Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

# Advancements and Challenges in Visual Perception for Autonomous Driving

Haotang Wang *

Beijing YuYing School, Beijing, 100036, China

* Corresponding Author Email: Wht070324@outlook.com

**Abstract.** As autonomous driving has become more popular, research on the technology has proliferated. There are many schemes to realize autonomous driving, including visual perception, lidar, and radar. Among them, visual perception implements advanced neural networks and transfer learning to achieve excellent performance. This article mainly discusses the advantages of visual perception, the challenges it faces, and possible solutions, while comparing it with other schemes from various aspects at the same time. The approaches to the study mainly involve literature review and comparative analysis. The article concludes that visual perception will thrive and dominate the autonomous driving industry in the future. The prospect of development is also investigated, and some potential solutions that can improve the performance of autonomous driving are provided in the passage. This comprehensive analysis underscores the pivotal role of the visual perception in advancing autonomous driving technology, highlighting its cost-effectiveness, scalability, and future potential in transforming transportation.

**Keywords:** Visual perception; Autonomous driving; Transfer Learning Survey.

## 1. Introduction

Autonomous driving, in recent years, has been receiving increasing attention for its potential to relieve drivers' burdens and improve the safety of driving [1]. Autonomous driving can enhance transportation safety and reduce the possibility of traffic accidents. What's more, autonomous driving can help people who have problems driving travel independently like old people or disabled. Autonomous driving will be a trend in future car development, and because of the fast growth of the electric car industry, the realization of it has become easier. Many car companies have already implemented the technology in their new products nowadays, and the technology has facilitated people's transportation experience.

The realization of autonomous driving involves various operations, ranging from gaining the information of the environment to the algorithm or mechanism to process the information and make a reaction. There are many ways to process information. Traditional rule-based approaches and advanced machine-learning methods are all involved [2]. The autonomous driving solution involves a visual solution, V2X, and lidar [3]. The advantages of the visual solution are as follows. The most significant one is its low cost. While others all need many devices launched around the car, the visual solution only needs several cameras. Another benefit is that due to the singular input information, conflicts between data from different sensors are minimized. A visual solution can recognize words on the road sign which is not able to be done by all other approaches. Lastly, the trained model of visual solution can also be used to build other things such as robots which will also rely on the input of such information. Visual solution deeply relies on the model and algorithm that processes the information obtained from the camera. The build of the model involves a neural network and deep learning. This poses a significant challenge for many car companies, as they struggle to develop a model capable of safely processing information for transportation applications. Moreover, this solution still faces many challenges such as its low performance under low visibility and inability to accurately detect further targets.

This article primarily delves into the visual solution for autonomous driving, juxtaposing it with alternative approaches. Employing comparative and analytical methodologies, the article scrutinizes the advantages and challenges inherent in the visual solution, with the ultimate objective of identifying avenues for enhancing its performance.

The research gap of visual solutions is as follows. Firstly, cameras are not able to detect things very close to the vehicle or things that are uncommon. Secondly, it would be harder for cameras to extract depth information than lidar, as the model may get deceived for some reason.

## 2. The Principle of Visual-Based Autonomous Driving

### 2.1. Role of Visual Perception in Autonomous Vehicles

Visual perception plays an important role in autonomous vehicles. The vehicle uses cameras launched in different directions of the car to obtain information about the external environment and convert it into digital signals. After processed in the neural network, the digital signals are converted into specific information like the direction and the essence of the information in the photo. The conventional autonomous driving solution always uses many sensors launched around the car such as lidar and radar and has the assistance of a global positioning system (GPS) to do the process and create a point cloud of the surroundings thus deciding driving [4]. The lidar can directly get the information of the distance and the direction of the surrounding object while the computer vision needs to integrate the information of cameras in different directions of the car to get the Bird's Eye View (BEV) thus achieving the same effect. This process significantly depends on computational accuracy and efficiency but can reduce overall costs.

### 2.2. Sensors and Algorithms in Visual-Based Autonomous Driving

Autonomous vehicles primarily use cameras as their main sensors, often supplemented by millimeter-wave radar. The most critical sensor is the camera, which is placed at various positions around the car. Above the front windshield, the front view camera is placed to detect the traffic signal and warn of the front car collision. The front view camera can be single or double while two of them can perform better in the task of distance detection but cost more. An autonomous vehicle also needs to contain several surround cameras to provide information to meet the needs of gaining overall information about the environment and give the reverse image.

The decision method of autonomous driving can be mainly divided into pipeline framework and end-to-end framework. This article mainly discusses the pipeline framework, which includes the intelligent driver model (IDM), Predictive driver model (PDM), and reinforcement learning methods. The rule-based model does not need the high computational complexity and may make faster responses. However, it is not flexible, and is hard to upgrade and adapt to new conditions. The other two models mentioned above all rely on neural networks and deep learning and thus need great computational resources and advanced algorithms. They can both learn from the feedback of the decisions they made, and upgrade quickly can adapt to new environments more rapidly and are more flexible.

## 3. Research Comparison Analysis

### 3.1. The Implementation of Occupancy Network in the Realm Of 3D Detection

Occupancy Networks are a type of neural network architecture used in 3D scene understanding, particularly in the context of computer vision tasks like 3D reconstruction and object detection. They were introduced to address the challenge of efficiently and accurately representing 3D geometry from 2D images or point clouds [2].

The Occupancy Network initially transforms the multi-view input information into a 3D feature space. Subsequently, it employs a deep neural network to learn the acquisition probability of this feature space and make decisions based on it. Compared to conventional algorithms that generate 3D space, the Occupancy Network can produce a more accurate representation of the object. Additionally, it can predict the occupancy of unknown objects.

### 3.2. Transfer Learning in Autonomous Driving Algorithms

Transfer learning serves as a valuable tool in autonomous driving algorithms, enabling existing models to achieve higher performance without requiring training from scratch. This approach allows models to adjust themselves based on new datasets and tasks, facilitating rapid adaptation to diverse environments. Additionally, transfer learning reduces the computational resources and dataset sizes required for training, making it particularly advantageous in the realm of autonomous driving where efficiency is paramount [4]. For instance, researchers leverage transfer learning to adapt pre-trained deep learning models, enhancing road scene understanding and traffic sign recognition in self-driving cars. Moreover, transfer learning aids pedestrian detection systems by enabling adaptation to diverse environments for improved performance. Additionally, sim-to-real transfer techniques facilitate training in simulation environments and subsequent adaptation to real-world scenarios, bridging the gap between simulated and real-world driving conditions. These applications highlight the versatility and efficacy of transfer learning in improving the generalization and adaptation capabilities of autonomous driving algorithms.

### 3.3. Comparison with Other Solutions from Many Aspects

In comparing visual perception, V2X communication, and lidar technology for autonomous driving, it is essential to consider their respective advantages and drawbacks across factors such as information perception, cost, safety, and computational complexity. The comparison of those aspects is shown in Table 1.

**Table 1**. The comparison of three schemes

| | Visual perception | V2X | lidar |
|---|---|---|---|
| Advantages | Information perception: visual perception can provide rich information about the surrounding environment compared to other schemes, including road signs, lane markings, traffic lights, and obstacles.<br>Low cost: comparing to other sensor technologies like radar or lidar, cameras are relatively inexpensive.<br>High resolution: modern cameras can capture high-resolution images, which contributes to the detailed analysis of the environment. | Safety: V2X enables vehicles to exchange real-time information about their status, intentions, and the surrounding environment.<br>Traffic efficiency: V2X allows vehicles to share information about traffic conditions, which can lead to reduced congestion and shorter travel time.<br>Optimized driving behavior: by obtaining the overall information of the vehicle and surrounding environment, V2X can help optimize driving behavior to improve fuel efficiency, reduce emissions, and enhance overall vehicle performance. | High-resolution 3D mapping: lidar sensors can provide high-resolution, three-dimensional maps of the vehicle's surrounding environment. This detailed mapping allows for accurate detection and tracking of objects, including vehicles, pedestrians and obstacles.<br>Long Range and Wide Field of View: lidar sensors can detect objects at long ranges and over a wide field of view, enabling early detection of potential hazards and improved situational awareness for autonomous vehicles.<br>Precise Object Detection and Localization: Lidar sensors provide precise measurements of the distance to objects, |

| | | | |
|---|---|---|---|
| | | | allowing for accurate object detection, localization, and tracking, even in complex urban environments with multiple obstacles. |
| Drawbacks | Performance under adverse conditions: the performance of visual perception will be greatly influenced under adverse weather conditions such as rain, snow, fog, or low-light environment, making the system hard to make accurate decisions.<br><br>Vulnerable when environment changes: changes in lighting conditions, shadows, reflections, or variations in object appearance can affect the performance of visual perception.<br><br>Computational complexity: Processing images requires significant computational resources. | Infrastructure dependency: V2X deeply relies on a robust infrastructure including roadside units and communication networks. The performance of V2X solutions can be limited in areas with poor infrastructure or limited network coverage.<br><br>Privacy issues: V2X communication involves the exchanges of location and vehicle data, raising concerns about privacy and data protection.<br><br>Cost and deployment challenges: deploying V2X infrastructure and equipping vehicles with V2X technology can be costly and time-consuming. The widespread adoption of V2X solutions may require significant investment in infrastructure upgrades and industry cooperation | Cost: Lidar sensors are currently more expensive than other sensor technologies such as cameras and radar, which can increase the overall cost of autonomous driving systems and limit their widespread adoption, especially in consumer vehicles.<br><br>Perception of Transparent or Low-Reflectivity Objects: Lidar sensors may struggle to detect transparent or low-reflectivity objects such as glass, certain types of plastics, or soft materials, which can pose challenges for autonomous vehicles navigating urban environments with diverse obstacles.<br><br>Complex Data Processing: Lidar sensors generate large volumes of point cloud data that require complex processing algorithms to extract meaningful information about the environment. This can increase the computational and processing requirements of autonomous driving systems, potentially impacting their real-time performance and scalability. |

## 4. Discussion

### 4.1. Advantages

Visual perception has many advantages compared to traditional approaches. And its strength mainly depends on the advancement of computer technology of algorithms and neural networks [5]. Firstly, it can provide with a relatively high accuracy of the recognition of the vehicle's surroundings. Secondly, the ability to extract meaningful information allows for robust performance in diverse conditions. The visual perception solution is also a cost-effective solution. Compared to other complex sensors, the cost is much lower. This feature will also make it easier to spread across different vehicles and markets. For future development, visual perception also has more advantages. The experience of real-world practice can help to refine and optimize the neural network architecture, enabling it to achieve better performance in the future [6]. The development of visual perception will also have benefits in other realms due to its scalability and generalization across the domains. The models or algorithms used in autonomous driving can be directly extended to robotics, surveillance, or augmented reality [7].

### 4.2. Challenges and limitations

Despite all the advantages mentioned above, visual perception still faces many challenges and limitations. First the performance under complex weather conditions and low visibility, the visual perceptron cannot always perform well. Visual perception is also vulnerable to adversarial attacks, while maliciously crafted input can deceive neural networks easily and lead to wrong predictions [8].

### 4.3. Advise and Application Prospect

There are several ways to improve the performance from several aspects. The most important and most important aspect would be the improvement of algorithms and deep learning models. The adjustment can reflect several improvements including the ability to learn complex features, computational efficiency, and robustness. Techniques, like skip connections, bottleneck layers, and network pruning, can all contribute to the improvement of the efficiency of convolutional neural networks. As most of the modes of autonomous driving are transformer-based, the improvement on the transformer would help improve the performance. Vision transformer models can replace the convolutional layers in traditional CNNs with self-attention layers. Hybrid architectures may also be helpful as they can combine different architectures thus leveraging the strength of both [9]. Directly use the model and data on the robotRemain some necessary rules to enhance safety and efficiency

To achieve fully autonomous driving, much work still needs to be done. More real-time test needs to be done to find the potential risks and help the deep learning model get more data to achieve a higher performance. Despite the advancement of computer technology and artificial intelligence, human still needs to be involved in many crucial decision-making. So, making sure to have a human-machine interaction design is still important. As most decision-making patterns were the imitation of human behavior, this makes take more computational resources and may confuse the computer. Remaining more rules would be more efficient and more stable.

## 5. Conclusion

This article mainly discusses the advantages and potential improvement of the visual perception solution to autonomous driving. Compared to other schemes, visual perception only uses cameras as the sensor and deeply relies on deep learning models to process the information. The difference in sensors will result in the cost-effective feature of a visual perception solution. Through visual perception, the vehicle can navigate complex environments, identify obstacles, and make decisions to ensure safety and improve efficiency during transportation. The article also mentioned the challenges visual perception is facing now including its ability to detect the distance of objects and performance under poor visibility. In the end, several possible improvements in visual perception from different aspects were provided. The achievement of fully autonomous driving depends on the development of artificial intelligence and deep learning systems, and it will bring revolution to the realm of transportation and reshape the future of mobility.

## References


[1] Jiageng Mao, Shaoshuai Shi, Xiaogang Wang, Hongsheng Li. 3D object detection for autonomous driving: a comprehensive survey. International Journal of Computer Vision, 2023, 131(8): 1909-1963.

[2] Juan Zhong, Zheng Liu, Xi Chen. Transformer-based Models and Hardware Acceleration Analysis in Autonomous Driving: A Survey. arXiv preprint arXiv:2304.10891, 2023.

[3] Si Liu, Chen Gao, Yuan Chen, Xingyu Peng, Xianghao Kong, Kun Wang, Runsheng Xu, Wentao Jiang, Hao Xiang, Jiaqi Ma, Miao Wang. Towards vehicle-to-everything autonomous driving: a survey on collaborative perception. arXiv preprint arXiv:2308.16714, 2023.

[4] Keli Huang, Botian Shi, Xiang Li, Xin Li, Siyuan Huang, Yikang Li. Multi-modal sensor fusion for auto driving perception: a survey. arXiv preprint arXiv:2202.02703, 2022.

[5] Hrag Harout Jebamikyous, Rasha Kashef, Autonomous vehicles perception (avp) using deep learning: modeling, assessment, and challenges. IEEE Access, 2022, 10: 10523-10535.

[6] Shifeng Zhang, Longyin Wen, Zhen Lei, Stan Z. Li. RefineDet++: Single-shot refinement neural network for object detection. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(2), 674-687.

[7] Karthik Allam. Big data analytics in robotics: unleashing the potential for intelligent automation. EPH-International Journal of Business & Management Science, 2022, 8(4), 5-9.

[8] Ibrahim Sobh, Ahmed Hamed, Varun Ravi Kumar, Senthil Yogamani. Adversarial attacks on multi-task visual perception for autonomous driving, 2021, Electronic Imaging, 34.

[9] Luigi Bonati, Yueyu Zhang, Michele Parrinello. Neural networks-based variationally enhanced sampling. Proceedings of the National Academy of Sciences, 2019, 116(36), 17641-17647.