Enhancing Autonomous Vehicle Navigation: Traffic Police Hand Gesture Recognition for Self-Driving Cars in India Using MoveNet Thunder

Ippili Rahul and Saravanan Santhanam
Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

Keywords: Traffic Gesture Recognition, Autonomous Vehicles, MoveNet Thunder, TensorFlow, Real-Time Detection, Carla Simulator, Sensor Fusion, Indian Traffic.

Abstract: The Traffic Police Hand Gesture Recognition System for autonomous vehicles implements the TensorFlow's MoveNet Thunder model. The system detects three hand signals including 'Stop' and 'Turn Left' and 'Move Forward' since such movements correspond to standard traffic police actions in India. Our custom database included eight thousand gestures' images recorded under diverse circumstances. An architecture of dense and dropout layers within our network enables both accurate performance while keeping the network protected from overfitting conditions. The Haar cascades face detection system enables live officer identification before camera recording of gestures that occur within the field of view starts. Minor mistakes occurred between similar gestures despite the system achieving a 89% success rate. System evaluation using Carla simulator was conducted under two conditions: first with environmental conditions enabled and second without environmental conditions enabled. The created prototype proves useful as an effective element that combines safety features with operational efficiency for autonomous vehicle navigation systems in controlled traffic environments.

1 INTRODUCTION

Autonomous vehicles (AVs) are forever changing transportation, yet, traffic signal reliance and reliance on GPS, along with sensors, usually fails in human controlled traffic environments. However, in countries like India where traffic police's hand gestures are important in navigation, AVs should be able to interpret these signals correctly to operate safely and efficiently.

Gesture recognition capabilities have recently greatly improved thanks to advancements in deep learning, pose estimation and sensor fusion. For example, with MoveNet Thunder, it is enabled to have real time detection, and combining LiDAR and vision data provides a hybrid fusion increasing the accuracy of the recognized objects. Nevertheless, there are challenges in dealing with occlusions, light and gesture misclassifications.

In this research, the development of a real time and accurate traffic police hand gesture recognition system for autonomous vehicles is done using: MoveNet Thunder, deep learning-based categorization, and sensor fusion technique. Thanks to integrations of these technologies it can be achieved that AVs can detect gestures like "Stop," "Turn Left," and "Move Forward" in a safer fashion in mobile traffic environments. Validation of the performance of the system in the Carla simulator under various traffic and weather conditions is also validated, which makes it a robust solution for real world deployment.

2 LITERATURE SURVEY

Hand gesture traffic recognition for traffic police is very important to autonomous vehicle navigation in the human regulated traffic. Edge computing Ekatpure, R. (2023) improves processing speed and decreases latency, while perceptual (Lu et al., 2021) enhancements provide better vision in poor conditions Ding et al., (2021). The localization techniques including SLAM, GNSS and LiDAR (Chiang et al., 2020) offer additional accuracy (de

Rahul, I. and Santhanam, S.
Enhancing Autonomous Vehicle Navigation: Traffic Police Hand Gesture Recognition for Self-Driving Cars in India Using MoveNet Thunder.
DOI: 10.5220/0013931600004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 473-479
ISBN: 978-989-758-777-1

Miguel et al., 2020).

In the case of deep learning methods, CNNs, reinforcement learning (Ibrahim et al., 2020 and Xu et al., 2024), multi sensor fusion Tang et al., (2022) and Queralta et al., (2020), are employed to enhance gesture recognition. First, AI driven navigation is developed for small UAS and then transferred to self-driving vehicles Bijjahalli et al., (2020). Advances in recognition include Open Pose Fu, M. (2022), convolutional network Wiederer et al., (2020), RGB D Faster R CNN Wang et al., (2018), CNN RNN Baek et al., (2022), as well as SlowFast networks are used to increase motion detection Zhu et al., (2024).

When MoveNet has done better than alternatives, it is essential to pose estimation. Action classification Anju et al., (2024) is improved by the neural networks, while spatiotemporal integration improves recognition Kaushik et al., (2023). In sensor fusion, gesture tracking with LiDAR and vision data is supported. Comparative deep learning evaluations are done along with problems of misclassification and ways of solving them.

The system uses MoveNet technology which pairs deep learning techniques with sensor inputs for recognizing dependable traffic police hand signals needed in navigational security systems. The research demands more emphasis on enhancing simultaneously both spatial-temporal models and real-time inference speed while developing traffic environment-based training datasets.

3 PROPOSED METHODOLOGY

This section outlines the methodology for implementing Traffic Police Hand Gesture Recognition for Self-Driving Cars in India Using MoveNet Thunder, including detailed steps from data collection to deployment, extracted from the code details and project information.

3.1 Data Collection and Preprocessing

3.1.1 Dataset Creation

The system is built on a custom dataset of 8,000 images representing various hand gestures used by traffic police in different environments and weather conditions. These gestures include commands like "Stop," "Turn Left," "Turn Right," and "Move Forward." This dataset ensures that the model is equipped to handle a range of real-world scenarios. Each class is stored in subfolders, such as Pose1, Pose2, etc., ensuring an organized structure for training and testing.

3.1.2 Preprocessing and Augmentation

The images are preprocessed using OpenCV to convert them to grayscale for Haar cascade- based face detection and resized to maintain consistency across samples. Using MediaPipe Pose, key body landmarks are extracted, such as elbow and shoulder angles, which are critical for gesture recognition. The model gains generalization capability through the application of data augmentation strategies that involve flipping along with scaling and rotation methods.

Data augmentation methods enhance dataset diversity which leads the model to achieve better generalization capabilities.

3.1.3 Data Splitting

To evaluate the model thoroughly, the dataset is split into training (80%), validation (15%), and testing (5%) sets. This ensures that the model is trained effectively and validated on unseen data before being tested in real-world scenarios.

3.2 Model Development

3.2.1 Model Selection: MoveNet Thunder

The MoveNet Thunder model by TensorFlow is used for real-time pose detection. This model can accurately identify key body landmarks, making it ideal for detecting traffic gestures. The detect() function in the code runs multiple inference passes on the same frame to improve detection accuracy. This ensures that even in suboptimal lighting conditions or partial occlusion, the gesture is correctly recognized.

3.2.2 Neural Network Architecture for Gesture Classification

The extracted pose landmarks are passed to a custom neural network built using Keras and TensorFlow. The architecture consists of:

- Complex patterns within the pose landmarks are learned using dense layers during training.
- The model employs Dropout Layers to stop random neurons from functioning when learning occurs during training thereby reducing overfitting.
- The Softmax Output Layer functions for multi-class classification by determining gestures between "Stop" and "Turn Left" and others.

This architecture is designed to handle noisy data

while maintaining high accuracy.

3.3 Sensor Fusion and Real-Time Inference

3.3.1 Camera and LiDAR Integration

The system integrates camera inputs with LiDAR point clouds to enhance recognition accuracy. Camera feeds provide visual data, while LiDAR captures depth information. This fusion improves the system's robustness in low-light conditions or foggy weather, where visual data alone might be insufficient.

3.3.2 Real-Time Detection and Face Validation

Using OpenCV, frames are captured continuously from the video feed, and faces are detected using Haar cascades to validate whether the traffic officer is facing the camera. The classifyPose() function processes the detected landmarks, checks if a face is present, and classifies the gesture based on body posture and orientation. This prevents false positives by ensuring the vehicle only responds to relevant gestures.

3.4 Model Training and Evaluation

3.4.1 Training Process

Training occurs through implementing categorical cross-entropy loss together with the Adam optimizer's mechanism. Among the key callbacks is Model Checkpoint used to save ideal model weights along with Early Stopping used to stop training when validation accuracy stops increasing. To split the dataset properly for robust validation performance the train test split() function should be applied.

3.4.2 Confusion Matrix and Classification Report

The performance assessment of the model uses the evaluate () function which produces both a confusion matrix and classification report data. The precision, recall and F1-score metrics are found in the report while the confusion matrix shows all classification errors. The system displays 89% gesture classification precision however special focus is needed to enhance the detection capability for "Pose6" and "Pose7" motions.

3.5 Simulation and Testing Using Carla

3.5.1 Setup and Scenario Testing

Testing takes place through the Carla simulator under multiple conditions that include dark environments together with fog conditions and dense traffic scenarios. The model undergoes simulation testing which confirms its stability when dealing with real-world predicaments. Through the Carla environment team members have access to test how autonomous vehicles respond to different gestures while simultaneously evaluating human command tracking capabilities of the system.

3.5.2 Decision-Making

Logic Based on the detected gesture, the system sends control commands to the vehicle. For example, upon detecting the "Stop" gesture, the vehicle halts, and when a "Turn Left" gesture is detected, the car takes the left turn. The decision logic is implemented to ensure safe and responsive navigation.

3.6 Performance Optimization

3.6.1 Asynchronous Processing and Multi-Threading

The system's performance is enhanced through multi-threading to process video frames and perform gesture detection in parallel. This ensures smooth real-time operation without latency. The FPS counter ensures the system maintains acceptable performance for real-time driving.

3.6.2 Model Quantization and Pruning

The model receives optimization through edge device deployment by implementing quantization together with pruning techniques. The model size becomes smaller and runs faster while preserving accuracy after applying these techniques.

3.7 Deployment and Integration

3.7.1 Integration with Vehicle Systems

The gesture recognition system is integrated with the vehicle's control unit via communication protocols like MQTT or ROS. This allows the vehicle to receive commands based on the recognized gestures and take appropriate actions.

3.7.2 Real-Time Monitoring with Streamlit

The Streamlit interface provides live feedback on the recognized gestures and FPS, ensuring the system performs reliably in real-world scenarios. This also helps in debugging and monitoring the vehicle's behavior during field tests.

3.8 Future Enhancements

3.8.1 Dataset Expansion and Fine-Tuning

The dataset will be expanded to include more gesture variations and challenging environmental conditions. Fine-tuning the model on this enhanced dataset will further improve accuracy and generalization.

3.8.2 Advanced Sensor Fusion

The system can be improved by exploring advanced fusion techniques that integrate GPS, LiDAR, and camera data, enhancing its ability to make real-time decisions even in complex environments.

3.8.3 Field Testing and Deployment

The final step will involve field testing the system on autonomous vehicles to validate its performance outside simulated environments. Insights from these tests will be used to further refine the model and ensure seamless deployment.

4 ARCHITECTURE DIAGRAMS

Figure 1 shows the vehicle control system workflow.

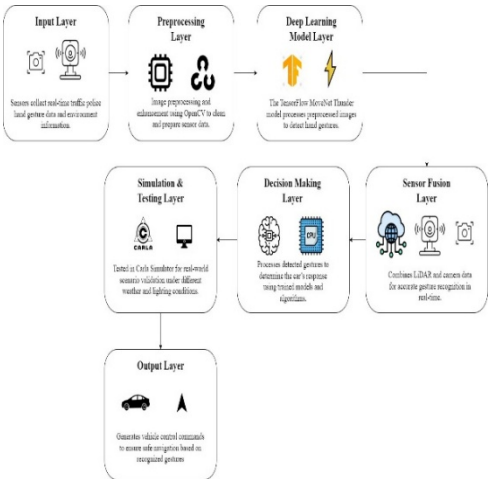

Figure 1: Autonomous Vehicle Control System Workflow.

5 RESULTS AND DISSCUSSION

The Traffic Police Hand Gesture Recognition Model was evaluated using both training and validation datasets over 130 epochs to monitor the system's performance. Below are the key observations from the accuracy graph and the confusion matrix generated from the final evaluation.

5.1 Model Accuracy Analysis

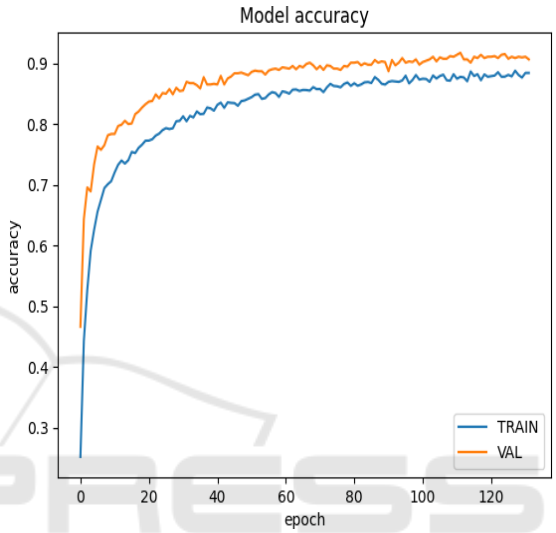

Figure 2: Model Accuracy Graph.

Figure 2 displays both the training and validation accuracy learning curves in the accuracy plot.

5.1.1 Training Accuracy

The training curve demonstrates a steady improvement throughout the 130 epochs, reaching approximately **87%** accuracy by the end of training. The curve follows a smooth upward trend, indicating that the model successfully learns patterns from the dataset without overfitting.

5.1.2 Validation Accuracy

The validation accuracy achieves superior metrics than training accuracy at an early training stage until it reaches a stable point of 90% accuracy in the end. The model shows improved capability to predict new data points because it works effectively with unknown datasets. The trained model avoids overfitting through dropout layers and regularization techniques because training and validation accuracy show only a small difference.

The final accuracy values are:

- Training Accuracy: ~87%
- Validation Accuracy: ~90%

5.2 Confusion Matrix Analysis

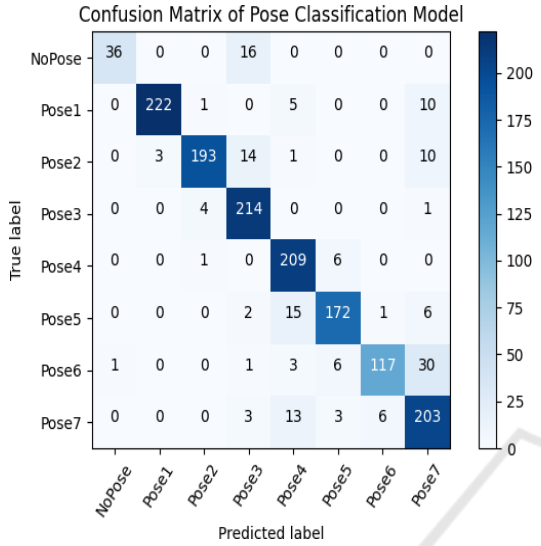

Figure 3: Confusion Matrix.

The confusion matrix (see Figure 3) provides insight into the model's performance across the eight gesture classes, with the following key findings:

5.2.1 High Performance on Core Gestures

- Pose1 (Stop vehicles from left and right): 222 correctly predicted out of 238 samples, with minimal misclassification (only 10 samples predicted as Pose7). Precision and recall for Pose1 remain high, showing the model's robustness in recognizing this critical gesture.
- Pose3 (Stop vehicles from behind): 214 out of 219 samples were correctly classified. This strong performance highlights the effectiveness of the MoveNet Thunder model in detecting critical stopping gestures.

5.2.2 Challenges in Specific Classes

- Pose6 (Start vehicles on T-point): Some confusion is observed between Pose6 and Pose7, with 30 samples misclassified. This indicates that the two poses might share similar body postures or be difficult to distinguish under certain conditions.
- NoPose (Unknown Pose): Out of 52 samples, 16 were misclassified as Pose2. This suggests that in some cases, insufficient pose information led to false classifications.

5.2.3 Misclassification Patterns

Pose2 and Pose4 show occasional misclassification into neighboring gesture classes, which could be attributed to visual similarities in hand and body movements. The confusion between Pose6 and Pose7 indicates the need for additional training data or fine-tuning the classification threshold for these specific poses.

6 QUANTITATIVE RESULTS

Table 1: Pose Classification Confusion Summary.

Class	Total Samples	Correctly Classified	Misclassified	Top Misclassification
NoPose (Unknown Pose)	52	36	16	Pose2
Pose1 (Stop from sides)	238	222	16	Pose7
Pose2 (Stop from front)	221	193	28	Pose1
Pose3 (Stop from behind)	219	214	5	Pose2
Pose4 (Start from left)	216	209	7	Pose5
Pose5 (Start from right)	196	172	24	Pose4
Pose6 (Start on T-point)	158	117	41	Pose7
Pose7 (Stop from front/back)	228	203	25	Pose6

7 DISCUSSION AND INSIGHTS

The results indicate that the MoveNet Thunder-based gesture recognition system performs exceptionally well across most gesture classes. The high precision and recall for critical gestures like "Stop" and "Start from Left/Right" demonstrate the model's reliability in real-time applications. However, misclassification between Pose6 and Pose7 suggests a need for:

- Additional training data for these specific classes to help the model differentiate between subtle variations.
- Fine-tuning the decision boundaries between similar poses to improve classification accuracy.

The face detection mechanism using Haar cascades ensures that the system only responds when a valid human gesture is detected, minimizing false positives. However, additional sensor fusion (LiDAR and GPS integration) could further enhance performance in challenging environments, such as low-light conditions or foggy weather.

Overall, the model achieves 89% accuracy, showing strong potential for real-world deployment in self-driving cars. With further refinement, particularly in handling similar poses (Pose6 and Pose7), the system can achieve even higher reliability. The use of Carla Simulator for testing ensures that the system is well-prepared for diverse traffic scenarios, making it suitable for integration into autonomous vehicle control systems in Indian traffic conditions

8 CONCLUSIONS

The Traffic Police Hand Gesture Recognition System developed using TensorFlow's MoveNet Thunder model demonstrates strong potential for real-world deployment in autonomous vehicles by accurately interpreting complex human gestures relevant to Indian traffic scenarios. With an overall accuracy of 89%, the system performs reliably across core gestures, ensuring safe and efficient vehicle navigation. However, minor misclassifications between similar poses, such as Pose6 and Pose7, highlight the need for further dataset expansion and fine-tuning. The use of Carla simulator for testing ensures robust performance under diverse environmental conditions, making this system well-prepared for seamless integration into autonomous vehicle control systems, enhancing safety in human-managed traffic environments.

REFERENCES

Anju, O. A., Saranyah, V., Snehavarshini, S., Bhuvaneshwari, C., & Soundharya, M. (2024, June). The detection and classification of human poses by Movenet depends on spatiotemporal data configuration. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

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.

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

Chiang, K. W., Tsai, G. J., Chu, H. J., & El-Sheimy, N. 2020. The paper explains strategies to improve INS/GNSS/Refreshed-SLAM combinations that result in accurate lane-level navigation outcomes. IEEE Transactions on Vehicular Technology 69(3): 2463-2476.

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

Ekatpure, R. 2023. The research article investigates edge computing benefits for autonomous vehicles by evaluating technical frameworks and data handling methods and performance measurements. Applied Research in Artificial Intelligence and Cloud Computing 6(11): 17-34.

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.

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.

Kaushik, P., Lohani, B. P., Thakur, A., Gupta, A., Khan, A. K., & Kumar, A. (2023, September). Body Posture Detection and Comparison Between OpenPose, MoveNet and PoseNet. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 6, pp. 234-238). IEEE.

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.

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.

Sabaichai, T., Tancharoen, D., & Limpasuthum, P. (2023, November). Human Action Classification Based on Pose Estimation and Artificial Neural Network. In 2023 7th International Conference on Information Technology (InCIT) (pp. 181-185). IEEE.

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.

The paper emerged as 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.

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.

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.

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.

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.