

# Upgrading an Automated Vehicle Research Platform for Enhanced Perception and Distributed Computing

Haoran Ding<sup>\*</sup>, Nithish Kumar Saravanan, Hengcong Guo,  
Nathan Fahner, Pranav Ramesh Bidare, Jeffrey Wishart,  
Junfeng Zhao

<sup>\*</sup> Arizona State University

**Abstract:** This paper presents significant upgrades to our automated vehicle research platform, where we focus on improving the perception system and distributing computational workloads for improved system efficiency. We expanded the original setup from a single front camera to a system with six cameras providing a full 360-degree view of the vehicle. To do so, we carefully mounted and calibrated the cameras with existing LiDAR sensors to perform sensor fusion. In addition to sensor integration, we upgraded the vehicle’s computing architecture by distributing computational tasks between a Nuvo computer and a Jetson edge computer. These enhancements have resulted in an overall increase in perception performance and real-time data processing capabilities, making the platform more robust for advanced automated driving tasks.

**Keywords:** automated Vehicles, Sensor Fusion, Perception Systems, Computer Vision, Distributed Computing, Object Detection

## 1. INTRODUCTION

Automated vehicles (AVs) have emerged as a leading research topic in transportation with the potential to revolutionize travel, improve road safety, and reduce traffic congestion. Vision-based self-driving technology, in particular, has gained more attention in recent years due to its cost-effectiveness and ability to provide rich semantic information about the environment. Camera-focused tasks such as object detection, classification, lane detection, and vision-based systems are becoming prominent in AV perception pipelines. When integrated with complementary sensors such as LiDAR and radar, these systems enable a robust and accurate understanding of complex driving environments.

As vision-based perception methods emerge as a leading trend in automated vehicle research, the development and evaluation of camera systems and a reliable testing platform has become a hot topic. These systems offer researchers a controlled and scalable environment to design, assess, and refine new algorithms and hardware configurations. Camera integration and distributed computing are vital to develop such platforms, as they are essential to test camera-only perception and actual deployment on AV. This paper presents an upgraded AV research platform to address these challenges and discusses potential innovations in vision-based automated driving systems.

The original vehicle set-up, specific to our lab as described in Guo et al. (2024), consisted of a fully electric 2022 Ford Mustang Mach-E integrated with Autoware (Autowarefoundation (2024)) and the DataSpeed Drive-by-Wire (DBW, Bodell (2024)) system. The system included a front-facing Lucid RGB camera with a 60-degree field

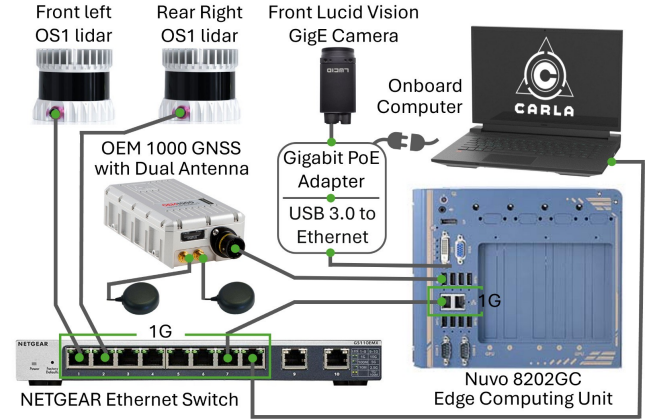


Fig. 1. Original Vehicle Setup

of view (FOV), two Ouster OS1-64 LiDAR sensors, a Conti long-range radar unit, and a OxTS GNSS GPS unit with two antennas. The DBW system facilitated precise longitudinal and lateral control through a brake-throttle emulator and steer-shift interface module. At the same time, a high-performance Nuvo computer managed sensor data processing and autonomy tasks. Autoware provided perception, planning, and control modules while our lab provided a vehicle interface package that bridges Autoware’s control commands with the DBW kit.

This setup formed a research platform for automated driving tasks, which works with various modules along the automated driving pipeline, such as sensing, mapping, localization, planning, and control. However, reliance on a single front-facing camera and centralized computing posed significant limitations. The narrow FOV of one single camera is insufficient for perception in complex envi-

ronments where obstacles and vehicles could appear from multiple directions. Furthermore, the centralized architecture imposed a heavy computational burden on the Nuvo computer, leading to bottlenecks in real-time performance as sensor data become heavy and the complexity of the perception algorithm increases.

To enhance the capability of this research platform and tailor it for vision-based perception in our setup, we expanded the system to include six cameras for 360-degree FOV perception coverage, which significantly improved the system’s sensing capabilities. Moreover, we upgraded the vehicle’s computational architecture by distributing tasks between the previous Nuvo computer and a newly added Jetson Orin edge computer. The Jetson computer handles GPU-heavy perception tasks, while the Nuvo computer focuses on the CPU-heavy portion perception and previous modules.

The key contributions of this work are:

- A well-designed six-camera vehicle system that provides full 360-degree perception coverage, enhancing the detection capabilities and safety of automated driving.
- The integration of multisensor hardware, including cameras and LiDAR calibration, into the Autoware pipeline to enhance perception capabilities and support sensor fusion.
- A distributed computing architecture that effectively partitions tasks between edge and central units, reducing latency and improving system efficiency.

These advancements establish a more robust and scalable platform for automated vehicle operations and provide a reference methodology for similar upgrades in other automated systems.

## 2. RELATED WORK

Recent developments in AV datasets and platforms have highlighted the growing interest in purely vision-based approaches for perception systems. Over a decade ago, the KITTI dataset (Geiger et al. (2012)) was released, and until today it remains a classic and widely recognized dataset. KITTI equipped its vehicle with two colored cameras and two grayscale cameras, all front-facing, one LiDAR, and a GPS/IMU localization unit. The primary goal of the original KITTI dataset was to support research in autonomous driving, specifically in urban and highway driving environments. These scenarios predominantly required forward-facing data for object detection, tracking, and lane-following tasks. The need for 360-degree data became apparent as research in autonomous driving and robotics evolved, requiring a more comprehensive understanding of the environment for tasks like full-scene semantic mapping, multi-agent interaction, and robust localization in complex scenarios. The KITTI team recently released a new dataset in response to evolving research needs. KITTI360 (Liao et al. (2022)) is a new project that provides richer sensor information and 360° annotations. The new KITTI360 vehicle had one 180° fisheye camera on each side and a 90° perspective stereo camera (baseline 60 cm) on the front. The upgraded KITTI360 has dense semantic and instance annotations for 3D point clouds and 2D images.

While the KITTI dataset helped validate many perception algorithms, the KITTI team designed it as a multipurpose dataset. Real-time perception tasks in autonomous vehicles would require a more explicitly optimized platform.

The most widely used AV research dataset would be the nuScenes dataset (Caesar et al. (2020)), which had six cameras, five radars and one LiDAR, providing a 360-degree field of view. This dataset has been instrumental in advancing multi-camera perception and sensor fusion techniques in autonomous driving and has inspired many AV research platforms. They have also collected data in various environments, including roads, lighting, and weather conditions. The nuScenes vehicle has an efficient and reliable sensor system for on-road automated vehicles. Various advanced fusion algorithms based their results on the dataset, such as BEVFusion (Liu et al. (2023)), BEVDet (Huang et al. (2021b)), and many more. The reliability of the nuScenes perception results led to the upgrade of our sensor system.

As AVs become more complex, the challenge of handling the significant increase in sensor data flow becomes more pronounced. For example, with its numerous sensors, the Waymo Open Dataset (Sun et al. (2020)) emphasizes the development of scalable perception systems through multi-camera setups and LiDAR integration. This dataset, with its extensive multi-sensor data collection, provides a foundation for research into sensor fusion techniques. The Waymo vehicle, with its four LiDARs and ten cameras, generates an unprecedented data flow of sensor data, ensuring accurate real-time perception. The results demonstrate how a highly complicated and powerful sensor system can achieve high accuracy, particularly in challenging urban environments. However, the challenge of handling the massive data flow from such a complicated sensor system has also become a question that has inspired us to develop a distributed computing system to address this challenge.

In addition, some datasets focus on other directions. The H3D dataset (Patil et al. (2019)) underscores the importance of data collection under diverse and complex environmental conditions. The H3D vehicle is equipped with 1 LiDAR and three cameras but resulted in a high density of 3D bounding boxes due to the area of data collection. ApolloScape (Huang et al. (2018)) focused on LiDAR perception, with their vehicle equipped with 2LiDAR and six cameras. They provided a survey-grade dense 3D point cloud for static objects, with a rendered depth map associated with each image, creating the first pixel-annotated RGB-D video for outdoor scenes. The Oxford RobotCar dataset (Maddern et al. (2017)) focuses on environmental changes. The RobotCar dataset focused more on environmental changes along the same route. The RobotCar vehicle traversed the same routes at different times, providing rich environmental data such as changes in lighting, weather variations, and dynamic obstacles. These datasets, with their unique focus on diverse environmental conditions, LiDAR perception, and temporal changes, provide valuable add-ons to AV research.

In addition to research domains, the industry has also shown a trend toward vision-based perception. Take the most successful commercialized AV brand Tesla as



Fig. 2. ZED X ONE cameras

an example. As demonstrated by Tesla’s Vision system (BelfastEV (2023)), they aim to rely solely on cameras for vehicle perception without needing more costly lidar sensors. Tesla has an eight-camera setup that provides full 360-degree coverage and enables the platform to perform high-accuracy perception using only cameras. Based on these datasets and the development trend of AVs, we decided to upgrade our vehicle. The success of the Nuscens dataset and Tesla has emphasized the importance of multi-camera perception for autonomous driving, and the sensor system upgrade aimed to benchmark our platform against research and industry standards.

As powerful Waymo vehicles have shown, upgrading the computing system becomes crucial as AVs become more complicated and data-heavy. Alabyad et al. (2024) also highlighted the evolution towards data-heavy computing in AV systems, discussing the integration of deep learning methods such as convolutional neural networks (CNNs) to improve obstacle and lane recognition using vision sensors. The high workload of these networks further supports our decision to improve system efficiency with distributed computing. The challenge in this upgrade, as mentioned in Shahian Jahromi et al. (2019), lies in implementing high-performance fusion systems on embedded edge computers. In search of an efficient modular framework, we improved the platform’s data transfer efficiency and the ability to split complex perception and control tasks. In conjunction with the sensor system upgrade, these enhancements allow the platform to investigate the feasibility of vision-based purely perception systems.

### 3. SYSTEM UPGRADES

The initial setup of our system relied on a single front-facing Lucid camera, which limited the perception capabilities to a narrow field of view. To overcome this limitation, we upgraded the system with six additional cameras to achieve 360-degree coverage using ZED X ONE cameras with GMSL2 connections (ZED (2024a)). The ZED X

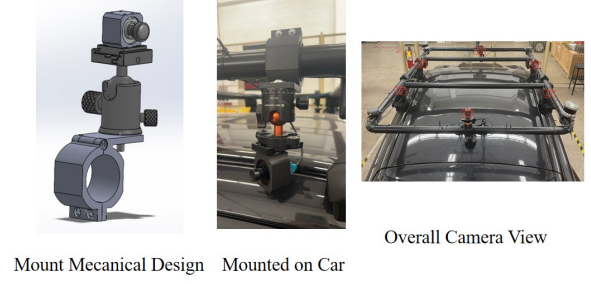


Fig. 3. Camera Mount and its final placement on vehicle

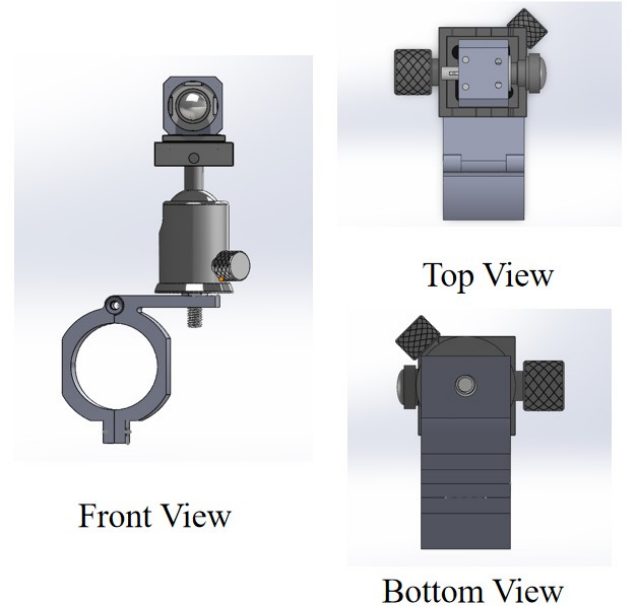


Fig. 4. Camera Mount Design

ONE cameras are compact, lightweight, and designed for high-performance perception in automated driving systems. Each camera features a resolution of 2.3 megapixels and a frame rate of up to 60 Hz, ensuring precise and smooth object detection even in dynamic environments. In our final setup, as shown in Figure 5, five cameras with 78-degree horizontal FOV were placed around the vehicle front and sides, while one last rear-facing camera with a 120-degree horizontal FOV covered the back. This setup enables an uninterrupted 360-degree FOV around the vehicle.

As can be seen from Figures 3 and 4, custom mounts were designed for the ZED X ONE cameras to integrate them into the system securely. These mounts are firmly attached to the vehicle and provide stable support even under rugged conditions. At the same time, the mounts remain flexible, allowing the cameras to rotate. This structure allows for future camera adjustments, ensuring better coverage and calibration tailored to evolving perception outcomes.

ZED X ONE cameras were chosen for the upgrade for various reasons. The ZED X ONE cameras have advanced sensors that perform well in various lighting conditions,

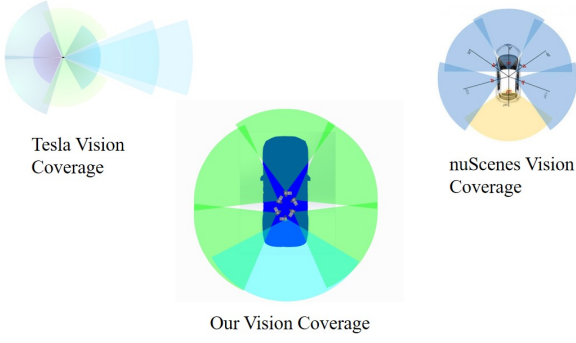


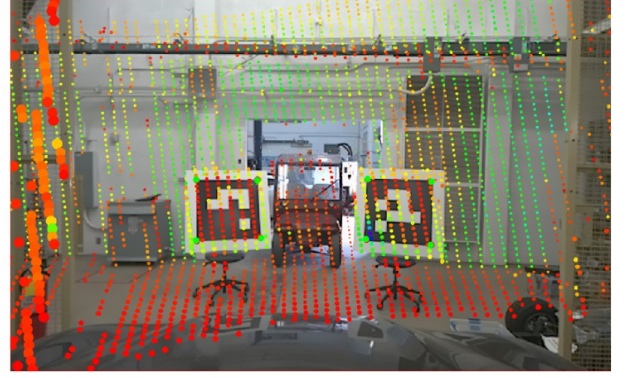
Fig. 5. Vision Coverage Comparison

making them reliable in diverse environments. The GMSL2 interface facilitates high-speed, low-latency sensor data transfer between cameras and the onboard computing system, enabling real-time image processing. The solid design of the cameras, which feature IP68-rated water and dust resistance, ensures reliable operation under challenging environmental conditions (ZED (2024b)). These cameras will significantly improve our vehicle’s perception capabilities, aligning it with research and industry benchmarks.

#### 4. CAMERA-LIDAR CALIBRATION

Integrating the spherically distributed cameras around the vehicle’s roof with LiDAR sensors required a precise ROS2 transform (tf) to align the coordinate frames of the sensors. While exploring the correct method of calibration, three approaches were considered. The manual calibration provided by DataSpeed (Dataspeed (2023)) involved adjusting the tf to match the pointcloud clusters with the corresponding objects in the camera image. Although this approach could be a temporary solution, it was highly time-consuming and prone to human errors. The other two methods were based on point matching between the pointcloud and the image; one is manual point selection, and the other is automated. The fully automatic **tag\_based\_pnp\_calibrator** (tier4 (2024)) was finalized to minimize manual errors that could propagate and amplify during later calibration stages. This approach minimized manual intervention while achieving high calibration accuracy through advanced techniques.

The **tag\_based\_pnp\_calibrator** aligns 2D points detected in camera images with their corresponding 3D points from LiDAR scans, using fiducial markers called AprilTags, as seen in Figure 6. The calibration process utilizes two key detection tools for critical points for matching: **apriltag\_ros** (Olson (2011)) and **lidartag** (Huang et al. (2021a)). **apriltag\_ros** identifies AprilTags in the camera images, extracting 2D positions and unique identifiers from the marker. **lidartag** detects corresponding markers in



● : Detected Corner from LiDAR pointcloud  
● : Detected Corner From Camera Photo

Fig. 6. Final Calibration Method: The point cloud and camera image are well-aligned at the center, demonstrating accurate calibration. The slight misalignment of points towards the edges of the image is due to lens distortion and limitations in the camera’s field of view, which is a common occurrence in wide-angle cameras and does not significantly impact overall perception accuracy.

the LiDAR point cloud, outputting their 3D coordinates. These tools ensure an accurate correspondence between the camera and LiDAR points by linking the 2D and 3D representations of the same physical markers.

The calibration then computes the transformation matrix (rotation and translation) that aligns the two sensor frames. This method employs the Sequential Quadratic Programming Perspective-n-Point (SQPnP) algorithm, an advanced optimization technique based on the original PnP method. SQPnP iteratively refines the transformation parameters to achieve sub-pixel reprojection accuracy, even in noise or minor distortions. The calibration process iteratively collects correspondences and continues until convergence criteria are met. In terms of convergence, a sufficient number of detection pairs are required (**calibration\_convergence\_min\_pairs**), and the method converges using minimal error metrics (e.g., reprojection error below a set threshold).

Other techniques were also implemented to reduce disturbance and improve results. A Kalman filter tracks AprilTag detections across multiple frames, reducing noise and stabilizing the detected positions of the tags. This temporal filtering ensures consistency in calibration results, even when markers are partially occluded or affected by motion. Tag Size Compensation was also used to compensate for the difference in detection sizes between **apriltag\_ros** and **lidartag**, ensuring accurate correspondence between the camera and LiDAR data. This process minimizes the reprojection error, defined as the pixel difference between the observed and projected positions of the markers. Our experiments achieved an average reprojection error of 3.28 pixels, significantly improving from the 28.79 pixels before calibration.

This calibration pipeline achieved precise and consistent alignment of ZED X ONE cameras and LiDAR sensors.

Table 1. ZED Camera Specifications

Specification	Value
Resolution	1920 x 1200 @ 15 FPS
Shutter Type	Global Shutter
Sensor Format	16:10
Sensor Resolution	2.3M Pixels
Pixel Size	3 Microns
Sensor Physical Size	1/2.6"

Raw tf (transform) Source Error VS.  
calibrated tf Error:  
28.79 vs. 3.28

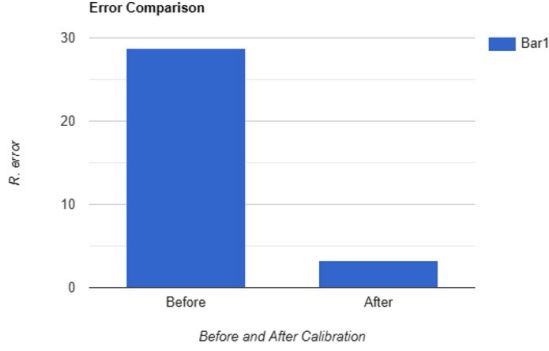


Fig. 7. Reprojection Error comparison: before and after calibration



Fig. 8. Calibration on road validation: The pointcloud and photo aligns well.

Automation also eliminates manual input, drastically reducing the time required for calibration and ensuring reproducibility. The combination of AprilTag detection, SQPnP optimization, and Kalman filtering enabled robust multi-sensor fusion, forming the foundation for reliable perception.

## 5. DISTRIBUTED COMPUTING SYSTEM

Our automated vehicle platform initially relied on a single Nuvo computer for all computational tasks, including perception, planning, and control. This centralized architecture created several performance bottlenecks, particularly in the perception module. The high volume of sensor data from cameras and LiDARs imposed a substantial processing burden, which resulted in a noticeable lag in real-time detection tasks, especially with LiDAR-based centerpoint detection. The Nuvo computer struggled to manage the simultaneous processing demands, leading to delays that affected the overall performance of the perception pipeline. As the complexity of the automated driving tasks grew, it became evident that a single-computer architecture was inadequate for real-time, high-fidelity perception and planning.

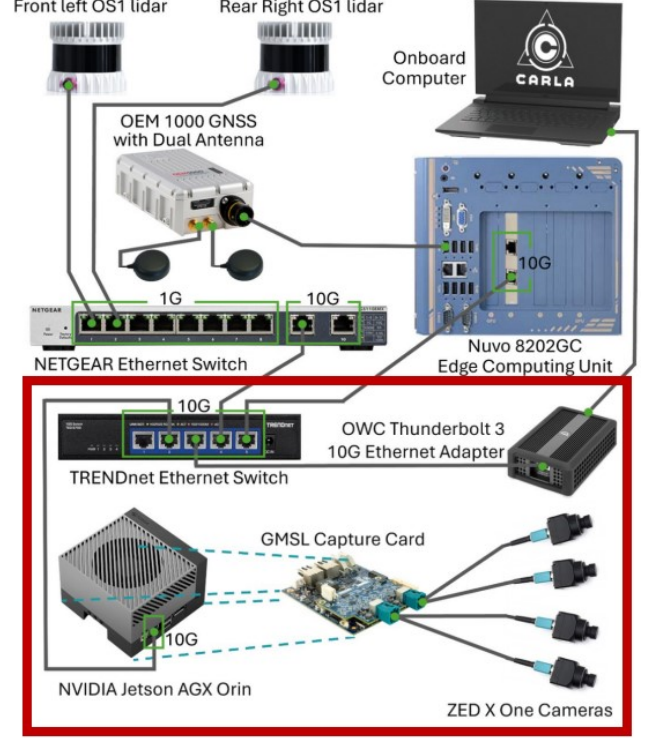


Fig. 9. New Setup With Distributed Computing, within the red box are the upgrades demonstrated in this paper

To address the limitations of the original architecture, we upgraded the system by integrating Jetson edge computers to handle distributed perception computations. The setup divides responsibilities between the Nuvo computer and the Jetson edge devices. The Nuvo computer, which now focuses on all the CPU-heavy operations, remains a crucial part of the system, handling tasks such as control and planning. On the other hand, the Jetson Edge computer handles GPU-heavy perception tasks. This division of roles reduces the computational load on the Nuvo computer, ensuring smoother and more efficient processing for control and planning tasks.

Furthermore, the vehicle network infrastructure was expanded to support 10G data communication between the computers and sensors, allowing high-speed data exchange across the edge and Nuvo systems. This enhanced bandwidth is crucial for maintaining the low-latency requirements for real-time perception and decision-making. Most of the perception module is now distributed to the Jetson edge computer, which handles most of the Autware perception stack, consuming less CPU and GPU for the Nuvo computer while maintaining a consistent update rate for overall perception outputs. This setup significantly alleviates the LiDAR detection workload that previously caused lag on the Nuvo computer. By distributing the computational load and leveraging edge computing, the platform is now better equipped to handle the demands of advanced automated driving tasks.

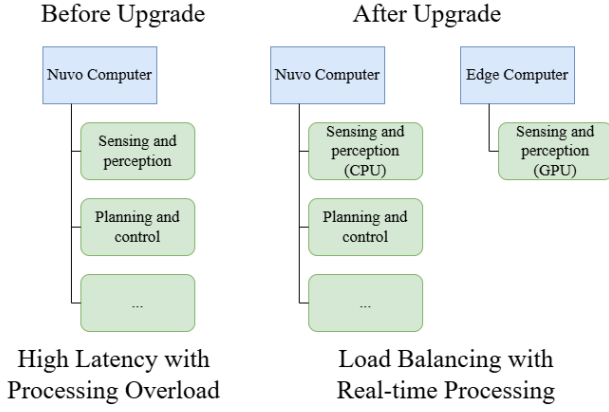


Fig. 10. System Upgrade Overview

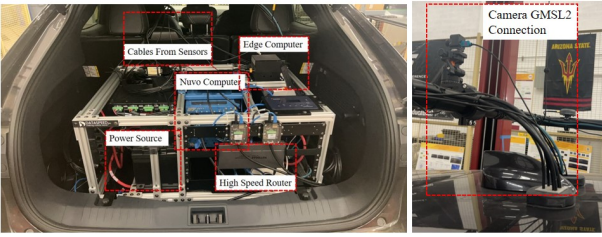


Fig. 11. The over all connection view of the computers and sensors

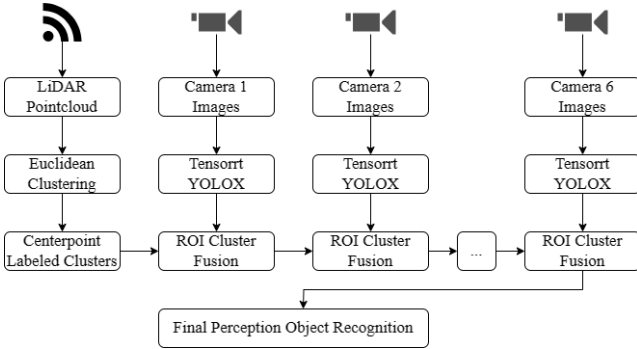


Fig. 12. Multi Camera-LiDAR Fusion Network

## 6. FUSION VALIDATION AND EVALUATION

The camera and LiDAR data were fused using an improved Autoware-based Region of Interest (ROI) cluster fusion method. Specifically, the fusion process involved object detection on each of the camera's images using the `autoware.tensorrt.yolox` module (Autoware (2024)), which leverages the YOLOX model (Ge (2021)) to detect objects such as cars, trucks, bicycles, and pedestrians. The `autoware.euclidean_cluster` package segments the LiDAR point cloud into clusters representing possible objects in the scene. The `roi_cluster_fusion` package then performed the fusion by projecting the LiDAR clusters onto each of the cameras' image planes. If the clusters' ROIs overlap with those identified by the camera detector, the YOLOX-based detection labels overwrite the original label from the center point. The Intersection over Union (IoU) metric determined the overlap between ROIs.



Fig. 13. Road Test Fusion Example Frames

The fusion approach was validated during a 20-minute vehicle test conducted around the ASU Polytechnic campus. The route encompassed diverse road conditions, including intersections, pedestrian crossings, straight roads, and curved segments, representing typical real-world driving scenarios. The tests demonstrated that camera-LiDAR fusion effectively refined detection labels, significantly improving perception accuracy and stability. These results underscored the fusion's potential to enhance situational awareness, providing a strong foundation for future optimization and scalability.

During the tests, 10,000 frames of ROS messages were processed from both the centerpoint detection and fusion result topics. The comparison of extracted objects for the front camera yielded the following observations: the LiDAR centerpoint algorithm detected a total of 5,040 objects, of which 1,161 objects had their labels corrected through fusion with the front camera, accounting for 23% of the total bounding boxes (corrected by YOLOx results). Due to the manufacturer's current driver limitations, launching all six cameras simultaneously is currently impossible. However, once these limitations are addressed, enabling all six cameras to operate concurrently, the system is expected to achieve nearly 100% coverage of all LiDAR-detected objects through the multi-camera fusion pipeline.

Table 2. Fusion Label Change Examples

Ground Truth	Label Before	Label After
Car	Unknown	Car
Car	Bus	Car
Bus	Unknown	Bus

In Table 2, it is evident that the integration of camera and LiDAR data significantly enhanced the system's perception capabilities. LiDAR excels at clustering bounding boxes but often struggles with accurate labeling. On the other hand, camera images provided rich semantic information for precise object classification, while LiDAR clusters contributed accurate spatial data for positioning and scaling (Geiger et al. (2012)). Many labels were corrected during the fusion process, proving the advantages of vision-based perception. Moving forward, we aim to explore the feasibility of camera-only perception. As multi-camera setups approach complete coverage of detected objects, reducing the dependence on LiDAR is possible, paving the way for cost-effective, camera-only perception systems. (Karpathy (2019))

## 7. CONCLUSION

The platform's upgraded sensing and computing system presents a promising future, surpassing the limitations of the original setup. The integration of six cameras into a 360-degree camera-LiDAR fusion framework has not only enhanced situational awareness but also improved object detection and classification accuracy, paving the way for a more robust and responsive platform for handling complex driving environments. The offloading of perception tasks to the Jetson edge computer has not only significantly reduced processing latency but also lightened the computational burden on the Nuvo computer, especially in managing perception tasks.

While the system has significantly improved perception accuracy and resource allocation through distributed computing, specific challenges remain. Synchronization issues between sensors and computing units still impact data consistency, and the current network protocols require optimization to enhance overall efficiency. Additionally, achieving real-time multi-camera fusion remains challenging, particularly for supporting seamless 360-degree perception coverage. To address the current limitations and further improve system performance, we plan to focus on the following areas for future work:

- **360-Degree Multi- Camera Fusion:**  
Optimize the fusion pipeline to support real-time processing with all six cameras, enabling consistent 360-degree perception coverage without performance degradation. Investigate and implement more advanced algorithms to improve perception ability.
- **Camera-Only Perception:**  
Continue exploring the feasibility of camera-only perception to reduce reliance on costly LiDAR sensors. We also plan to explore vision-based depth estimation techniques to improve 3D object detection and localization accuracy.
- **Standardized Dataset Creation:**  
Develop a standardized dataset leveraging the new setup to support benchmarking and further validation of perception methods.

The successful integration of distributed computing and enhanced perception capabilities marks a significant milestone. This work lays a robust foundation for advancing autonomous vehicle research and system upgrades. By addressing these challenges, we aim to push the boundaries of automated perception, driving progress toward safer and more efficient autonomous driving systems.

## REFERENCES

- Alabyad, N., Hany, Z., Mostafa, A., Eldaby, R., Tagen, I.A., and Mehanna, A. (2024). From vision to precision: The dynamic transformation of object detection in autonomous systems. *2024 6th International Conference on Computing and Informatics (ICCI)*, 332–344. doi:10.1109/icci61671.2024.10485026.
- Autoware (2024). Autoware tensorsrt yolox.
- Autowarefoundation (2024). Autowarefoundation/autoware: Autoware - the world's leading open-source software project for autonomous driving. URL <https://github.com/autowarefoundation/autoware>.
- BelfastEV (2023). Tesla vision.
- Bodell, H. (2024). Drive-by-wire kit.
- Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., and Beijbom, O. (2020). Nuscenes: A multimodal dataset for autonomous driving. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr42600.2020.01164.
- Dataspeed (2023). Dataspeed toolkit.
- Ge, Z. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*.
- Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3354–3361.
- Guo, H., Li, J., Saravanan, N.K., Wishart, J., and Zhao, J. (2024). Developing an automated vehicle research platform by integrating autoware with the dataspeed drive-by-wire system. *SAE Technical Paper Series*. doi:10.4271/2024-01-1981.
- Huang, J.K., Wang, S., Ghaffari, M., and Grizzle, J.W. (2021a). Lidartag: A real-time fiducial tag system for point clouds. *IEEE Robotics and Automation Letters*, 6(3), 4875–4882.
- Huang, J., Huang, G., Zhu, Z., Ye, Y., and Du, D. (2021b). Bevdet: High-performance multi-camera 3d object detection in bird-eye-view. *arXiv preprint arXiv:2112.11790*.
- Huang, X., Wang, P., Cheng, X., et al. (2018). The apolloscape open dataset for autonomous driving and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, TBD(TBD), TBD. doi:10.1109/TPAMI.2018.2872487.
- Karpathy, A. (2019). Tesla autonomy day presentation. <https://www.tesla.com/autonomy>. Accessed: January 2025.
- Liao, Y., Xie, J., and Geiger, A. (2022). Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3), 3292–3310.
- Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D.L., and Han, S. (2023). Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In *2023 IEEE international conference on robotics and automation (ICRA)*, 2774–2781. IEEE.
- Maddern, W., Pascoe, G., Linegar, C., and Newman, P. (2017). 1 year, 1000 km: The oxford robotcar dataset. In *The International Journal of Robotics Research (IJRR)*, volume 36, 3–15. doi:10.1177/0278364916679498.
- Olson, E. (2011). Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*, 3400–3407. IEEE.
- Patil, A., Malla, S., Gang, H., and Chen, Y.T. (2019). The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. In *2019 International Conference on Robotics and Automation (ICRA)*, 9552–9557. IEEE.
- Shahian Jahromi, B., Tulabandhula, T., and Cetin, S. (2019). Real-time hybrid multi-sensor fusion framework for perception in autonomous vehicles. *Sensors*, 19(20), 4357. doi:10.3390/s19204357.

Sun, P., Kretzschmar, H., d'Ascoli, S., et al. (2020). Scalability in perception: The waymo open dataset. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2446–2454. Available online at <https://waymo.com/open/>.

tier4 (2024). Tier4/calibrationtools: Sensor calibration tools for autonomous driving and robotics. URL <https://github.com/tier4/CalibrationTools>.

ZED (2024a). Get started with zed x one.

ZED (2024b). Zed x one.