Safe Driving Across Domains

Products of Research

This project involves two primary data sources. The first is driving logs from TIER IV and Waymo (open source), which include LiDAR and camera sensor data as well as location information from their deployed vehicles in 50+ Japanese cities and 25 cities in the US. These data will be utilized to validate dynamic optimization algorithms developed in the project, providing real-world scenarios and sensor inputs essential for testing and refining autonomous vehicle safety models. The second data source is camera sensor data, which focus on tracking trajectories for driving styles and vehicle-vehicle interactions. Both data sources and derivatives will be documented and stored on secure servers at Carnegie Mellon University (CMU) or CMU's Box cloud service account with regular backups, ensuring data integrity and accessibility for future research. For larger datasets, they will be stored on a 4TB SSD and the Pittsburgh Supercomputing Cluster. Detailed metadata will be provided to facilitate reproducibility and usability by other researchers, and all datasets, models, and software will be shared through CMU's institutional repository and public repositories like GitHub.

Data Formats and Standards

The data from this project will be stored in different formats to ensure accessibility and usability. The LiDAR data provided by TIER IV will be stored in the American Society for Photogrammetry and Remote Sensing (ASPRS) LAS format, which is a standard format for LiDAR data, and in a binary format for efficient storage and processing. Camera data will be stored in mp4 format for video files and jpeg format for image files, both widely used and supported by standard software platforms. If any data are initially stored in unusual or not generally accessible formats, they will be converted to more accessible formats as needed, ensuring they can be shared with and utilized by interested parties. In cases where existing standards are inadequate, documentation will be provided along with proposed solutions to ensure data accessibility with minimal added cost. All data will be securely stored on CMU's servers, with regular backups to maintain data integrity and availability for future research and collaboration.

Dissemination, Access and Sharing of Data

The driving log data from TIER IV, including LiDAR and camera sensor outputs, will be stored in American Society for Photogrammetry and Remote Sensing (ASPRS) LAS format for LiDAR and mp4 and jpeg formats for camera data. This data, along with pertinent metadata such as sensor calibration, location, environmental conditions, and timestamps, will be made available through public repositories like the CMU Kilthub Repository. Citizen science data from educational activities, focusing on human-vehicle interactions, will be stored in CSV format and include metadata on driving habits, interaction contexts, and participant demographics. A dedicated project website will be maintained by the PI's research group to provide ongoing access to the data, software, and documentation. This website will be hosted by Carnegie Mellon's Heinz College and maintained for at least five years post-project completion. The software and code developed as part of the project, including dynamic optimization algorithms, will be made available as open-source under a permissive license (e.g., MIT License) via

GitHub. Data will be released before formal publication whenever possible to encourage early collaboration and feedback from the research community. Collaboration with industry partners like TIER IV and research centers such as CMU's Safety21 will ensure that data and analytical tools are shared and managed effectively, supporting broader impacts and sustained use beyond the project duration. Publication delay policies will be clearly communicated to all stakeholders to manage expectations and ensure transparency.

Re-Use, Re-Distribution and Production of Derivatives

The data provided through general access or sharing will be accompanied by comprehensive metadata necessary for its re-use, including detailed descriptions of the data collection methods, analytical procedures, and any processing steps undertaken. Specifically, data from the driving logs (LiDAR and camera sensors) collected from TIER IV will include information about the location, environmental conditions, sensor calibration, and timestamps to ensure the reproducibility of the dynamic optimization algorithms being tested. Citizen science data collected from educational activities will include metadata about the driving habits observed, the context of human-vehicle interactions, and demographic information of the participants to inform dynamic optimization modeling for pre- and post-crash vehicle maneuvers. To protect privacy, confidentiality, security, and intellectual property, all sensitive data will be anonymized before sharing. Users accessing the data will be required to adhere to specific terms and conditions outlined on the project website, including the proper citation of the data source in any publications or products. Disclaimers regarding the limitations and appropriate use of the data will be clearly stated to prevent misuse. Additionally, the rights and obligations of users, including restrictions on redistribution and the requirement to acknowledge the original creators, will be explicitly communicated. This approach ensures that the data is both accessible and responsibly used by the broader research community, fostering collaboration and furthering advancements in the field of transportation safety and automation.

Archiving of Data

All data collected during the project, including driving logs from TIER IV and citizen science data, will be archived digitally. Digital data will be stored on secure servers managed by Carnegie Mellon University, with backup copies in distributed locations to prevent loss due to disasters.

Data will be periodically transferred to new storage media to address technological obsolescence. An easily accessible index on the project website will document the locations of all archived data and provide access instructions. For datasets lacking a suitable repository, metadata will be prepared and made publicly available online via the project website and the CMU KiltHub Repository. Partnerships with Safety21 will further support data preservation and sharing, ensuring long-term accessibility for researchers and stakeholders.