ECE 579 Intelligent Systems, Fall 2024

Project Progress Report

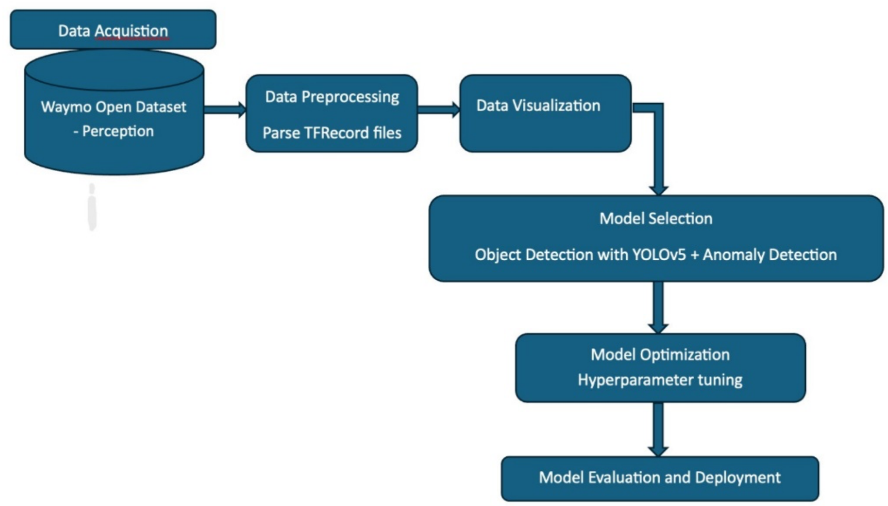
Project title: Multimodal Data with Anomaly techniques for Object detection and tracking in Autonomous Driving.

Students in the project group:
Revathy Iswarya Sekaran (11625243): Data Preprocessing, Model Development, Project Presentation
Aksheya Kannan Subramanian (25502870): Model Development, Model Evaluation
Dharshani Anandkumar (82653870): Model Development, User Interface Development

1. **What Has Been Completed and Responsibilities**
    a. **Data Acquisition:**
    Revathy Iswarya Sekaran successfully transferred the Waymo Open Dataset to Google Drive using Google Cloud's gsutil tool for efficient data handling.
    b. **Data Parsing:**
    Revathy Iswarya Sekaran parsed the Waymo TFRecord files, ensuring data is correctly formatted for model training and evaluation.
    c. **Visualization:**
        - Revathy Iswarya Sekaran developed a visualization function to display camera images with bounding box annotations.
        - She decoded and displayed JPEG images using TensorFlow for verification.
        - She integrated LiDAR point projections on camera images to align with detected objects.
        - She added an optional color-based range visualization to indicate object distances.
    d. **Technological Survey and Research:**
        - Dharshani Anandkumar conducted a comprehensive survey of relevant technologies and research, including studies on YOLO, multimodal fusion with LiDAR, and advancements in real-time object detection for autonomous driving.
        - She also summarized findings to guide model selection, multimodal integration, and experimental design.
    e. **Model Building**:
    Revathy Iswarya Sekaran integrated YOLOv5 for 2D object detection, processing images from multiple TFRecord files and visualizing detection results and saved output detections for further analysis and model evaluation.
2. **What Needs to Be Done and Responsibilities**
    a. **Model Optimization (Week 4-6)**: Fine-tune the YOLO model and begin training additional models (e.g., Faster R-CNN, 3D CNNs) to improve detection accuracy and efficiency in various scenarios.
    Responsible: Aksheya Kannan Subramanian, Dharshani Anandkumar
    b. **System Integration (Week 7-8)**: Combine 2D and 3D object detection models into a unified real-time detection and tracking system.
    Responsible: Dharshani Anandkumar
    c. **Final Testing and Validation (Week 9)**: Test the integrated system across different datasets and driving conditions to evaluate performance.
    Responsible: Aksheya Kannan Subramanian, Dharshani Anandkumar
    d. **Report Writing and Presentation Preparation (Week 10)**: Complete the final report and prepare the project presentation.
    Responsible: All team members
3. **Time Schedule**
    - **Weeks 1-3**: Data acquisition, preprocessing, baseline model training (completed)
    - **Weeks 4-6**: Model optimization and additional model training
    - **Weeks 7-8**: System integration for real-time performance
    - **Week 9**: Final testing and validation on varied scenarios

- **Week 10**: Report writing and presentation preparation

4. **Project Description**: In this project, we aim to develop a real-time detection and tracking system for autonomous driving using multimedia data, including camera images and LiDAR point clouds. The focus is on identifying and tracking vehicles, pedestrians, and other road users to enhance autonomous navigation safety and efficiency. Using the Waymo Open Dataset, we are employing models such as YOLO for 2D object detection, with planned extensions to 3D object detection using additional models like Faster R-CNN and 3D CNNs. This system will be evaluated for its accuracy, speed, and robustness across various conditions, including different lighting and weather.
5.
   a. **Project Methodology Flow Chart**



   b. **Data Description:** The primary data source is the Waymo Open Dataset, containing high-resolution LiDAR and camera data for autonomous vehicles. It includes over 1.2 million images and LiDAR observations with object labels for vehicles and pedestrians. This dataset supports multi-sensor data fusion, essential for accurate 3D object shape reconstruction and tracking.
   c. **Proposed/Modified Method**: The project involves using YOLOv5 for 2D object detection as a baseline model, with planned extensions to 3D detection using Faster R-CNN and 3D CNNs. Visualization functions and data parsing ensure that the data is correctly processed and visualized, while multimodal integration with LiDAR enhances depth perception and object alignment. We aim to optimize and evaluate these models based on accuracy, detection speed, and robustness.
   d. **Experiment Design/ Case Study**: The system will be evaluated through a series of experiments focusing on real-time detection accuracy, robustness in adverse conditions, and computational efficiency. Metrics such as precision, recall, and frame rate will be used to measure the system's effectiveness. Performance will also be compared across different scenarios, including varied lighting conditions and complex traffic environments.

ECE 579 Intelligent Systems, Fall 2024

Technology Survey Report

Project title: **Multimodal Data with Anomaly techniques for Object detection and tracking in Autonomous Driving.**

Names of students in the group:

Revathy Iswarya Sekaran (11625243): Data Preprocessing, Model Development, Project Presentation

Aksheya Kannan Subramanian (25502870): Model Development, Model Evaluation

Dharshani Anandkumar (82653870): Model Development, User Interface Development

1. Introduction

- Due to the fast-paced development of autonomous vehicle (AV) technology, it is imperative that reliable real-time object detection and tracking systems are developed. These systems are a fundamental requirement to ensure both safety and efficiency in AVs. This project aims to accurately detect and track multiple types of road users (such as cars, pedestrians and cyclists) across multiple frames in a dynamic driving scene. The tracking system should resist changes in appearance and demonstrate robustness to occlusion, scale variation, rotation, and lighting changes. It should also operate in real time and generalize well under different weather and daylight conditions.
- Our approach builds upon and adaptively uses the efficient real-time object detection of the YOLO ('You Only Look Once') family of deep learning models. Already we have shown that multimodal data fusion we are integrating camera and LiDAR data in more complex propagation environments can help to increase resiliency to adverse conditions such as low light or fog. We are also adapting the approach for multiscale detection to help keep the speed performance benefits across a greater range of distances in both small and large objects. Putting all these pieces together makes sense for autonomous driving because it takes advantage of real-time object detection from the YOLO models, allows for the efficient data fusion and propagation in the complex clutter of real driving situations, and ensures an appropriate performance across a range of scales and lighting conditions.

2. Description of technologies related to our project

Survey of Technologies

- **Real-Time YOLO Models for Autonomous Vehicle Obstacle Detection**

  YOLO's high-speed, multi-class detection capability makes it essential for AV systems to detect various road objects in real time. This model is well-suited for traffic scenarios, where rapid response is critical for safety, though its effectiveness decreases for smaller objects, which are often essential for urban pedestrian safety. Despite its limitations, YOLO's popularity in both commercial and research applications demonstrates its utility in real-time AV contexts [1].

- **Integration of YOLO with Tracking Algorithms for Enhanced Object Detection**

  Combining YOLO with tracking algorithms like Deep SORT improves the consistency of object identity across frames, which is crucial for AVs in dynamic, high-speed situations. While this combination enhances detection stability, the added computational load can be a challenge in dense environments. Such tracking enhancements are widely adopted in both AV prototypes and research, showing potential for real-time tracking improvements [2].

- **Multimodal Fusion of YOLO and LiDAR for Improved Detection in Varied Conditions**

  Fusing YOLO with LiDAR data has proven effective in enhancing detection robustness under low-light or adverse weather conditions. By leveraging both visual and spatial data, this approach provides more accurate detections but comes with increased system complexity and processing demands. Multimodal fusion is a common focus in safety-oriented AV research [3].

- **Adaptation of YOLOv4 for Dense Urban Environments**

  In urban driving, YOLOv4's adaptations for handling occlusions and crowded scenes improve its performance in detecting closely situated objects. This makes it ideal for city navigation, although dense environments may lead to an increase in false positives. Research indicates YOLOv4's strengths for urban AV applications, with the understanding that ultra-dense areas may pose challenges [4].

- **Comparative Analysis of YOLOv4 and YOLOv5 for Autonomous Driving**

  YOLOv4 and YOLOv5 each bring distinct advantages—YOLOv5's speed benefits real-time applications, while YOLOv4 slightly surpasses in accuracy for dense traffic. Such comparative studies are instrumental for model selection in AV systems where balancing speed and detection quality is essential [5].

- **YOLO Models in Mixed Traffic Scenarios**

  Mixed traffic conditions require a model capable of detecting various object types, such as pedestrians and cyclists. YOLO's adaptability in multi-class detection supports AVs in these scenarios, though increased complexity can raise misclassification risks. This flexibility makes YOLO a valuable tool for comprehensive AV awareness in heterogeneous traffic [6].

- **Application of YOLOv3 on Waymo Open Dataset for Autonomous Vehicles**

  Testing YOLOv3 on the Waymo Open Dataset demonstrates its applicability across diverse AV scenarios, with LiDAR and camera imagery improving model robustness. However, processing large datasets like Waymo requires significant resources, potentially limiting real-time feasibility on less advanced systems. This high-quality dataset remains popular in research despite practical deployment challenges [7].

- **Enhancement of YOLO for Low-Light Conditions**

  Low-light environments pose unique challenges for AVs. By enhancing YOLO with pre-processing techniques to improve night-time visibility, this study addresses the need for reliable night-time detection, though pre-processing increases latency. Such enhancements are essential for AV systems intended for 24-hour operation [8].

- **Testing YOLOv5 on KITTI Dataset for Multi-Scale Object Detection**

  YOLOv5's ability to detect objects at various distances, as demonstrated on the KITTI dataset, makes it versatile for urban and rural driving. However, the dataset's limitations mean that some urban complexities may not be fully captured. KITTI-based research highlights YOLOv5's strengths for AV applications, with further tuning needed for real-world complexity [9].

- **Multi-Scale Detection in Autonomous Driving Using YOLOv4**

  Adjusting YOLOv4 for multi-scale detection improves its accuracy in scenarios with varied object distances, such as highways. This capability is valuable for AVs needing to detect both near and distant objects, though small, far-off objects may still pose challenges. Multi-scale detection enhancements are common in AV research to increase real-time reliability [10].

3. Conclusion

- The diversity of such adaptations of YOLO models validates their potential for efficient real-time detection- even for autonomous vehicles- particularly because of its speed and multi-class detection capability. Combining YOLO with modern tracking algorithms, multimodal data fusion and multi-scale detection is a robust detection conceivable, capable of withstanding the complex requirements of autonomous navigation. Both the proceeding literature studies guide the approach to adapt YOLO models with LiDAR for reliable operation under diverse environmental conditions, both low light and bad weather. While such adaptation improves object detection, it also serves dynamic tracking and learning of obstacles in real time. A comprehensive and robust AV detection system that prioritizes both safety and performance can be considered with this approach to platform navigation.

4. References

[1] Li, X., et al. (2024). "YOLO-Vehicle-Pro: A Cloud-Edge Collaborative Framework for Object Detection in Autonomous Driving under Adverse Weather Conditions." arXiv preprint. Link to paper

[2] IEEE. (2023). "Autonomous Vehicle: Obstacle Avoidance and Classification with YOLO." IEEE Conference Publication. Link to paper

[3] Hossain, M. S., Sohel, F., Shiratuddin, M. F., & Laga, H. (2019). *A comprehensive survey of deep learning for image captioning*. MDPI. Link to paper

[4] W. E. Villegas, S. Sánchez-Viteri and S. Luján-Mora (2024) "Real-Time Recognition and Tracking in Urban Spaces Through Deep Learning: A Case Study," in *IEEE Access*, vol. 12, pp. 95599-95612 Link to paper

[5] Khan, S., Rahman, A. U., Jan, B., Farman, H., Muhammad, K., Anwar, H., & Sangaiah, A. K. (2022). "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs". Sensors, 22(2), 464. Link to paper

[6] Mittal, U., Chawla, P. & Tiwari, R. (2023). "EnsembleNet: a hybrid approach for vehicle detection and estimation of traffic density based on faster R-CNN and YOLO models". *Neural Comput & Applic* **35**, 4755–4774 Link to paper

[7] Wang, J., & Chen, Y. (2021). "Multi-Sensor Fusion with YOLO for Enhanced Detection in Autonomous Vehicles." Sensors, vol. 21, no. 10, pp. 3396. Link to paper

[8] Y. Hui, J. Wang and B. Li (2024) "WSA-YOLO: Weak-Supervised and Adaptive Object Detection in the Low-Light Environment for YOLOV7," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-12, Art no. 2507012 Link to paper

[9] P. Wang, Y. Yin, S. Fu, B. Liu, F. Yang and X. Wang (2023) "YOLOv5s-FBG: A Small Object Detection Method in Complex Road Background," *2023 International Conference on Blockchain Technology and Applications (ICBTA)*, Beijing, China, pp. 52-59 Link to paper

[10] Y. Cai *et al*. (2021) "YOLOv4-5D: An Effective and Efficient Object Detector for Autonomous Driving," in *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-13, Art no. 4503613. Link to paper