Proceedings of the 4th International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.34046

Deep Learning in Autonomous Driving: Current Status, Challenges and Future Outlook

Zishen Wang

School of Information Science and Technology, Beijing University of Technology, Beijing, China
zishenw@outlook.com

Abstract. Recently, autonomous driving (AD) has become a popular technological frontier, mainly driven by the integration of Deep Learning (DL). This paper offers a thorough review of DL applications in the AD field. It analyzes how some core DL architectures are applied in AD, such as convolutional neural networks (CNNs), transformers, and end-to-end (E2E) models. These developments have helped move AD from theoretical research to real-world use. Even with notable progress, though, the way to fully autonomous driving (Level 5) still faces major challenges. The paper points out and talks about the main limitations of current systems, like difficulties generalizing in unusual situations (for example, bad weather), the "black-box" characteristics of neural networks that impact interpretability and safety certification, and computational limits that get in the way of real-time processing in embedded systems. The paper then brings these points together to provide a roadmap for future research, looking into trends such as Explainable AI (XAI) and neuromorphic computing to build more intelligent and reliable autonomous systems.

Keywords: Deep Learning, Autonomous Driving, decision control

1. Introduction

Recently, autonomous driving (AD) has emerged as a trending topic in both academia and the automotive industry [1]. People are aiming to develop a kind of autonomous, or self-driving vehicles, which can significantly minimize human drivers' efforts, enhance road safety, and optimize traffic efficiency. So far, the biggest advancement in the area has been the integration with artificial intelligence (AI). AI has solved the incompetence problem of traditional AD algorithms, especially in real-time perception and decision making [2]. To be precise, at the heart of this technological innovation lies deep learning (DL), a subset of AI. Due to its outstanding performance in processing tasks with large amounts of data and complex rules, DL has become the critical technology in AD [3].

Many studies pointed out the progress in AD achieved by DL technology in recent years. For example, convolutional neural networks (CNNs) act as the core algorithm in autonomous vehicles' perception modules, making considerable contributions to object detection, semantic segmentation, and depth estimation [4]. As technology continues to develop, other strategies like transformers also play a crucial role in object detection [5]. Meanwhile, sequential networks have enhanced sequence modeling for trajectory prediction and behavior planning [6]. Another fundamental technology is

DOI: 10.54254/2755-2721/2026.34046

end-to-end models. It can reduce error accumulation between modules, thus improving the model's efficiency [7]. These breakthroughs have promoted the industry from theoretical research to real-world deployment.

However, despite these achievements, the path to fully autonomous driving (Level 5) remains a challenge. One of the main challenges lies in the safety and reliability of autonomous vehicles in edge cases. Current deep learning models often face difficulties in generalizing across different weather conditions, such as rare lighting scenarios and other uncommon edge cases [8]. Additionally, deep neural networks (DNNs) have a "black-box" nature, which erfers to its lack of interpretability. This issue raises concerns about transparency of the reasoning process behind action, which is critical for safety certification and public trust [9]. Moreover, the high computational complexity of deep learning models limits the real-time processing performance and energy efficiency of autonomous driving systems, typically in embedded devices [10].

The main contribution of this article is to present a comprehensive review of DL applications in AD and analyze both the current landscape and unresolved issues. With this purpose in mind, this paper begins by analyzing the core architectures of perception, decision-making and control systems. After that, the paper identifies key challenges of the existing AD system. Finally, the paper explores emerging trends such as explainable AI, neuromorphic computing, offering insights into the future orientation of this field. By integrating these elements, the study aims to not only popularize these frontier technologies and products, but also indicate some possible future improvements for researchers to work towards more intelligent autonomous systems.

2. Background

2.1. Deep Learning

Deep Learning (DL) is a core subfield of machine learning, which aims to let machines learn from data and to be as intelligent as humans. DL is based on deep neural networks and developed on the basis of algorithmic models such as artificial neural networks and statistical machine learning. As for neural networks, from a micro view, it belongs to the category of connectionism. The main proposition of this school of thought is attaining artificial intelligence by simulating the mechanism of the human brain, or how neuron works. Scientists' study on this topic has lasted for a long time since the last century. After scientists had completely figured out the working mechanism of neurons, in the 1940s, expert F.Rosenblatt first introduced the perceptron model, the simplest neural network composed of only one neuron. In the 1980s, Hinton et al. proposed the back-propagation(BP) algorithm and multi-layer perceptron (MLP) model based on the sigmoid activation function. Since a neural network is fully connected, it requires powerful computing power. Besides, its training process requires massive amounts of data. These aspects have blocked neural networks from implementation until the development of high-performance GPUs and the emergence of ImageNet, a large-scale annotated dataset in the 2010s. From then on, neural networks and DL usher in a period of rapid growth.

The most important feature of DL is its ability to extract features automatically. These extracted features are also called deep features. Contract with manually designed features, deep features are more robust. Moreover, DL gets rid of the reliance on manual feature engineering and can process raw, unstructured data. These advantages enable DL to surpass traditional methods by a wide margin when handling complex tasks [11]. The fundamental DL technology in AD is convolutional neural networks (CNNs). CNNs extract local features through convolutional layers, pooling layers, and fully connected layers, leveraging structural advantages such as parameter sharing and local

DOI: 10.54254/2755-2721/2026.34046

connectivity. This enables efficient processing of grid-based data (e.g., images and speech) without relying on additional feature engineering.

2.2. Autonomous Driving

Over the past decade, the number of traffic accidents worldwide has risen significantly. According to the latest report from the World Health Organization, the eighth leading cause of death globally goes to traffic accidents. Approximately 1.3 million lives died in traffic accidents every year and 20 to 50 million of them are non-fatal injuries. In order to reduce traffic accidents and enhance road safety, the scientific community and the automotive industry are committed to developing some innovative technologies. Among the innovative technologies, AD system takes place [1].

The concept of autonomous driving can date back to 1925, when Francis Houdini invented a prototype vehicle controlled remotely. The development of AD has gone through several phases. The eureka project PROMETHEUS, conducted from 1987 to 1995, was one of Europe's earliest large-scale autonomous driving research initiatives, it paved the way for Daimler-Benz on highways. After that, the "Grand Challenge" and "Urban Challenge" competitions launched by defense advanced research projects agency (DARPA) spurred technological innovation. Although no team completed the 2004 race, five teams successfully finished the off-road course in 2005, and six teams completed the ADS challenge in an urban environment in 2007. Additionally, autonomous racing platforms (such as F1/10, Formula Student) provide platforms for technological improvement.

To standardize this field, the Society of Automotive Engineers (SAE) established the "SAE J3016" standard, classifying levels of driving automation into six tiers ranging from Level 0 (fully manual) to Level 5 (fully automated) [12] shown in Figure 1. Currently, the scientific and industrial communities are focusing on Level 5 automation, because it refers to fully driverless operation and is viewed as a key milestone in technological progress.

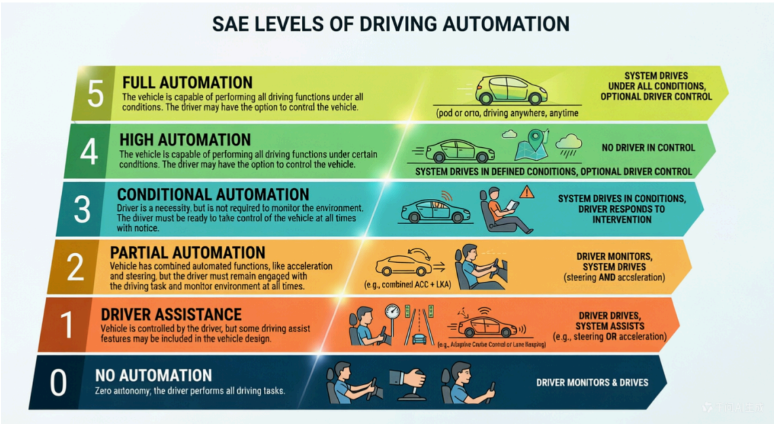

Figure 1. SAE 6 levels of driving automation [12]

3. Applications for Deep Learning in Autonomous Driving

3.1. Perception

3.1.1. Object detection

In order to build AD system, one of the primary abilities a vehicle should be equipped with is recognizing the vehicle's situation. For example, vehicles should obtain data such as amount, shape,

DOI: 10.54254/2755-2721/2026.34046

position and distance of surrounding object, otherwise the car is "blind" and unable to make choice for next movement. Moreover, vehicles should recognize traffic lights, traffic signs, crosswalk, lane marking, etc shown in Figure 2. In complex urban scenario. These small and delicate signs raised the bar of AD for object detection ability.

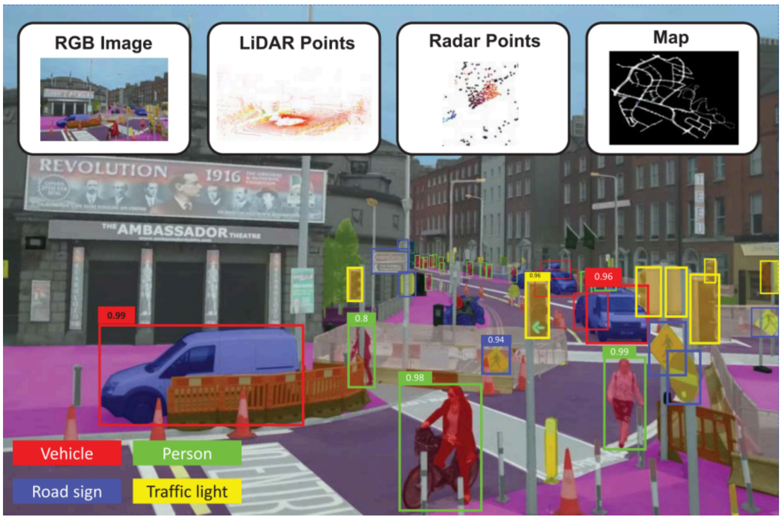

Figure 2. The driver-less car uses different signals to detect objects, such as RGB camera images, Radar points, LiDAR points, and map information [4]

Therefore, the fundamental task of object detection of AD is identifying and localizing multiple objects within a scene. Typically, models identify object categories by calculating classification probabilities and use bounding boxes to determine their specific locations. Object detection algorithm often uses DL methods based on CNN. These methods have established performance benchmarks on mainstream datasets. Recently, mainstream DL object detection networks mainly follow two workflows: two-stage detection or one-stage detection [4].

In two-stage object detection, the detection mode is divided into two stages. In the first stage, the algorithm generates a series of object candidates according to the vehicle's scene, which are referred to Regions of Interest (ROI) or Region Proposals (RP). In the second stage, put these candidate regions into further verification and classification, and optimize them based on classification scores and location information. As for specific applications, OverFeat and R-CNN were pioneering works in the field. In respective approaches, candidates were first generated using a sliding window method or selective search, then fed into a regional CNN for feature extraction to perform object classification and bounding box regression. Fast R-CNN gain improvements by applying a larger CNN to the entire image to generate global feature maps, and then directly extracting features for individual regions. By introducing a small fully connected network called region proposal network (RPN), Faster R-CNN standardized the detection workflow. RPN replaced the selective search process in R-CNN to generate bounding box candidates. It slides over high-level CNN feature maps to directly generate high-quality ROIs.

On the other hand, One-stage object detection directly conducts intensive sampling on the input image, and the data transfer through single network. It can directly map feature maps to bounding boxes and classification scores using a unified CNN model. MultiBox is an early attempt at one-stage object detection. It predicts a binary mask from the entire input image and infers bounding boxes in subsequent steps. YOLO (You Only Look Once) presents a more mature approach. It's a unified detector that directly regresses bounding boxes from a CNN model. Single shot multiBox

DOI: 10.54254/2755-2721/2026.34046

detector (SSD), on the other hand, uses small convolutional filters to perform regression on feature maps at multiple resolutions, which enables the prediction of bounding boxes at different scales and handling objects of varying sizes. Comparing to two-stage object detection, one-stage approach sacrifices detection accuracy. In return, the model complexity decreases and reasoning speed accelerates.

3.1.2. Segmentation

Scene analysis is a key step in AD system. It constructs a complete picture of the vehicle's surroundings and helps vehicles to detect objects. Scene analysis is done by aggregating the contextual information extracted through scene understanding. There are various information forms and sources that can be used for AD's scene understanding. Among them, visual data captured by visual sensors is regarded as the most reliable. Thus, an AD system needs computer vision (CV) technology to process visual figures. From a machine learning perspective, visual data is typically processed by deep convolutional neural networks. Before, these information streams were extracted in isolation through separate CV algorithms. Recently, these algorithms have been replaced by segmentation mechanisms based on CNN. Segmentation is a mechanism that delineates the boundaries of various object types and assigns different colors to each pixel that belongs to a distinct object. Pixel labeling can be divided into semantic segmentation and instance segmentation. Instance-level segmentation assigns a different color to each object, even within the same category. For example, it would label different people with different colors. In contrast, pixel-level semantic segmentation assigns the same color label to objects of same categories, meaning different people share the same color.

Specifically, there are three representative semantic segmentation approaches, which are fully convolutional networks (FCNs), deep fully convolutional neural network architectures (known as SegNet), and the DeepLab approach. Each network employs different mechanisms. As for FCN, it adopts an encoder-decoder structure to extract deep discriminative features for subsequent instance localization and segmentation tasks. The encoder consists of standard convolutional layers and down-sampling layers used in CNN classification problems. The decoder uses transposed convolutional layers to up-sample the coarse feature maps from the architecture's bottleneck layer. The up-sampling process bypasses the traditional final layer output, instead it passes through transposed convolutional layers, which helps generate a prediction map with the same size as the input frame. On the other hand, SegNet is built on a series of deconvolutional layers. It changes the extracted features into class-scoring prediction maps as output with same frame size as input. The SegNet network comprises two functional modules. The first uses a CNN to extract features from the input frame, the second constructs a prediction map with class scores through a series of transposed convolutions and unpooling layers, and produces instance-level segmentation results at last. The DeepLab approach for semantic segmentation utilizes convolutional layers with up-sampling filters known as atrous convolutions. To obtain a prediction map of the same size as the input frame, DeepLab combine atrous convolutions with bilinear interpolation. Currently, a large number of latest semantic segmentation methods for visual scene understanding have also achieved impressive results [13].

3.1.3. Attention

Some other methods of AD's detection module also use attention model, such as transformer structure. Although CNNs have long been the workhorses of image processing, transformer has

DOI: 10.54254/2755-2721/2026.34046

showed the potential to outperform CNNs in 3D object detection. Transformer is originally designed for natural language processing using the self-attention mechanism. Self-attention does not distinguish between text and images as input, but focuses merely on the relationships between elements. In image or 3D detection, images can be divided into small patches or points in a point cloud, which can be treated as sequential elements. Without being constrained by local convolutional kernels as CNNs are, the transformer extracts feature by calculating the correlations between these elements. Consequently, the transformer is gradually surpassing CNNs in 3D detection in several aspects [5].

First, attention mechanism has a wider field of view. In CNN cases, to view the whole picture, many layers must be stacked, leading to high computational costs. In AD, this means that CNNs struggle to directly understand the complex spatial relationships of a "whole intersection". In contrast, an attention mechanism allows any two points in the network to interact directly. The transformer can instantly capture their relationship no matter how far apart two objects are in the image. This feature of attention is crucial for understanding the spatial layout of 3D scenes. Additionally, the attention mechanism has flexible feature interactions. In a transformer, there is a concept of "Queries". Models such as DETR3D and BEVFormer generate a set of "object queries". These queries can actively "search" for matching information within image features, enabling flexible Query-Key-Value interactions. While feature fusion in CNNs is typically fixed, it only supports simple addition or concatenation. Attention mechanisms also possess an advantage in handling heterogeneous data. Autonomous driving requires simultaneous processing of camera images (2D) and LiDAR point clouds (3D), yet CNNs struggle to directly fuse these two types of data, since they have completely different structures. While transformers can map data from different sources into a single "feature space", for example image features, point cloud features, and even radar data. Transformers can then perform deep fusion through attention mechanisms [5].

3.2. Decision-making and control systems

3.2.1. Trajectory prediction

Trajectory prediction shown in Table 1 is the task of forecasting the future positions, velocities, and directions of moving agents based on their past or current motion. DL methods have been a prominent technology in trajectory prediction. It surpassed traditional prediction techniques by effectively handling complex scenarios, long-term predictions, and crucially, the interaction factors between traffic participants. There are several different methods based on DL [6].

1) Sequential networks. Sequential networks are fundamental because they can extract features from historical trajectory data. Different kinds of sequential networks have been used in trajectory prediction. RNNs, particularly long short-term memory (LSTM) and gated recurrent units (GRU), are able to process temporal information and handle vanishing gradient problems in trajectory predictions. They are widely used in maneuver-based prediction and encoder-decoder frameworks. CNNs excel at processing spatial information. It often use rasterized bird's-eye-view images or semantic maps. Hybrid models combining CNNs for spatial feature extraction are common to capture both scene context and time-series data. Moreover, by mimicking human focus, Attention mechanisms allow for parallel computation and achieve superior performance in long-term prediction. It can handle missing data by filtering high-value information from large datasets.

2) Graph neural networks (GNNs). GNNs are suitable for non-Euclidean traffic data and modeling complex interactions. Unlike regular grids, traffic scenarios can be represented as irregular graphs, in which nodes denote vehicles or map elements, edges denote spatial or temporal

DOI: 10.54254/2755-2721/2026.34046

relationships. There are some other advanced networks based on GNNs. Graph convolutional networks (GCNs) extend convolution to graph data, and aggregates neighbor features to learn interaction-aware representations. Graph attention networks (GATs) further enhance this by assigning adaptive weights to different neighbors based on their importance.

3) Generative adversarial networks(GANs). In true applications for AD, multiple plausible futures exist. To address the uncertainty of future trajectories, generative models are employed. GANs use a generator to create trajectories and a discriminator to evaluate their realism, capturing social interactions among pedestrians or vehicles. Conditional variational autoencoders (CVAEs) improves basic auto encoders. It learns from probabilistic latent space, enabling the generation of diverse and feasible trajectory samples rather than simply memorizing data.

Table 1. Different DL methods in trajectory prediction [6]

Classification	Methods	Year	State Encoder	Context Encoder	Interaction Module	Decoder	Description
RNN	MFP	2019	RNN	CNN	Radial Basis Function	RNN	Learn latent variables to model the multimodel trajectories
CNN	CoverNet	2020	CNN	CNN	-	Trajectory Set Generator	Apply the raster image as input
CNN	HOME	2021	1D-CNN, GRU	CNN	Self-Attention	CNN	Output 2D topview heatmap
CNN	TPCN	2021	PointNet++		Joint Learning	Displacement Prediction	Use point cloud learning-based methods
CNN and RNN	DESIRE	2017	GRU	CNN	Social Pooling	GRU	Use deep IOC framework to encode
CNN and RNN	CS-LSTM	2018	LSTM	-	Social Pooling	LSTM	Six LSTM decoders to generate distributions of six specific maneuvers
Attention Mechanism	MHA-JAM	2021	LSTM	CNN	Attention Head	LSTM	Each attention head to generate a distinct future trajectory to address multimodality
Attention Mechanism	mm Transformer	2021	Transformer	VectorNet	Transformer	MLP	Stacked Transformers to refine a set of fixed proposals
GNN	VectorNet	2020	PointNet		GNN	MLP	Operate on the vectorized HD maps and trajectories
GNN	DenseTNT	2021		VectorNet		Goal Set Predictor	Directly output a set of trajectories from dense goal candidates
Generative Model	TS-GAN	2020	LSTM	-	Auto-Encoder Social Convolution	LSTM	Incorporate GAN into modeling spatial and temporal information

DOI: 10.54254/2755-2721/2026.34046

Table 1. (continued)

Generative Model	PRIME	2021	1D-CNN, LSTM	LSTM	Self-Attention	Model-based Generator	Model-based generator and learning-based evaluator

3.2.2. End-to-end Autonomous Driving

Current AD systems mainly use a modular pipeline consisting of interconnected modules for perception. While this approach has some significant drawbacks. For instance, errors like misclassification may occur in early stages, potentially leading to unsafe behaviors. Besides, each module is trained separately for specific outputs, which can leads to redundant computations. To cope with this issue, end-to-end(E2E) driving emerged. E2E aims to avoid these issues by optimizing the entire pipeline as a single learning task. Instead of discrete steps, the system learns a direct mapping from raw sensor data (inputs) to vehicle maneuvers (outputs). During the input section, the model processes raw data from cameras, LiDAR, GPS, or vehicle dynamics. In the output section, E2E models generate various types of outputs like waypoints, cost function, direct control and acceleration to control the vehicle and assist output. There are two main learning methods for the E2E model. The first is imitation learning (IL), and the second is reinforcement learning (RL).

IL is a learning method involving learning from expert demonstrations. One way is behavioral cloning (BC). It's a supervised learning approach where the model minimizes the difference between its actions and the expert's actions. While this method has a drawback regarding distribution Shift, which means the model fails in scenarios not covered by the training data because it hasn't learned the underlying causality. To address this issue, direct policy learning (DPL) collects additional data where the policy fails, which improves the policy but requires continuous expert access. Instead of copying actions, inverse reinforcement learning (IRL) tries to learn the reward function underlying the expert's behavior. This helps in understanding why an expert takes a certain action.

On the other hand, RL aims to maximize cumulative rewards over time without expert data. The overall mechanism of RL is to interact with the environment and receives rewards for safe driving and penalties for collisions. In the field of RL, there are algorithms including value-based (DQN), actor-critic (DDPG, PPO), and maximum entropy models (SAC). Recent trends combine RL with human expertise, which can ensure safer exploration.

4. Discussion

4.1. Challenges

Despite significant progress, several critical challenges continue to impede the deployment of fully AD systems powered by DL. According to researches [8, 9, 14], these challenges are listed as follows.

1) Applicability. Current DL models for AD are typically trained within specific designed domains, such as clear-weather roads or highway environments. While in some rare edge scenarios, like adverse weather conditions (e.g., heavy rain, snow, fog, unusual lighting), AD's performance substantially degrades. This lack of robust generalization is a fundamental barrier for AD systems now to achieve fully automated standard, since Level 5 autonomy requires reliable operation across

DOI: 10.54254/2755-2721/2026.34046

all environments and conditions. Improvements in AD system to address this challenge concerning advancements in such as domain adaptation and continual learning. The development of more diverse and representative datasets also presents a way [8].

2) Interpretability. Deep neural networks are often perceived as "black boxes", meaning that it's unable to explore how the internal system works from outside. In AD, it's difficult to understand the reasoning behind the decisions. This lack of interpretability poses a threat to safety certification. That's because in case the system occurs a mistake like the vehicle makes an unexpected maneuver, engineers and regulators need to understand why, otherwise the reasoning of AD system is untrustworthy. Current methods for model interpretability are often insufficient for establishing causal reasoning. The absence of transparent decision-making processes undermines public trust, and can create ambiguity in traffic accidents. Developing an explainable AI (XAI) frameworks provide a solution, it may let humans understand the justifications in real time [9].

3) Computational constraints and real-time processing. DL models, particularly large transformer-based architectures, demand substantial computational resources. The resources include high-end GPUs and broad memory bandwidth. Deploying these models on embedded automotive platforms is struggling because embedded systems have limited computation powers. Thermal, strict power, and latency constraints in embedded chips remain a major challenge. Also, to ensure safe vehicle operation, AD requires real-time processing. Perception and planning systems must operate at some frequency stringently. Some solutions like model compression techniques, such as pruning and quantization often come at the cost of low accuracy. Therefore, striking the optimal balance between model performance and computational efficiency is an ongoing area for researches [14].

4.2. Future prospects

Looking ahead, several emerging trends and technologies hold the potential to address current limitations and accelerate progress toward fully AD.

Explainable AI (XAI) for Autonomous Driving. The integration of XAI techniques into AD systems is expected to be critical in gaining public trust and getting regulatory approval [9]. It's necessary for future systems to incorporate natural language explanations. It may describe the vehicle's reasoning process to passengers and safety operators, or structured causal models that provide analysis of driving decisions. Passengers can then get evidence of the autonomous vehicle's reason for its action. To build inherently interpretable AD systems, approaches combining the pattern recognition capabilities of neural networks with the logical reasoning of symbolic AI will be promising, like concept-based explanations and neuro-symbolic.

Neuromorphic Computing. Neuromorphic computing architectures, which mimic the structure and function of biological neural networks using event-driven spiking neural networks (SNNs), offer the potential for dramatically reduced power consumption and latency compared to conventional GPU-based processing. These architectures are particularly well-suited for processing data from event cameras, a sensor modality that captures pixel-level brightness changes asynchronously rather than full frames at fixed intervals. Therefore, the combination of neuromorphic hardware and event-driven perception could enable surprisingly low latency and efficiency in energy acquiring, which is superior in resource-constrained embedded platforms.

5. Conclusion

So far, the integration of DL has propelled the AD industry from theoretical concepts to tangible reality in many ways. In AD's perception model, DL beard crucial responsibility. Models based on

DOI: 10.54254/2755-2721/2026.34046

CNNs has become the "eye" of autonomous vehicles, achieving great standard in detecting surroundings. Segmentation technology played an important role, helping vehicles identify different objects. In AD's control and decision-making model, Attention and other latest technologies have made AD systems more powerful, and eventually fueled its realization. These achievements is generally obtained on account of DL's outstanding qualities. By automatically extracting deep features from raw data, DL can deal with real-time perception and decision-making jobs. While standard CNNs and advanced architectures like transformers have enhanced the system's ability to understand 3D spatial layouts and heterogeneous sensor data. The shift towards end-to-end learning and the utilization of high-fidelity simulations have streamlined development and improved driving strategies.

However, there are also many improvements that need to be made in this field. Current systems are not yet robust enough to handle the infinite variability of the real world, like in rare edge cases and adverse conditions. Besides, the lack of transparency in deep neural networks also poses a barrier to legal approval and public trust. The "black-box" nature in decision-making process is insufficient in strict safety applications.

Looking forward, the future aspects of this field are dynamic. Although different levels of automatic functions have been added into driving systems of vehicles, and cars with autonomous driving abilities have gone to mass production, fully autonomous vehicles haven't gone into use yet. That's partly because of the technical barrier to more efficient hardware and more advanced algorithms. To further develop intelligent algorithms, there may be improvements including but not limited to research to XAI and adoption of neuromorphic computing. By considering together these potential means, researchers and practitioners can pave the way for the next generation of AD systems that can not only be fully autonomous but also become safer and trustworthy.

References

[1] Yurtsever, E., Lambert, J., Carballo, A., & Takeda, K. (2020). A survey of autonomous driving: Common practices and emerging technologies. IEEE access, 8, 58443-58469.

[2] Sahoo, L., & Varadarajan, V. (2025). Deep learning for autonomous driving systems: technological innovations, strategic implementations, and business implications - a comprehensive review. Complex Engineering Systems.

[3] Huang, Y., & Chen, Y. (2020). Autonomous driving with deep learning: A survey of state-of-art technologies. arXiv preprint arXiv: 2006.06091.

[4] Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., ... & Dietmayer, K. (2020). Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. IEEE Transactions on Intelligent Transportation Systems, 22(3), 1341-1360.

[5] Mao, J., Shi, S., Wang, X., & Li, H. (2023). 3D object detection for autonomous driving: A comprehensive survey. International Journal of Computer Vision, 131(8), 1909-1963.

[6] Huang, Y., Du, J., Yang, Z., Zhou, Z., Zhang, L., & Chen, H. (2022). A survey on trajectory-prediction methods for autonomous driving. IEEE transactions on intelligent vehicles, 7(3), 652-674.

[7] Chib, P. S., & 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.

[8] Rahmani, S., Rieder, S., de Gelder, E., Sonntag, M., Mallada, J. L., Kalisvaart, S., ... & Calvert, S. C. (2024). A systematic review of edge case detection in automated driving: Methods, challenges and future directions. arXiv preprint arXiv: 2410.08491.

[9] Atakishiyev, S., Salameh, M., Yao, H., & Goebel, R. (2024). Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions. IEEE Access, 12, 101603-101625.

[10] Barabas, I., Todoruţ, A., Cordoş, N., & Molea, A. (2017). Current challenges in autonomous driving. In IOP conference series: materials science and engineering (Vol. 252, No. 1, p. 012096). IOP Publishing.

[11] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

DOI: 10.54254/2755-2721/2026.34046

[12] Taxonomy, S. (2016). Definitions for terms related to driving automation systems for on-road motor vehicles (j3016). Soc. Automot. Eng., Warrendale, PA, USA, Tech. Rep. J3016_201806.
[13] Muhammad, K., Hussain, T., Ullah, H., Del Ser, J., Rezaei, M., Kumar, N., ... & De Albuquerque, V. H. C. (2022). Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks. IEEE Transactions on Intelligent Transportation Systems, 23(12), 22694-22715.
[14] Muhammad, K., Ullah, A., Lloret, J., Del Ser, J., & De Albuquerque, V. H. C. (2020). Deep learning for safe autonomous driving: Current challenges and future directions. IEEE Transactions on Intelligent Transportation Systems, 22(7), 4316-4336.