Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/93/20240986

Research on image recognition and processing application technology of unmanned vehicle based on deep learning

Junji Duan

School of Computer Science, McGill University, Montreal, Canada

junji.duan@mail.mgill.ca

Abstract. Image recognition technology is critically important in various fields, including the rapidly advancing sector of autonomous vehicles. As one of the core components enabling driverless cars to perceive their environment and make informed decisions, image recognition has seen significant advancements due to deep learning. This paper focuses on the application of deep learning in image recognition for self-driving cars and explores its implications for the future of autonomous driving technology. To begin with, this paper examines the empirical evaluation of deep learning models in highway driving scenarios. By employing Convolutional Neural Networks (CNN), these models achieve high detection rates and superior accuracy in recognizing vehicles and lanes. The robustness of these models is tested under varying weather conditions and times, demonstrating their effectiveness compared to classical computer vision techniques. Next, the paper discusses radar-camera fusion technology, highlighting different fusion strategies such as data-level, feature-level, object-level, and hybrid-level. The findings suggest that while feature-level fusion excels in detecting small objects in complex scenes, hybrid-level fusion is optimal for diverse driving situations. This section provides valuable insights into the integration of multimodal data for improved object recognition and semantic segmentation. After discussing fusion technologies, the paper finally reviews the security challenges posed by adversarial attacks on deep learning-based unmanned systems.

Keywords: Autonomous Vehicles, Image Recognition, Radar-Camera Fusion, Deep Learning, Adversarial Attacks.

1. Introduction

Image recognition technology has become a pivotal element in many advanced systems, particularly in autonomous vehicles. The rapid development of autonomous driving technologies is significantly driven by advancements in deep learning. This technology has shown remarkable success in processing and interpreting visual data. This success is largely attributed to the capabilities of deep learning models, especially Convolutional Neural Networks (CNNs), to learn and identify intricate patterns in large datasets [1]. In particular, the significance of image recognition in autonomous vehicles lies in its ability to enable these systems to understand their surroundings and make informed decisions, thereby ensuring safe and efficient navigation.

The evolution of image recognition technology can be traced back to the early days of computer vision, which relied heavily on handcrafted features and classical machine learning algorithms. However, these methods often struggled with the complexity and variability of real-world driving

DOI: 10.54254/2755-2721/93/20240986

scenarios. The advent of deep learning, particularly CNNs, revolutionized the field by offering a way to automatically learn hierarchical feature representations directly from raw images. This paradigm shift has been crucial in addressing the challenges faced by traditional computer vision techniques and has paved the way for more robust and accurate image recognition systems in autonomous vehicles.

As autonomous driving technology advances, the integration of multimodal sensor data, such as radar and camera fusion, has become increasingly important. These multimodal systems leverage the complementary strengths of different sensors to enhance the overall perception capabilities of autonomous vehicles. For instance, radar sensors are adept at measuring distances and velocities, while cameras provide rich contextual information about the environment. The fusion of these data streams at various levels-data-level, feature-level, and object-level-has been shown to significantly improve the detection and recognition performance of autonomous driving systems, particularly in complex and dynamic driving scenarios.

Additionally, the security and robustness of deep learning models in autonomous vehicles are critical concerns. Adversarial attacks, which involve subtly altering input data to mislead deep learning models, pose significant threats to the reliability and safety of autonomous driving systems. Research has shown that these attacks can severely impact the performance of image recognition systems, highlighting the need for robust defense mechanisms and real-time monitoring solutions to mitigate such risks.

In conclusion, by examining these subtopics, this paper aims to provide a comprehensive understanding of the current state and future potential of deep learning in enhancing the image recognition capabilities of autonomous vehicles.

2. Security and technology in driverless car image recognition

2.1. Overview of image recognition technology for driverless cars

Image recognition technology is one of the core elements of many automated devices today to realize their capabilities, and emerging driverless cars are no exception. In recent years, deep learning has made significant progress in image recognition. As a result, it has provided a strong impetus to drive the rapid development of driverless technology. In this paper, this paper will outline several key research findings that highlight the progress and achievements of deep learning in the application of image recognition for self-driving cars.

Huval et al. reported an empirical evaluation exploring the potential of deep learning technology in highway driving scenarios [2]. The research team extensively collected data from highway scenarios, and used Convolutional Neural Networks (CNN) to achieve effective recognition of vehicles and lanes. The study reveals that the tuned Overfeat model performs well in recognizing lanes and vehicles, especially at the high detection rate of 44 frames per second. In addition, the study compares the performance differences between deep learning strategies and classical computer vision techniques, further confirming the superiority of deep learning in dealing with complex driving scenarios. Importantly, the team collected and labeled a wide range of highway data covering variable weather conditions and different time periods to build a solid data base for testing the robustness of the neural network under various driving conditions. The study also demonstrates how to improve detection accuracy and efficiency by adapting the CNN architecture, such as changing the central region of the detection labels to reduce the number of bounding box predictions. Through a series of quantitative evaluation metrics (e.g., accuracy, recall, and F1 score), the study validates the high accuracy of the improved Overfeat model in lane and vehicle detection.

Yao et al. reported an all-round overview of the radar-camera fusion technology in the field of autonomous driving for object recognition and semantic segmentation [3]. The study carefully distinguishes four fusion strategies, namely, data-level, feature-level, object-level, and hybrid-level, and discusses the advantages and limitations of each. The experimental results show that feature-level fusion is more robust when dealing with tiny object detection and complex scenes, while hybrid-level fusion is optimal in variable driving situations. The paper provides the research community with

DOI: 10.54254/2755-2721/93/20240986

valuable insights into radar-camera fusion by thoroughly analyzing the construction, target task, sensor configuration, and data representation format of the different datasets. Through in-depth analysis of datasets such as nuScenes, CARRADA, and RADIATE, similarities and differences in terms of mission scope, sensor configurations, and data size are summarized. In particular, feature-level fusion techniques (e.g., SAF-FCOS and BIRANet) are effective in object detection and semantic segmentation tasks, and are especially robust to the challenges of small objects and complex environments. As shown in Figure 1, Hybrid-level fusion technology, on the other hand, highlights its excellent performance in complex driving scenarios, which indicates its promising future in practical applications.

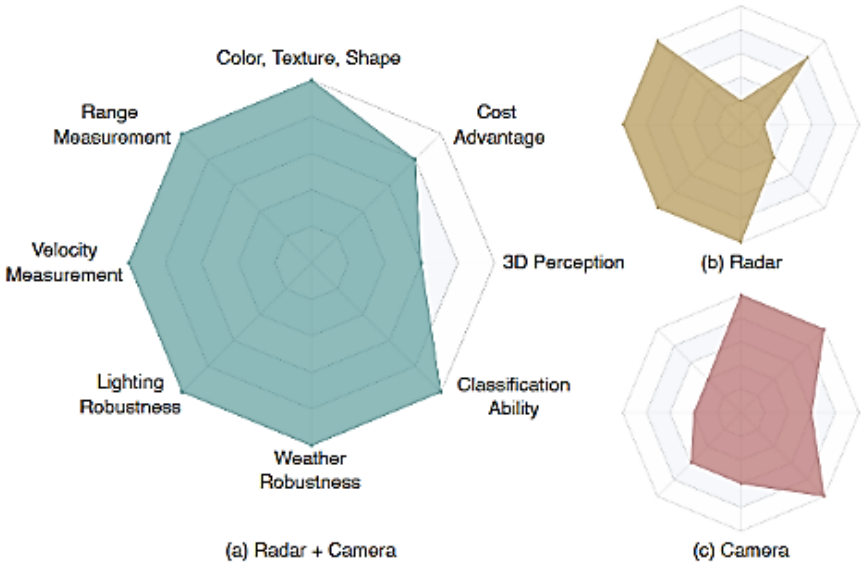

Figure 1. Comparison of radar and camera characteristics [3].

Moreover, A Survey of Deep Learning Techniques for Autonomous Driving provides a comprehensive overview of the use of deep learning techniques in autonomous driving [4]. The study covers multiple dimensions such as driving environment perception, path planning strategy, behavioral decision making and motion control, and explores in detail the specific examples of Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Deep Reinforcement Learning (DRL) in the application of autonomous driving. It is found that CNN-based object recognition and semantic segmentation techniques have achieved outstanding results on benchmark test platforms such as ImageNet; meanwhile, deep reinforcement learning tools facilitate the optimization of driving strategies in simulation scenarios, which directly enhances the effectiveness and safety of real-vehicle driving. The literature provides a clear overview of the underlying deep learning theories and analyzes the implementation and potential of these techniques in the field of autonomous driving. For example, CNNs excel in image information processing, RNNs are uniquely suited for time-series data analysis, and DRL is dedicated to optimizing driving strategies. After comparing and evaluating multiple methods, this study highlights the broad application scope and great development potential of deep learning technology in the field of autonomous driving, and lays a solid knowledge base and direction guide for subsequent research.

The above study reveals the broad application prospects of deep learning technology in image recognition for autonomous vehicles, especially with refined multi-sensor data fusion and optimization strategies, which greatly improve system robustness and accuracy. In the future, this paper should continue to focus on deepening the practice of deep learning technology in a variety of driving

DOI: 10.54254/2755-2721/93/20240986

scenarios in order to continuously improve the safety, security and reliability of autonomous driving technology.

2.2. Research on the security of driverless car image recognition based on deep learning

As deep learning is widely used in driverless cars, security issues are becoming more prominent. Therefore, this paper reviews several important researches to explore the threats of adversarial attacks on driverless systems and their defense strategies.

Frist, Deng et al. reported a detailed analysis of the multiple types of attacks against deep learning-based unmanned systems, as shown in Figure 2, including physical attacks, cyber-attacks, and learning-based adversarial attacks [5]. Physical attacks involve jamming sensors or faking signals, cyber-attacks attempt to infiltrate the autonomous driving system without authorization, and learning-based adversarial attacks mislead the model by modifying the input data. The literature summarizes existing defense mechanisms and suggests future research directions, with a particular emphasis on improving model robustness and security through diverse datasets and enhanced training methods. The research also explores the importance of real-time monitoring and defense systems in real-world applications, suggesting the possibility of implementing anomaly detection on cloud/edge servers. Through a systematic review of the existing literature, the study provides an important reference for future unmanned safety research.

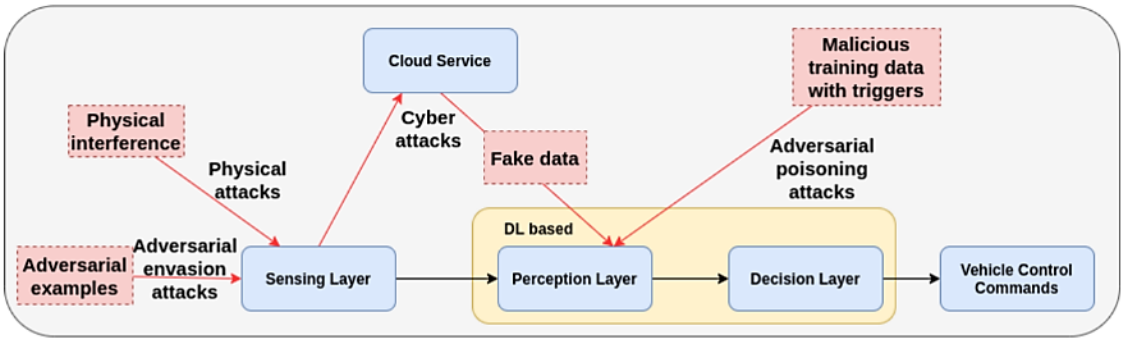

Figure 2. Overview of attacks on each part in an ADS [5].

Second, Boltachev et al. explored the potential cyber threats to unmanned models, especially the adversarial attacks on deep learning-based models [6]. The literature analyzes several major types of adversarial attacks in detail, including data poisoning (Poisoning) attacks and Evasion attacks. Data poisoning attacks disrupt the entire learning process by injecting malicious samples into the training data, while evasion attacks tweak or manipulate the samples to circumvent the system. The study points out that the computer vision system of an autonomous driving system is a main target for counterattacks. By slightly tweaking traffic sign images, an attacker can cause a driverless car to misrecognize traffic signs, raising potential safety risks. The literature also summarizes existing defense strategies and highlights the limitations of these strategies in practical applications, suggesting that more efficient defense methods need to be further explored in the future. The study reveals the serious threat of adversarial attacks on unmanned systems through experiments and literature review, and highlights the practical application challenges that need to be considered when developing more effective defense strategies.

Finally, Mądry et al. proposed a robust optimization framework to improve the robustness of deep learning models through methods such as adversarial training and defensive distillation [7]. Specifically, the literature treats adversarial training as a saddle-point problem involving an inner maximization problem and an outer minimization problem, by finding adversarial samples that maximize the model's loss through adversarial training, and then training the model to minimize that loss. Experimental results show that the model's resistance to adversarial attacks can be significantly improved by optimizing the CNN structure and employing a multi-step approach (e.g., Projected

DOI: 10.54254/2755-2721/93/20240986

Gradient Descent, PGD). The study also emphasizes the importance of model capacity in adversarial training and proposes to further enhance the model's resistance to attacks by exploring the details of the loss surface. Through a series of quantitative evaluations, the studies validate the effectiveness of the robust optimization framework in improving model robustness, providing guidance for future applications in complex driving environments.

Overall, these studies reveal the serious threat of adversarial attacks on unmanned systems and propose some effective defense strategies. Future research should focus on defense mechanisms in practical applications to ensure the safety and reliability of unmanned systems. In particular, when developing and deploying unmanned technologies, various attack threats need to be considered comprehensively to design a more comprehensive and effective security protection system.

2.3. Research on the security of driverless car image recognition based on deep learning

As deep learning techniques continue to evolve, their potential for application in the field of self-driving vehicles is growing. Consequently, this paper summarizes and analyzes several recent studies that provide insights into the possible directions of deep learning technology's advancement in future autonomous driving technologies.

Baltrušaitis et al. reported a comprehensive review of the recent developments in the field of multimodal machine learning, introducing a novel classification system and providing insights into five core challenges: characterization, transformation, alignment, fusion, and cooperative learning [8]. Among them, the characterization challenge focuses on the inherent heterogeneous nature of multimodal data, which involves significant inconsistencies among different modalities in terms of feature attributes, noise distribution, and missing data. The transformation challenge is in how to implement effective data transformation or mapping between modalities, especially in situations where inter-modal associations are open or subjective. The alignment problem focuses on establishing direct associations between cross-modal subcomponents. The fusion problem is concerned with integrating multi-source modal information to enhance prediction efficacy. Collaborative learning aims at facilitating inter-modal knowledge transfer and sharing. The study emphasizes that the robustness and prediction accuracy of unmanned technology can be enhanced by the joint and collaborative representation approach, which can cope with multimodal data processing more effectively. Looking into the future, the paper points out that enhancing representation learning capabilities, improving transformation models, optimizing fusion strategies, and advancing collaborative learning techniques are the main research trends in this field.

In addition, Samadi et al. introduced an innovative Generative Adversarial Network (GAN) technique that utilizes saliency graphs to create a more elucidating and thorough Counterfactual Example (CF) [9]. Counterfactual Examples (CFs) using saliency graphs to create more elucidating and thorough examples as a way to enhance the interpretability of deep neural network (DNN)-based systems. By incorporating the guidance of saliency maps in the generation of confrontation samples, it ensures that changes to the input data are tightly focused on the most conspicuous feature regions, thus producing confrontation instances that are both highly explanatory and transparent. The experimental sessions are conducted on the BDD100k dataset, and the validation results show that the SAFE strategy exhibits significant advantages over existing models in terms of effectiveness and streamlining dimensions in generating adversarial instances. These research results highlight the outstanding contribution of the SAFE strategy in promoting DNN interpretability and transparency, and pave the theoretical and practical foundation for the future development of applications in complex driving scenarios.

Finally, Chen et al. introduced an innovative multimodal large-scale language model (LLM) architecture that enhances the performance of DNNs by integrating object-level vector modalities with pre-trained LLMs, as shown in Figure 3 [10]. By integrating object-level vector modalities with pre-trained LLMs, the structure enhances the environmental context and decision-making effectiveness of autonomous driving systems. An original object-level multimodal framework is proposed in this paper, which realizes the integration of object-level quantitative data modalities with pre-trained LLMs,

DOI: 10.54254/2755-2721/93/20240986

aiming to improve the insight and interpretation performance of driving contexts. In addition, the paper publishes a driving-scenario-specific dataset containing 160,000 question-answer pairs and establishes an evaluation criterion for Driving Domain Question and Answer (Driving QA). Furthermore, the experimental data demonstrates that the pre-trained model performs superiorly on perceptual tasks and behavioral predictions, and scores significantly better on the Driving-related QA assessment. This study shows that the fusion of multimodal data can effectively improve the stability and recognition accuracy of the system, laying an important foundation for subsequent research and practical applications.

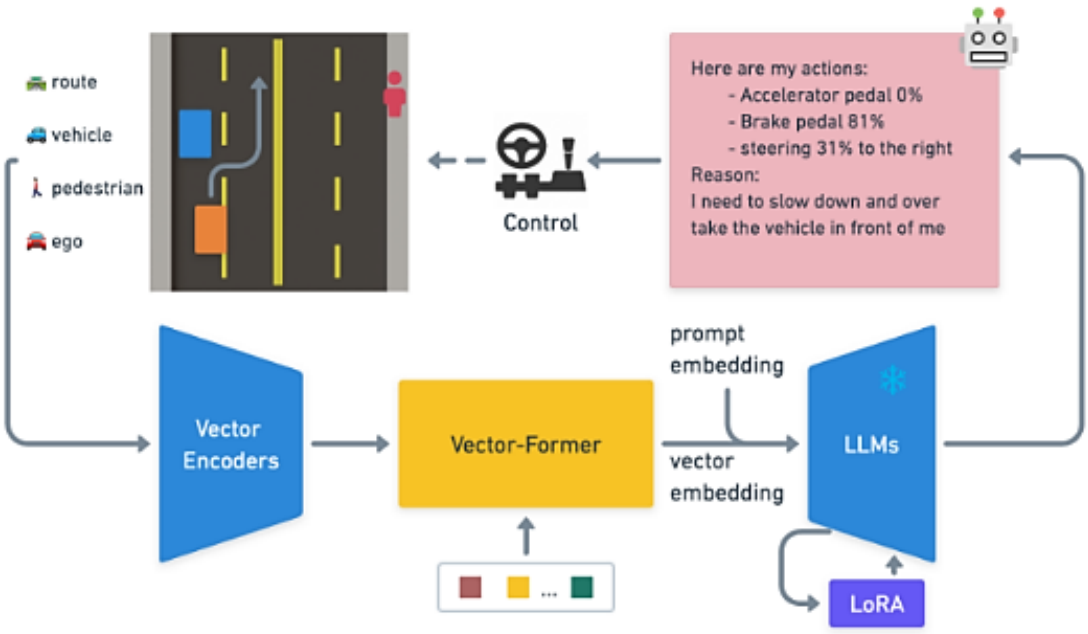

Figure 3. An overview of the architecture for Driving with LLMs [10].

In conclusion, the key path for the future development of deep learning technology in the field of driverless cars focuses on multimodal fusion technology, the increase of interpretability and the strengthening of the ability to resist attacks. Through a continuous optimization and innovation process, these technological elements will lay a solid foundation for the evolution of unmanned systems and expand the breadth of their applications. Future research should focus on the exploration of multi-modal machine learning, generative adversarial networks, and large-scale language models in the field of unmanned applications, with the aim of continuously improving the safety attributes, reliability, and interpretability of the system, and promoting the expansion of the boundaries of unmanned technology and its maturation process.

3. Conclusion

This paper has explored the critical role of deep learning in advancing image recognition technology for autonomous vehicles. Various aspects such as the application of Convolutional Neural Networks (CNNs), the integration of radar-camera fusion technology, and the security challenges posed by adversarial attacks are examined, highlighting the significant progress made in this field. Building on the previous discussion, the empirical evaluation of deep learning models in highway driving scenarios demonstrates their superior accuracy and detection rates, proving their effectiveness in recognizing vehicles and lanes under varying conditions. The robustness of these models, tested across different weather conditions and times, underscores the reliability of deep learning over traditional computer vision techniques. Furthermore, the exploration of radar-camera fusion strategies reveals that feature-

DOI: 10.54254/2755-2721/93/20240986

level and hybrid-level fusion techniques offer substantial improvements in detecting small objects and handling complex driving environments. Security remains a paramount concern for autonomous vehicles, as adversarial attacks can significantly undermine the safety and reliability of these systems. The study of different attack types and defense mechanisms underscores the necessity for enhancing model robustness and implementing real-time monitoring to safeguard against potential threats. This study underscores the importance of developing comprehensive security strategies to ensure the safe operation of unmanned systems.

Looking ahead, the future of deep learning in autonomous driving is promising, with several key areas poised for further development. The integration of multimodal data will continue to enhance the perceptual and decision-making capabilities of autonomous vehicles, making them more adaptable to diverse driving scenarios. Advances in generative adversarial networks (GANs) and large-scale language models (LLMs) will contribute to improving the interpretability and transparency of deep learning models, thereby fostering greater trust and adoption of autonomous driving technologies.

In conclusion, deep learning has proven to be a transformative force in the field of image recognition for autonomous vehicles. Its ability to process and interpret complex visual data with high accuracy and reliability is pivotal for the future of autonomous driving. As the technology continues to evolve, it will undoubtedly play a crucial role in realizing the vision of fully autonomous and safe driving systems, thereby paving the way for a new era in transportation.

References

[1] Bojarski M, Del Testa D, Dworakowski D, Firner B, Flepp B, Goyal P, et al 2016 End to end learning for self-driving cars arXiv:1604.07316

[2] Huval B, Wang T, Tandon S, Kiske J, Song W, Pazhayampallil J, et al 2015 An empirical evaluation of deep learning on highway driving arXiv:1504.01716

[3] Yao S, Guan R, Huang X, Li Z, Sha X, Yue Y, et al 2023 Radar-camera fusion for object detection and semantic segmentation in autonomous driving: A comprehensive review IEEE Trans. Intell. Veh.

[4] Grigorescu S, Trasnea B, Cocias T, Macesanu G 2020 A survey of deep learning techniques for autonomous driving J. Field Robot. 37 362-386

[5] Deng Y, Zhang T, Lou G, Zheng X, Jin J, Han Q L 2021 Deep learning-based autonomous driving systems: A survey of attacks and defenses IEEE Trans. Ind. Inf. 17 7897-7912

[6] Boltachev E 2023 Potential cyber threats of adversarial attacks on autonomous driving models J. Comput. Virol. Hack. Tech. 1-11

[7] Madry A, Makelov A, Schmidt L, Tsipras D, Vladu A 2017 Towards deep learning models resistant to adversarial attacks arXiv:1706.06083

[8] Baltrušaitis T, Ahuja C, Morency L P 2018 Multimodal machine learning: A survey and taxonomy IEEE Trans. Pattern Anal. Mach. Intell. 41 423-443

[9] Samadi A, Shirian A, Koufos K, Debattista K, Dianati M 2023 In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC) 5655-5662

[10] Chen L, Sinavski O, Hünermann J, Karnsund A, Willmott A J, Birch D, et al 2023 Driving with llms: Fusing object-level vector modality for explainable autonomous driving arXiv:2310.01957