Autonomous Driving Core Tasks: Object Detection and Path Planning

Junlin Chen[1*], Hanmo Wang[2]

[1]*School of Cyber Security, Guangdong Polytechnic Normal University, Guangzhou, China*
[2]*School of Automation, Beijing Information Science & Technology University, Beijing, China*
*Corresponding Author. Email: chen1067530243@outlook.com

Abstract. With the continual improvement of artificial intelligence, sensing, and computational technologies, autonomous vehicles have become an important part of intelligent transportation systems research. Environment perception being accurate and safe trajectory creation is the basis of autonomous driving, and object detection and path planning are the two must-have basic tasks. A detailed review of two key Autonomous Driving technologies, Object Detection and Path Planning, is given by this article. On objects, it looks at the YOLO-based single stage, the Transformer-based global approach, and two-stage based or multi-modal approaches. It discusses each model's principles, advantages and disadvantages for path planning, traditional algorithms, end to end deep learning, and deep reinforcement learning methods will be introduced in terms of features and problems. The study points out that YOLO has high real-time, but is lacking on robustness in some cases; the transformers model brings in global context but adds much more computational cost; two-stage, multi-modal models provide high accuracy, however, they have difficulty in deployment. Path Planning traditional algorithms are more steady than not adaptive, deep leaning algorithms are adaptive than data dependant. Future works: Lightweighting the model, Multi-modal model, Online learning, and Interpretable model to improve safety, stability and practical application.

Keywords: autonomous driving, object detection, path planning, YOLO, deep learning

1. Introduction

Autonomous drive, which is a key part of intelligent transportation system, has made great progress recently on perceiving, deciding, and controlling. Its main goal is for cars to use sensors to see the world and drive themselves around safely and effectively based on what they sense. The key elements of an autonomous driving system have to satisfy safety and stability of the vehicle when performing object detection and path planning. Object detection identifies vehicles, pedestrians, traffic signs, etc., in the driving environment, supplying valid inputs for path planning and decision-making. Path planning generates path from vehicle state and environment information, so as to ensure smooth and safe travel under different conditions [1].

These two aspects are used throughout this study to make a systematic review of autonomous driving technology. First, object detection is divided into YOLO-based detection algorithms and

Transformer-based global modeling methods and some two-stage or multi-mode detection based on CNN and RNN [2]. Second, on the path planning, traditional algorithms such as A*, RRT are analyzed based on its characteristics, advantages, disadvantages, and future development. Based on end-to-end and reinforcement learning, deep learning is also studied [3]. Comparing all these, this paper points out that there is a trade-off among real-time performance, accuracy, and robustness, which can serves as a reference for the next step.

The significance of this review is to introduce the main kinds of object detection and path planning methods and trends, so as to offer some suggestions and inspirations for scholars and engineers Identifies limits in current methods of complex traffic cases, points out directions for further improvement and application.

2. Autonomous driving object detection

Object detection belongs to the basic perception module of autonomous driving system, which can correctly identify and locate cars, pedestrians, traffic signs and other obstacles under complicated road environment. And then those detections provide reliable pieces for making plans and choices. In recent years, the great progress on deep learning has greatly accelerated the iteration and optimization of object detection method and they become widely used in autonomous driving context On the whole, object detection methods are mainly classified into YOLO-based one-stage methods, transformer-based methods, two-stage detection methods, and multi-modal based methods [4].

2.1. YOLO-based object detection

The YOLO series belong to a typical category of single-stage object detection algorithms, which feature end-to-end training as well as fast speed. YOLOv1 first introduced the idea of directly predicting bounding boxes and class probabilities using a single neural network architecture and removing the laborious region proposal step, greatly accelerating detection performance. Following versions including YOLOv3, YOLOv4, YOLOv5 came out with improvements on network architecture and feature extraction strategies as well as training techniques, better performance on small object and increase both accuracy and robustness. YOLOv7 and YOLOv8 improved speed and accuracy further and can be deployed on vehicles and embedded systems [5].

In autonomous driving scenarios, YOLO methods can complete object detection within millisecond time, this is very important for real time perception. Simple structure, easy to train, quick inference. But YOLO models can still miss objects or make mistakes due to things like tricky weather, different lighting, or lots of cars. Distant and small objects must be detected for good performance. Thus, YOLO-based systems often employ data augmentation, multi-scale feature fusion, etc., for better robustness [6].

Dong Hongzhao et al. systematically summarized the development and application of YOLO in traffic object detection. They put traffic targets in two categories: "pedestrians-vehicles-road infrastructure" and looked at the architectural enhancements and performance upgrades done on varying YOLO versions. Key techniques: Multi-scale feature fusion, attention mechanism, lightweight network. The authors have shown that applying cross-layer feature fusion and special attention module to the backbone of YOLO has increased the performance of detection of the small and occluded object in traffic scenarios of different situations. From experiments they know that it improved the accuracy and robustness on various lighting and occlusions conditions [6]:

Mao Zhihui et al. reviewed YOLO based object detection on autonomous driving and gave some improvements. Reviewing the different versions on YOLO in terms of the use of feature extraction, anchor boxes and training strategies was performed, which proposed a combined use of data amplification, multi-scale prediction and attention. These experiment results on the public car traffic data sets showed that joint has big increase in mean Average Precision(mAP) and detection at low light compared to dark, and has more less miss of small objects. And it also stresses the need for balancing the modeled complexity and computational resources of the embedded platform in order to attain the best trade off between speed and accuracy [7].

2.2. Transformer-based object detection

Transformer architectures, which were first developed for natural language processing, are really good at sequence modeling and long-range dependencies. and then I was introduced to computer vision task. DETR (Detection Transformer) is the first one to use Transformers, which replace the traditional region proposal methods and instead uses sequence modeling to output bounding boxes and class probabilities directly. Transformer performs much better as compared to CNNs as far as capturing global scenery info and modeling interactions among multiple objects in a complex traffic scenario [8].

Autonomous driving, it is more accurate for autonomous driving cars to use transformer-based methods to identify cars, people, and traffic signs in crowded traffic scenes. Jing Rongrong et al. used an interactive perception and trajectory prediction based on a Transformer, with the aim of improving prediction precision and stability by modelling dynamic interaction between different traffic participants through multi-head self-attention [8]. Moreover, Deformable DETR proposed deformable attention, which improved the training convergence and small-object detection. However, Transformers are still facing issues regarding computationally high consumption of resources and inferring speed, both key things for low-latency autonomous driving. and the research will be more of lowering the calculation needed but keep it accurate [9].

2.3. Other object detection methods

Other than the YOLO and Transformer methods, the Two-stage approach is still an integral part of autonomous driving. typical examples include Faster R-CNN, it combined a region proposals network with classification and regression networks realizing precises detections. Two-stage methods are strong when it comes to accuracy but slow, so they cannot be used in real time much, so they are just best for cases where it needs extremely accurate precision.

Some of the more enhanced convolution neural nets such as SSD and RetinaNet join in to spot the objects for an autonomous drive as well. These ones are speedy and right, special for those middle-sized things. Multi-sensor fusion methods have attracted attention as the same time. Combining LiDAR points with camera images for low light or poor weather robustness can use multi-modal detection It increases perceived reliability; It's a big safety safeguard in autonomous driving. but the high cost of sensors and the arduous task of annotating data is too much for the real world [10].

From the whole, autonomous driving has developed object detection in several directions: YOLO methods take charge because they are fast, Transformer methods are coming on strongly thanks to being able to model the whole world, two-stage and multi-modal methods continue to be important when you need to be really good at it. each of these has their upsides and downsides, and future

work will endeavor to improve these approaches and fuse them with algorithms so as to achieve accuracy and swiftness that remains firm across various applications.

3. Path planning for autonomous driving

Path planning belongs to a key part of autonomous driving system which can generate feasible, safe and effective driving routes from vehicle's current location to target point according to surrounding condition information. A good path planning is directly related to whether driving is safe, stable, and passengers are comfortable. It has a significant impact on the overall intelligence of the autonomous driving system. According to their technical properties, path planning is divided into two categories: traditional path planning algorithms and deep learning path planning algorithms.

3.1. Traditional path planning methods

Traditional path planning algorithms mainly apply graph search, sampling, and optimization approaches; they have distinct mathematical structures and strong determinism. Early AD mainly uses these methods in research/application, they're stable in a static environment.

3.1.1. Graph search-based planning

Search on the graph consists of Dijkstra and A* algorithms. The Dijkstra algorithm has an overall optimal result by expanding nodes step-by-step in accordance with their distances, its computing efficiency is low and can't be used for real-time. A* is an improved form of Dijkstra, adding a heuristic function to direct the search toward the goal, improving computational efficiency and doing well in fixed and structured settings, but in complicated moving situations or multiple vehicles interacting, frequently updating the route is necessary, raising computational cost and making response less swift [11].

Lin Y. proposed an improved path planning method that incorporates Dijkstra with A*. This is done in order to enhance navigation in complicated environments. Add a multi-level heuristic on the A* algorithm, different environment states have different weightings, it can maintain global optimality and greatly reduce the unnecessary search amount. Experiments found it could achieve high path smoothness and security under a complex road environment. It proves that the hierarchical A* - based model could actually implement an effective and real - time global optimal path [11].

3.1.2. Sampling-based planning

Sampling methods can also be flexible in unstructured environment. The Rapidly-exploring Random Trees and its variants are typical representatives. RRT is good at exploring feasible paths in high-dimensional spaces, its path can meet on irregular roads and deal with some troublesome road obstacles. Path generated by RRT are optimized at the same time to be smooth and near-optimal. Although these stochastic methods are flexible, the non-deterministic nature can effect path stability and control precision in high-speed dynamic scenarios [12].

Chen R. et al. put forward a hybrid RRT-Dijkstra Path Planning Strategy to improve path search efficiency and global optimality in complicated environments. Firstly use RRT to quickly generate feasible path for high-dimensional space, then use Dijkstra algorithm to search for global optimization for the path, smooth out the trajectory and reduce the cost as well. In dynamic obstacle conditions, there is about 30% acceleration in the planning speed, about 12% reduction of path length, and it also largely improves the trajectories continuity and control preciseness. It shows the

possibilities and the advantages of combining RRT and RRT+Dijkstra in an unstructured environment [12].

3.1.3. Optimization-based local planning

Local path planning is often done with vehicle dynamics constraints. The DWA searches feasible velocity space and selects optimum velocities, which generates a local path. The DWA makes sure feasible paths with considering of vehicle dynamics. But it is still dependent on the environment modelling, so there might be local optimum in complicated traffic scenarios [13].

Liu B.et al. proposed a DWA-based path planning algorithm to improve the local planning's real-time performance and safety. It creates smooth, controllable local paths via optimization w.r.t feasible velocities, and accelerations. And they have environmental cost function which will judge where the obstacles are so that vehicles can avoid from getting into collision and they can change their way as well. The path is smoother, the planing time takes shorter with more stable and flexible trajectory and environment conditions [13].

Briefly speaking, traditional path planning methods are theoretically mature, interpretable, and certain. But they are poor in dynamic, complex, and large-scale environment, and poor in real-time response and generalization.

3.2. Deep learning-based path planning methods

In the deep learning advancing process, more and more work applies it to autonomous driving path planning These methods will be mainly used end-to-end learning method which is to map the perception inputs directly onto vehicle control commands or path outputs, or combining reinforcement learning to make the decision.

3.2.1. End-to-end learning

End-to-end methods train deep neural networks to generate path or the control command directly from sensor data. take convolutional neural networks (CNNs) for example which can recognize road boundaries and line as well as objects from camera pictures and then connect layers gives steering angle of acceleration/speed from speed. It can avoid complex modeling and improve flexible and real-time path generation. However, end-to-end ones need lots of different, good training material, and they're not very good at dealing with things not seen before [14].

Ge B et al. proposed an end-to-end Deep Reinforcement Learning based Vehicle Path Planning method. The CNN takes in an image and outputs road boundaries, lane lines, and obstacles, and finally sends out commands on how to turn, accelerate, and move, automating path creation. From experiments can see this methods gets workable routes in complex roads faster than depending on old models and laws when making plans although, it's generalization is bounded by the quality and amount of training data [14].

3.2.2. Deep Reinforcement Learning-based methods

Deep Reinforcement Learning (DRL) turns path planning into a Markov Decision Process (MDP), where vehicles interact with the virtual surroundings to gain optimal policies. DQN, DDPG, etc. are popular algorithms. DRL can do real - time decisions for path choosing and obstacle avoiding in dynamic traffic, and get more long - term rewards. DRL has strong adaptability but needs abundant

interaction data, plenty of computational resources, and it also has slow convergence and training instability problems [15].

Huang D et al. introduced a DRL based three dimensional approach for an autonomous robot to navigate and avoid obstacles within a three dimensional space. modeling planning as an MDP, and using DQN in simulation, learning optimal strategies in real time Reward function contains path length, energy usage and risk of colliding. In experiments, there is a 25% improvement in planning success rate as well as a decrease of 18% on the average path length in high-dimensional dynamic scenario, validating DRL is efficient for non-structured environment [15].

3.2.3. Hybrid path planning methods

Hybrid methods are made up of the best parts of traditional planning and deep learning. for instance, a simple approach is that, with deep learning predicting a global path, optimization provides a smooth, feasible local path [12]. It achieves a balance between accuracy, stability, and real-time performance, thus having a higher feasibility in complex environments. Ref [16].:

Pavel et al. combined deep learning and traditional planning to form a vision-driven autonomous driving framework. CNNs predict global paths; then optimize-based smoothing of local feasible paths. In the experiment, less movement paths and system response time reduces with 15% in urban traffic, system stability and security improved [16].

In summary, deep learning based path planning is better at dealing with dynamic and complex as well as uncertain environment. However, there are some limitations such as performance constrained by datasets, generalization capacity, and computational overhead. Safety has to be thoroughly verified before real-world application.

4. Current limitations and future prospects

Even though autonomous driving development has made advancements in target detection and path planning, practical adoption still comes with challenges like inadequate robustness, restricted ability to generalize, hefty computational cost, system safety limits. These challenges stop it from being widely and fully adopted.

4.1. Current limitations

4.1.1. Insufficient robustness in object detection

Object detection is related to driving safety. The YOLO series and transformer-based models get good accuracy for experiments, but they are still bad at complex traffic situations. And some situations like it's night time driving, or rain or snow of fog, occlusion, a lot of targets can cause a missed detection or false detect. and far or small targets detection needs to improve, high speed environment delays or wrong judgments, it has safety risk [6,9].

4.1.2. Limited generalization in path planning

The traditional algorithm like A* and RRT can work stable in static and structured environments, but they don't have flexibility and adaptability when the situation is dynamic or unstructured, so it's hard to react quickly to unexpected events [11]. Deep learning-based methods get better at adapting, but they depend a lot on the different kinds and how good the training material is. When the cases

differ a great deal from the training set, it may result in an output with a poor route or a dangerous route [14].

4.1.3. High computational cost and hardware requirements

Transformer Models and DRL methods need a lot of data and computing power. High accuracy and real time on embedded vehicle system is still tough. Even though have great hardware, inference latency and computational usage can limit how good the overall system is [9].

4.1.4. Limited system safety and explainability

Deep learning models usually work as a black box so it is tough to ascertain how they arrived at a decision. When there are misjudged or accidental occurrences, lacking traceability escalates regulatory trouble and diminishes public confidence [14]. End-to-end and multi-fusion have less standard safety verification which makes them even more difficult to apply.

4.2. Future development directions

4.2.1. Model lightweighting and efficiency

Using techniques like network pruning, quantization, knowledge distillation, and designing efficient architectures to keep the computational complexity and power consumption down but still getting the same detection accuracy and path planning. They make it so embedded car things can give answers in just tiny fractions of a second, which is close enough to real time for what people drive on.

4.2.2. Multi-modal perception and decision-making

Integration of vision, LiDAR and millimeter-wave radar, combined with high- definition maps improves robustness a lot. Multiple sensors will better the objects detecting, the path planning and more stable under circumstances such as night, fog etc [10].

4.2.3. Adaptive and online learning mechanisms

Vehicles can keep learning, improve online, and constantly change their mathematical model parameters to match changes in the environment and unpredictable car actions through continuous learning and online improvement. In this way it not only improves generalization, but makes less dependent on the offline large-scale annotations, more flexible.

4.2.4. Safety and explainability research

Future works will be more towards the interpretable detection and path finding model making sure that have traceable decisions [17]. Formal verification, redundancies, anomalies can improve a system's safety in an environment combined.

4.2.5. Interdisciplinary integration and standardization

To achieve autonomous driving development we need to merge it with traffic engineering, control science and ethics and also legal. Setting up common industry rules, experiments, and checks will

assist with regulation of the tech applications, risk reduction in operation, and making commercial deployment easier.

5. Conclusions

This paper focuses in the key technology of autonomous driving, focusing mainly on object detection and path planning. Object detection, this paper systematically introduced the above research on the single-stage method represented by the yolo series and the transformer-based object detection method, as well as the two-stage and multimodal detection method. The YOLO performs better when it comes to real-time but it is not so good and reliable on complex scenarios. Transformer-based models could acquire global information, but it cost much computation. Two-stage and multi-modal detect methods are high accurate but slow and expensive.

For the problem of path planning, the traditional algorithm has good determinacy and stability, but it is difficult to handle the dynamic environment. Based on deep learning method, it has strong adaptability, depends heavily on data and it is not flexible with generalization. In a nutshell, autonomous driving technology still has trade off between accuracy/real time and robustness.

Future work will mainly deal with model light weight, multi-modal sensors fusion, online learning, and system explainability. Those kinds of enhancements will lead to even greater safety, stability, and practicality for autonomous driving systems, meaning such systems will get used far more in the real world.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Wu, Y., & Cai, J. (2025). A review of 3D object detection methods based on deep learning for autonomous driving. Journal of Engineering Technology II · Information Science, Nanjing University of Aeronautics and Astronautics.

[2] Vijayakumar, A., & Vairavasundaram, S. (2024). YOLO-based object detection models: A review and its applications. Multimedia Tools and Applications, 83(35), 83535–83574.

[3] Ye, F., Zhang, S., Wang, P., & others. (2021). A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles. In 2021 IEEE Intelligent Vehicles Symposium (IV) (pp. 1073–1080). IEEE.

[4] Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets, and applications. Multimedia Tools and Applications, 82(6), 9243–9275.

[5] Hussain, M. (2023). YOLO-v1 to YOLO-v8: The rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines, 11(7), 677.

[6] Dong, H., Lin, S., Yu, Y. (2024). Research Progress of YOLO Detection Technology for Traffic Targets. Journal of Basic Science and Engineering Technology II · Information Science, Zhejiang University of Technology, Institute of intelligent Transportation Systems.

[7] Mao, Z., Zhu, J., Wu, X., & others. (2022). A review of object detection for autonomous driving based on YOLO. Journal of Computer Engineering & Applications, 58(15).

[8] Jing, R., Wu, L., & Zhang, K. (2023). Trajectory prediction for interactive perception in autonomous driving based on Transformer. Science Technology & Engineering, 23(26).

[9] Shah, S., & Tembhurne, J. (2023). Object detection using convolutional neural networks and Transformer-based models: A review. Journal of Electrical Systems and Information Technology, 10(1), 54.

[10] Arkin, E., Yadikar, N., Xu, X., & others. (2023). A survey: Object detection methods from CNN to Transformer. Multimedia Tools and Applications, 82(14), 21353–21383.

[11] Lin, Y. (2024). Dijkstra and A* algorithms in automated vehicle driving. In International Conference on Mechatronic Engineering and Artificial Intelligence (MEAI 2023) (Vol. 13071, pp. 584–590). SPIE.

[12] Chen, R., Hu, J., & Xu, W. (2022). An RRT-Dijkstra-based path planning strategy for autonomous vehicles. Applied Sciences, 12(23), 11982.

[13] Liu, B., Lan, Y., Huang, W., & others. (2024). UAV dynamic path planning algorithm based on dynamic window approach. Journal of System Simulation, 36(8), 1843.

[14] Ge, B., Tian, W., Xia, C., & others. (2024). Solving the capacitated vehicle routing problem based on end-to-end deep reinforcement learning. Application Research of Computers (Jisuanji Yingyong Yanjiu), 41(11).

[15] Huang, D., Jiang, C., & Han, K. (2020). 3D path planning algorithm based on deep reinforcement learning. Journal of Computer Engineering & Applications, 56(15).

[16] Pavel, M. I., Tan, S. Y., & Abdullah, A. (2022). Vision-based autonomous vehicle systems based on deep learning: A systematic literature review. Applied Sciences, 12(14), 6831.

[17] Huang, T., Xiang, M., & Shao, S. (2025). A review of road scene semantic segmentation based on Transformer. Journal of Basic Science and Engineering Technology II · Information Science, Southwest Minzu University.