

End-to-end system architectures in autonomous driving: Comparative analysis against modular design and technological exploration

Lisha Yuan

School of Automotive Studies, Tongji University, Shanghai, 201804, China

2051922@tongji.edu.cn

Abstract. In the context of accelerated technological advancement, artificial intelligence is becoming an increasingly pivotal enabler of autonomous driving. With the integration of machine learning into autonomous driving systems, the traditional modular architecture has been disrupted, evolving into a more advanced end-to-end architecture. The end-to-end architecture has the advantages of simple structure and no cumulative error, especially under the current situation where a large amount of data and functions are continuously inputted, it demonstrates superior performance compared to the traditional modular system. However, the implementation of end-to-end is still mostly confined to academia and has not been widely applied to actual vehicles; the mainstream on actual vehicles is still the modular system architecture. As a result, the two system architectures are compared in this paper to ascertain their relative complexity, the extent to which errors are accumulated, their interpretability, the security they afford, and the extent of their data dependency. This is achieved through a comprehensive literature review. Furthermore, the paper delineates the deployment of the end-to-end system and outlines potential avenues for addressing the challenges that emerged during the implementation phase. The results indicate that the end-to-end system exhibits a certain degree of sophistication. However, a series of challenges pertaining to interpretability and security must be addressed to fully realize the potential of the end-to-end system.

Keywords: Artificial Intelligence, Autonomous Driving, End-to-End System Architecture, Modular System Architecture.

1. Introduction

Nowadays, the exponential advancement of artificial intelligence has positioned the domain of intelligent autonomous driving as a pivotal area of technological advancement, in which end-to-end, a pioneering architectural approach, has emerged as a disruptive force, challenging the conventional modular system architecture. It is founded upon machine learning and extensive model datasets, continuously generalizing the acquired patterns to novel contexts. This architectural style offers several advantages, including a streamlined system structure, the absence of accumulated errors, robust generalization capabilities, and high adaptability [1][2][3]. Nevertheless, a considerable proportion of the recent developments in end-to-end research remain at the theoretical level and modular architectures continue to represent the prevailing approach. This paper therefore describes the technical characteristics and advantages and disadvantages of end-to-end and modular system architectures,

respectively, and focuses on comparing the aspects of complexity, error accumulation, interpretability, security, evaluation testing, and data dependency. Furthermore, it provides a concrete reflection of the advancement of end-to-end solutions. The paper facilitates the pursuit of further technical research in a focused manner and indicates the direction of future development.

2. System Architectures in Autonomous Driving

2.1. End-to-end System Architecture

The ALVINN (Autonomous Land Vehicle In a Neural Network) self-driving test vehicle, introduced back in 1988, is based on the end-to-end architecture, which can achieve up to 70km/h. The end-to-end architecture models the entire autonomous driving function into an overall network. In recent years, the concept of end-to-end has been extended to a greater extent, and its core definition has been updated to “lossless transmission of sensory information, which can be achieved Global Optimization of Autonomous Driving Systems.” As shown in Figure 1, the system directly maps raw sensor inputs to control command outputs without explicitly distinguishing each stage. To illustrate, deep learning models are employed to derive the steering angle of the vehicle directly from the sensed image information [4].

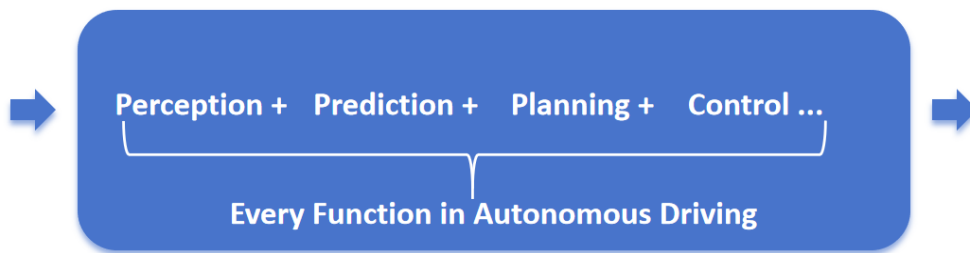


Figure 1. End-to-end system architecture

2.2. Modular System Architecture

The structure of the architecture divides the entire function into distinct modules, each with a unique set of functions. The modular system architecture is distinguished by the characteristics of “structure layering and function superposition,” as each module is developed independently. In this architecture, each module has its own input and output. The output of a certain module can be used as the input of other modules. Notably, the output modality of each module is designed to be compatible with the input modality of subsequent modules in the system, in order to ensure correct propagation of information through the modular system [1][5].

The system is comprised of four modules, as shown in Figure 2. The first module is perception, which captures different types of information through various sensors and then processes and analyzes it through built-in perception algorithms to achieve the effect of understanding the surrounding environment. The second module is prediction, which combines state information, road structure information, high-precision map information, interaction information between targets, semantic information of the environment, and intention information of the target, to predict the perceived information, optimizing the planning effect and avoiding possible dangerous situations in advance. The third module is planning. The decision algorithm can select the best action strategy based on this input information, including path planning, speed control, lane keeping, obstacle avoidance, etc. The fourth module is control. This system often includes vehicle power system control, steering control, vehicle stability control, etc. Through the cooperation of various controllers, the vehicle can ultimately drive safely and stably at the target speed and path.

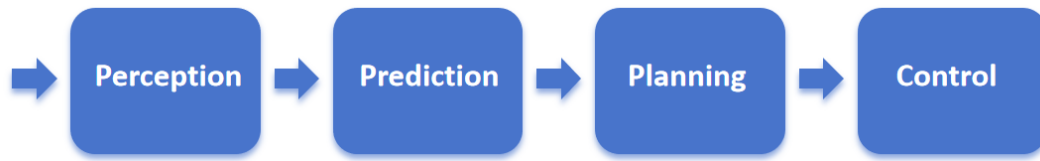


Figure 2. Modular system architecture

3. Comparison between End-to-End and Modular System Architectures

3.1. Complexity

In autonomous driving system architecture, the complexity is mainly reflected in the number of models in the system, because each model requires separate training, optimization and iteration. The higher the complexity, the more parameters will increase during the continuous evolution of the model, which will ultimately lead to more investment in R&D personnel and costs. In a modular system architecture, functions are divided into multiple modules, and each of them is composed of smaller models. To illustrate, the extensive module of perception encompasses positioning, tracking, and classification. As autonomous driving functions become more sophisticated, the modular system architecture becomes more complex, and system response times increase. In the end-to-end system architecture, the functions of the aforementioned multiple models are implemented through an overall model, thereby providing a more streamlined solution. For example, the UniAD model proposed by Hu, et al., is an autonomous driving model that integrates the principal tasks of the full stack into a deep neural network [6].

3.2. Error Accumulation

In autonomous driving, certain errors will occur in each module during execution due to various factors, and the size of the overall system error will determine the upper limit of the system's performance potential. In a modular system architecture, the occurrence of errors in the output of a previous level of the model will result in the transmission and accumulation of errors at the subsequent level, leading to the emergence of cascading errors. As the number of links increases, these cascading errors will become more significant. Concurrently, the disparate optimization objectives of the constituent models will result in an overall effect that is compromised by a lack of coordination between tasks. In the end-to-end system architecture, since the control signal can be directly output after inputting the sensor signal, the probability of cascading errors is greatly reduced, thereby greatly increasing the upper limit of the system's performance potential. At the same time, in some systems, it is also proposed to design each link for the final task. For example, the UniAD model is designed for planning purposes, with each link taking the planning effect as its optimization goal and thereby achieving superior final results [6].

3.3. Explainability

Explainability refers to whether the specific operating logic in the model can be explained. Explainability includes two aspects: interpretability and completeness. The former refers to the understandability of the explanation, and the latter refers to the detailed definition of the model through explanation [1]. The quality of interpretability directly determines whether engineers can repair system defects in advance and whether they can debug later. Good interpretability can avoid some possible problematic results in advance, and can quickly locate the problem point when running problems occur, so that engineers can better debug the system and provide better performance guarantee [7]. The interpretability of the modular system architecture is relatively good, because the input and output of each module can be directly obtained. We can not only understand the specific functions of each module, but also quickly locate the corresponding module for debugging when a problem occurs. The interpretability of the end-to-end system architecture is very low. This is because it is an entire large system, equivalent to a black box model, with only the initial input and final output, bringing great

difficulties to later error reporting and debugging, and making it impossible to predict the lower limit of its performance.

3.4. Safety

Safety is the most important factor to consider in an autonomous driving system, and is the premise and foundation of all other performance. In the modular system architecture, the modular driver stack has added some specific safety-related rules to the prediction or planning module to constrain and optimize, and in some cases, enforce them, thereby integrating Security constraints and rule implementation guarantee security [8]. In the end-to-end system architecture, since there are no relevant modules added, the low interpretability as a “black box problem” also makes it inherently lack some precise mathematical guarantees about security. In addition, in multi-frame training methods, causal confusion may occur that may lead to serious consequences. These security-related issues need to be improved and resolved [9].

3.5. Data Reliance

In a modular system architecture, the algorithmic operations of the modular system are primarily based on the rules that have been defined in advance. In the absence of such a framework, the system is susceptible to failure. This also implies that the dependence on the quantity and quality of data is relatively limited, and the situation of the data will not have a significant impact on the final outcome. In the end-to-end system architecture, the algorithmic process is driven by the data, which enables the system to continuously iterate through extensive and diverse data sets. However, the efficacy of the algorithm is contingent upon the quality of the input data, with the ultimate learning outcome largely determined by the data itself. To illustrate, the selection of a specific architectural style is predominantly influenced by the requirements and constraints inherent to the pertinent data [10].

4. Application of End-to-End Autonomous Driving Technology

End-to-end autonomous driving technology represents a state-of-the-art solution in autonomous driving systems, whereby raw sensor data is directly mapped to vehicle control commands, thereby simplifying the traditional modular architecture. This technology has made significant progress in the industry, with various companies and tech firms actively exploring and implementing it. Tesla’s FSD V12 system is a notable example, demonstrating the potential and challenges of end-to-end network architecture in Bird’s Eye View (BEV) space planning in real-world environments [10]. To further explore the practical applications and development trends of end-to-end autonomous driving technology, this section will introduce how major Original Equipment Manufacturers (OEMs) integrate end-to-end technology into their vehicles, and discuss how leading autonomous driving systems and algorithm companies are advancing the technology through innovative solutions. Through these analyses, a more comprehensive understanding of the role of end-to-end autonomous driving technology in the current market can be gained.

4.1. Original Equipment Manufacturers (OEMs)

On May 20, 2024, Xpeng Motors made a notable advancement in autonomous driving technology with the announcement that its end-to-end large model would be integrated into its vehicles. It is anticipated that this development will enhance both the safety and efficiency of driving. However, it also presents technical challenges related to data processing and real-time responses. Additionally, Huawei has made a noteworthy contribution to the field with the release of an updated version of its intelligent driving system, QianKun ADS 3.0, on April 24, 2024. This update exemplifies Huawei’s technological advancement and may serve as a foundation for further advancements in end-to-end architectures. The system is designed to integrate big data and artificial intelligence in order to provide more accurate driving decision support. NIO has announced plans to introduce end-to-end active safety features, thereby demonstrating its commitment to enhancing vehicle safety. It seems probable that NIO’s

strategy will involve the utilisation of advanced sensors and data fusion technology with a view to improving the real-time responsiveness of its active safety functions. Furthermore, Zero One Auto has set a goal of implementing end-to-end autonomous driving technology by the end of 2024. This timeline demonstrates the company's accelerated pace of technological advancement and testing, which has the potential to propel advancements in the autonomous driving industry as a whole [11][12].

4.2. Autonomous Driving System and Algorithm Companies

Pony.ai has made significant progress in simplifying its system architecture by integrating conventional autonomous driving modules into an end-to-end model, which was applied to both L4 autonomous taxis and L2 assisted driving passenger cars as of August 2023. This integration not only streamlines the system but also enhances its compatibility and scalability, allowing for more effective handling of complex driving scenarios and diverse traffic conditions. In addition, Nvidia revealed its development plans for autonomous driving, including breakthroughs in the L2++ system and the integration of large language models (LLM) and visual language models (VLM) into vehicles, which indicates that Nvidia is investigating sophisticated AI technologies to enhance the system's intelligence and adaptability. However, this also presents significant challenges in data processing and algorithm optimization. SenseTime, for its part, demonstrated innovation in this area with the launch of its end-to-end autonomous driving solutions, UniAD and DriveAGI, which, as a "modular end-to-end" solution, may have advantages in terms of flexibility and maintainability, while DriveAGI, as a "single-mode end-to-end" solution, may have an overall performance and consistency have advantages. The comparison and application of these two approaches will have a significant impact on the future development of autonomous driving technology [11][12].

5. Future Prospects

5.1. Optimizations for Explainability and Testing

In order to ensure the interpretability of end-to-end automated driving systems, it is necessary to implement interpretation methods that are both local and global in nature. A local explanation provides an account of the fundamental principles that underpin the model. In contrast, global explanations seek to achieve a comprehensive understanding of the model's behavior, thereby enhancing the interpretability of the system. This is achieved by elucidating the underlying knowledge content. There is not much overall research on global explanation, and future research needs to focus on this area [13]. The current boom in large models in language and vision can be combined with autonomous driving to solve some problems. For example, the DriveMLM large model proposed by SenseTime represents a promising avenue for further investigation [14]. In this large model, the introduction of LLM is not merely a matter of convenience; rather, it is a necessary step to enable the model to control the vehicle. However, LLM is not a suitable solution on its own, as it is designed to output language and lacks the capability to control the vehicle. DriveMLM is aligned with the decision-making state and serves to convert language signals into vehicle control signals. This large model enables closed-loop testing operations and the generation of comprehensive driving interpretations and aligned decisions at a reduced cost, thereby facilitating the effective resolution of complex scenarios.

5.2. Optimization of Security Issues

With regard to the input side, two significant challenges are occlusions and obstacles that fall beyond the sensing range. Nevertheless, V2X technology will prove an efficacious solution. By exchanging and sharing information with other road users, it is possible to supplement blind spots and correct the shortcomings of the sensing end itself. This provides a basis for further improvements. Complex information processing provides solutions [11]. Furthermore, the integration of sensor fusion enables the incorporation of supplementary sensors, thus augmenting the array of sensor capabilities and the number of parameters. It is possible to identify specific strategies for ensuring security in modular architectures. Supplementary "if-then" conditions can be integrated into the downstream control system

to impose constraints. As an alternative approach, reinforcement learning algorithms can be utilized to impose constraints, with the objective of limiting the final planning effect to a minimum. Huawei's recently released ADS 3.0 employs this methodology of integrating artificial rules with end-to-end modeling, which is regarded as a comparatively secure solution within the industry.

5.3. Optimization of Data Problems

In the context of automatic driving models, it is inevitable that some real-world scenarios will arise that are not reflected in the distribution of the training data. In order to cope with these challenges, it is essential to continuously optimize the end-to-end system in conjunction with the key techniques of zero-shot/less-shot learning [9]. As the end-to-end system necessitates a substantial quantity and quality of data, the iterative development of data and models can be facilitated by an autopilot data engine. The data engine automates the process of labeling high-quality perceptual labels with the assistance of large-scale perceptual models. Furthermore, it is capable of handling the generation and editing of some mining-difficult/extreme-case scenarios, which ultimately contributes to the diversity of the data set and the generalization capability of the models [9].

6. Conclusion

The investigation into the integration of artificial intelligence into autonomous driving systems has revealed that end-to-end architectures have the potential to exceed the capabilities of traditional modular approaches. The comparison shows that end-to-end is a highly sophisticated approach, offering significant advantages in terms of complexity and error accumulation. However, it also presents challenges in terms of interpretability, security, and data dependence, which have limited its broader adoption. The results indicate that these issues can be addressed through global interpretability studies and the use of large-scale fundamental models to improve the interpretability of the system, as well as by optimizing the information on the input side and supplementing the definitional rules to improve the safety of the system, paving the way for end-to-end systems to become the norm for autonomous driving in the future. It can be believed that end-to-end system architectures will have wider practical applications in vehicles in the future. Further research is required to address the remaining issues identified in this paper. This paper lacks further quantitative comparisons of the various characteristics of the two systems and requires the collection of extensive experimental data. In addition, additional research is required to develop a more comprehensive optimization scheme for the end-to-end system.

References

- [1] Chib, P.S. and Singh, P. (2023) Recent advancements in end-to-end autonomous driving using deep learning: A survey. *IEEE Transactions on Intelligent Vehicles*, 9(1): 103-118,
- [2] Hu, S.L., Hua, X.P., Dou, M., Fei, H.L. and Fu, Z.J. (2024) Current Status and Development Trends of End-to-End Autonomous Driving for Vehicles. *Times Automobile*, (13): 4-6+109.
- [3] Li, K., Dai, Y.F., et al. (2017) Development Status and Trends of Intelligent Connected Vehicle (ICV) Technology. *Journal of Automotive Safety and Energy Efficiency*, 8(1): 1-14.
- [4] Bojarski, M., Testa, D.W., Dworakowski, D., et al. (2016) End to End Learning for Self-Driving Cars. *ArXiv*, abs/1604.07316.
- [5] Singh, A. (2023) End-to-end autonomous driving using deep learning: A systematic review. *arXiv preprint arXiv:2311.18636*.
- [6] Hu, Y., Yang, J., Chen, L., et al. (2023) Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17853-17862.
- [7] Zablocki, É., Ben-Younes, H., et al. (2022) Explainability of deep vision-based autonomous driving systems: Review and challenges. *International Journal of Computer Vision*, 130(10): 2425-2452.
- [8] Brüdigam, T., Olbrich, M., et al. (2021) Stochastic model predictive control with a safety guarantee for automated driving. *IEEE Transactions on Intelligent Vehicles*, 8(1): 22-36.

- [9] Chen, L., Wu, P., Chitta, K., Jaeger, B., Geiger, A. and Li, H. (2023) End-to-end autonomous driving: Challenges and frontiers. arXiv preprint arXiv:2306.16927.
- [10] Zhao, J., Zhao, W., et al. (2023) Autonomous driving system: A comprehensive survey. *Expert Syst. Appl.*, 242: 122836.
- [11] Research In China (2024) A Comprehensive Study on End-to-End Autonomous Driving in 2024. <http://www.pday.com.cn/Htmls/Report/202404/24543611.html>
- [12] Cherish Capital, Autonomous Driving Branch to Shanghai Alumni Association of Nanjing University, and JiuZhang-AI. (2024) End-to-End Autonomous Driving Industry Research Report. <https://img.iduodou.com/images/docs/20240612/C0EC32EF-5BFA-4BA4-9DAA-8FDA4BF4BCA7.pdf>
- [13] Tian, Y., et al. (2018) DeepTest: Automated Testing of Deep-Neural-Network-Driven Autonomous Cars. 40th International Conference on Software Engineering, Gothenburg, Sweden, 303-314
- [14] Wang, W., Xie, J., Hu, C., et al. (2023) Drivemlm: Aligning multi-modal large language models with behavioral planning states for autonomous driving. arXiv preprint arXiv:2312.09245.