ISSN 2616-5775 Vol. 8, Issue 7: 27-32, DOI: 10.25236/AJCIS.2025.080704

A Review of the Development of BEV Perception Algorithms for Autonomous Driving Applications

Wen Su[1,a], Guifang Guo[1,b]

[1] Xizang Minzu University, Xianyang, Shaanxi, 712082, China
[a]1622553823@qq.com, [b]379284030@qq.com

Abstract: *With the advancement of driver assistance technology and intelligent vehicle technology, navigation-assisted driving (NOA) technology has become a popular research and development and commercially sought-after technology hotspot in the automotive industry in order to make driving easier and more convenient when driving on urban roads.NOA is an automated driving technology that enables vehicles to drive on urban roads automatically with the help of high-precision navigation technology and automated driving technology, and its emergence is attributed to the development and progress of BEV technology solutions. The rise of NOA is due to the development and advancement of BEV technology, which can unify multimodal data into a single feature space in the automatic driving perception task module, and has better development potential than other perception learning models. This paper focuses on the application and importance of BEV perceptual algorithms in autonomous driving, and also analyses the challenges faced by BEV perceptual algorithms in complex and dynamic traffic environments.*

Keywords: *Autonomous Driving; BEV; Multimodal Fusion; Application*

1. Introduction

With the deepening awareness of environmental protection and the development of new energy vehicles, China has ushered in a once-in-a-lifetime opportunity to realise industrial overtaking with the help of new energy vehicles. Electric vehicles, known as one of the vehicles of the new century, are favoured by researchers at home and abroad because they have no seasonal emissions, no noise pollution, and do not require frequent maintenance[1]. As the world enters the era of new energy vehicles, autonomous driving will undoubtedly be the jewel of the world.

Compared to previous front and perspective views in the 2D domain, BEV (Bird's-Eye-View) has significant advantages in that it does not have the occlusion or scaling issues that are common in 2D tasks, and allows for better identification of occluded or intersecting vehicles. Representing objects or road elements in this form facilitates the development and deployment of subsequent modules. However,many BEV perception models currently face problems such as high computational overhead, high model complexity, and long inference time in practical applications, which affects their large-scale landing applications in real scenarios[2]. Therefore, it is necessary to explore more deeply how to optimise the performance of BEV models, reduce the computational cost, and improve the real-time and stability of the models in the research, which is of great significance to promote the commercial application of autonomous driving technology.

2. BEV Perception Algorithm

2.1 Fundamental Principle

BEV, refers to viewing the target object from a top-down perspective, like a bird looking down on ground objects from the air[3].The functional implementation of BEV is based on deep learning. Firstly, the powerful extraction ability of convolutional neural network(a diagram of the convolutional neural network architecture is shown in Fig. 1) is used to extract the point and cloud data information in the original picture and learn the spatial and semantic information of the target, and then the coordinates in the three-dimensional space are transformed to the two-dimensional plane coordinates in the bird's-eye-view perspective through the designed transformation matrix, so as to unify the data from different sensory sources; finally, the fused feature data are again subjected to convolutional processing, so as to obtain the environmental information required for the application of self-driving decision-making. Finally, the fused feature data is convolved again to obtain the environmental information needed for the application of automatic driving decision.

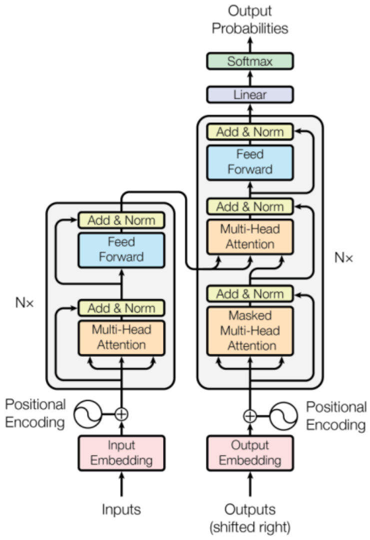

Figure 1 Convolutional Neural Network Architecture Diagram

In BEV space, time series information can be easily fused, and the algorithm can predict whether there is an object in the occluded area based on prior knowledge[4]. Although the 'brainstormed' objects are certainly 'imaginary', they are still of great benefit to the subsequent control module. Currently, typical BEV algorithms (typical BEV algorithms and technical features are shown in Fig. 2) are BEVFormer, BEVDepth, BEVFusion, LSS, and Occupancy Networks.

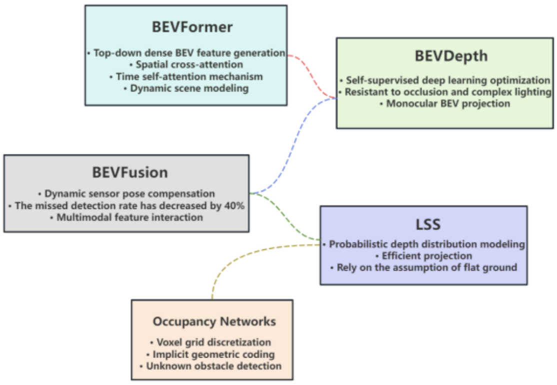
BEVFormer
• Top-down dense BEV feature generation
• Spatial cross-attention
• Time self-attention mechanism
• Dynamic scene modeling
BEVDepth
• Self-supervised deep learning optimization
• Resistant to occlusion and complex lighting
• Monocular BEV projection
BEVFusion
• Dynamic sensor pose compensation
• The missed detection rate has decreased by 40%
• Multimodal feature interaction
LSS
• Probabilistic depth distribution modeling
• Efficient projection
• Rely on the assumption of flat ground
Occupancy Networks
• Voxel grid discretization
• Implicit geometric coding
• Unknown obstacle detection

Figure 2 Typical BEV Algorithm

The BEV solution makes the perception of autonomous driving stronger, through the collection of data from various sensors, combined with 360-degree information of the entire vehicle surroundings[5], the situation around the vehicle will be displayed to the automatic driving system, so that the vehicle can know the road conditions, the location of obstacles, the status of pedestrians and the status of other cars, and ultimately provide a set of correct decision-making reference for the vehicle.

2.2 Multimodal Fusion Module

Multimodal fusion technology (the development process of multimodal fusion technology is shown in Figure 3) is the core part of BEV perception algorithm, which is one of the research hotspots of artificial intelligence, and its purpose is to fuse a single set of features when possessing multimodal information from several different sources, so that the complex world can be more fully and accurately felt and understood[6]. Multimodal fusion is mainly achieved by deep neural network fusion, convolutional neural network fusion, RNN fusion, Transformer fusion and other methods to achieve feature fusion and semantic fusion between the information, which mainly consists of preprocessing technology, feature extraction technology, multimodal fusion technology, decision-making technology. The main fusion methods about multimodal fusion technology include: early fusion method[7], late fusion method and hybrid fusion method.

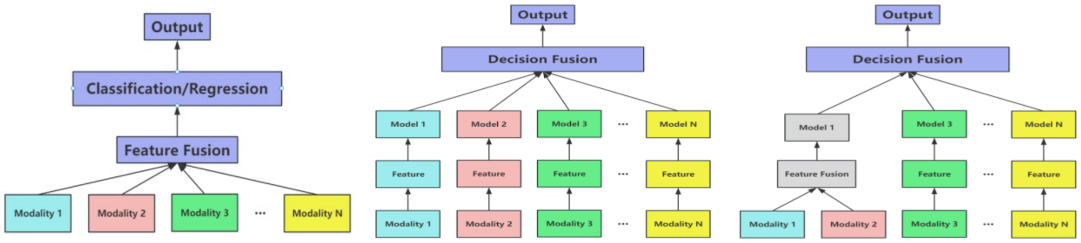

Figure 3 Multimodal Fusion Technology Development Process

(a) The early stage adopts the data-level fusion paradigm, which directly uses the original data of multiple modalities for unified modelling and maintains the original information connection between modalities, which is more effective for modalities with similar quality similarity between modalities, but has high feature alignment requirements and high computational costs.

(b) Late use of decision-level fusion paradigm based on each modality through the network for feature extraction and coarse decision-making[8], and later unified integration of decision-making, the modularity of integration is highly modular, expandable, and has advantages for the modal heterogeneity and limited resources, but the fusion approach will lose the fine inter-modal interaction information.

(c) Hybrid fusion method is a more advanced fusion method, the fusion architecture is hierarchical fusion, for the strong correlation modalities for the pre-fusion (mainly visual, language-related modalities), capture the fine inter-modal interaction information, for the weak correlation modalities are not processed (mainly text, audio-related modalities), and finally the gating or attention weights are used for decision-making fusion, which achieves a very good performance and efficiency in terms of computation. and performance has reached a good balance, which is the mainstream of current research.

2.3 Research Status

In the current BEV perception research, according to the different sensor types, it can be divided into three major directions: image-based BEV perception, LIDAR-based BEV perception and multimodal fusion-based BEV perception. Among them, image-based BEV perception, also known as the pure vision technology route, takes a single 2D monocular image or multiple 2D surround view images as inputs, and since the cost of the camera is much lower than that of LIDAR, this direction has become the frontier and hotspot of current international research[9]. In non-pure visual perception, LIDAR-based BEV perception takes point cloud data as input, and it is easier to learn robust and generalisable BEV feature representations from point clouds because the point clouds themselves have 3D characteristics. The fusion of image and point cloud data can take advantage of both the rich colour and texture features in the image, as well as the accurate depth information and geometric features in the point cloud, and has a wide range of applications in both depth estimation and BEV perception. BEV perception based on multimodal fusion receives both image and point cloud inputs, and is therefore higher in performance than the previous two classes of methods.

3. Application Overview

In autonomous driving, vehicles need the ability to make safe and effective decisions in a timely manner based on information about their surroundings[10]. A car driving on a busy city street must be able to detect pedestrians in front of it, vehicles on the side, and traffic lights in the distance, etc., to provide detailed and comprehensive environmental information for the car's next decision-making planning module, thus effectively reducing the probability of traffic accidents caused by self-driving cars.The main applications of BEV sensing technology in self-driving include target identification[11], tracking, path planning, road detection, and environment modelling. tracking, path planning, road detection and environment modelling.

3.1 Application of Target Recognition and Trajectory Tracking

The scene perception system based on the BEV perspective improves the speed of object detection and tracking a lot. Based on the deep learning method, the advanced target detection and tracking algorithm is applied, which can accurately detect the relevant moving vehicles, pedestrians, and traffic signs, etc. on the road; for the vehicle tracking, it calculates the vehicle in each frame of the image, and combines the current position of the vehicle with an algorithm to predict the next position of the vehicle, and then the next position of the vehicle. The algorithm is used to predict the position information of the vehicle at the next moment, and the prediction data is used as an important reference for traffic flow analysis and intelligent driving[12]. It is very helpful for road safety and road traffic efficiency.

3.2 Application of Route Planning and Decision Control

The use of BEV perception technology can greatly improve the safety of the vehicle in the process of road travel and the efficiency of the driver's travel, only need to use multiple high-precision sensors and various types of complex algorithms, you can do at any time to clearly understand the situation around you, coupled with the use of the calculation of the depth of the learning algorithms to make the corresponding driving probability of driving a safe route to make driving decisions at a millisecond level, with the speed of milliseconds. With the use of deep learning algorithms to make the appropriate driving decisions on safe driving routes that are likely to occur, it can respond to the driving situation in milliseconds by choosing the safest and most efficient way of driving to avoid accidents[13]. At the same time, automatic driving can also optimise the driving path to achieve the best way to avoid the phenomenon of traffic jams, and will not be blocked during peak hours resulting in a variety of traffic accidents, making the process of driving the road safer, more appropriate, but also to ensure that the overall road traffic efficiency is higher, bringing a better driving experience for the user.

3.3 Application of Road Detection

In traditional lane line detection methods, the use of colour queue value to simply detect yellow and white lane lines; the use of edge detection and Hough's transformation; the detection of traditional detection methods based on the fitting of the poor robustness of the detection method, for the human workload demand is greater. And the Hough transformation method cannot bend detection[14]. In recent years deep learning detection methods have been widely recognised in academia and industry, including: segmentation-based methods, anchor-based methods and parameter-based methods. Compared to traditional detection methods, deep learning detection methods do not require manual design of feature extraction rules, have strong generalisation ability, are adaptable and highly accurate .

3.4 Application of Environment Modelling

With the technical support of Telematics, BEV environment modelling can achieve information sharing and cooperative sensing among multiple vehicles, which is suitable for vehicle-road cooperative scenarios. The vehicle-road collaboration system consists of a roadside system, a vehicle system, a cloud system and a wireless communication device[15]. The roadside system includes sensors, car parks, road environment, weather, traffic conditions, routing, DSRC/5G/LTE_V, and intelligent decision-making platform and other components, which can detect road conditions and traffic signal signs and other information, and the data collected by roadside sensors are processed by BEV modelling and sent to the self-driving vehicle to provide it with more comprehensive information about the environment; the in-vehicle system consists of a V2X vehicle, multi-source heterogeneous sensors, and cloud system. The in-vehicle system is mainly composed of V2X vehicles, multi-source heterogeneous sensors and an intelligent decision-making platform, which is responsible for controlling the vehicle's own state information and sensing the surrounding environment; the cloud system is composed of an intelligent decision-making platform, a command centre, a cloud infrastructure, and a core intelligent network, which realizes the interaction between the road-side equipment and the in-vehicle unit.

4. Summary and Outlook

4.1 Summary

Perception technology as the core technology of automatic driving has always been the key direction of continuous innovation and deep optimisation by researchers, from 2D perception, 3D perception, multi-eye perception and perception based on the attention mechanism to the development of today's BEV perception, automatic driving perception technology has experienced a development process from local to global, from static to dynamic, and from a single to a fusion, and the eventual birth of BEV perception has become the key breakthrough in the field of technological innovation. BEV perception has become a key breakthrough in the field of technological innovation. Early perception technology mainly relies on 2D perception, using the camera to detect and identify the target object to get its coordinate position in two-dimensional space, to achieve simple classification and localisation of the object; followed by 3D perception using LIDAR to obtain 3D point cloud data of the surrounding environment and attempting to use multiple cameras to collect images from different angles of the multi-purpose perception technology came into being; the subsequent generation of perception technology based on the attention[16]. Subsequently, attention-based perception techniques have not fundamentally solved the problems of multi-sensor data fusion and global domain perception, and it was not until the emergence of the BEV perception algorithm that the automatic driving perception technology was completely revolutionised, because the BEV perception algorithm can unify the multi-modal data fusion into a feature space in the automatic driving perception task module, which has a better potential for development than other perception learning models.

4.2 Outlook

In response to the above inquiry, BEV perception can more accurately and efficiently handle more complex scenarios and large datasets based on the continuous improvement of deep learning and computational hardware during the development of the future parts. In the future, the focus may be on how to fuse different sensor data more efficiently; and how to combine BEV perception technology with Telematics to enable self-driving cars to communicate more accurately and safely with other vehicles around them or with traffic infrastructure and keep them working together[17]. Afterwards, when the technology matures, the relevant legal system will be gradually improved, so as to make the laws, regulations and standards related to autonomous driving clearer and, to a certain extent, regulate the behaviour of various companies and restrain the reckless development of the industry.

5. Conclusion

The BEV solution makes the perception of automatic driving stronger, through the collection of data from various sensors, combined with the entire 360-degree information around the vehicle, the situation around the vehicle will be displayed to the automatic driving system, so that the vehicle can know the road conditions, the location of obstacles, the status of pedestrians and the status of other cars,and ultimately provide a set of correct decision-making reference for the vehicle. Automatic driving BEV perception algorithm is one of the key technologies to achieve urban NOA[18], after inputting a large amount of inaccurate data can make the correct vehicle travelling decision and control instructions, so the advantages and disadvantages of the BEV perception algorithm will directly affect the good or bad of the automatic driving system, so it is vital to the research and development of BEV perception algorithm, which is important for the promotion of the development of China's automatic driving technology.

References

[1] Xu J ,Song C ,Shi C ,et al.UncertainBEV: Uncertainty-aware BEV fusion for roadside 3D object detection[J].Image and Vision Computing,2025(159):105567-105569.

[2] Chen J , Chen Z B, Tan G.Visual planimetry localisation based on BEV perception[J].Computer Science,2025,83(11):564-565.

[3] Liu X ,He Y ,Wang H ,et al.Research on BEVFormer-based Underwater Object Detection and Localization Model Trained with Migrated Dataset[J].Journal of Physics: Conference Series, 2024, 2891(03):32025-32025.

[4] Qiang Hu.Exploration of optimisation strategies for automotive automatic sensing systems[J]. Automotive Test Report,2024,25(17):152-154.

[5] Wang B L ,Yang J ,ZHANG L ,et al.Cooperative sensing method based on multi-sensor fusion[J]. Radar Journal,2024,13(01):87-96.

[6] Guo X L .Analysis of key technologies of perception system for self-driving cars[J].Special Purpose Vehicle,2023,53(10):60-62.

[7] Xu J ,Li Z Q .Optimisation study of perception system for self-driving cars[J].Automotive Test

Report,2023,19(12):22-24.

[8] Chen Y P.The future road of automatic driving perception system[J].Automotive Maintenance and Repair,2023,16(05):71-73.

[9] Ma K.The future direction of automatic driving from the perception system[J].Auto Zongxiang, 2023, 32(02):87-90.

[10] Chen Y P.Analysis of automatic driving perception system[J].Automotive Maintenance and Repair, 2022,104(15):74-75.

[11] Li Z, Wang W, Li H, et al. BEVFormer: Learning Bird's-Eye-View Representation From LiDAR-Camera Via Spatiotemporal Transformers.[J].IEEE transactions on pattern analysis and machine intelligence,2024,214(13):1388-1401.

[12] Liu X ,He Y ,Wang H ,et al.Research on BEVFormer-based Underwater Object Detection and Localization Model Trained with Migrated Dataset[J].Journal of Physics: Conference Series, 2024, 2891(3): 32025-32025.

[13] Huo R ,Chen J ,Zhang Y ,et al.3D skeleton aware driver behavior recognition framework for autonomous driving system[J].Neurocomputing,2025,5(613):128743-128743.

[14] Islam T ,Sheakh A M ,Jui N A ,et al.A review of cyber attacks on sensors and perception systems in autonomous vehicle[J].Journal of Economy and Technology,2023,17(10):1242-258.

[15] AfsharM ,Shirmohammadi Z ,Ghahramani A S ,et al.An Efficient Approach to Monocular Depth Estimation for Autonomous Vehicle Perception Systems[J].Sustainability,2023,15(11):7812-7815.

[16] Rana K ,Gupta G ,Vaidya P ,et al.The perception systems used in fully automated vehicles: a comparative analysis[J].Multimedia Tools and Applications,2023,12(31):1-23.

[17] Anders C ,Carl B ,Martin C ,et al.On Perception Safety Requirements and Multi Sensor Systems for Automated Driving Systems[J].SAE International Journal of Advances and Current Practices in Mobility,2020,2(06):3035-3043.

[18] Jiang J ,Wei C ,Cang S ,et al.Road Context-aware Intrusion Detection System for Autonomous Cars[J]. Future Transportation,2024,8(16):176-182.