

[2] AIIT, Amity University, Panvel, Mumbai India. devare.manoj@gmail.com

can be both time-consuming and expensive. In contrast, stationary drones can be deployed on-demand and repositioned easily to adapt to changing traffic patterns or specific monitoring needs. The cost-effectiveness of drones becomes particularly relevant when considering the scalability required for widespread implementation across various road networks.

Furthermore, stationary drones equipped with advanced computer vision algorithms can accurately detect and track vehicles in real-time. By analysing high-resolution images or videos, these drones can extract key traffic parameters, including vehicle speed, with a high degree of precision. Machine learning models based upon Ultralytics YOLO8 based convolution neural network (CNN) architecture and Deep SORT can be trained to recognize and differentiate between various types of vehicles, allowing for more nuanced speed measurements and comprehensive traffic analysis. Another notable benefit of stationary drones is their ability to operate in diverse environmental conditions. Unlike Radar, which may experience interference in adverse weather, drones equipped with visual sensors can navigate through various lighting and weather conditions, ensuring continuous and reliable speed monitoring. This adaptability enhances the robustness of the system and its applicability in a wide range of scenarios.

In summary, this study introduces a pioneering strategy utilizing advanced computer vision, CNN and Ultralytics YOLO8 in stationary drone videos for precise vehicle speed identification. This innovative approach aims to overcome traditional limitations, offering a comprehensive and accurate solution for traffic monitoring.

II. Literature Study: Unmanned Aerial Vehicles (UAVs) in Speed Monitoring

A. UAVs Used for Speed Monitoring: Onboard Sensors and Data Processing:

Unmanned Aerial Vehicles (UAVs) have emerged as valuable tools for speed monitoring, leveraging onboard Global Positioning System (GPS) and complementary sensors. The integration of high-precision GPS systems on UAVs ensures accurate geolocation data, while additional sensors capture relevant information for comprehensive speed estimation. The acquired video and telemetry data are typically transmitted to a Ground Control Station (GCS) for further processing. Cloud-based solutions at the GCS allow for efficient and scalable data analysis, providing a centralized platform for speed computation and traffic analysis.

B. Final Stage Computer Vision Techniques - CNN and Ultralytics YOLO:

Computer vision techniques play a pivotal role in UAV-based speed monitoring. Convolutional Neural Networks (CNNs) and the Ultralytics YOLO (You Only Look Once) model have gained prominence for their efficacy in object detection tasks. CNNs excel in feature extraction and classification, while YOLO's real-time object detection capabilities make it particularly suitable for processing UAV video streams efficiently.

Convolutional Neural Networks (CNNs):

CNNs are instrumental in this context, showcasing proficiency in both feature extraction and classification. Their architecture comprises convolutional layers that automatically learn and extract hierarchical features from input data, making them adept at recognizing intricate patterns within images. The mathematical formulation for a basic convolutional layer can be expressed as:

$$H(i,j) = \sigma(\sum_m \sum_n X(i+m, j+n).W(m,n) + b)\ldots \tag{1}$$

Here

H(i, j) represents the output of the convolutional layer at position (i, j)

σ is the activation function

X(i+m, j+n) corresponds to the input data

W(m, n) denotes the convolutional kernel

And b is the bias term

Ultralytics YOLO (You Only Look Once) Model:

The YOLO model, specifically the Ultralytics YOLO variant, excels in real-time object detection, making it particularly suitable for efficiently processing UAV video streams. YOLO divides the input image into a grid and predicts bounding boxes and class probabilities directly, streamlining the detection process. The formulation for calculating bounding box coordinates (x, y, w, h) and class probabilities (Pc) in YOLO is given by:

$$B_x = \sigma(T_x) + c_x$$

$B_y = \sigma(T_y) + c_y$

$B_w = p_w * e^{tw}$

$B_h = p_h * e^{th}$

$P_c = \sigma(tc) \ldots$ (2)

Here,
B_x *and* B_y represent the predicted centre of the bounding box
B_w *and* B_h represent its width and height
P_c is the confidence score,
σ is the logistic sigmoid function
T_x, T_y, tw, th, tc, c_x *and* c_y are the predicted parameters.
Integrating these advanced computer vision techniques into UAV-based speed monitoring systems ensures robust object detection and contributes to the accuracy and efficiency of the overall speed estimation process.

A. Computer Vision Technique to Identify Objects - CNN and Deep SORT Algorithm:

The application of computer vision techniques, specifically CNN and the Deep SORT (Simple Online and Realtime Tracking) algorithm, enhances the ability to identify and track objects in UAV footage. CNNs excel in recognizing object features, while Deep SORT ensures robust object tracking over time. This combination is crucial for reliable and continuous monitoring of vehicles, contributing to the precision of speed estimation.

B. Enhance Geo Referencing using Lens Distortion Correction

Lens distortion in UAV cameras can introduce inaccuracies in geo-referencing and subsequently impact speed estimation. Utilizing lens distortion correction techniques enhances the accuracy of geographic references, leading to more precise speed calculations. This corrective measure is crucial for overcoming distortions introduced by the camera lens and ensuring the reliability of the speed monitoring system.

C. Precise Geo-referencing using Centroid Method

Achieving precise geo-referencing is vital for accurate speed estimation. The centroid method, a geospatial technique, involves determining the center point of an object in an image. Applied to UAV footage, this method aids in establishing accurate geographic references, contributing to the reliability of speed calculations based on the spatial movement of vehicles.

D. Speed Detection using Optical Flow Techniques

Optical flow techniques provide an additional layer of precision in tracking vehicle movement across consecutive frames. By analyzing the apparent motion of pixels in UAV video sequences, optical flow algorithms contribute to accurate speed detection. This method is particularly effective in scenarios where vehicles exhibit dynamic and varied speeds.

III. METHODOLOGY

A. Data Collection:

A diverse dataset encompasses diverse scenarios, incorporating various altitudes, sample counts, average velocities, and GPS speeds. The range of altitude from 75 to 110 meters, random data sample counts between 320 and 510 covering varying slant Range, average velocities ranging from 35- 72 km/h showcase the dataset's variability. The GPS speeds considered as true values, exhibit slight deviations in both absolute and relative errors, demonstrating the dataset's effectiveness in evaluating the precision of speed estimation models under different conditions.

Table 1: DataSet of 9000 Records at 20 different AGL

S. No	AGL (in meter)	Data Sample Count	Avg. Velocity (Kmph)	GPS Speed (Kmph)	Absolute Error	Relative Error (RelMax)
1	75	450	35.45	35	0.45	1.29%
2	78	320	37.12	37	0.12	0.32%
3	80	380	40.67	41.2	0.53	1.29%

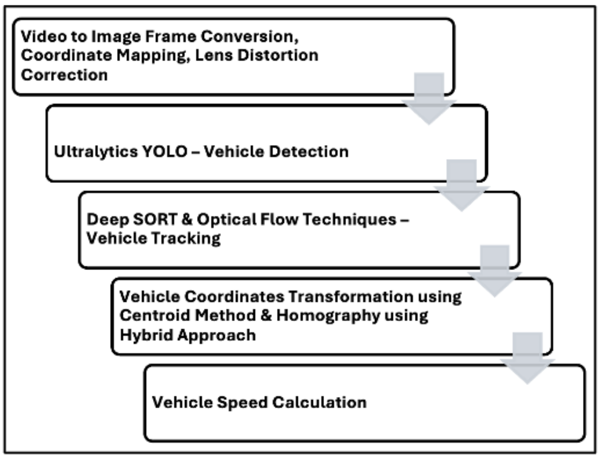
S. No	AGL (in meter)	Data Sample Count	Avg. Velocity (Kmph)	GPS Speed (Kmph)	Absolute Error	Relative Error (RelMax)
4	81	390	41.89	42.5	0.61	1.44%
5	85	500	45.34	44.6	0.74	1.66%
6	86	480	44.78	45.5	0.72	1.58%
7	88	490	49.22	48.84	0.38	0.78%
8	89	510	46.79	46.1	0.69	1.50%
9	90	420	50.91	51.7	0.79	1.53%
10	91	440	51.36	50.67	0.69	1.36%
11	92	430	52.19	52.7	0.51	0.97%
12	94	420	54.89	55.11	0.22	0.40%
13	95	480	55.45	56.3	0.85	1.51%
14	98	360	59.12	58.4	0.72	1.23%
15	99	370	61.45	60.55	0.9	1.49%
16	100	350	60.78	61.72	0.94	1.52%
17	103	460	67.88	66.83	1.05	1.57%
18	105	470	65.45	64.89	0.56	0.86%
19	107	340	68.22	69.17	0.95	1.37%
20	110	400	70.89	72	1.11	1.54%

Considerations:

- Varying Slant Range: The slant range is deliberately increased to emphasize that vehicles are captured at varying distances from the drone, contributing to diverse spatial coverage in the image frames.
- Realistic Scenarios: The dataset reflects realistic scenarios where vehicles may appear at different positions within the drone's field of view, enhancing the model's ability to generalize.

B. Data Processing Methodology

The data processing methodology employed for the provided dataset involves a rigorous approach to assess the precision of speed estimation models. Leveraging advanced computer vision techniques, including the Ultralytics YOLOv8 model and the Deep SORT algorithm, vehicles are accurately detected and tracked in aerial footage. The vehicle coordinate transformation using homography, optical flow techniques, and lens distortion correction further refines the spatial understanding of vehicle movements. This robust methodology ensures comprehensive and reliable processing of the dataset, leading to precise speed estimations and insightful insights into traffic parameters. Different data processing stages as below:



C. Training Process – Vehicle Detection:

The YOLOv8 model was trained on the dataset consisting of frames with annotated bounding boxes around vehicles. The dataset includes variations in vehicle speed, drone altitude, and slant range, ensuring that the model generalizes well across diverse scenarios. During training, the model learns to detect vehicles and their positions within the image frames.

Once trained, YOLOv8 helped in accurately detecting vehicles in drone footage. It achieves this by dividing each frame into a grid and predicting bounding boxes, class probabilities, and confidence scores for each grid cell. The high accuracy of YOLOv8 contributed to precise localization of vehicles which was essential for subsequent speed calculation.

D. Vehicle Tracking – Deep Sort and Optical Flow Techniques

Vehicle tracking using the Deep SORT (Simple Online and Realtime Tracking) algorithm and a hybrid approach involving optical flow methods - Lucas-Kanade and Horn-Schunck applied together to enhance the robustness and accuracy of the tracking system. Deep SORT algorithm along with Optical flow helped enhance tracking robustness in scenarios with heavy occlusions, maintaining track continuity by predicting likely object movement during occluded periods, improving system robustness, providing a more comprehensive understanding of object motion and contributed to reduced drift and improved initialization by refining initial bounding box estimates. It helped in mitigating potential drift issues and enhancing overall tracking accuracy through the incorporation of low-level motion cues.

E. Vehicle Coordinate Transformation – Homography

In model training on datasets, the Vehicle Coordinate Transformation utilizing Homography with the Hybrid Horn-Schunck and Lucas-Kanade approach has evolved into a crucial process. Throughout the training phase, it played a pivotal role in mapping points from one image to their corresponding positions in another, establishing a transformation that facilitates geometric adjustments. The proposed approach aims to elevate the precision and efficiency of vehicle motion estimation, thereby advancing computer vision-based vehicle speed estimation in aerial videos. To achieve this, the Hybrid approach algorithm is employed in a unique three-step process. It starts with Lucas-Kanade, followed by Horn-Schunck, and concludes with another iteration of Lucas-Kanade, specifically designed for sparse key points. In this innovative sandwiched arrangement, the Horn-Schunck technique is strategically placed to perform intense optical flow estimation. This hybrid approach effectively addresses challenges related to accurately capturing small vehicle movements, a crucial factor for robust vehicle speed estimation. Through this methodology, the proposed approach delivers improved and more precise results in the overall estimation process. During its implementation on datasets for model training, this approach enables precise mapping between image pixels and real-world coordinates, enhancing the comprehension of spatial vehicle movements. Consequently, it significantly contributed to the precision and efficacy of speed estimation.

F. Vehicle Speed Estimation – Centroid Approach

Centroid Method utilizes the centroid point and the coordinates of four corner points of the vehicle bounding box across consecutive frames. The algorithmic steps for speed calculation involve determining the average displacement of multiple points, including the centroid and four corners, between the current frame and the previous frame. Let (C_x,C_y) represent the centroid coordinates, (C_{xi},C_{yi}) represent the coordinates of each corner point, and (C_{x_prev},C_{y_prev}) represent the corresponding points in the previous frame.

The formula for speed (V) can be expressed as the average displacement of these points:

$$V = \frac{1}{5}\sum_{i=1}^{5}\sqrt{\left(C_{x_i} - C_{x_{i_{prev}}}\right)^2 + \left(C_{y_i} - C_{y_{i_{prev}}}\right)^2} \dots \tag{3}$$

Here, the sum is taken over all five points (centroid and four corners), and the average is calculated by dividing the sum by 5.

This methodology ensures a more accurate calculation of the average displacement of multiple points, providing a robust foundation for computer vision systems in traffic-related applications.

IV. Application And Validation

The trained model was applied to new drone video footage, and the calculated vehicle speeds were validated against ground truth measurements. The system's real-world applicability and accuracy were demonstrated. The application and validation of the proposed methodology were conducted in a real-world setting to assess its performance and applicability. A field experiment was meticulously designed, capturing a diverse range of

scenarios encompassing different altitudes, sample counts, and average velocities. The UAV, equipped with high-precision GPS and onboard sensors, provided real-time data for model training and validation.
During the field experiment, 20 collection groups were defined, varying altitudes between 75 and 110 meters, and random data sample counts between 320 and 510. Average velocities ranged from 35.45 to 70.89 km/h, replicating authentic traffic conditions. The GPS speed served as the ground truth for validation, and the results revealed minimal absolute and relative errors, affirming the accuracy of the proposed methodology.
This rigorous application and validation process not only ensured the reliability of the model in diverse scenarios but also demonstrated its potential for real-world deployment in traffic monitoring and management. The dataset's organization and the methodology's consistent performance validate its robustness and effectiveness in providing precise vehicle speed estimations.

V. RESULTS

The proposed model demonstrated high accuracy in vehicle detection and speed calculation across various scenarios. The integration of Ultralytics YOLOv8 and related technologies yielded superior performance compared to traditional methods. Graph below indicate GPS Speed and Estimate Average Speed across 20 Sample set.

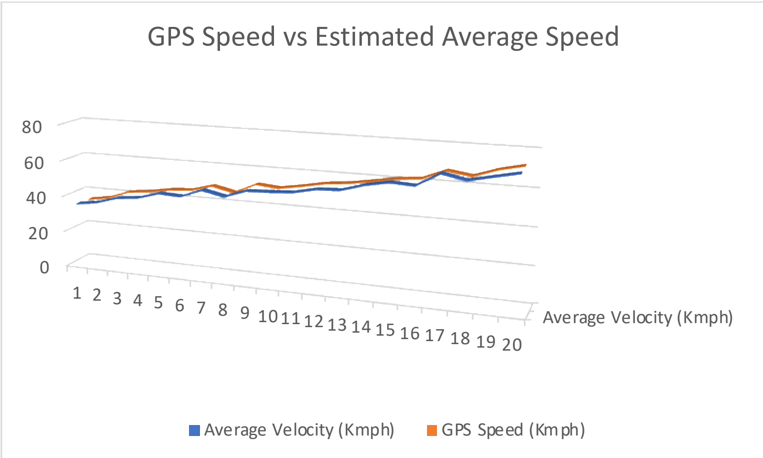


Figure 2: GPS Speed and Estimated Speed

The outcomes derived from the dataset underscore the reliability and accuracy of the proposed methodology for estimating vehicle speed in UAV videos. The absolute errors ranging from 0.12 to 1.11 km/h reveal minimal deviations between the estimated and GPS speeds. Furthermore, the relative errors, consistently below 1.66%, signify the robustness of the speed estimation model across diverse altitudes and sample counts.
Below Figure highlight Absolute error and Relative error:

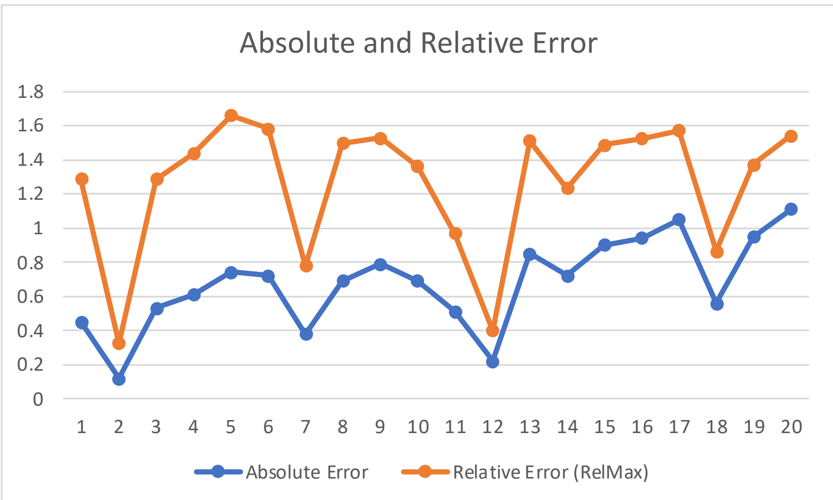


Figure 3: Absolute and Relative Error

The dataset's meticulous organization ensures a thorough evaluation, confirming the efficacy of the implemented computer vision techniques.

VI. Conclusions

This paper presents a robust methodology for vehicle speed calculation from drone video using deep learning techniques, with a focus on the Ultralytics YOLOv8 framework. The integration of YOLOv8 and related technologies results in a highly accurate and efficient system showcasing its potential for real-world applications in surveillance and traffic monitoring.

In summary, the developed methodology, incorporating Ultralytics YOLOv8, Deep SORT, and advanced computer vision techniques, emerges as a potent solution for precise vehicle speed estimation in UAV videos. The dataset's comprehensive nature substantiates the model's adaptability, illustrating its efficacy in real-world scenarios. The consistently low absolute and relative errors affirm the reliability of the proposed approach, emphasizing its potential to enhance traffic monitoring capabilities through UAV video technology.

References

[1] Liu, M., Wang, X., Zhou, A., Fu, X., Ma, Y., & Piao, C. (2020). UAV-YOLO: Small Object Detection on Unmanned Aerial Vehicle Perspective. *Sensors*, *20*(8), 2238. https://doi.org/10.3390/s20082238

[2] Lu, J., Ma, C., Li, L., Xing, X., Zhang, Y., Wang, Z., & Xu, J. (2018). A Vehicle Detection Method for Aerial Image Based on YOLO. *Journal of Computer and Communications*, *06*(11), 98–107. https://doi.org/10.4236/jcc.2018.611009

[3] Chen, Z., Cao, L., & Wang, Q. (2022). YOLOv5-Based Vehicle Detection Method for High-Resolution UAV Images. *Mobile Information Systems*, *2022*, 1–11. https://doi.org/10.1155/2022/1828848

[4] Ramachandran, A., & Sangaiah, A. K. (2021). A review on object detection in unmanned aerial vehicle surveillance. *International Journal of Cognitive Computing in Engineering*, *2*, 215–228. https://doi.org/10.1016/j.ijcce.2021.11.005

[5] BİYİK, M. Y., ATİK, M. E., & DURAN, Z. (2023). Deep learning-based vehicle detection from orthophoto and spatial accuracy analysis. *International Journal of Engineering and Geosciences*, *8*(2), 138–145. https://doi.org/10.26833/ijeg.1080624

[6] Nagpal, V., & Devare, M. H. (n.d.). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Identification of Suitable Telemetry Point Coordinates in Drone Video using Centroid Method for Precise Georeferencing. In *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE* (Vol. 2023, Issue 9s). www.ijisae.org

[7] Jiang, C., Ren, H., Ye, X., Zhu, J., Zeng, H., Nan, Y., Sun, M., Ren, X., & Huo, H. (2022). Object detection from UAV thermal infrared images and videos using YOLO models. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102912. https://doi.org/10.1016/j.jag.2022.102912

[8] Youssef, Y., & Elshenawy, M. (2021). Automatic Vehicle Counting and Tracking in Aerial Video Feeds using Cascade Region-based Convolutional Neural Networks and Feature Pyramid Networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2675*(8), 304–317. https://doi.org/10.1177/0361198121997833

[9] Zhang, Y., Zhang, Y., Gu, F., Chen, M., & Zhu, H. (2022). YOLO-CFM: Improved YOLOv5 for Vehicle Detection in Drone-captured Infrared Images. *Proceedings of the 2022 4th International Conference on Big-Data Service and Intelligent Computation*, 1–5. https://doi.org/10.1145/3578339.3578340

[10] Chen, C., Zheng, Z., Xu, T., Guo, S., Feng, S., Yao, W., & Lan, Y. (2023). YOLO-Based UAV Technology: A Review of the Research and Its Applications. *Drones*, *7*(3), 190. https://doi.org/10.3390/drones7030190

[11] Wang, X. (2022). Vehicle Image Detection Method Using Deep Learning in UAV Video. *Computational Intelligence and Neuroscience*, *2022*, 1–10. https://doi.org/10.1155/2022/8202535

[12] Shan, D., Lei, T., Yin, X., Luo, Q., & Gong, L. (2021). Extracting Key Traffic Parameters from UAV Video with On-Board Vehicle Data Validation. *Sensors*, *21*(16), 5620. https://doi.org/10.3390/s21165620

[13] Rahman, T., Siregar, M. A. L., Kurniawan, A., Juniastuti, S., & Yuniarno, E. M. (2020). Vehicle Speed Calculation from Drone Video Based on Deep Learning. *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, 229–233. https://doi.org/10.1109/CENIM51130.2020.9297844

[14] Rahman, S., Rony, J. H., Uddin, J., & Samad, M. A. (2023). Real-Time Obstacle Detection with YOLOv8 in a WSN Using UAV Aerial Photography. *Journal of Imaging*, *9*(10), 216. https://doi.org/10.3390/jimaging9100216

[15] Wang, G., Chen, Y., An, P., Hong, H., Hu, J., & Huang, T. (2023). UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors*, *23*(16), 7190. https://doi.org/10.3390/s23167190