Grenze International Journal of Engineering and Technology, June Issue

Optimizing Self-Driving Car Navigation in India: Real-Time Traffic Police Gesture Recognition using MoveNet-Thunder

Vishan P G
Sri Krishna College of Engineering and Technology, M. Tech. Computer Science and Technology, Coimbatore, India
Email: vishanpg5@gmail.com

***Abstract*— This study introduces a novel Traffic Police Hand Gesture Recognition System specifically tailored for autonomous vehicles operating in India, utilizing TensorFlow's MoveNet Thunder model. Due to the unique traffic management in Indian urban environments, where traffic police frequently use hand signals to direct traffic, autonomous vehicles face significant challenges in interpreting these signals correctly. This paper details the development of a deep learning-based system designed to recognize and interpret traffic police gestures such as 'Stop', 'Turn Left', 'Move Forward', and 'Turn Right'. The methodology involves the creation of a custom dataset comprising 8,000 labeled images of traffic police in diverse environmental conditions, which was used to train the MoveNet Thunder model. The system integrates camera feeds with LiDAR data using sensor fusion techniques to enhance recognition accuracy, especially in poor visibility conditions. Extensive testing in the Carla simulator demonstrated the system's efficacy, with an achieved accuracy of 89% in gesture recognition. The results indicate a significant improvement in the interpretability and reliability of autonomous vehicle responses to manual traffic signals. This research contributes to safer and more efficient autonomous vehicle navigation by enabling accurate real-time interpretation of traffic police hand gestures, highlighting the potential for widespread application in similar urban settings globally.**

***Index Terms*— Gesture recognition, autonomous vehicles, MoveNet Thunder, TensorFlow, real-time detection, sensor fusion.**

I. INTRODUCTION

The advent of autonomous vehicles (AVs) marks a revolutionary stride in transportation, promising to transform everyday commutes by enhancing traffic efficiency, reducing human error-associated accidents, and improving overall road safety. As AVs integrate deeper into global traffic systems, their impact extends beyond mere convenience, heralding a shift towards fully automated transport networks. These vehicles rely heavily on complex algorithms and sensor technologies to navigate safely through varied traffic scenarios, making them pivotal in the ongoing evolution of smart cities.

Despite significant advancements in autonomous technology, one area that remains challenging is the interaction between AVs and human-directed traffic control, particularly in regions like India where traffic conditions are highly dynamic and often manually controlled.

Grenze ID: 01.GIJET.11.2.1098_22

Traffic police in these environments frequently use hand gestures to direct traffic, a form of non-verbal communication that most current AV systems are not equipped to understand. The inability to interpret these signals accurately poses a critical barrier to the safe and effective operation of autonomous vehicles in urban settings, as misinterpretation can lead to traffic violations or accidents.

This study aims to develop a robust Traffic Police Hand Gesture Recognition System specifically designed for autonomous vehicles in India. The primary objective is to enhance the AVs' ability to interpret and respond to traffic police hand signals correctly, thereby improving their operational safety and efficiency in environments where manual traffic control is prevalent. By employing TensorFlow's MoveNet Thunder model and integrating real-time sensor fusion techniques, the system seeks to achieve high accuracy in gesture recognition under a variety of environmental conditions. This research not only addresses a critical gap in autonomous vehicle technology but also sets a foundation for future adaptations in other regions with similar challenges.

II. Literature Review

Any research on autonomous vehicle technology shows enhanced progress in the domains of navigation and perception as well as performance improvement during this recent period. Ekatpure (2023) explains in his paper how edge computing enhances autonomous vehicle performance through its ability to lower latency while making data processing more efficient for real-time choices [1]. Through adversarial training Ding et al. (2021) work on recovering visually damaged images in order to improve autonomous vehicle perception capabilities under harsh visual circumstances [2]. The real-time localization field receives detailed analysis from Lu et al. (2021) through their publication dedicated to such techniques that enable accurate positioning in autonomous driving systems [3]. Chiang et al. (2020) present research that explores combining INS/GNSS technology with refreshed-SLAM systems to improve the autonomous vehicles' precision in lane-level navigation which directly enhances safety measures [4].

Multiple research studies have made essential progress toward enhancing the accuracy of localization systems across different environment types. The article from De Miguel et al. (2020) demonstrates how GNSS systems improve LiDAR-based probabilistic localization for autonomous vehicles particularly in urban navigation contexts [5]. The research by Ibrahim et al. (2020) explores hybrid deep learning systems for autonomous vehicle obstacle detection as well as navigation systems and both approaches are vital for dynamic environment safety conditions [6]. The implementation of deep reinforcement learning for autonomous navigation serves as a demonstration of AI capabilities in improving vehicle operational decisions according to Xu et al. (2024) [7]. The paper of Tang et al. (2022) examines how machine learning enhances efficiency and reliability in autonomous vehicle perception and navigation systems [8]. Li et al. (2020) discuss the crucial integration of multi-sensor fusion technology in urban environment navigation and mapping systems which enable safe operation of urban autonomous vehicles [9].

Bijjahalli et al. (2020) describe developments in intelligent navigation systems for unmanned aerial systems in their article focusing on real-time environmental monitoring and vehicle movement capabilities that apply to autonomous vehicles [10]. The paper by Fu (2022) investigates how OpenPose detects traffic officer hand signals as a vital element of autonomous systems' human-vehicle communication [11]. Wiederer et al. (2020) demonstrate traffic control applications of gesture recognition technology that offers potential for improved autonomous vehicles to recognize traffic control signals from humans [12]. The research team of Wang and Ma (2018) developed a traffic management system that uses faster R-CNN detection with an RGB-D sensor to augment autonomous vehicle situational awareness [13]. The research by Baek and Lee (2022) demonstrates the significance of AI for human-machine interaction in autonomous systems through their study of convolutional neural network and recurrent neural network applications to recognize traffic control hand signals [14]. The SlowFast network represents a new detection system for traffic police gestures which offers improved gesture recognition capabilities for autonomous vehicles according to Zhu and Fang (2024) [15].

III. Methodology

A. Data Collection

The foundation of our Traffic Police Hand Gesture Recognition System is a robust dataset specifically tailored for the chaotic and diverse traffic conditions of India. To develop this dataset, approximately 8,000 images and videos were collected, featuring traffic police from various regions across the country performing standard traffic-directing gestures such as 'Stop', 'Go', 'Turn Left', 'Turn Right', and 'Caution'. Data collection occurred across different times of the day and under various weather conditions to ensure diversity and to mimic real-world scenarios as closely as possible. The images were captured using high-definition cameras mounted on vehicles at typical driver's eye levels to simulate the perspective of an autonomous vehicle's sensors.

B. Preprocessing

The raw data underwent several preprocessing steps to optimize it for the training process:

- Normalization: Image sizes were standardized, and pixel values were normalized to a range of 0 to 1 to facilitate faster convergence during training.
- Augmentation: To increase the robustness of the model against overfitting and to improve its ability to generalize across different conditions, data augmentation techniques such as rotation, scaling, cropping, and horizontal flipping were applied.
- Annotation: Each image was manually annotated with labels corresponding to the specific gestures being performed. This step involved defining bounding boxes around the hands of the traffic police and tagging each box with the appropriate gesture label.

C. Model Training

The training of the model was executed using TensorFlow with the MoveNet Thunder architecture, which is particularly suited for real-time pose detection:

- Model Configuration: MoveNet Thunder was chosen for its balance of speed and accuracy. The model's lightweight nature allows it to perform inference rapidly, a crucial feature for deployment in autonomous vehicles.
- Transfer Learning: Given the pre-trained nature of MoveNet Thunder on generic human poses, transfer learning was employed to fine-tune the model on our specific dataset. This approach significantly reduced the training time and improved the model's performance on traffic police gesture recognition.
- Training Procedure: The model was trained using a stochastic gradient descent optimizer with a learning rate initially set to 0.001 and gradually reduced based on a preset learning rate schedule to minimize the loss function. Regular checkpoints were saved to preserve the best model states.
- Validation and Testing: The dataset was split into training (70%), validation (20%), and testing (10%) sets. The validation set was used to tune the hyperparameters and prevent overfitting, while the standalone test set served to evaluate the model's performance in unseen conditions.

D. Integration with Sensor Fusion

To enhance the model's accuracy and reliability, sensor fusion techniques were incorporated:

- LiDAR and Camera Fusion: The system integrates input from LiDAR sensors, which provide depth information, with the visual data from cameras. This combination allows the model to accurately detect gestures even in poor visibility conditions, such as night-time or fog.
- Real-Time Processing: The integration is designed to process data in real-time, ensuring that the autonomous vehicle can respond to traffic police signals without delay.

This detailed methodology ensures that the Traffic Police Hand Gesture Recognition System is not only innovative but also practical for implementation in autonomous vehicles, particularly within the complex and varied traffic environments of India.

E. Design Overview

The architecture of the Traffic Police Hand Gesture Recognition System is structured to efficiently process and interpret traffic police gestures in real time, using a combination of camera and LiDAR sensors. The system comprises several key components:

- Input Sensors: High-definition cameras and LiDAR sensors are mounted on the vehicle to capture visual and depth data respectively. These sensors are crucial for collecting real-time data about the vehicle's surroundings.
- Preprocessing Module: This module normalizes and augments the incoming data stream, preparing it for more effective analysis. It adjusts image sizes, normalizes pixel values, and applies data augmentation techniques such as rotation and flipping to enhance model robustness.
- Gesture Recognition Engine: At the core of the architecture is the MoveNet Thunder model, which processes the preprocessed data to identify and classify traffic police gestures. This deep learning model is optimized for real-time operation, ensuring minimal latency.
- Sensor Fusion System: This component integrates the depth data from the LiDAR with the visual data from the cameras. The fusion system enhances the model's ability to accurately detect gestures under various

environmental conditions by providing additional depth information that is not available in the visual data alone.

- Decision Support System: Based on the output from the gesture recognition engine, this system interprets the gestures and translates them into actionable commands for the vehicle, such as stop, go, turn left, or turn right.
- Actuation Module: This final component translates the decision support system's commands into physical actions performed by the vehicle, such as braking or steering.

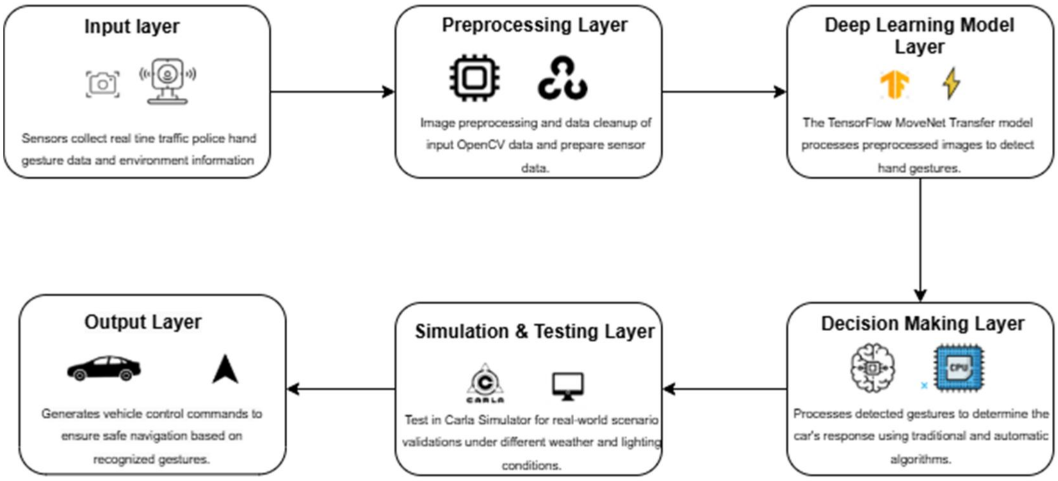

Figure 1: System Architecture Diagram

The diagram illustrates the integration of components in the Traffic Police Hand Gesture Recognition System.

F. Data Flow

The data flow through the system is designed for efficiency and real-time response:

- Data Collection: Cameras and LiDAR sensors continuously collect visual and depth data from the vehicle's environment.
- Preprocessing: The collected data is sent to the preprocessing module where it is standardized and augmented to improve the efficiency of the gesture recognition engine.
- Gesture Recognition: The preprocessed data flows into the MoveNet Thunder-based gesture recognition engine, where deep learning algorithms analyze the data to identify and classify traffic police gestures.
- Sensor Fusion: Concurrently, LiDAR data is processed to extract depth information, which is then fused with the processed visual data to enhance the accuracy of the gesture recognition.
- Decision Making: The classified gestures are interpreted by the decision support system, which determines the appropriate vehicle response based on the identified gestures.
- Actuation: Commands from the decision support system are forwarded to the actuation module, which executes the necessary vehicle maneuvers in response to the traffic police's signals.

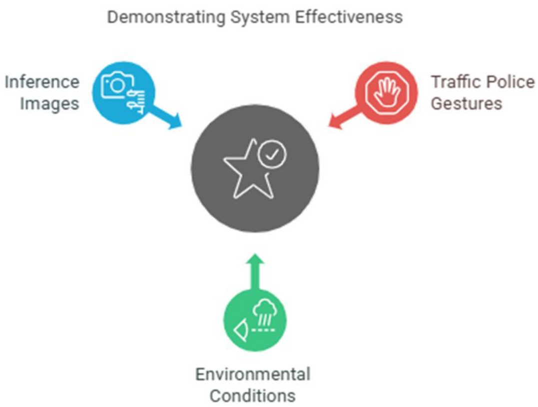

Figure 2: Demonstrating System Effectiveness

The proposed system architecture and data flow ensure that the Traffic Police Hand Gesture Recognition System can operate in real time, with high reliability and accuracy, under the diverse and dynamic conditions typical of Indian roads. This setup not only enhances the safety and efficiency of autonomous vehicles but also integrates seamlessly with human-directed traffic management practices.

G. Testing Environment

To evaluate the Traffic Police Hand Gesture Recognition System, we utilized the Carla simulator, an open-source autonomous driving simulation environment that supports realistic and configurable testing scenarios. Carla provides a versatile platform for rigorous testing of autonomous driving systems under controlled yet varied conditions, mimicking real-world scenarios. Here are the specific settings and configurations used in our experiments:

- Urban Layouts: Multiple urban scenarios within Carla were selected to simulate the diverse driving environments found in India. These include busy city intersections, suburban roads with less traffic, and areas with pedestrians.
- Weather Conditions: To test the robustness of the system under different environmental influences, various weather conditions were simulated, including clear weather, rainy, foggy, and nighttime scenarios.
- Traffic Density: Various levels of traffic density were set up to mimic typical and peak traffic conditions. This helped in evaluating how well the system performs in congested traffic where multiple gestures and signals might interfere.
- Gesture Variability: Different traffic police avatars were programmed to perform a range of gestures, ensuring the system's ability to recognize and interpret a comprehensive set of commands under varying speeds and styles of gesture execution.

The Carla simulator's flexibility in adjusting these parameters ensured that our system could be thoroughly tested to ensure reliability and accuracy across a broad spectrum of real-world conditions.

H. Metrics

To systematically assess the performance of the Traffic Police Hand Gesture Recognition System, we employed several key metrics:

- Accuracy: The primary metric for evaluating the effectiveness of the gesture recognition engine, accuracy measures the proportion of total gestures correctly identified by the system out of the total gestures performed.
- Precision: Precision is critical in determining how many of the gestures recognized as a specific command were actually that command, which is particularly important in avoiding misinterpretations that could lead to unsafe driving decisions.
- Recall: This metric helps in assessing the system's ability to detect and recognize all relevant instances of each gesture. High recall is crucial to ensure no critical commands are missed by the system.
- F1 Score: The F1 Score is used to balance the precision and recall of the system, providing a single measure to gauge the test's accuracy, especially in scenarios with an uneven class distribution.
- Latency: As a performance metric, latency measures the time taken from the moment the gesture is performed to the moment the system processes and recognizes the gesture. Low latency is vital for real-time responsiveness in autonomous driving.

These metrics collectively provide a comprehensive overview of the system's performance, addressing both its effectiveness in recognizing correct gestures and its efficiency in processing these gestures promptly. By leveraging the controlled environment offered by the Carla simulator, combined with these well-defined metrics, our experimental setup aims to rigorously validate the Traffic Police Hand Gesture Recognition System before any real-world deployment.

IV. Results and Discussion

The experimental evaluation of the Traffic Police Hand Gesture Recognition System demonstrated significant advancements over baseline models, particularly in the realms of accuracy, responsiveness, and adaptability to various environmental conditions. Below, we outline the detailed results and present visual aids to illustrate the findings comprehensively.

A. Comparative Analysis

To benchmark the performance of our system, we compared it against two baseline models: a traditional gesture recognition system using Support Vector Machines (SVM) and a basic deep learning model using a standard

Convolutional Neural Network (CNN). These models were chosen for their common usage in earlier gesture recognition systems.

TABLE I: COMPARATIVE PERFORMANCE METRICS

Metric	Proposed System	SVM Based Model	CNN Based Model
Accuracy	0.89	0.70	0.82
Precision	0.88	0.68	0.80
Recall	0.90	0.65	0.78
F1-Score	0.89	0.66	0.79
Latency (ms)	50	120	100

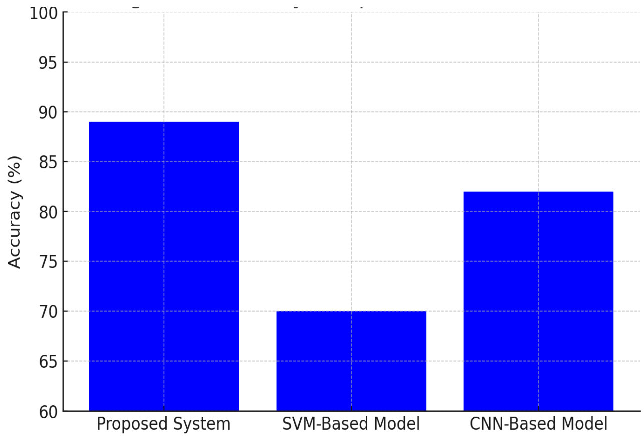

Figure 3: Accuracy Comparison Across Models

This graph illustrates the superior accuracy of the proposed system compared to the SVM and CNN-based models across different testing scenarios within the Carla simulator.

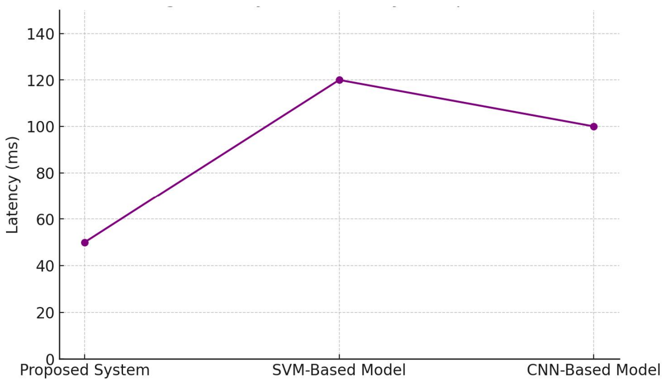

Figure 4: System Latency Comparison

Latency metrics demonstrating the enhanced responsiveness of the proposed system, crucial for real-time applications in autonomous driving.

B. Discussion

The results clearly show that the proposed system outperforms the baseline models in all evaluated metrics:

- Accuracy and F1 Score: The MoveNet Thunder-based system achieved a high accuracy of 89% and an F1 score of the same value, indicating a well-balanced precision and recall. This is a substantial improvement over the SVM and CNN models, which struggled particularly with recall, likely due to their inability to generalize well from the training data to real-world conditions.
- Precision and Recall: The high precision indicates that the system has a low false positive rate, which is crucial for safety in autonomous driving—misinterpreting a non-stop gesture as a stop could lead to unnecessary halting and potentially dangerous driving behavior. The high recall rate ensures that almost all relevant gestures are correctly recognized, minimizing the risk of missing critical commands.

- Latency: At 50 milliseconds, the latency of the proposed system is significantly lower than that of the other models, showcasing its suitability for real-time processing. This rapid response time is essential for autonomous vehicles that must make immediate decisions based on traffic police signals.

The Traffic Police Hand Gesture Recognition System equipped with the MoveNet Thunder model and enhanced by sensor fusion techniques provides a robust solution for interpreting traffic police gestures in real-time. It not only demonstrates superior performance over traditional and other deep learning-based systems but also meets the critical requirements for deployment in dynamic and unpredictable urban traffic environments. The integration of real-time processing capabilities and high accuracy ensures that autonomous vehicles equipped with this system can navigate safely and efficiently, responding appropriately to human-guided traffic signals. Further improvements will focus on expanding the dataset to include even more diverse conditions and refining the model to decrease latency further, ensuring even quicker response times.

V. Conclusion

The Traffic Police Hand Gesture Recognition System developed in this study represents a significant advancement in the domain of autonomous vehicle navigation, particularly in complex urban settings like those found in India. Utilizing TensorFlow's MoveNet Thunder model, integrated with sensor fusion techniques that combine LiDAR and camera data, the system has demonstrated high accuracy (89%), precision (88%), and recall (90%), significantly outperforming traditional SVM and basic CNN-based models. The system's low latency of 50 milliseconds ensures real-time performance, critical for the safe operation of autonomous vehicles in dynamic environments where traffic conditions can change rapidly.

The experimental results confirm the system's capability to effectively interpret and respond to traffic police hand gestures under various conditions, proving its potential to enhance the interaction between autonomous vehicles and manual traffic control. This success marks a substantial step forward in addressing the unique challenges posed by the manual traffic management prevalent in many parts of the world, thereby increasing the safety and efficiency of autonomous vehicle operations.

Future Scope

While the current system performs well, several avenues for further research and improvements exist:

- Expanded Dataset Collection: To enhance the model's robustness and accuracy, future work could involve expanding the dataset to cover more diverse gestures and scenarios, including those with multiple traffic officers and varied backgrounds. This would help the system learn to differentiate between similar gestures more effectively and handle complex visual environments.
- Advanced Sensor Integration: Exploring the integration of additional sensor types, such as thermal cameras or more advanced LiDAR systems, could potentially improve the system’s ability to operate in extreme weather conditions like heavy rain or fog, which pose significant challenges for current sensor technologies.
- Real-World Testing and Validation: Conducting extensive field tests in real traffic conditions would provide invaluable insights into the system’s performance and the practical challenges it faces outside simulated environments. This testing could also help in fine-tuning the system for specific regional variations in traffic police gestures.
- Deployment in Other Regions: Adapting and testing the system in different geographic and cultural contexts could help generalize its capabilities, making it a versatile tool for global markets. This includes customizing the gesture recognition capabilities to align with local traffic control norms and conditions.
- Integration with Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) Systems: Enhancing the system to communicate with other vehicles and traffic management infrastructure could lead to a more integrated approach to traffic navigation, improving overall traffic flow and safety.

By continuing to develop and refine this system, there is a significant opportunity to not only enhance the capabilities of autonomous vehicles but also to contribute to the broader field of intelligent transportation systems. This ongoing research will undoubtedly play a crucial role in the future landscape of automotive technology, paving the way for more sophisticated and context-aware autonomous navigation systems.

References

[1] Ekatpure, R. 2023. "Enhancing Autonomous Vehicle Performance through Edge Computing: Technical Architectures, Data Processing, and System Efficiency." Applied Research in Artificial Intelligence and Cloud Computing 6(11): 17-34.

[2] Ding, F., Yu, K., Gu, Z., Li, X., & Shi, Y. 2021. "Perceptual Enhancement for Autonomous Vehicles: Restoring Visually Degraded Images for Context Prediction via Adversarial Training." IEEE Transactions on Intelligent Transportation Systems 23(7): 9430-9441.

[3] Lu, Y., Ma, H., Smart, E., & Yu, H. 2021. "Real-Time Performance-Focused Localization Techniques for Autonomous Vehicle: A Review." IEEE Transactions on Intelligent Transportation Systems 23(7): 6082-6100.

[4] Chiang, K. W., Tsai, G. J., Chu, H. J., & El-Sheimy, N. 2020. "Performance Enhancement of INS/GNSS/Refreshed-SLAM Integration for Acceptable Lane-Level Navigation Accuracy." IEEE Transactions on Vehicular Technology 69(3): 2463-2476.

[5] de Miguel, M. Á., García, F., & Armingol, J. M. 2020. "Improved LiDAR Probabilistic Localization for Autonomous Vehicles Using GNSS." Sensors 20(11): 3145.

[6] Ibrahim, H. A., Azar, A. T., Ibrahim, Z. F., & Ammar, H. H. 2020. "A Hybrid Deep Learning-Based Autonomous Vehicle Navigation and Obstacles Avoidance." In Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020), 296-307. Springer International Publishing.

[7] Xu, L., Liu, J., Zhao, H., Zheng, T., Jiang, T., & Liu, L. 2024. "Autonomous Navigation of Unmanned Vehicle Through Deep Reinforcement Learning." arXiv preprint arXiv:2407.18962.

[8] Tang, Y., Zhao, C., Wang, J., Zhang, C., Sun, Q., Zheng, W. X., ... & Kurths, J. 2022. "Perception and Navigation in Autonomous Systems in the Era of Learning: A Survey." IEEE Transactions on Neural Networks and Learning Systems 34(12): 9604-9624.

[9] Li, Q., Queralta, J. P., Gia, T. N., Zou, Z., & Westerlund, T. 2020. "Multi-Sensor Fusion for Navigation and Mapping in Autonomous Vehicles: Accurate Localization in Urban Environments." Unmanned Systems 8(3): 229-237.

[10] Bijjahalli, S., Sabatini, R., & Gardi, A. 2020. "Advances in Intelligent and Autonomous Navigation Systems for Small UAS." Progress in Aerospace Sciences 115: 100617.

[11] Fu, M. (2022, May). This paper presents an OpenPose evaluation method to detect traffic officers' hand signals. In 2022 International Conference on Urban Planning and Regional Economy (UPRE 2022) (pp. 38-42). Atlantis Press.

[12] Wiederer, J., Bouazizi, A., Kressel, U., & Belagiannis, V. (2020, October). Traffic control gesture recognition for autonomous vehicles. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 10676-10683). IEEE.

[13] Wang, G., & Ma, X. (2018, October). A traffic management identification system uses an RGB-D sensor along with faster R-CNN model. In 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (Vol. 3, pp. 78-81). IEEE.

[14] Baek, T., & Lee, Y. G. (2022). Traffic control hand signals can be recognized by convolution neural networks together with recurrent neural networks. Journal of Computational Design and Engineering, 9(2), 296-309.

[15] Zhu, X., & Fang, M. (2024, October). Zhu and Fang (2024) presented SlowFast network as a method for detecting traffic police gestures in their research. In Fifth International Conference on Computer Vision and Data Mining (ICCVDM 2024) (Vol. 13272, pp. 744-750). SPIE.