Current State of Autonomous Driving Applications Based on Distributed Perception and Decision-Making

Baoming Wang¹,*, Han Lei², Zuwei Shui³, Zhou Chen⁴, Peiyuan Yang⁵

¹Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
²Computer Science Engineering, Santa Clara University, Santa Clara, USA
³Information Studies, Trine University, Phoenix, USA
⁴Software Engineering, Zhejiang University, Hangzhou, China
⁵Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
*Corresponding Author, wangbm0215@gmail.com

Abstract: This article reviews the key role of distributed cloud architecture in autonomous driving systems and its integration with intelligent computing networks. By spreading computing resources across multiple geographic locations, the distributed cloud enables localized processing and storage of data, reducing latency and improving real-time decision making in autonomous vehicles. The article points out that the combination of distributed cloud technology and intelligent computing network provides a powerful solution to meet the challenges of autonomous driving technology. By dynamically allocating computing resources and deeply integrating cloud, network, and chip technologies, distributed cloud gives autonomous driving systems enhanced data processing capabilities to ensure stable and reliable performance in a variety of driving scenarios. Finally, the paper highlights that the synergy of distributed cloud and intelligent driving technology marks an important milestone for intelligent transportation systems, heralding the accelerated adoption of distributed cloud solutions in the automotive industry, driving the pace of innovation and transformation.

Keywords: Distributed cloud architecture; Autonomous driving system; Data localization processing; Intelligent perception and decision making; Real-time decision capability

1. INTRODUCTION

As an essential facet of future transportation, autonomous driving technology is encountering unparalleled challenges in computational power. Autonomous vehicles must adeptly discern intricate environmental cues such as pedestrians, vehicles, and traffic signs while navigating at high speeds, necessitating real-time decision-making and control. This mandates that the autonomous driving system possess robust data processing capabilities and computational power support. [1-3] According to authoritative forecasts, the global self-driving car market is poised to exceed $100 billion by 2025. Industry statistics reveal a staggering surge in the requisite computational power for self-driving cars. An advanced self-driving vehicle may necessitate computing power equivalent to thousands of conventional servers combined. Confronted with such immense computational demands, the traditional cloud computing model struggles to suffice. Traditionally, this model centralizes data storage in remote data centers for processing, introducing challenges of data transmission latency and potential risks to data security and privacy protection.

2. RELATED WORK

2.1 Distributed cloud

Distributed cloud is different from distributed computing, and the concept of distributed cloud is still from the 2021 strategic technology trend forecast released by Gartner at the end of 2020, one of the most important concepts is distributed cloud, which is Gartner's second time to mention distributed cloud as a strategic technology. Let's take a look at how Gartner defines a distributed cloud.

Distributed cloud: [4] The distribution of public cloud services (often including the necessary hardware and software) to different physical locations (i.e., the edge), while the ownership, operation, governance, updating, and development of the service remains the responsibility of the original public cloud provider.
Gartner also predicts that by 2025, more than 50 percent of organizations will use distributed cloud in their chosen locations, enabling a transformational business model. [5]In literal terms, distributed cloud refers to the redistribution of public cloud services in physical locations, but operations remain unified. Our careful analysis shows that this is actually consistent with the premise of the multi-cloud concept. The distributed cloud embodies the principle of proximity of computing power, which in fact solves the needs of customers to keep cloud computing resources close to the physical location of data and business activities.

Distributed cloud has two key points in the architecture, one is distributed or decentralized, and the other is centralized or unified.[6]Decentralized is the bearer of different business attributes and regional requirements, centralized is the management and standards.

Distributed cloud Features:

1. High availability

A distributed cloud architecture avoids the risk of a single point of failure by spreading resources across multiple locations. [7]Even if one node fails, other nodes can still provide services to ensure business continuity and reliability.

2. Scalability

The distributed cloud can flexibly adjust the scale of resources based on service requirements. When the service load increases, you can add more nodes to expand the computing and storage capabilities to meet user requirements. On the contrary, when the service load is reduced, the number of nodes can be appropriately reduced to save resources and costs.

3. Flexibility and flexibility

The distributed cloud is characterized by elasticity and flexibility. It can automatically adjust resources on demand to accommodate traffic peaks and valleys. [8]This flexibility enables the distributed cloud to cope with sudden changes in business demands and traffic, providing stable performance and user experience.

4. Data locality

The distributed cloud allows data to be stored on the node closest to the user, reducing latency and bandwidth consumption in data transmission. This advantage of data locality can improve the efficiency and responsiveness of data access, especially for application scenarios and cases distributed in different regions of the world

2.2 Distributed systems and autonomous driving

With the continuous improvement of the performance of AI chips and sensors such as lidar, millimeter-wave radar, and cameras, coupled with the gradual reduction of costs, the line and mooring integrated program gradually began to be mass-produced. In the future, there will be a large number of new cars equipped with an integrated solution on the road, which can collect more valuable data, and then promote the continuous improvement and evolution of high-level intelligent driving functions such as NOA and AVP through a data-driven way. The distributed cloud provides an entirely new solution. [9-11]By decentralizing cloud computing resources across multiple geographic locations, the distributed cloud enables localized processing and storage of data. According to industry experts, through the distributed cloud architecture, the autonomous driving system can achieve localized processing, reduce data transmission delay, and improve real-time. [12]Taking Kehua Data's solution as an example, its distributed intelligence center can reduce data processing latency to the millisecond level, ensuring that the autonomous driving system can make rapid and accurate decisions in a variety of complex environments.

In addition, the distributed cloud also has high reliability and high scalability. By building multiple redundant data centers and computing nodes, the distributed cloud can ensure stable and reliable computing support for autonomous driving systems in any situation. At the same time, with the continuous expansion of the autonomous driving market and the continuous progress of technology, distributed cloud can also be flexibly expanded and upgraded according to demand.
By dynamically allocating computing resources, the computing power network can meet the requirements of the autonomous driving system in different scenarios. Through the deep integration of cloud, network and chip technology, Kehua Data realizes the efficient construction of computing power network, and provides strong computing power support for the autonomous driving industry. [13] Taking Kehua Data's distributed intelligent computing center as an example, its computing power network has been successfully applied to a number of autonomous driving projects. [14-16] According to practical application cases, the computing power network can increase the processing speed of the autonomous driving system several times, while reducing power consumption and cost. This successful case not only proves the practical application value of distributed cloud and computing power network in the autonomous driving industry, but also provides valuable experience and reference for other enterprises.

In summary, distributed cloud and computing power networks have become the biggest demand of the autonomous driving industry. By providing powerful data processing and computing support, they provide a solid guarantee for the innovative development and application of autonomous driving technology. In the future, with the further popularization and application of autonomous driving technology, the demand for distributed cloud and computing power networks will be more vigorous. [17][18] We look forward to more companies joining this field and jointly promoting the prosperity of the autonomous driving industry.

2.3 Autonomous driving perception and decision making

Figure 1 below gives a clear picture of how an autonomous vehicle system is structured. In this diagram, you'll see two main parts: the perception system and the decision system, each represented by different colored modules.

The perception system is like the eyes and ears of the autonomous vehicle. It gathers information from various sensors onboard the vehicle, such as [19] LIDAR, radar, cameras, [20] GPS, [21] IMUs, and odometers. These sensors help the vehicle understand its surroundings, including detecting objects, understanding road conditions, and keeping track of its own position and movement. The perception system also relies on prior knowledge about sensor capabilities, road networks, traffic rules, and vehicle behavior.

On the other hand, the decision system is like the brain of the autonomous vehicle. It takes the information provided by the perception system and uses it to make decisions about how the vehicle should navigate through its environment. [22] This includes tasks like planning the vehicle's route, avoiding obstacles, obeying traffic laws, and ensuring passenger safety and comfort.

Overall, this architectural block diagram illustrates how an autonomous vehicle combines perception and decision-making capabilities to navigate safely and efficiently in its environment.

![Figure 1. Overall operational framework of automatic driving](Image)
The decision system guides the car from its starting point to a destination defined by the user. It considers the car's condition, the environment, road rules, and passenger comfort. [23] To navigate effectively, it needs to know where the car is in its surroundings. This is where the locator module comes in. It estimates the car's position, speed, and direction using a map of the environment. This map is created before the car operates autonomously, often using its own sensors. Sometimes, manual adjustments are needed, like marking crosswalks or removing non-static objects. The positioning module takes offline maps, sensor data, and odometer readings as inputs and outputs the car's status. [24] While GPS can assist, it's not always reliable in urban areas due to interference. The mapper module then uses offline maps and current data to create online maps. These online maps combine static map information with real-time sensor data. They're crucial for decision-making and can include a Moving Object tracking module to detect and remove dynamic elements.

In this part, we discuss key methods for autonomous vehicle perception systems, like locating the vehicle, mapping obstacles, roads, tracking moving obstacles, and recognizing traffic signals. The positioning module estimates the car's position and direction relative to the map or road, often using GPS [25]. However, GPS isn't always reliable in urban areas. Other methods, like [26] LiDAR, LiDAR plus cameras, or just cameras, offer alternatives. LiDAR provides accurate measurements but can be costly. Combining LiDAR with cameras reduces costs while maintaining accuracy. Camera-based methods are cheaper but may be less precise.

3. METHODOLOGY

3.1 Experimental design

Data closed loop plays a crucial role in the development of autonomous driving technology. Understanding its composition is critical to advancing the development of autonomous driving technology. The following are the main components of a data closed loop:

1. Data collection: Data collection is carried out by professional collection vehicles, test vehicles or production vehicles. These vehicles are equipped with a variety of sensors and data recording devices to capture various scenarios and situations in the real world [27].

2. Data return: The collected data is processed locally, including classification, desensitization, compression and packaging, and then uploaded to the cloud server through 4G/5G and other networks. This ensures the safe transmission and storage of data.

3. Data annotation: Annotate the collected data on the cloud server. Annotation refers to giving semantic information to data, such as identifying vehicles, pedestrians, traffic lights, etc. [28] This process is crucial for training and validating autonomous driving systems.

4. Model training: Marked data is sent to the training platform for training machine learning models. These models include parts such as perception, decision-making, and are used for all aspects of autonomous driving systems.

5. Simulation test: After model training, simulation test is required. Simulation testing is to simulate real world situations through virtual scenarios to verify and evaluate the performance and safety of the trained model.

By closing the loop on these key links, autonomous driving technology can continuously collect data from the real world, train models, and optimize and improve its performance through simulation tests. This circular feedback process is an important guarantee for the realization of high-level intelligent driving perception and decision-making.

3.2 Data acquisition

After the equivalent production vehicle reaches a certain scale, the amount of data collected will be large. Some professionals estimate that if the total amount of data collected in accordance with 100,000 vehicles is estimated for 300 days a year, the total amount of data faced by car companies in the future will reach ZB level. If so much data is uploaded to the cloud without effective screening, it will also bring great pressure on transmission bandwidth, data storage and processing in the cloud. [29] Therefore, how to select valuable data from massive data for efficient return is the key to affecting the iteration speed of high-level intelligent driving system.
There are several common intelligent data acquisition methods:

A. Set the Trigger layer at the car end

According to model failure analysis and model decision boundary analysis, the scene to be collected is set in advance and the collection logic is formulated. Then, the trigger layer (data return trigger) is set at the end of the vehicle, and the required scene data set is automatically obtained according to the scene algorithm detection.

B. Intelligently label data for specific scenarios

For some scenarios that need to actively accumulate data in the cloud for learning, such as complex working conditions such as tunnels, roundabout islands, unprotected left turns, etc., developers can upload pictures that need to be obtained by the vehicle, issue instructions through the cloud, and the vehicle will take a similar way of "image search" to automatically intercept similar scenes.

3.3 Data Sending

After collecting the data, the data needs to be classified, desensitized, compressed and packaged, and uploaded to the cloud server through 4G/5G[30]. However, the uplink and downlink of data transmission is relatively long, and the link of the car network is usually not stable, and the car may cross different base stations in the process of fast driving, there is a 4G-5G switch, which will lead to the loss and disorder of the signal in the transmission process. So, how to ensure the data transmission speed and quality of each link node?

In view of some problems in data transmission, some enterprises have adopted a vehicle-cloud integrated transmission scheme. For example, Zhixhuitong combines data slices through SDK vCloud of cloud data management, adopts a verification and retransmission mechanism for slices, retransmits lost data slices, and stores them after data retransmission is complete, ensuring the integrity of data transmission.

3.4 Data Annotation

After the car end data is returned to the cloud, it also needs to be marked with data to become sample data that can be used by the algorithm model. [31-32]Labeling is the process of assigning coded values to raw data, thus using the labeled data as training data for AI to practice cognition. Encoded values include, but are not limited to, assigning class labels, drawing bounding boxes, and marking object boundaries.

With the continuous iteration of intelligent driving technology, intelligent driving system has higher and higher requirements for the accuracy of perception model. Therefore, to improve the accuracy of vehicle perception models, large-scale and high-quality data sets are needed for training. Traditional manual annotation has been unable to meet the demand for massive data sets in model training in terms of efficiency and cost, and a new annotation method is needed to improve the annotation quality and efficiency.

The pre-labeling algorithm can greatly reduce the time required for each box of data to be labeled. Baidu autonomous driving cloud technology experts have said: "Before the annotation, we will first use the algorithm to do a pre-annotation, which can greatly improve the efficiency of the tagger single frame annotation." In the tagging process, we introduce a lot of intelligent algorithms to assist our taggers. For example, when doing area segmentation, we will learn from the edging algorithm similar to photoshop to achieve better fitting effect and improve the tagging efficiency."

3.5 Model Training

After the completion of data annotation, it is necessary to train the labeled data and adjust parameters. The training of large models used in advanced intelligent driving has a high demand for computing power. Some car companies have specially created their own intelligent computing centers, such as Tesla's Dojo, Geely's Star Rui intelligent computing Center, Xiaopeng's "Fuyao", and the Snow Lake • OASIS (MANA OASIS).

Model training is to explore and analyze data through analytical means and methods to find causality, data logic and business laws. In the model training, we can design a set of automatic training engine with the help of Auto ML and other tools to automate part of the model training.
In addition, from the perspective of data operation, the efficiency and quality of model training can also be improved. Baidu has taken the following measures:

Help customers build the most effective data set in the model training process - help customers to plan what data needs to be marked, what categories to mark, and what its distribution is in the early stage. For the marked data, Baidu Intelligent Cloud will help enterprises evaluate their models according to the existing huge evaluation sample set, and discover the bad cases or shortcomings of the current model. In view of the shortcomings of the current model, it is necessary to supplement enough training sets to help the enterprise improve its model indicators and optimize the model.

3.6 Requirements for cloud access for perceptual model training

The training of automatic driving perception model requires large data volume, high algorithm accuracy and training efficiency, which is suitable for cloud, and adopts cloud service for data processing. CNN is a common deep learning model in intelligent driving perception, and its design is basically aimed at INT8. BF16 is the most suitable format for Transformer model architecture. Some industry insiders admitted: "Transformer belongs to violent aesthetics, compared with CNN, its model is larger, the number of participants is one billion billion, one billion trillion is not rare, the number of layers is thousands of layers, it is not the old data training center can support."

Among them, for L2 and L3 levels, if you do L2 and L3 level Demo, you only need millions of pictures and several GPU cards to train. Due to the small amount of data and limited scope of use, the requirements are not very high from the perspective of compliance, and it probably needs several million dollars of investment. If you do L2 and L3 level mass production, you need hundreds of millions of pictures and 100+ GPU cards for training scheduling, and the capital investment can reach tens of millions.

For L4 level, if you do L4 level Demo, the amount of data required to store more than 1PB, 100+ GPU cards are required for training scheduling, and the investment scale reaches 50 million to 200 million. If you want to achieve L4 level mass production, the amount of data required to store will exceed 50PB, 500+ GPU cards are required for training scheduling, and the related investment will exceed 500 million levels.

3.7 Distributed Cloud Simulation

The traditional single-machine simulation test presents some problems, such as insufficient computing power and unable to realize accelerated testing, such as long test cycle and low efficiency. Cloud platform simulation, with its distributed architecture and parallel accelerated computing capabilities, can greatly improve the efficiency of simulation testing, and is an effective solution to achieve large-scale simulation scenarios for autonomous driving. Although cloud simulation can effectively improve the efficiency of simulation test, the industry still has the following questions about simulation test: whether the simulation scene is true, the simulation scene is incomplete, the iteration speed is not fast, the simulation evaluation is accurate or not.

For these problems, Baidu Cloud simulation platform has carried out targeted improvements and solutions:

In terms of the reality of the scene, in accordance with compliance standards, the integration of high-precision maps, 1:1 depiction of the physical world of the road topology. For dynamic traffic participation elements, mining is done based on real road data, and the interaction of dynamic elements is accurately described.

In the scene generation mode, the combination of manual scene editing mode and scene mining based on real road data has covered 98% of the scenes (including cities, highways, parking lots, closed parks, etc.), according to Baidu insiders.

In terms of iteration speed, relying on the technical support and computing power advantages of Baidu Intelligent Cloud, Baidu's cloud simulation platform can realize the concurrent operation of hundreds of thousands of tasks, and achieve a daily simulation mileage of tens of thousands of kilometers.

In terms of simulation evaluation standards, Baidu has summarized more than 200 evaluation indicators in six categories after several years of experience. In addition to safety and traffic regulations, comfort and intelligence are also added to the evaluation criteria through rules.
4. CONCLUSION

Based on the comprehensive analysis of distributed cloud architecture and its application in autonomous driving systems, it is evident that distributed cloud technology plays a pivotal role in enhancing the efficiency, reliability, and scalability of autonomous driving functions. By decentralizing computing resources across multiple geographic locations, distributed cloud enables localized processing and storage of data, thereby reducing latency and improving real-time decision-making capabilities in autonomous vehicles.

Furthermore, the integration of distributed cloud with intelligent computing networks presents a robust solution to address the evolving challenges in autonomous driving technology. Through the dynamic allocation of computing resources and the deep integration of cloud, network, and chip technologies, distributed cloud empowers autonomous driving systems with enhanced data processing capabilities, ensuring stable and reliable performance in various driving scenarios.

In conclusion, the synergy between distributed cloud, intelligent computing networks, and autonomous driving technology marks a significant milestone in the advancement of intelligent transportation systems. As the demand for autonomous driving technology continues to grow, the adoption of distributed cloud solutions is poised to accelerate, driving innovation and transformation across the automotive industry.

ACKNOWLEDGEMENT

I would like to extend my sincere gratitude to the authors of two influential studies that have significantly inspired and guided my research. First, my thanks go to Zhou, Yanlin, et al. for their insightful paper, [5] "Utilizing AI-Enhanced Multi-Omics Integration for Predictive Modeling of Disease Susceptibility in Functional Phenotypes," published in the Journal of Theory and Practice of Engineering Science. Their innovative approach to AI-enhanced multi-omics integration has provided a valuable foundation for my work.

Additionally, I am deeply appreciative of the contributions made by Wang, B., He, Y., Shui, Z., Xin, Q., and Lei, H. in their study on [6] "Predictive Optimization of DDoS Attack Mitigation in Distributed Systems using Machine Learning." Their research has greatly influenced the development of my methodologies and has been a key source of inspiration for my project.

Thank you for your groundbreaking research and the significant impact it has had on my work.

REFERENCES


