

Application of Deep Learning in Autonomous Driving

Mingkai Yang

College of International Collaboration, Dalian Maritime University, Dalian, China
yangmingkai@dlnu.edu.cn

Abstract: Autonomous driving technology represents a critical area of innovation within the contemporary technological landscape. Its implications extend beyond the realm of transportation, significantly contributing to economic growth and the boarder process of social development. Mature autonomous driving technologies can significantly lower traffic accident rates, enhancing safety for pedestrians and drivers. Deep learning is a key component of modern AI research, with diverse applications in image processing. Recent advancements in deep learning have enabled its integration with autonomous driving systems. This paper reviews the theoretical and practical applications of deep learning in autonomous driving, presenting an overview of current vehicle applications and a detailed discussion of deep learning concepts and their evolution. It examines key algorithms and models used in this field, including Convolutional Neural Networks, Recurrent Neural Networks, and Deep Reinforcement Learning, summarizing their applications and relevant details. Lastly, the paper addresses future challenges for deep learning in autonomous driving and suggests potential development trends, aiming to provide insights into the current landscape and future directions of deep learning in vehicle technologies.

Keywords: Autonomous driving technology, Deep learning, Convolutional neural network, Recurrent neural network, Deep reinforcement learning

1. Introduction

Autonomous driving technology integrates AI, computer vision, sensor technology, and big data, serving as a key driver of innovation. Recent research emphasizes enhancing environmental perception through advanced sensors and AI algorithms for real-time traffic analysis. Efforts also focus on developing robust decision-making frameworks to ensure safety and reliability in diverse scenarios. However, perception capabilities are limited in complex situations due to algorithm generalization issues, impacting system stability. Ethical decision-making remains crucial, especially in potential accident scenarios. This paper employs a literature review to summarize advancements in deep learning techniques for autonomous driving and to outline future development trajectories. It serves as a comprehensive resource for researchers and industry professionals, facilitating a deeper understanding of deep learning applications and advancements in autonomous driving technologies while enhancing the theoretical knowledge base in the field.

2. Theoretical Basis of Deep Learning

The article investigates the theoretical foundations of deep learning by defining core concepts, analyzing key algorithms and models, and summarizing essential components. Prior to delving into

deep learning, it is crucial to establish a foundational understanding of machine learning (ML), a vital subfield of artificial intelligence that has become synonymous with AI, highlighting its pivotal role in enhancing intelligent systems' capabilities and applications. Partial researchers admit that ML is the process by which machines learn patterns from large amounts of historical data through algorithms, enabling them to intelligently recognize new samples or make predictions about the future [1].

Many scholars argue that Professor Geoffrey Hinton's 2006 work [2] catalyzed deep learning research in academia and industry. The article posits that deep neural networks possess remarkable feature learning abilities, effectively tackling training challenges via layer-wise initialization strategies, including unsupervised learning. This method fosters a more fundamental data representation, improving their applicability in visualization and classification tasks. After 2010, with the title of Hubel-Wiesel [3] and other models, the human brain's cerebral nervous system was shown to have a rich hierarchical structure. This inspired relevant researchers to explore the principles of deep learning from a bionic perspective. Until today, deep learning research in fields other than bionics is still in its infancy, but still shows a high degree of applicability.

The limitations of shallow learning networks have been well-documented in the literature [4], with evidence suggesting their inability to effectively represent a significant class of mathematical functions. This highlights the inherent constraints of certain implicit mechanical learning models.

3. Main Algorithms and Models

The development of numerous deep learning models has been driven by the cumulative efforts of multiple generations of researchers and practitioners within the industry. This article aims to present an overview of three seminal models that have significantly shaped the field: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Deep Reinforcement Learning (DRL).

4. Convolutional Neural Networks (CNN)

As early as 1995, Lecun and his students started to publish work [5] on CNN. Subsequently, Lecun et al. designed and trained LeNet-5 [6] based on BP algorithm. As a classical CNN structure, it has served as the foundation for numerous subsequent advancements, demonstrating significant success, particularly in the domain of pattern recognition.

After combining with bionics, the basic structure of the CNN was finalized. It contains input layer, convolutional layer, pooling layer, fully connected layer and output layer. In most cases, convolutional layers and pooling layers are implemented in multiple iterations within the network architecture (Figure 1).

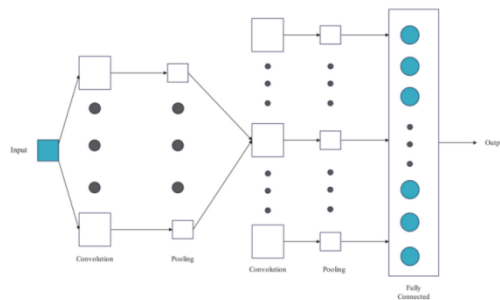


Figure 1: The network structure of Convolution Neural Network

The convolutional layer consists of multiple Feature Map, each consisting of multiple neurons. Usually, the convolutional layer is considered as the core component of CNNs. The size (i.e., number

of neurons) N of each output Feature Map of each corresponding convolutional layer in CNNs satisfies the following relationship [7]

$$N = \left(\frac{(x-y)}{z} + 1 \right) \quad (1)$$

Where: x denotes the size of each input Feature Map, y denotes the size of the convolution kernel, and z denotes the sliding step of the convolution kernel in its previous layer.

In traditional CNN structure, the output of the convolution operation usually adopts saturating nonlinear functions such as sigmoid, tanh and non-saturating nonlinear functions like ReLU. The formula for the ReLU function is shown below [8]

$$R(x) = \max(0, x) \quad (2)$$

ReLU and analogous activation functions effectively address vanishing gradients and gradient explosion, improving the stability and efficiency of deep neural networks. The pooling layer, subsequent to the convolutional layer, comprises multiple feature maps and performs secondary feature extraction through max and average pooling. Overall, CNNs, with their local connectivity and pooling operations, utilize fewer connections and parameters than conventional pattern recognition algorithms, leading to enhanced training performance and efficacy. In autonomous driving, CNNs are widely employed for lane recognition and scene classification.

4.1. Recurrent Neural Network (RNN)

RNNs enhance traditional feed-forward neural networks, creating a category of self-recurrent networks in deep learning. A standard RNN comprises an input layer, output layer, and a hidden layer with recurrent connections, where activation states are influenced by previous states, making them suitable for time-series tasks. The Long Short-Term Memory (LSTM) network, developed by Hochreiter and Schmidhuber, is a key architecture in this field, demonstrating outstanding training performance when combined with gradient-based learning algorithms.

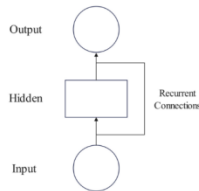


Figure 2: The network structure of Recurrent Neural Network

If an input sequence $x = (x_1, x_2, \dots, x_T)$ is given in the RNN network, the recurrent update equations of RNNs can be expressed as follows:

$$h_t = g(Wx_t + Uh_{t-1}). \quad (3)$$

Where: g represents an activation function (e.g., sigmoid or tanh), W denotes the weight matrix associated with the current input, and U represents the weight matrix governing the transition from the previous time step to the current one.

In autonomous driving technology fields, making accurate decisions in real time based on the dynamically changing environmental information is critical. Consequently, RNNs are frequently employed in such scenarios to extract and store relevant information, enabling a more comprehensive understanding of the environmental state.

4.2. Deep Reinforcement Learning (DRL)

Deep reinforcement learning represents the integration of advancements from the fields of deep learning and reinforcement learning. The relationship among the three is illustrated in Figure 3.

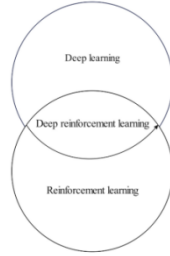


Figure 3: Relationship among DL, RL and DRL

Reinforcement learning employs a "trial and error" approach for optimizing action strategies via environmental interaction. While adept at decision-making, it struggles with complex feature learning. Conversely, deep learning excels in abstract feature extraction but is less effective in decision-making. Deep reinforcement learning integrates both methodologies, providing a robust framework for complex decision-making challenges. It leverages state and observation data, processes it through deep reinforcement learning algorithms, and generates corresponding actions and their values. Presently, most deep reinforcement learning algorithms are value-based.

Mnih et al. [10] introduced the Deep Q-Network (DQN) by integrating CNNs with reinforcement learning.

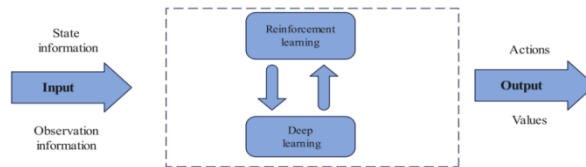


Figure 4: The diagram of DRL

In DQN, CNNs are employed as function approximators, with the network weights θ representing the Q-network, as shown below [11]:

$$Q^*(s_t, a_t, \theta_i) = E[r(s_t, a_t) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta_{i-1}) | s_t, a_t]_{a_{t+1}} \quad (4)$$

5. Application of Deep Learning in Autonomous Driving

5.1. Environment Awareness

Environmental perception is vital for autonomous driving, underpinning decision-making. It serves as the interface between the vehicle and its surroundings, ensuring safety and reliability through accurate position assessment and environmental analysis. Key tasks include obstacle detection and lane recognition[11]. Obstacle detection is critical, focusing on identifying potential driving obstructions. With advancements in deep learning, binocular camera-based vision sensors are increasingly replacing traditional radar technologies[12]. Two-dimensional obstacle detection with

binocular cameras can be categorized into single-stage and two-stage methods. The single-stage method processes input images directly, using regression analysis to predict target coordinates and confidence probabilities, primarily employing the You Only Look Once (YOLO) algorithm, introduced by Redmon et al. in 2016. This innovative approach utilizes CNNs for direct detection, enabling real-time obstacle detection. While the YOLO algorithm achieves real-time performance, it compromises detection accuracy [13].

Two-stage algorithms, such as R-CNN proposed by Girshick et al. [14], demonstrate superior accuracy compared to single-stage algorithms, albeit at reduced speed. This method identifies candidate regions, iteratively merges and regresses to optimize regions of interest (ROIs), resizes them uniformly, and feeds them into a CNN for feature extraction. Target object classification is performed using clustering techniques like Support Vector Machines (SVM).

5.2. Simultaneous Positioning and Map Construction

SLAM technology is essential for autonomous driving, facilitating real-time vehicle localization and environmental mapping for passenger safety. It is divided into LiDAR-based and vision-based SLAM. LiDAR-based SLAM, while less affected by lighting variations, incurs higher development costs, leading to a preference for vision-based approaches. A vision-based SLAM system comprises five critical components: the visual sensor, front-end module (Visual Odometry), back-end module, loop closure module, and mapping module. The visual sensor captures environmental imagery, forming the basis for processing. The front-end module estimates the camera's motion trajectory by analyzing sequential image frames, while the back-end module refines these estimates for a globally consistent environmental map [15]. The loop closure module enhances accuracy by identifying previously visited areas. Lastly, the mapping module generates and refines the map for localization using the optimized trajectory and feature point data.

5.3. Path Planning

Path planning ensures that an autonomous vehicle may safely reach its destination by finding the best route. This method considers safety, efficiency, and practicality, making it essential to autonomous driving. Traditional path planning algorithms like PSO [16] and GA [17] have been created and used by researchers. Conventional methods often fail autonomous vehicles in complicated and dynamic surroundings. Recent advances in deep learning have helped feature extraction advance. Deep learning is being considered a possible way to improve path planning algorithms.

Currently, deep learning techniques are predominantly utilized in local path planning, also referred to as dynamic planning. The primary focus of this research lies in end-to-end path planning based on sensors, target objectives, and mapping [11]. Deep reinforcement learning, as it closely resembles human cognitive processes and enables trial-and-error in complex environments, has become the primary learning paradigm for such algorithms. For instance, Xiao Hao and colleagues [18] developed a global path planning method for autonomous vehicles based on Deep Q-Learning and Deep Predictive Network Technologies. Compared to traditional algorithms such as Dijkstra and A*, this approach achieves an approximate 18% reduction in travel time.

6. Conclusion

Autonomous driving technology has rapidly evolved due to advancements in deep learning, moving towards practical implementation and large-scale production. This survey reviews recent developments in deep learning for autonomous vehicles, starting with an overview of innovations in both fields. It details commonly used neural network architectures relevant to autonomous systems and outlines the current research landscape. The paper also addresses challenges faced by deep

learning-based autonomous vehicles and forecasts future trends. Deep learning is poised to be a key focus in autonomous driving, requiring ongoing research and validation. The trend towards intelligent connectivity, driven by the Third Industrial Revolution, underscores the vast potential of autonomous vehicles. Deep learning is expected to enhance their autonomy and operational capabilities, offering solutions to challenges like precision and robustness that have limited traditional technologies. This study acknowledges limitations due to a lack of precise data for stronger conclusions, with future research planned to adopt a more quantitative approach to address these gaps.

References

- [1] Yu, K., Jia, L., Chen, Y., et al. (2013). *The past, present, and future of deep learning*. *Computer Research and Development*, 50(09), 1799-1804.
- [2] Hinton, G. E., & Salakhutdinov, R. R. (2006). *Reducing the dimensionality of data with neural networks*. *Science*, 313(5786), 504-507.
- [3] Hubel, D. H., & Wiesel, T. N. (1977). *Ferrier lecture—Functional architecture of macaque monkey visual cortex*. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 198(1130), 1-59.
- [4] Braverman, M. (2011). *Poly-logarithmic independence fools bounded-depth boolean circuits*. *Communications of the ACM*, 54(4), 108-115. <https://doi.org/10.1145/1924421.1924446>
- [5] LeCun, Y., & Bengio, Y. (1995). *Convolutional networks for images, speech, and time series*. In *The Handbook of Brain Theory and Neural Networks* (pp. 3361-3368).
- [6] LeCun, Y., Bottou, L., Bengio, Y., et al. (1998). *Gradient-based learning applied to document recognition*. *Proceedings of the IEEE*, 86(11), 2278-2324. <http://dx.doi.org/10.1109/5.726791>
- [7] Jin, L. P. (2016). *Research on ECG classification methods for clinical applications [Doctoral dissertation, Suzhou Institute of Nano-tech and Nano-bionics, Chinese Academy of Sciences]*.
- [8] Li, Y. D., Hao, Z. B., & Lei, H. (2016). *A survey of convolutional neural networks*. *Computer Applications*, 36(09), 2508-2515+2565.
- [9] Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory*. *Neural Computation*, 9(8), 1735-1780. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>
- [10] Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2013). *Playing Atari with deep reinforcement learning*. *arXiv: Learning*.
- [11] Duan, X. T., Zhou, Y. K., Tian, D. X., et al. (2021). *A survey on deep learning applications in autonomous driving*. *Unmanned Systems Technology*, 4(06), 1-27.
- [12] Cao, L., Wang, C., & Li, J. (2015). *Robust depth-based object tracking from a moving binocular camera*. *Signal Processing*, 112, 154-161.
- [13] Redmon, J., Divvala, S., Girshick, R., et al. (2016). *You only look once: Unified, real-time object detection*. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1-15). <http://dx.doi.org/10.1109/cvpr.2016.91>
- [14] Girshick, R., Donahue, J., Darrell, T., et al. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation*. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-15). <http://dx.doi.org/10.1109/cvpr.2014.81>
- [15] Liuz, M., Malise, M., & Martinet, P. (2022). *A new dense hybrid stereo visual odometry approach*. In *Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 6998-7003). *IEEE*.
- [16] Ni, J., Wu, L., Shi, P., & Yang, S. X. (2017). *A dynamic bioinspired neural network-based real-time path planning method for autonomous underwater vehicles*. *Computational Intelligence and Neuroscience*, 2017, 9269742. [CrossRef]
- [17] Ni, J., Yang, L., Wu, L., & Fan, X. (2018). *An improved spinal neural system-based approach for heterogeneous AUVs cooperative hunting*. *International Journal of Fuzzy Systems*, 20, 672-686. [CrossRef]
- [18] Xiao, H., Liao, Z. H., Liu, Y. Z., et al. (2020). *Path planning for unmanned vehicles based on deep Q learning in real environments*. [Accessed December 9, 2020, February 7, 2021]. <http://kns.cnki.net/kcms/detail/37.1391.T.20201015.1717.002.html>