

World models for autonomous driving

Yuanzhe Chen

School of Electronic and Information Engineering, Tongji University, Shanghai,
200092, China

yuanzhechen@tongji.edu.cn

Abstract. Advancements in autonomous driving have been achieved due to the increasing use of artificial intelligence. The existing autonomous driving technology is insufficient to handle intricate traffic conditions. In autonomous driving, the world model is a crucial and innovative technology. Utilizing sensors and pre-existing knowledge, a world model can enhance the autonomous driving system's comprehension of the surroundings, offer essential data for future judgments, and enhance the system's resilience. The paper employs literature analysis and review methods to investigate the research and implementation of world models in autonomous driving, including environment perception and modeling, path planning, decision-making, and safety. This study examines the use of artificial intelligence technology in autonomous driving and analyzes the research and application of the world model in this field. It offers new insights for addressing challenging scenarios in autonomous driving and enhancing the safety of the system.

Keywords: autonomous driving, artificial intelligence, world model, security.

1. Introduction

Artificial intelligence technology has enabled computers to achieve significant advancements in natural language processing, computer vision, and various other domains previously considered unimaginable. Human beings have always aspired to achieve autonomous driving. Autonomous driving technology is a prominent subject in artificial intelligence, garnering increasing interest and being extensively applied in vehicles, unmanned aerial vehicles, and various other sectors. Waymo's study suggests that autonomous driving technology can be more reliable than human drivers up to a certain point [1]. Autonomous driving technology is still far from being fully developed, particularly in handling intricate traffic situations. The world model is a multi-modal machine learning model that can make decisions and predict future outcomes by gathering and assimilating information from the external environment [2]. In the realm of autonomous driving, the world model is a developing multimodal artificial intelligence technology that aids the autonomous driving system in precisely perceiving and comprehending the surroundings, offering essential information for future judgments.

This paper uses the research methods of literature analysis and review to analyze the application of world model in the field of autonomous driving, such as environment perception and modeling, path planning and decision-making, safety and so on. Tesla presented a technical report at CVPR 2023 [3], while Wayve discussed the research progress of the world model in autonomous driving, highlighting its latest applications in their autonomous driving systems [4-5]. This study examines the use of

artificial intelligence technology in autonomous driving and explores the application of the world model in this field. It offers new ideas and methods to address challenging scenarios in autonomous driving and enhance the safety of autonomous driving systems.

2. Introduction of basic concepts

2.1. World Model

The world model is an innovative machine learning model designed to gather information from the environment, make decisions, and forecast the future by mimicking the human brain's structure. Prior to the development of the world model, several machine learning models were limited to learning only one specific form of input, such as solely spoken language or images. Yet, these strategies often struggle to address all scenarios in the intricate real world, encountering corner situations and requiring enhanced generalization. In the realm of autonomous driving, crucial for life safety, we must seek enhanced ways to enhance safety. When humans drive, their brain utilizes superior language skills to acquire semantic information from images, aiding in making accurate decisions. The world model can acquire knowledge on human driving by analyzing images and associated driving behaviors from sensors in the dataset. This enables it to develop into a skilled driver, enhance its capability to handle intricate scenarios, and improve the reliability and safety of the autonomous driving system.

2.2. Autonomous driving technology

Autonomous driving technology utilizes sensors and algorithms to perceive the surrounding environment, make driving decisions, and control the vehicle through automatic control algorithms. Recently, advancements in artificial intelligence have led to significant progress in autonomous driving through the use of machine learning and computer vision technologies, particularly deep learning [6-10]. Waymo, a startup specializing in autonomous driving, has introduced self-driving cabs primarily through the utilization of multi-sensor fusion techniques. Tesla, an electric vehicle business, primarily follows a pure vision approach and continuously conducts research to enhance its autonomous driving system [3, 11].

3. World Model in autonomous driving

3.1. Environment Perception and modeling

Autonomous driving vehicles commonly use cameras, millimeter wave radar, ultrasonic sensors, and some sophisticated models even incorporate LiDAR technology. Cameras employ machine learning algorithms to identify lane lines and objects, while millimeter wave radar, ultrasonic sensors, and LiDAR gauge distances and detect impediments. These sensors function as the visual perception system of self-driving cars, aiding in environmental detection and playing a crucial role in autonomous driving technology. The sensor data from autonomous vehicles can supply the world model with external traffic environment information, including time and place, serving as a crucial data source for the world model. Waymo, an autonomous driving firm, has introduced the Waymo Open Dataset. It includes data from different sensors like cameras and LiDAR, offering top-notch datasets for research and development in autonomous driving.

3.2. Path Planning and Decision-making

While driving a car, we must follow fundamental traffic rules and navigate through several intricate traffic situations, such as a row of traffic cones and a construction vehicle suddenly materialize in front of the road. When presented with this situation, the human driver can use their accumulated knowledge and experience to make a wise decision and select the vehicle that avoids the construction. The world model may learn decision behavior directly from driving behavior videos in the dataset using end-to-end training, acquiring specific prior knowledge. When the autonomous driving system faces a real-world scenario, it can efficiently make optimal decisions and seamlessly transition from

picture recognition to action execution and perception to decision-making. Wayve's Vision-Language-Action Models (VLAMs) can enhance the explainability of machine learning in autonomous driving by utilizing natural language to elucidate driving behaviors based on the training data of human drivers.

3.3. Security

World models can allow autonomous driving systems to make better judgments and decisions when faced with a variety of real-world scenarios by learning from the experience of human drivers. At the same time, the autonomous driving system based on the world model also has some semantic reasoning ability through the comprehensive learning of images and driving behaviors. According to the autonomous driving company Wayve, through the world model [4-5], the autonomous driving system can not only realize the end-to-end effect from visual input to driving action output, but also use natural language to explain the reasons for the corresponding driving action in view of the current external environment and make certain analysis. To a certain extent, it enhances the interpretability of the machine learning "black box" model and contributes to enhancing the robustness and safety of the autonomous driving system.

However, it is important to acknowledge that the present machine learning models, mostly relying on mathematical statistical methods, still have specific constraints. The world model's ability to acquire human drivers' driving experience from the dataset is closely linked to the dataset's quality. When a dataset is sourced from skilled and accurate human drivers, the prior knowledge about traffic and driving acquired by the trained world model from the dataset tends to be more accurate and dependable. If the data set includes flaws like risky driving behaviors such as distracted driving, the pre-existing knowledge of traffic and driving acquired by the trained world model from the data set can have biases and vulnerabilities, posing significant safety hazards when applied on the road. Thus, meticulously filtering accurate and high-quality driving behavior data is crucial to guarantee the safety of the world model.

To enhance the quality of the world model, we must attain data closure. Data closure involves utilizing freshly acquired high-quality data to enhance the training of the model, enabling continual iteration and improvement of the model. Implementing data loop closure can enhance the diversity of corner cases in the dataset, hence enhancing the resilience and security of the autonomous driving system. Tesla, an electric vehicle manufacturer, has continuously improved its autonomous driving system by utilizing data loop closure, leading to successful outcomes [3].

To guarantee safety when the end-to-end world model generates driving actions from visual input, it is necessary to incorporate a set of explicit and understandable safety rules to limit the model's behavior within safe parameters. When the output of the world model conforms to the artificially added safety logic and rule bottom line, we can believe that compared with the results of the traditional perception-decision separation model, the output of the world model may be closer to the driving behavior of human drivers, and we may have a better driving experience. However, if the output results of the world model are contrary to the safety logic and rule bottom line, we need to let the artificially added safety logic and rule bottom line play a dominant role and take over the decision-making of the autonomous vehicle to avoid the danger caused by the wrong judgment of the world model on corner cases.

4. Challenges and future development of the World Model

The world model has numerous benefits. The world model can lower dataset annotation costs using self-supervised learning techniques and enhance the interpretability, generalization, and resilience of the model. Nevertheless, implementing the world model in autonomous driving has numerous hurdles. The industry suggests utilizing the generative world model to create a virtual environment for testing autonomous driving systems. Nevertheless, this approach has issues with the unknown quality and inconsistency of the generated outcomes with real-world scenarios. Secondly, the topic of security has been extensively discussed in the preceding part, hence the author will not reiterate it here. The issue of

computer power is also a concern. The generative world model may necessitate a significant amount of computational resources, and there is potential for further enhancement and streamlining.

The world model, as an emerging multimodal machine learning technique, offers vast development potential and promising application prospects in the future. The world model gains knowledge from sensing the external environment, resembling the learning process of the human brain. World models are anticipated to advance AI to a higher level of intelligence, enabling the resolution of intricate tasks like autonomous driving, enhancing the safety of autonomous driving, and promoting human well-being.

5. Conclusion

According to the analysis in this paper, we can make the following preliminary conclusions: World model, a new multimodal machine learning device, shows great potential for study and implementation in the autonomous driving industry. The world model uses sensors to perceive the external physical environment and learns prior knowledge from it, enhancing the intelligence of the AI model. This improves environment perception, modeling, path planning, decision-making, and safety in autonomous driving. Implementing high-quality data sets, data loop closure, and incorporating a set of known and interpretable safety logic and rule bottom lines can help minimize corner cases, limit the world model within safety boundaries, enhance the performance of autonomous driving vehicles in complex scenarios, and ultimately boost the robustness and safety of the autonomous driving system. Tesla and Wayve, among other autonomous driving startups, have shown the most recent outcomes of implementing world models in autonomous driving. The demonstration videos from both companies show that the world model has made significant progress in various areas of autonomous driving, including environment perception, path planning, decision-making, safety, and creating simulation environments for future use. The outlook for the future is positive. This study also has some shortcomings, such as the lack of some surveys, insufficient collected data, and incomplete research, etc. In the future, it will be continuously improved, and further research will be done in the application of the world model in the field of autonomous driving.

References

- [1] Johan Engstrom, Shu-Yuan Liu, Azadeh Dinparastdjadid, Camelia Simoiu. Modeling road user response timing in naturalistic settings: a surprise-based framework. arXiv preprint, 2023.
- [2] Anna Dawid, Yann LeCun. Introduction to Latent Variable Energy-Based Models: A Path Towards Autonomous Machine Intelligence. arXiv preprint, 2023.
- [3] Tesla Research Team. Foundation Models for Autonomy. Conference on Computer Vision and Pattern Recognition (CVPR 2023 Workshop).
- [4] Anthony Hu, Lloyd Russell; Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, Gianluca Corrado. GAIA-1: A Generative World Model for Autonomous Driving. arXiv preprint, 2023.
- [5] Wayve Research Team. Introducing GAIA-1: A Cutting-Edge Generative AI Model for Autonomy. <https://wayve.ai/thinking/introducing-gaia1/>, 2023.
- [6] Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Deep learning[J].Nature,2015,Vol.521(7553): 436-444.
- [7] He, Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun, Jian. Deep Residual Learning for Image Recognition[J].2016 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR),2016,: 770-778.
- [8] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection[A].2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C],2016.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar. Attention is all you need[A].NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems[C],2017.
- [10] An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [A]. ICLR 2021.

- [11] Waymo Research Team. Progress towards Scalable Deployment of Autonomous Driving. Conference on Computer Vision and Pattern Recognition (CVPR 2023 Workshop).
- [12] Pei Sun; Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, Dragomir Anguelov. Scalability in Perception for Autonomous Driving: Waymo Open Dataset[A].2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)[C],2020.
- [13] Wayve Research Team. LINGO-1: Exploring Natural Language for Autonomous Driving. <https://wayve.ai/thinking/lingo-natural-language-autonomous-driving/>, 2023.