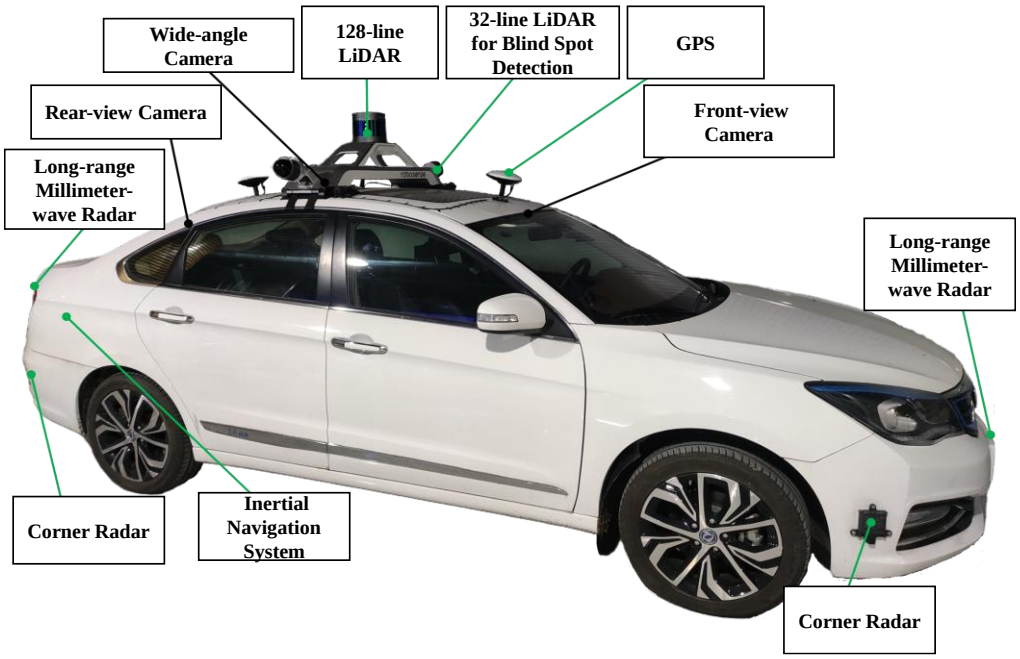
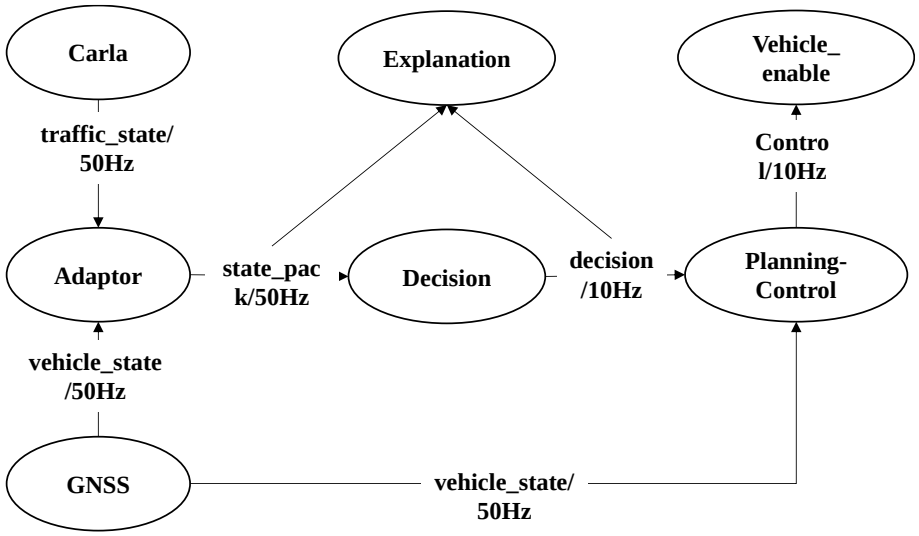
Test platform



Revision-Fig 1 Hardware Configuration Diagram of the Test Vehicle

The experimental vehicle, as depicted in Revision-Fig 1, is equipped with a camera, GNSS (Global Navigation Satellite System), and IMU (Inertial Measurement Unit) integrated positioning system. The lateral control interface of the vehicle is the steering wheel angle, while the longitudinal control interface includes deceleration commands and drive torque.

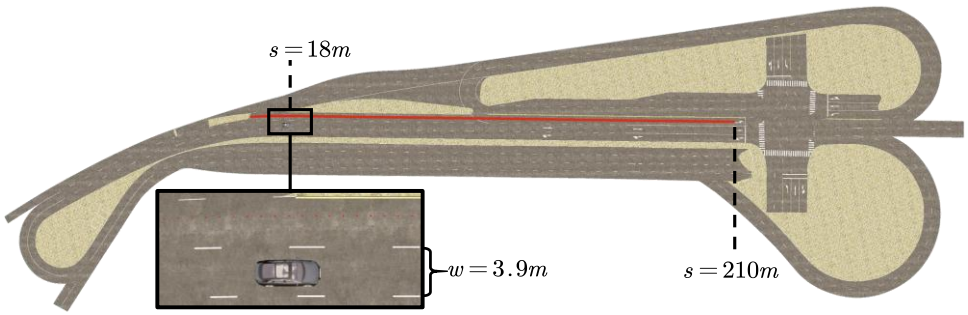
In the experiment, the communication system adopts ROS primarily due to its modular design, allowing independent development and testing of various functional modules, which are then integrated into a unified system. Secondly, ROS's communication mechanism facilitates efficient data exchange between different nodes for real-time transmission and processing of large volumes of data from both real vehicles and virtual environments. The ROS node design in this study is illustrated in Revision-Fig 2. The GNSS node publishes real-world autonomous vehicle speed, position, and heading information at a frequency of 50Hz. The Adapter node subscribes to GNSS messages and virtual environment nodes, converting real-world vehicle states from the GNSS node to corresponding vehicle states in the virtual environment coordinate system. It packages and publishes these states along with the status of traffic participants in the virtual environment at a frequency of 50Hz. The Decision module subscribes to Adapter messages and generates decision commands based on the main vehicle and traffic vehicle states, specifically lateral target lane commands and longitudinal target velocity commands at a frequency of 10Hz. The Planning Control module subscribes to decision messages and GNSS real vehicle states, generating desired steering wheel angles and drive torque or deceleration commands at a frequency of 50Hz. At the vehicle end, the Autonomous Driving Enable node ensures that subscribed vehicle steering wheel angles and longitudinal control quantities are transmitted to the vehicle and entered into autonomous driving mode only when both chassis enable signals and self-driving mode enable signals are satisfied. Finally, the Explanation node subscribes to Adapter and Decision messages, generating explanations.



Revision-Fig 2 ROS Communication Architecture

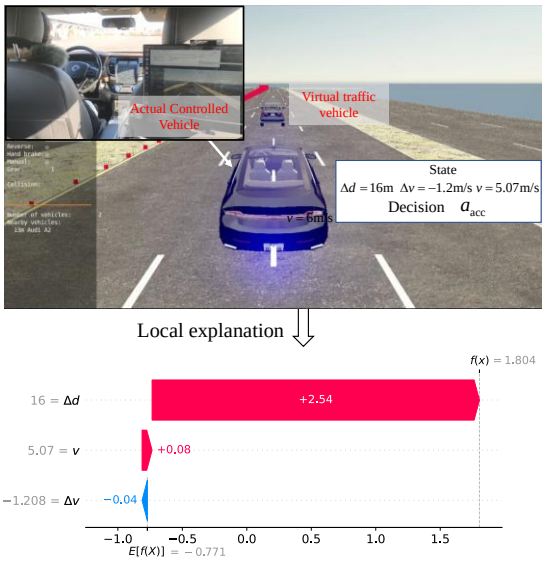
Test result

The experiment was conducted at the East District Test Site of the Tongji University Intelligent and Connected Vehicle Testing Base, as shown in the Revision-Fig 3. A test route with a total length of 210 meters was selected, with the vehicle starting at a longitudinal position of 18 meters. The lane width was 3.9 meters.



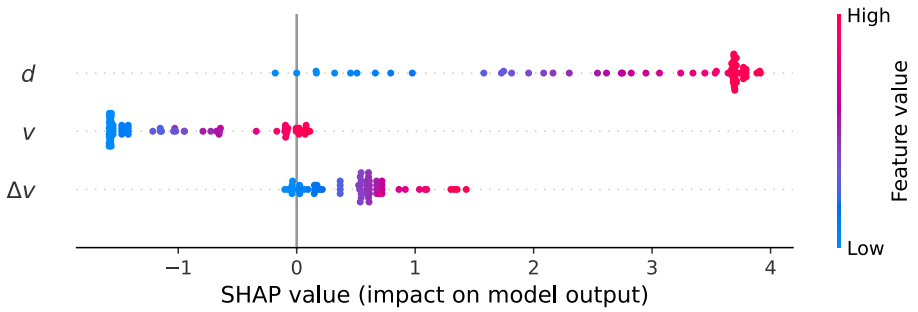
Revision-Fig 3 Tongji University's Intelligent and Connected Vehicle Testing Experimental Field

Local explanations are provided to show the contributions of features to a given instance because features with large Shapley values are important. It follows from the positive SHAP scheme that the sum of the feature contributions is always non-negative and equal to the bias between the raw prediction of the model and the base value. The local explanation approach implements Deep SHAP, which calculates the feature contributions with respect to the Q-functions. As shown in Revision-Fig 4, when the relative distance is 16, the contribution is +2.54, indicating a positive impact on the a_{acc} decision. A large relative distance implies that the vehicle is far from the vehicle in front, prompting the model to accelerate to achieve a higher speed. In contrast, since relative distance has the most significant impact, the other two features—relative velocity and speed—have a comparatively smaller influence on the current acceleration decision.



Revision-Fig 4 Local explanation for a_{acc} decision

Global interpretation is used to help researchers investigate the most important features and their impact on the model in a post-hoc manner. The SHAP value distribution summary is a global interpretation plot that shows the relationship between the feature value and its impact on the prediction. By combining feature importance with feature effects, each point in the model summary is the Shapley value of a feature or an instance. As shown in Revision-Fig 5, the positions on the y-axis and x-axis are determined by the feature and Shapley value, respectively. The color represents the feature value ranging from low to high. Overlapping points are stacked on the y-axis. From the figure, it can be observed that as the relative distance increases, the model tends to favor acceleration, which is indicated by the positive SHAP values. Similarly, as the relative speed increases, the model also tends to favor acceleration. These tendencies align with common driving behaviors, where drivers are more likely to accelerate when there is a greater distance to the vehicle in front and when the relative speed difference is larger. However, it can be observed that as the vehicle's speed increases, the model becomes more inclined to accelerate. This behavior seems counterintuitive when compared to general driving principles, and it likely results from the complex interaction between relative velocity and speed.



Revision-Fig 5 SHAP value distribution for a_{acc}