Executive Summary

Comprehensive Autonomous AI-Driven Driving System

Infosys Springboard Internship 5.0

Under the guidance of

Mr. Nitig Singh

Abstract

This research develops a modular AI-driven framework for autonomous vehicles, integrating state-of-the-art models for object detection (SSD MobileNet), lane detection (YOLO-based), semantic segmentation (U-Net), and traffic sign recognition (YOLOv5). The system demonstrates high accuracy and real-time performance in urban and highway environments, with SSD MobileNet achieving a 20ms inference time, U-Net with a high pixel-wise accuracy, and YOLOv5 detecting traffic signs within 10ms per frame. Limitations include reduced performance in low-light and adverse weather conditions. The findings contribute to the development of scalable autonomous systems for safer transportation.

Objective

The objective of this research is to design and implement an AI-based autonomous driving system capable of real-time obstacle detection, lane maintenance, hazard warnings, and traffic sign recognition. By integrating cutting-edge AI models, the goal is to enhance vehicle safety and navigation efficiency in diverse driving environments.

Introduction

The advent of artificial intelligence has ushered in new possibilities in autonomous driving, transforming conventional vehicular systems into intelligent, adaptable agents capable of navigating complex environments. The integration of AI-driven models into autonomous systems has led to improved object detection, lane tracking, and environmental awareness, critical for real-time decision-making. Despite advancements, the implementation of robust systems that balance accuracy and computational efficiency in real-world scenarios remains a challenge. This research addresses these issues by developing a modular framework where each module is tailored for a specific task, ensuring the system's adaptability and scalability across various driving conditions. This study focuses on evaluating the performance of cutting-edge AI models integrated into an autonomous driving framework and their effectiveness in urban and highway environments [4].

Past Work

Recent years have seen considerable progress in the domain of autonomous driving, with AI models being employed for tasks such as object detection, semantic segmentation, and traffic sign recognition. Convolutional Neural Networks (CNNs) [2], such as SSD MobileNet, have gained prominence for their lightweight architecture and real-time processing capabilities, making them suitable for object detection in resource-constrained environments [3]. YOLO models, known for their speed and accuracy, have been extensively applied in real-time lane and object detection tasks [1]. U-Net, originally developed for medical imaging, has been repurposed for semantic segmentation in autonomous driving due to its ability to produce detailed pixel-wise classifications. However, the integration of these models into a cohesive system poses challenges, including the alignment of inference times and ensuring computational efficiency without compromising accuracy. This research builds on previous studies by not only employing these models but also addressing their integration within a modular architecture.

Dataset

The dataset used in this study is the Cityscapes dataset, which provides a diverse range of real-world scenarios to simulate complex urban driving conditions. The Cityscapes dataset comprises high-resolution images captured in urban environments under varying conditions, including daylight, twilight, and adverse weather. It includes detailed annotations for semantic segmentation, object detection, and lane markings, ensuring a comprehensive representation of real-world challenges. To further enhance model performance and generalizability, data augmentation techniques such as flipping, cropping, and noise addition were applied to the dataset. The data was divided into training, validation, and testing subsets, allowing for an unbiased evaluation of the system's performance across different tasks.

Methodology

The proposed framework follows a modular architecture where individual tasks are addressed by specialized models. Object detection is performed using SSD MobileNet due to its lightweight design and real-time inference capabilities. Lane detection employs a customized YOLO-based approach, optimized for identifying curved and broken lane boundaries. Semantic segmentation is handled by U-Net, which provides high pixel-wise accuracy necessary for understanding complex road layouts. For traffic sign recognition, YOLOv5 is employed due to its superior detection speed and accuracy. Each module was trained and optimized independently before integration into the system. The modules were interconnected through a central decision-making algorithm, which synthesized the outputs to generate navigational commands.

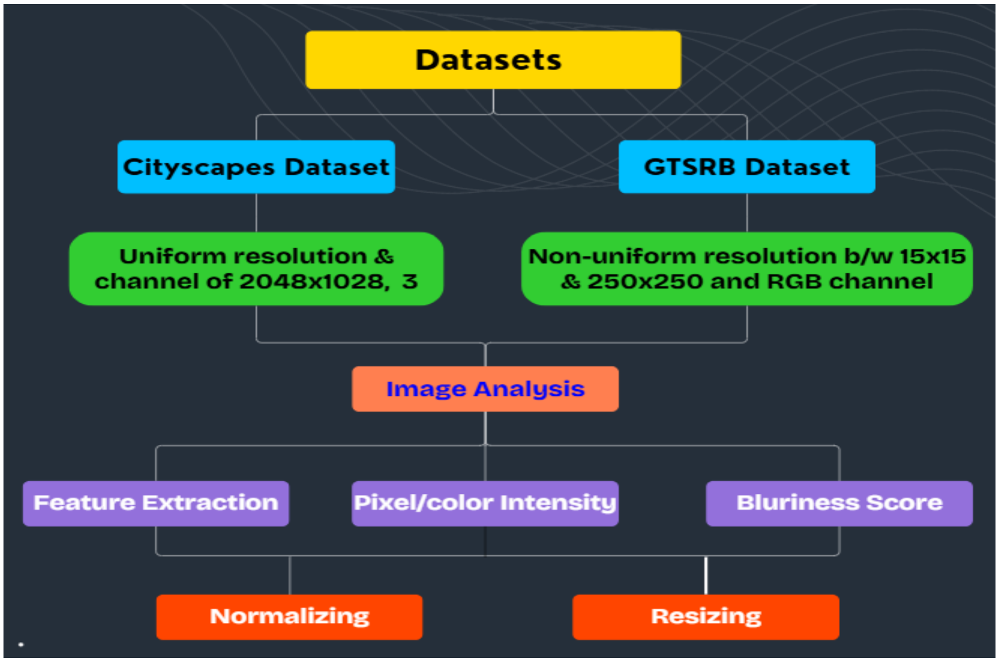

Fig 1: Flow Chart of the Project

Result and Discussions

The system's performance was evaluated through the integration of task-specific models selected for their balance between accuracy and computational efficiency. Each module underwent rigorous testing, demonstrating its effectiveness in fulfilling the requirements of the autonomous driving framework. The table below presents the final model selection for each module, along with key performance metrics such as inference time and accuracy:

Table 1: Results

Module	Model	Key Features	Inference Time	Accuracy
Object Detection	MobileNet SSD	Lightweight, efficient, real-time detection	20ms/frame	~88.6%
Lane Detection	YOLOv8	Advanced version of YOLO, robust for curves	25ms/frame	~90%
Semantic Segmentation	U-Net	Pixel-wise accuracy, optimized for road layouts	30ms/frame	~80%

Traffic Sign Recognition	YOLOv5	High-speed detection with reliable accuracy	10ms/frame	~93%

Each model contributed to a modular system that effectively performed its respective task in diverse urban and highway environments. MobileNet SSD provided efficient detection of objects such as vehicles and pedestrians, maintaining real-time performance. YOLOv8 was selected for lane detection due to its robustness in identifying lane boundaries, including curved and discontinuous markings. U-Net excelled in semantic segmentation, achieving detailed pixel-level classifications of road scenes. YOLOv5 reliably detected and classified traffic signs, ensuring accurate navigation and adherence to traffic rules.

These results underscore the importance of carefully tailoring model selection to meet the specific demands of autonomous driving tasks, balancing speed, and precision to enable real-time operational capabilities.

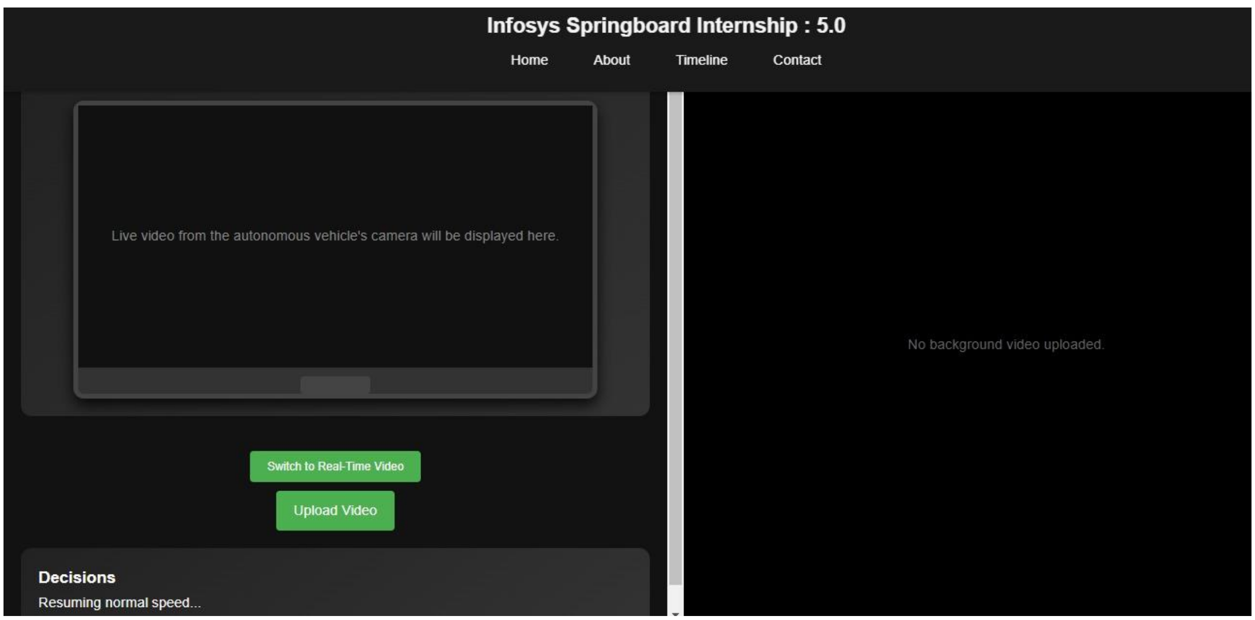

Fig 2: Home page of the website.

Conclusion

This research demonstrates the potential of a modular AI-driven framework for autonomous driving, achieving high accuracy and real-time performance in diverse scenarios. By integrating task-specific models, the system balances computational efficiency and robustness, addressing key challenges in autonomous navigation. However, the findings underscore the need for further enhancements, such as incorporating sensor fusion techniques and expanding datasets to cover more diverse scenarios. Future work will explore the use of reinforcement learning for adaptive decision-making and the development of hybrid architectures to improve performance under challenging conditions. This study contributes to the ongoing efforts to develop scalable, real-world-ready autonomous driving systems, paving the way for safer and smarter transportation solutions.

Future Scope

Future enhancements of the system will include the integration of **drowsiness detection** to monitor driver alertness and introduce **sophisticated decision-making algorithms** to make the system more adaptive to real-world complexities. Additionally, reinforcing decision-making through **reinforcement learning** and improving lane detection and object recognition under varying conditions will further strengthen the system's robustness.

Acknowledgements

We would like to express our gratitude to **Infosys Spingboard**, for their valuable support throughout the project. Special thanks to our mentors, leads, co-leads, and team members whose expertise and collaboration made this research possible. Their guidance and encouragement were crucial in the successful completion of this project.

References

1. **Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.** (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. https://doi.org/10.1109/CVPR.2016.91

2. **Long, J., Shelhamer, E., & Darrell, T.** (2015). Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

3. **Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L.** (2014). Microsoft COCO: Common Objects in Context. *European Conference on Computer Vision (ECCV)*, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

4. **González, J. D., Valcarce, A., & Béron, P.** (2018). A survey of autonomous vehicle systems: An analysis of challenges and opportunities. *IEEE Access*, 6, 16758-16770. https://doi.org/10.1109/ACCESS.2018.2812648