1. Introduction

1.1 Background

Driver distraction is a significant issue in road safety, contributing to a large percentage of traffic accidents worldwide. According to the World Health Organization (WHO) and the National Highway Traffic Safety Administration (NHTSA), driver inattention is one of the leading causes of collisions. Distractions such as mobile phone usage and improper handling of the steering wheel can lead to delayed reaction times and reduced situational awareness, increasing the risk of accidents.

Traditional driver monitoring systems rely on complex hardware and computationally intensive AI models, often requiring powerful GPUs for real-time processing. However, with the advancements in Tiny Machine Learning (TinyML), it is now possible to implement efficient and low-power AI models on embedded devices. This project aims to develop a lightweight, real-time driver distraction detection system using TinyML and Nicla Vision, focusing on detecting whether the driver is holding the steering wheel and whether they are using a mobile phone while driving.

1.2 Project Objectives

The primary objective of this project is to enhance road safety by implementing a compact, low-power, and real-time driver distraction detection system. The system specifically focuses on:

- Detecting hands-off steering wheel behavior to identify when the driver is not actively controlling the vehicle.
- Detecting mobile phone usage while driving, which is a major cause of inattention and road accidents.
- Implementing an energy-efficient solution using TinyML to enable real-time inference on an embedded device.

This solution aims to be cost-effective, power-efficient, and deployable on a large scale, making it suitable for both individual drivers and fleet management applications.

2. Actions Taken to Carry Out the Project

To achieve the project objectives, the following steps were taken:

2.1 Data Collection and Model Training

- A dataset of driver hand positions and mobile phone usage scenarios was collected using the Nicla Vision's camera.
- Images were annotated and labeled to distinguish between:
    1. Hands on the steering wheel vs. Hands off the steering wheel
    2. No phone usage vs. Holding a mobile phone
- The dataset was used to fine-tune the FOMO (Faster Objects, More Objects) model, a lightweight object detection model optimized for edge devices.

2.2 Fine-Tuning the FOMO Model

- The FOMO model, designed for real-time object detection on low-power devices, was chosen due to its efficiency in detecting multiple small objects in an image.
- The model was fine-tuned using transfer learning, where it was trained on the collected dataset of driver behavior.
- The training process included:
    - Data augmentation (rotation, scaling, and brightness adjustments) to improve robustness.
    - Quantization to reduce model size and optimize it for TinyML.
    - Validation and testing to ensure high accuracy in distinguishing between safe and distracted behaviors.

2.3 Deployment on Nicla Vision

- The trained FOMO model was converted into a TinyML-compatible format and deployed onto the Nicla Vision board.
- The Nicla Vision processes video frames in real-time to classify driver behavior.

2.4 System Testing and Evaluation

- The system was tested in various lighting conditions and environments to ensure reliability.
- Performance metrics such as detection accuracy, inference time, and power consumption were recorded.
- The edge inference was optimized to achieve low latency and real-time feedback, ensuring the system could function effectively without cloud-based processing.

3. Results achieved

The TinyML-based Driver Distraction Detection System using Nicla Vision and the FOMO model achieved strong performance in detecting hands-off steering wheel and mobile phone usage in real-time. The model was evaluated on a validation set, with the following key performance metrics:

- **Precision: 0.82** – indicating that 82% of detected distractions were correctly classified.
- **Recall: 0.93** – showing that 93% of actual distractions were successfully identified.
- **F1 Score: 0.87** – reflecting a balanced trade-off between precision and recall.

These results demonstrate that the system is highly effective at detecting driver distractions while maintaining low computational requirements, making it ideal for real-world deployment on embedded devices. Future improvements could further enhance detection accuracy through data augmentation, model refinement, and additional sensor integration.

4. Problems Encountered and Solutions

During the development of our TinyML-based Driver Distraction Detection System, one of the major challenges we faced was deploying our YOLO model on the Nicla Vision board. YOLO (You Only Look Once) is a powerful object detection model, but it requires substantial computational resources and memory to process images efficiently. The Nicla Vision, being a low-power, resource-constrained embedded device, was unable to load and run the YOLO model due to insufficient memory and processing power.

This issue made it impossible to perform real-time inference on the Nicla Vision, rendering our initial approach unfeasible. To overcome the memory constraints, we re-evaluated our model selection and adopted a more lightweight solution:

1. **Switching to the FOMO Model (Faster Objects, More Objects)**
    - We replaced YOLO with the FOMO model, which is specifically designed for low-power devices like the Nicla Vision.
    - Unlike YOLO, FOMO does not require high computational resources, allowing it to run efficiently on embedded platforms.
    - FOMO is optimized for edge devices and maintains real-time object detection capabilities while using significantly less memory.
    - This change enabled our system to function smoothly and in real-time without requiring additional external hardware.
2. **Using Edge Impulse for Model Training and Deployment**
    - To further optimize our solution, we utilized Edge Impulse, an online platform designed for TinyML model development and deployment on edge devices.
    - Edge Impulse provided a streamlined pipeline for:
        - Data collection and labeling
        - Model training and fine-tuning
        - Quantization and memory optimization
        - Direct deployment to Nicla Vision
    - With Edge Impulse, we could train and optimize the FOMO model, ensuring that it met both memory constraints and real-time performance requirements.

5. Conclusion

This project successfully developed a TinyML-based Driver Distraction Detection System using Nicla Vision and the FOMO model to detect hands-off steering wheel and mobile phone usage in real-time. Overcoming memory constraints with YOLO, we optimized the model using Edge Impulse, enabling efficient, low-power, and accurate detection on an embedded system. The system is compact, cost-effective, and energy-efficient, making it ideal for real-world applications. Future improvements could include expanding detection capabilities and integrating mobile connectivity. This work highlights TinyML's potential in automotive safety, contributing to safer driving and accident prevention.