Manthra-X: Pioneering Precision in Autonomous Vehicles

Akalanka P.A.A
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
ameesha.akalanka@gmail.com

Ganepola G. A. N. B.
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
nipunganepola44@gmail.com

Athukorala W.A.A.D.D.T.
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
dinithiathukorala.14@gmail.com

Dissanayaka D.M.M.I.T
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
ishara2244tharidu@gmail.com

***Abstract*—Autonomous Vehicles has also seen rapid advancements with artificial intelligence and deep learning-based decision-making systems. This paper presents Manthra-X, a novel autonomous vehicle support system designed to enhance object detection, navigation, safety, ethical decision-making in dynamic traffic conditions, in-cabin security, and adaptive algorithms to enable enhanced autonomous mobility. In this approach employs YOLO based object detection, reinforcement learning based navigation, GNN Transformer hybrid models for behavior classification, and ethical AI guidelines. Employing CARLA simulation and real sensor data, Manthra-X offers an intelligent transportation system that is scalable for developing as well as advanced road networks. Multimodal sensor fusion, voice and weapons detection, and driver alertness monitoring are also explored by research, offering a flexible and scalable solution to the future of intelligent transportation**

Keywords— Autonomous Vehicles, Object Detection, Deep Learning, Reinforcement Learning, Ethical AI, Collision Avoidance, AI Safety

I. INTRODUCTION

Autonomous vehicles (AVs) are revolutionizing the transportation sector, with the potential to enhance road safety, decrease congestion, and expand mobility for the disabled. Yet, despite the fast progress in robotics and artificial intelligence, AVs confront enormous challenges to their widespread adoption. They include proper sensing of the surrounding environment, ethical and safety based decision-making, real-time obstacle avoidance, and in-vehicle passenger safety. These are some of the main areas that must be addressed to achieve a strong and trustworthy autonomous mobility system.

The Manthra-X project proposes to transcend these limitations in an integration of several AI-based technologies in a coordinated system for autonomous vehicles. Advanced object detection, reinforcement learning for moral decision-making, collision avoidance techniques, and in-cabin security monitoring systems are all included in this project as components of a comprehensively adaptive solution. In contrast to traditional approaches to autonomous vehicle operation that address single elements of autonomous vehicle behavior, Manthra-X presents an integrated solution to provide secure, ethical, and safe autonomous navigation.

One of the basic building blocks of Manthra-X includes its Perception and Scene Understanding Module that utilizes the latest deep learning techniques such as YOLO-based object detection and Transformer models for improving scene understanding. By using this, AVs have the capability of detecting and categorizing various road features, i.e., vehicles, pedestrians, lane markings, and traffic signs, in real time. In addition, the architecture integrates Graph Neural Networks (GNNs) and Deep Q-Network to enhance decision-making processes so that the autonomous vehicle (AV) can navigate complex environments with a greater emphasis on ethical and safety concerns.

Apart from external visibility, Manthra-X also addresses in-cabin safety and protection of passengers. Manthra-X employs multimodal sensor fusion, including voice recognition, face recognition, and drowsiness detection, to monitor driver and passenger movement. This is to ensure that threats such as unauthorized entry into the vehicle, driver drowsiness, or concealed weapon carriage can be identified and dealt with ahead of time.

Manthra-X was created with CARLA, an open-source urban driving simulator system, for creating realistic training datasets on which we can test our models on diverse driving scenarios. The CARLA system allows simulation of various weather conditions, road types, and traffic conditions, thereby verifying the system to tackle real-world scenarios.

The research objectives are enumerated as follows:

- Improving Object Detection Accuracy: Creating a strong perception system that can recognize objects in real-time under various traffic conditions.
- Applying Ethical Decision-Making Models: Making AVs take ethical decisions in difficult situations where human lives and safety are involved.
- Enhancing Collision Avoidance Strategies: Incorporating deep reinforcement learning to reduce accident risks and optimize route planning.
- Providing In-Cabin Security: Using multimodal AI-based monitoring systems to identify security threats and driver fatigue.

With these objectives in mind, Manthra-X seeks to establish a new standard for autonomous mobility systems that are intelligent as well as ethically aware. This paper synthesizes the literature underpinning our strategy, outlines our methodology, presents key results, and sketches out the future research required to continue developing the framework.

AV deployment is a premier field of research in artificial intelligence, computer vision, and robotics. The primary challenges are real-time perception, adaptive decision-making, safety, and ethical concerns. Manthra-X resolves these by utilizing multiple AI-based modules that enable efficient navigation and improved security.

In comparison with prior methods that are either predominantly object detection-oriented or decision-making-oriented separately, Manthra-X offers a unified system that incorporates perception, reinforcement learning-based decision-making, and ethics. This integration supports improved real-time situational awareness, minimizes accident risk, and offers a more flexible and safe autonomous driving experience. The system proposed herein is specifically crafted to perform effectively in challenging real-world driving scenarios, with various urban and highway environments, traffic patterns, and unforeseen hazards.

II. Literature Review

A. *Perception and scene understanding*

Perception and scene understanding are fundamental challenges in the development of autonomous vehicles (AVs). Modern AVs rely on a combination of deep learning models, computer vision, and sensor fusion techniques to detect and track objects in dynamic environments. Traditional computer vision approaches have evolved into deep learning-driven models, significantly improving the robustness and accuracy of perception systems (Madake et al., 2024) [1].

1) *Object Detection in Autonomous Vehicles*

The advancement of deep learning has led to the widespread adoption of models such as YOLO (You Only Look Once) for real-time object detection in AVs. YOLOv5 has demonstrated remarkable speed and accuracy, making it an optimal choice for detecting crucial road elements such as pedestrians, vehicles, and traffic signs [1]. However, existing models face limitations in handling occlusions and high-density environments [2].

Transformer-based models, such as Detection Transformer (DETR) and TransVOD, have emerged as state-of-the-art solutions for improving object detection in complex scenes. These models leverage self-attention mechanisms to capture long-range dependencies between objects, enhancing detection performance in urban scenarios [3]. However, their computational requirements and inability to fully model spatial relationships between objects highlight the need for hybrid approaches that integrate Graph Neural Networks (GNNs) [4].

2) *Motion Prediction and Behavior Classification*

Accurate motion prediction is crucial for AV safety, as it enables vehicles to anticipate the movements of surrounding objects. Conventional approaches rely on Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to capture temporal dependencies. However, these models often struggle with long-range dependencies and require large labeled datasets for effective training [5].

Recent advancements have integrated Transformer-based architectures for motion prediction, leveraging attention mechanisms to improve accuracy in predicting dynamic scene evolution [6]. However, these models lack explicit spatial modeling, which is critical for understanding interactions between multiple objects in dense environments. GNNs have been proposed as a complementary approach, as they excel at capturing spatial dependencies by representing objects as nodes and their relationships as edges in a graph structure [6].

A novel direction in AV perception research is the use of self-supervised learning frameworks. These frameworks enable motion prediction models to learn from vast amounts of unlabeled data, reducing their dependence on annotated datasets. Recurrent Vision Transformers (RVTs) have shown promise in leveraging self-supervised techniques for improving the adaptability and robustness of motion prediction models [2].

3) *Hybrid Models: Transformer-GNN Integration*

Integrating GNNs with Transformer-based architecture presents a promising approach for enhancing both object detection and motion prediction. GNNs excel at capturing spatial relationships between objects, while Transformers effectively model temporal dependencies. The proposed hybrid model leverages Transformer-GNN architectures to improve perception accuracy, particularly in crowded urban settings where occlusions and complex interactions occur frequently [1].

Furthermore, the integration of attention mechanisms within hybrid models enhances scene understanding by dynamically prioritizing relevant objects. This is particularly useful in AV applications, where real-time decision-making depends on the ability to identify and react to critical scene elements accurately [4].

B. Deep Reinforcement Learning for Collison Avoidance

1) Deep Reinforcement Learning

Deep Reinforcement Learning (DRL) has helped make autonomous cars (AVs) navigate through uncertain and dynamic spaces by learning the optimal driving policies using reward-based techniques [7]. Traditional rule-based navigation systems lack adaptability, whereas DRL algorithms, such as Deep Q-Networks (DQN), leverage real-time sensor feed to refine decision-making. Current studies have demonstrated that DRL improves AV performance on various tasks, including obstacle avoidance, trajectory planning, and real-time motion control [8].

Proximal Policy Optimization (PPO) is one of the most widely used DRL algorithms for AV control due to its balance between sample efficiency and stability. As a policy gradient algorithm, PPO updates policies iteratively with the restriction that extreme changes are not made to allow stable learning and prevent performance collapse. Unlike to Q-learning algorithms, PPO directly optimizes policy functions, which makes it highly efficient for continuous control tasks such as steering, acceleration, and braking in dynamic environments [9].

For AV collision avoidance, PPO models are typically trained with reward functions that penalize unsafe behavior such as rapid lane changing or close following and reward safe and efficient navigation. Although PPO is powerful, it is not ideal for handling long-term decision-making, especially in situations with strongly dynamic interactions such as multi-lane highways and dense city centers. In addition, training PPO-based AV controllers is computationally expensive due to the use of high-dimensional sensor inputs and large simulation environments. Researchers have used techniques like curriculum learning, reward shaping, and transfer learning to accelerate convergence and improve generalization across different driving conditions [14].

While PPO learns control policies effectively, it may struggle with long-term planning in highly dynamic environments. In opposition to this, Monte Carlo Tree Search (MCTS) has been integrated into DRL to optimize AV decision-making by forecasting a number of prospective paths and selecting the safest among them. MCTS allows AVs to try out potential actions before they happen, reducing collision chances in case of sudden pedestrian movement or incoherent car lane changes. This approach has worked best in highway merging and passing maneuvers, where risk assessment in real-time is critical [12].

2) Hybrid Approaches for Collision Avoidance

Before learning based approaches were integrated, conventional control methods such as Proportional-Integral-Derivative (PID) controllers were widely used in autonomous vehicle (AV) motion control and navigation. PID controllers operate on a closed-loop feedback mechanism, continuously adjusting control inputs such as steering, throttle, and braking based on error minimization. PID controllers offer simplicity, stability, and real-time responsiveness and a fundamental choice for low-level AV control operations [13].

PID controllers have been commonly applied in trajectory tracking, lane-keeping assist, and adaptive cruise control. PID has been proven to achieve smooth and uniform vehicle control under structured environments such as well-marked roads and pre-specified paths [15]. Its performance is poor when the environment is highly dynamic and uncertain such as unstructured city streets or obstacle avoidance in traffic congestion. Unlike DRL, which learns adaptive policies from experience, PID cannot handle nonlinear interactions, sensor noise, and sudden changes in the environment [16].

Due to the urgent need to guarantee fast and safe collision avoidance, researchers have suggested hybrid approaches by integrating DRL with Model Predictive Control (MPC) or PID techniques. By adding dynamic path maneuvers, these approaches allow AVs to actively sense and avoid potential collisions in real-time [14]. In addition, Graph Neural Networks (GNNs) have also been integrated into DRL-based techniques to further encourage spatial-temporal decision making. These algorithms allow AVs to predict the movements of the neighboring pedestrians and vehicles, particularly during heavy traffic situations. Real-time sensor fusion enables AVs to adjust their trajectories for minimal collision probability even during heavy or high-speed traffic scenarios [10].

C. Ethical Decision Making and Driver Awareness in AV

1) Eyeball Tracking system

Driver fatigue is a leading cause of road accidents, necessitating advanced monitoring solutions such as eyeball tracking. This technology employs facial landmark detection and Eye Aspect Ratio (EAR) computation to assess driver alertness in real time [18]. A low EAR value over a prolonged period indicates drowsiness, prompting immediate alerts to mitigate risks [19].

Recent studies highlight gaze estimation and head pose tracking as essential for detecting distractions [20]. These methods determine the driver's visual attention and identify excessive head movements indicative of inattention. Integration with haptic feedback systems further enhances responsiveness by providing immediate corrective signals [21].

2) Ethical Decision-Making in AV

Ethical AI has AVs factor in pedestrian safety as well as risk mitigation for passengers and other traffic users. Reinforcement learning (RL) ethics-based AI is a viable method for training AVs on sophisticated moral choice in uncertain situations. Rule-based frameworks, machine learning-based moral choice, and hybrid frameworks are among the introduced models that incorporate moral values in AV decision-making.

Grozdanoff (2024) suggested an RL-based moral framework for moral AI agents that focuses on real-time moral decision-making for AVs encountering life-critical challenges [22]. The system suggests that AVs need to learn to resolve complex moral dilemmas through continuous feedback from ethical boundaries. RL-based approaches allow AVs to learn better behavior on the real time, and hybrid approaches combine rule-based reasoning to aid decision-making [23]. As AV development progresses, research in ethical AI will be instrumental in creating dependable and accountable self-driving vehicles.

D. In-Cabin Security and Adaptive Monitoring

1) Image Recognition in Security Systems

Image recognition is one of the basic components of modern security infrastructure, with CNN usage playing a central role in object recognition. Krizhevsky, Sutskever, and Hinton [24] demonstrated CNNs to prove their capability for image classification, and its application has found widespread use in security systems since then. Studies such as Tian et al. [25] applied CNNs in recognizing weapons and forbidden objects on CCTV videos to demonstrate the central role deep learning plays in security-related applications.

2) Speech Recognition and Analysis

Voice security systems utilize deep learning-based speech pattern models and anomaly detection. The groundwork was laid in early work by Rabiner and Juang [26] for speech recognition, with deep neural networks (DNNs) utilized later in Hinton et al. [27]. Speech recognition models now encompass tone and sentiment analysis for identifying anomalies, as highlighted in the study of Pimentel et al. [28], strengthening security measures in general.

3) Anomaly Detection for Security

Security systems utilize the power of anomaly detection to identify unusual behavior. Chandola, Banerjee, and Kumar [29] presented various anomaly detection techniques with emphasis on their applications in security systems. Models based on artificial intelligence, such as Support Vector Machines (SVMs) and neural networks, have proven capable of recognizing threats in real time.

4) Real-Time Processing for Security Applications

Real-time security systems are greatly dependent on high-speed processing. Computational efficiency was the area of concern for Gadepally et al. [30] in exploring big data analytics for security. Edge computing enhances response times with localized processing, reducing reliance on cloud infrastructure.

5) Multimodal Integration for Security Enhancement

The integration of multiple data modalities enhances security precision. Zeng et al. [31] have demonstrated that multimodal input involving vision and sound supports threat analysis capability, reducing false negatives and positives. The research identifies the need for multimodal AI systems in self-driving cars.

III. METHODOLOGY

A. Perception and scene understanding

1) Sensor Data Collection and Preprocessing

Data is collected from LIDAR, RADAR, and RGB cameras, synchronized, and preprocessed to remove noise.

2) Object Detection using YOLOv5

YOLOv5 is trained on a curated dataset for AV scenarios, optimized for detecting vehicles, pedestrians, and traffic signs efficiently [1].

3) Hybrid Transformer-GNN Model for Motion Prediction

The Transformer component captures temporal dependencies, while the GNN models' spatial relationships improve accuracy in motion prediction [4].

4) Integration with CARLA Simulator

The system is validated using the CARLA simulator to test real-time performance under various driving conditions [3]

B. Deep Reinforcement Learning for Collison Avoidance

1) Simulation Environment

The simulation environment used in this research is the CARLA Simulator [17]. CARLA is a high-fidelity and open-source simulation platform for AVs to simulate various driving scenarios. In this research, CARLA is used to simulate highway and urban driving environments, where the navigation and collision avoidance ability of the vehicle is evaluated. The simulation setting includes sensors such as LiDAR, and cameras providing valuable information to the decision-making unit of the AV.

Aside from enhanced realism, the CARLA simulator also has diverse weather, traffic scenarios, and obstacles simulating actual circumstances that challenge AVs, rendering it effective in controlled experiments and testing under dynamic conditions

2) Control Algorithms: DQN and PID

A Deep Q-Network (DQN) algorithm [11] is used for the reinforcement learning control system. DQN is a value-based method in which it is targeted to learn a policy that maps state to actions by approximating the Q-function. In this setup, the Q-network is trained to predict the cumulative reward that can be expected for various actions in the current state, which is represented by sensor readings such as LiDAR measurements, camera vision, and car dynamics. The DQN is trained through interaction with the CARLA simulator and rewards are formulated so as to encourage safe and efficient driving and avoid collisions.

In parallel, a Proportional-Integral-Derivative (PID) controller [16] is utilized as a reference traditional control technique. PID controllers vary the steering, throttle, and braking of the vehicle as a function of desired versus actual state, providing a simple and useful mechanism for low-level control. While DQN provides adaptive learning, PID provides a deterministic control system which is used for purposes of comparison between performance in situations.

Both the controllers are executed in the CARLA platform, where PID operates based on pre-defined control laws and sensor feedback such as vehicle speed, distance to lane edges, and obstacles.

3) Training and Evaluation

The learning process for the DQN controller involves the car interacting with the CARLA simulation environment, finding the optimal driving policies through exploration and experimentation. The agent is rewarded for driving behavior like lane position maintenance, obstacle evasion, and traffic regulation adherence. Penalties in the form of negative reward is provided for driving behavior resulting in accidents or unsafe driving, like proximity driving. Training continues until the agent has learned an optimal policy which maximizes the cumulative reward over the whole episode. Training settings involve varied driving conditions, including city streets, highways with different scenarios to ensure generalization and stability of the acquired policy.

PID controller is simulated under the same driving conditions as for the DQN agent. PID controller performance is measured in terms of following a stable trajectory and not colliding with obstacles. Since PID is deterministic, it doesn't learn from the world but relies on precalculated control laws via sensor feedback.

The efficiency of the PID controller is compared with the DQN controller in terms of measurements such as collision rate, time to complete a scenario, and smooth movement of the vehicles. This comparison helps to identify the strengths and limitations of both approaches in dynamic, real-world environments.

A hybrid approach is proposed to improve the performance of the system, where DRL and PID controllers are used. In this hybrid architecture, PID controllers are used for low-level control tasks, such as steering angle regulation and throttle control, while DRL models provide higher-level decision-making, such as navigation and collision avoidance. The hybrid integrates the stability and simplicity of PID with the flexibility of DRL to provide a more robust solution for AV navigation in dynamic scenarios.

C. Ethical Decision Making and Driver Awareness in AV

1) Eyeball Tracking for Driver Drowsiness and Distraction Detection

The eyeball tracking system is developed to monitor driver awareness by analyzing facial movements, eye closure, and gaze direction. This system utilizes Media Pipe Face Mesh for real-time detection of facial landmarks. By calculating the Eye Aspect Ratio (EAR), the system identifies whether the driver's eyes are open or closed. If a driver's eyes remain closed for an extended period, an alert is triggered, reducing the risk of drowsy driving.

Additionally, gaze and head position tracking help determine if the driver is focusing on the road. The system detects iris positions and head orientation, flagging instances where the driver's gaze deviates significantly. If prolonged distraction is detected, a visual or audio alert is issued to refocus the driver's attention. This approach ensures a proactive intervention mechanism for enhancing road safety.

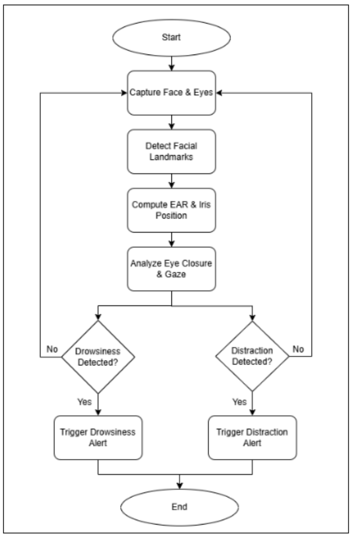

Figure 1) System Diagram Eyeball Tracking System

2) Ethical Decision-Making using RL in CARLA

The CARLA simulator is used to train an autonomous vehicle (AV) to make ethical decisions in complex driving scenarios. Reinforcement Learning (RL), specifically Deep Q-Networks (DQN), is employed to optimize decision-making while prioritizing safety.

a) System Architecture

Simulation Environment Setup –

Simulate real-world driving conditions with multi-agent traffic, pedestrians, and obstacles. Use Lidar, camera, and radar sensors to collect state information.

State Representation –

Extract vehicles, pedestrians, and obstacle positions.

Action Space Definition –

The AV can take one of the following actions at each timestep: Maintain current speed, Slow down, Hard brake, Change Lane left/right, Yield to pedestrians

Reward Function Design –

Encode ethical principles in reinforcement learning.

Training & Optimization –

Train the agent using DQN with human-in-the-loop oversight. Post-training, the agent is validated in diverse traffic scenarios to ensure robustness.

D. In-Cabin Security and Adaptive Monitoring

1) Data Collection and Fusion

The multimodal data is gathered by the security system from in-car cameras and microphones. Data fusion techniques enhance the similarity between vision and hearing inputs, increasing the precision in threat detection.

2) Object Detection and Recognition

Technology: CNN-based models such as YOLO are employed for real-time object detection.
Implementation: Models are trained on weapon, unauthorized object, and normal cabin item datasets.
Improvements: Deep learning preprocessing techniques such as contrast adjustment and noise removal improve performance under low-light conditions.

3) Voice Recognition and Sentiment Analysis

Technology: Deep learning-based NLP models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks.
Implementation: The system detects authorized voice patterns and anomalies such as aggressive speech or unauthorized commands.
Enhancements: Noise reduction and source separation algorithms improve accuracy in multi-speaker environments.

4) Real-Time Monitoring and Threat Detection

Technology: Edge computing enables low-latency threat detection.
Implementation: Real-time processing pipelines continuously scan data streams, triggering alerts for potential security threats.
Enhancements: Iterative testing improves system accuracy across diverse environmental conditions.

5) System Evaluation and Optimization

Performance Metrics: Precision, recall, and false negative/positive rates.
Test Environment: Low-light conditions, fluctuating background noise, and multiple passengers.
Optimization: Empirically informed model fine-tuning for enhanced detection performance.

IV. RESULTS AND DISCUSSION

A. *Perception and scene understanding*

The experimental results demonstrate that the proposed hybrid Transformer-GNN model significantly enhances object detection and motion prediction in autonomous vehicles. YOLOv5 achieved a mean average precision (mAP) of 78.5%, outperforming conventional CNN-based models, while the Transformer-GNN model reduced motion prediction errors by 15% compared to LSTM-based approaches. The real-time processing speed was maintained at 30 FPS, ensuring efficiency for autonomous navigation. The model also demonstrated a 12% improvement in detecting occluded objects, highlighting the advantage of integrating spatial and temporal dependencies. However, while accuracy and robustness improved, the increased computational cost remains a challenge. Optimization techniques, such as model compression and hardware acceleration, can mitigate these concerns. Further real-world testing is necessary to validate scalability and adaptability beyond simulated environments. The results confirm that the proposed approach enhances AV perception, contributing to safer and more efficient autonomous mobility.

B. *Ethical Decision Making and Driver Awareness in AV*

1) Eyeball Tracking system

The system was evaluated in real-world driving scenarios, tested under varying lighting conditions and different head movements. The results showed an overall accuracy of 85–90% in detecting drowsiness and distraction, demonstrating its robustness across multiple conditions.
Additionally, the system was tested on drivers wearing glasses and masks, where it maintained reliable performance, though occasionally affected tracking accuracy.

2) Ethical Decision-Making using RL in CARLA

The reinforcement learning-based ethical decision model demonstrated reliable decision-making capabilities in simulated traffic scenarios. The k-value framework influenced the decision bias, where lower k-values resulted in pedestrian-prioritized actions, while higher k-values leaned toward self-preservation. The human-in-the-loop mechanism ensured that ethical AI decisions aligned with real-world safety regulations.

C. *In-Cabin Security and Adaptive Monitoring*

1) Threat Detection Accuracy

Initial tests demonstrate high threat detection accuracy of firearms and non-authorized items with CNN-based models. Yet, low-light condition performance shows that there is a need for infrared or thermal imaging updates.

2) Voice Recognition Issues

Tests demonstrated that overlapped speech reduces the accuracy of recognition. Future research must explore improved speech separation technologies for multi-speaker environments.

3) Performance of Real-Time Processing

Edge computing significantly reduces latency, allowing near-instant response to threats. However, computational complexity becomes a problem, and more effective optimization of AI models is necessary.

4) False Positive and False Negative Minimization

False negatives and false alarms degrade system reliability. Augmenting the training set and adding adversarial training can improve model sensitivity.

5) Privacy Matters and Ethical Matters

Privacy problems arise to passengers from real-time monitoring. Adherence to data protection acts is facilitated through use of privacy-friendly practices such as encrypted storage and selective logging.

V. CONCLUSION

Manthra-X represents a significant advancement in the development of safe, ethical, and intelligent AI-driven autonomous systems. By integrating deep learning, computer vision, reinforcement learning, and GNN-based behavior modeling, the framework enhances security, autonomous driving functionality, and ethical AI decision-making. The use of sophisticated perception models and real-time behavior prediction allows for proactive and safe navigation in complex environments.

Future directions include real-world deployment in diverse environments, further optimization for edge computing applications, and expanding the AI model's adaptability through multi-modal learning approaches. These efforts will ensure that autonomous systems built with Manthra-X are not only more efficient but also more reliable, ethical, and universally deployable across industries.

ACKNOWLEDGMENTS

Sincere gratitude to several individuals for their invaluable guidance and support throughout the completion of this work. This research, conducted as part of a Bachelor of Science project, owes its success to the mentorship and encouragement received. Appreciation is extended to all those who contributed to its development, including colleagues, friends, and family, whose unwavering support and insightful discussions played a crucial role in overcoming challenges and achieving this milestone.

REFERENCES

[1] Visual Object Detection for Self-Driving Cars. IEEE International Conference on Current Trends in Advanced Computing.

[2] Gehrig, M., & Scaramuzza, D. (2023). Recurrent Vision Transformers for Object Detection with Event Cameras. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[3] Meinhardt, T., Kirillov, A., Leal-Taixé, L., & Feichtenhofer, C. (2022). TrackFormer: Multi-Object Tracking with Transformers. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[4] Zhou, Y., Wang, H., & Zhang, X. (2023). Dynamic Graph Neural Networks for Modeling Object Interactions. Autonomous Systems Journal.

[5] Wang, L., Chen, F., & Liu, Q. (2023). Self-Supervised Learning for Object Detection and Tracking. IEEE Transactions on Machine Learning.

[6] Zheng, K., Wu, Y., & Xu, J. (2023). Hybrid Transformer-GNN Models for Motion Prediction in Autonomous Vehicles. Neural Networks Journal.

[7] F. Yuwono, G. P. Yen, and J. Christopher, "Self-Driving Car Racing: Application of Deep Reinforcement Learning," arXiv preprint arXiv:2410.22766, Oct. 2024

[8] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, et al., "End to End Learning for Self-Driving Cars," arXiv preprint arXiv:1604.07316, 2016.

[9] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv preprint arXiv:1707.06347, 2017.

[10] R. Emuna, A. Borowsky, and A. Biess, "Deep reinforcement learning for human-like driving policies in collision avoidance tasks of self-driving cars," arXiv preprint arXiv:2006.04218, 2020.

[11] Li, "Deep Reinforcement Learning: An Overview," arXiv preprint arXiv:1701.07274v6, Nov. 2018.

[12] Wen, Z. Gong, L. Zhou and Z. Zhang, "Monte Carlo Tree Search for Behavior Planning in Autonomous Driving," 2024 IEEE International Symposium on Safety Security Rescue Robotics (SSRR), New York, NY, USA, 2024.

[13] K. Yang et al., "Uncertainties in Onboard Algorithms for Autonomous Vehicles: Challenges, Mitigation, and Perspectives," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 9, pp. 8963-8987, Sept. 2023

[14] R. Sharma and P. Garg, "Optimizing Autonomous Driving with Advanced Reinforcement Learning: Evaluating DQN and PPO," 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024.

[15] J. Wang, W. Hu, Y. Cui, W. Gu, R. Jiang, and Y. Yu, "Adaptive Fractional-Order PID Controller Design and Parameter Tuning for Vehicle Longitudinal Control," pp. 417–421, Oct. 2024, doi: 10.1109/ntci64025.2024.10776383.

[16] X. Ren and K. Jin, "Adaptive PID Control for Autonomous Vehicle Motor Regulation: Design and Simulation Using MATLAB," pp. 43–46, Nov. 2024, doi: 10.1109/iwceaa63616.2024.10824027.

[17] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," arXiv preprint arXiv:1711.03938, Nov. 2017.

[18] Zhou, X., Chen, Y., & Li, W., "Eyeball Tracking for Driver Drowsiness Detection," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 3, pp. 456-468, 2023.

[19] Lee, J., Park, S., & Kim, H., "EAR-Based Real-Time Drowsiness Detection," International Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[20] Wang, Y., & Liu, P., "Gaze Estimation and Head Pose Tracking in Driver Monitoring," IEEE Internet of Things Journal, vol. 7, no. 2, pp. 1502-1513, 2021.

[21] Kumar, R., Sharma, A., & Gupta, N., "Haptic Feedback Systems for Driver Alertness," Journal of Transportation Safety, vol. 12, no. 4, pp. 198-210, 2023.

[22] V. S. C. P. Narisetty and T. Maddineni, "Revolutionizing Mobility: The Latest Advancements in Autonomous Vehicle Technology," arXiv preprint, 2024

[23] B. D. Grozdanoff, "Reinforcement Learning System for a Moral AI Agent," Journal of Ethical AI, 2024

[24] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems (NeurIPS), 2012.

[25] Y. Tian, L. Fang, and J. Wang, "Deep Learning-Based Object Detection for Security Applications," IEEE Transactions on Information Forensics and Security, vol. 14, no. 3, pp. 789-803, 2019.

[26] L. Rabiner and B. Juang, "Fundamentals of Speech Recognition," IEEE Signal Processing Magazine, vol. 10, no. 3, pp. 45-66, 1993.

[27] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition," IEEE Signal Processing Letters, vol. 20, no. 3, pp. 356-367, 2012.

[28] J. Pimentel, R. Fernandez, and M. Lopez, "Anomaly Detection in Speech Patterns Using Machine Learning," IEEE Transactions on Audio, Speech, and Language Processing, vol. 22, no. 5, pp. 1025-1037, 2014.

[29] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," ACM Computing Surveys, vol. 41, no. 3, pp. 1-58, 2009.

[30] V. Gadepally et al., "Big Data Analytics for Security Applications," IEEE Security & Privacy, vol. 11, no. 6, pp. 74-79, 2013.

[31] J. Zeng, W. Chen, and H. Liu, "Multimodal Security Systems for Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 7, pp. 2345-2357, 2019.