Proceedings of the 4th International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/86/2025.20334

Research Advanced in Image Recognition under Autonomous Driving Scene Based on Deep Learning

Zhengwen Zhu[1,a,*]

[1]*Ira A. Fulton Schools of Engineering, Arizona State University, AZ, USA*
a. zzhu122@asu.edu
**corresponding author*

Abstract: Image recognition has always been a fundamental research task in the computer vision community, aimed at identifying the categories of objects in images and has been widely used in many fields, especially in autonomous driving. Early image recognition technologies were mostly based on machine learning, and their recognition speed and accuracy could not meet the application requirements of complex autonomous driving scenarios. With the great success of convolutional neural networks, image recognition technology based on deep learning has attracted increasing research interest. Taking the autonomous driving scenario as an example, this article introduces the latest research progress of image recognition technology, including representative methods and their basic pipelines. In addition, this paper also introduces the commonly used dataset in image recognition and discusses the existing problems of image recognition in autonomous driving tasks. Finally, this paper looks forward to the future development directions of this field, hoping bring some new insight to advance the further development of image recognition under autonomous driving scene.

Keywords: Image recognition, autonomous driving, deep learning

1. Introduction

As artificial intelligence and autonomous driving technologies advance quickly, the use of driverless cars has progressively moved out of the lab and into the real world. In the process of achieving fully autonomous driving, accurately identifying and understanding complex road environments is important in determining vehicle safety and intelligence. As one of the key technologies for autonomous driving, image recognition is responsible for acquiring and analyzing road information. Through visual sensors such as cameras, autonomous driving systems can detect objects on the road, traffic signs, pedestrians, vehicles, etc., in real-time, so as to make decisions and ensure driving safety.

In recent years, breakthroughs in deep learning technology have significantly improved the accuracy and efficiency of image recognition. Methods such as convolutional neural networks (CNNs) have performed outstandingly in tasks such as object detection, road sign recognition, and pedestrian recognition for autonomous driving. However, the complex real-world driving environment, such as different weather conditions, light changes, and irregular road scenes, still poses a serious challenge to image recognition technology. In addition, the high requirements of autonomous driving for system real-time performance and stability also impose higher standards on image processing technology.

DOI: 10.54254/2753-8818/86/2025.20334

The purpose of this review is to systematically sort out the key applications, existing technologies, and challenges of image recognition in autonomous driving, and discuss future research directions and technological breakthroughs. We will analyze from multiple perspectives, such as the core tasks of image recognition, mainstream technical methods, typical datasets, and difficulties in practical applications, to provide a reference for future research and development.

2. Recognition Tasks in Autonomous Driving

2.1. Main Task

The image recognition task in autonomous driving can be divided into multiple subtasks, each solving a different visual problem. Traditional methods often have difficulty directly addressing these complex tasks but applying deep learning has greatly improved their ability to solve them. The main subtasks include[1]:

(1) Image Verification. Image verification determines whether the input image matches a reference image by calculating the distance between their feature vectors. Traditional methods were used in tasks like fingerprint or face recognition, while deep learning improves accuracy with triplet loss functions.

(2) Object Detection. Object detection aims to locate objects of a certain category within an image. Traditional methods, such as Haar-like features with AdaBoost for face detection, have been surpassed by deep learning, which allows multi-class detection in a single network.

(3) Image Classification. Image classification assigns objects in an image to predefined categories. Methods like bag-of-features (BoF) were commonly used, but deep learning now excels in large-scale classification tasks, surpassing human-level accuracy.

(4) Comprehending the Scene. One of the main tasks of scene understanding is semantic segmentation, which aims to assign a category to each pixel in a picture. Deep learning has made this possible, but conventional approaches have had difficulty solving it.

(5) Specific Object Recognition. Specific object recognition focuses on identifying a particular object using features such as SIFT. Deep learning further improves performance with methods like LIFT.

2.2. Autonomous Driving Levels

The Society of Automotive Engineers (SAE) suggested a standardized language, known as SAE-J3016, in 2014 to address the inconsistent and confusing terminology used in the autonomous driving sector. SAE-J3016 specifies the Levels of Driving Automation from 0 to 5 [2]. Image recognition technology becomes more and more important as automation levels rise.

(1) L0-L2 (Assisted Driving). At levels L0 to L2, the vehicle primarily provides driver assistance, with the driver still responsible for vehicle control and safety. Features like Lane Departure Warning (LDW), Automatic Emergency Braking (AEB), and Adaptive Cruise Control (ACC) rely on image recognition technology as a supporting component. While image recognition primarily helps the driver at these levels, systems are now able to recognize objects and situations with greater accuracy thanks to the incorporation of deep learning and enhanced sensor precision.

(2) L3-L5 (High and Full Automation). At levels L3 and above, the system takes full control of the vehicle, especially at L4 and L5, where no human intervention is required in specific or all driving environments. At these levels, image recognition systems must not only handle real-time object detection and classification but also manage complex scene understanding and dynamic predictions to ensure safe navigation in complex traffic and environmental conditions. Deep learning has enabled breakthroughs in tasks such as object detection, pedestrian recognition, and scene segmentation by allowing models to learn complex data representations. As the level of autonomous driving increases,

DOI: 10.54254/2753-8818/86/2025.20334

image recognition not only needs to meet higher real-time and robustness requirements, but also be combined with other sensors (such as lidar, radar, etc.) to ensure recognition accuracy and system decision-making ability in complex environments.

3. Image Recognition Technology

3.1. Traditional Image Recognition Methods

Traditional image recognition methods [3] rely on machine learning algorithms and manual feature extraction. These approaches include Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees, which typically convert image data into one-dimensional vectors for processing. However, since images are inherently two-dimensional, this flattening process can lose important spatial relationships between pixels.

(1) SVM Algorithm. SVM is a classical machine learning algorithm used for binary classification. In image recognition tasks, SVM often applies kernel methods (such as the Radial Basis Function, RBF) to handle non-linear classification problems. However, transforming images into vectors for SVM can result in the loss of spatial features crucial for accurate image recognition.

(2) Feature Extraction and Template Matching. Techniques like histograms and shape detection are often used to classify images based on extracted features. However, this manual process requires significant expert knowledge and struggles to adapt to complex and varied environments.

Despite their extensive use, traditional machine learning methods struggle with large-scale and complex image data, especially in real-time applications like autonomous driving, where the system must process vast amounts of visual information quickly and accurately.

3.2. Deep Learning Methods

Recent years have seen a significant increase in the efficiency and accuracy of picture identification because to deep learning, namely Convolutional Neural Networks (CNNs). Unlike traditional methods, deep learning can automatically extract features from images without human intervention. CNNs are capable of capturing spatial hierarchies and identifying complicated patterns through the use of several convolutional layers, pooling, and fully linked layers.

(1) Convolutional Neural Networks (CNNs). CNNs are the backbone of deep learning in image recognition. They automatically extract local features through convolutional layers and reduce dimensionality via pooling layers, making the model both efficient and highly accurate. Classic models like LeNet and AlexNet have demonstrated excellent performance in tasks like handwritten digit recognition.

(2) Object Detection Techniques. Two algorithms are mainly used in this field. A real-time object identification method called YOLO can identify and categorize several items in a picture at the same time. Because of its efficiency and speed, it is perfect for use in autonomous driving, where it is necessary to identify items like cars, pedestrians, and traffic signals quickly. A region proposal network (RPN) is utilized by the R-CNN family, which includes Faster R-CNN, to produce areas of interest, which are then classified. This approach offers high accuracy, though it can be computationally expensive compared to YOLO.

3.3. Recent Advances in Image Recognition in Autonomous Driving

The YOLO (You Only Look Once) family of models, in particular, has made considerable strides in deep learning recently, which have improved the accuracy and speed of vehicle and pedestrian detection in autonomous driving. The improved YOLOv4 model [4], which makes use of the CSPDarknet45_G backbone network and adds a DBG module made up of the activation functions

DOI: 10.54254/2753-8818/86/2025.20334

for GELU, Batch Normalization (BN), and DarknetConv2D, is one significant advancement. This improved structure enhances the generalization ability of the model and increases the detection accuracy, particularly for small objects such as pedestrians and traffic lights, even under complex weather conditions. The modified model achieves an impressive mean average precision (mAP) of 90.45% and a recall of 94.37%, with a real-time processing capability of 50 frames per second. This balance of accuracy and speed is crucial for ensuring reliable and efficient image recognition in real-world autonomous driving scenarios.

Another area of advancement is the detection of lanes and traffic signs [5], which are essential for path planning and safety in autonomous vehicles. Deep learning-based lane detection systems have replaced more conventional computer vision approaches. Fully connected convolutional neural networks (CNNs) are one example of this. For example, models based on VGG-16 architecture have achieved lane detection accuracy of up to 98.58% when applied to datasets like the KITTI Road/Lane Detection Evaluation. In addition to lane detection, the German Traffic Sign Recognition Benchmark (GTSRB) dataset has been extensively used to train CNNs for traffic sign recognition, with models achieving an accuracy of over 95%. hese improvements in lane and traffic sign detection are critical for enabling self-driving cars to safely navigate through diverse driving environments.

3.4. Limitations and Challenges of Deep Learning

Despite its success, deep learning in image recognition faces several challenges:

(1) Dependency on Data. Large volumes of labeled data are usually needed for deep learning models to reach high accuracy. In autonomous driving, collecting and annotating diverse and high-quality datasets is time-consuming and costly.

(2) Computational Resources. Deep learning model training is resource-intensive and frequently time- and computational-consuming, which can be problematic for real-time applications such as autonomous driving.

Nonetheless, deep learning's ability to automatically extract and learn features from raw images has made it the dominant approach in modern image recognition, providing a strong foundation for visual perception in autonomous driving systems.

4. Datasets in Autonomous Driving Image Recognition

High-quality datasets are critical for training and evaluating image recognition models in autonomous driving. Below are some of the most commonly used datasets:

The KITTI dataset [6] is one of the most widely used datasets in autonomous driving research. It offers a variety of data formats spanning urban, rural, and highway settings, such as RGB photos, depth maps, and LiDAR point clouds. The dataset includes 389 stereo image pairs, optical flow estimation data, and annotations for 3D object detection tasks. Since the data is captured from real-world driving environments, KITTI is ideal for training deep learning models to handle complex road conditions.

Cityscapes [7] is focused on semantic segmentation in urban environments. It offers 20,000 roughly tagged photos and 5,000 carefully annotated photographs spanning a variety of traffic situations from 50 locations, mostly in Europe. Cityscapes has become the standard benchmark for pixel-level image labeling, particularly suited for training models to recognize pedestrians, vehicles, and traffic signs in complex city scenes.

Waymo Open Dataset [8] is a large-scale dataset providing high-resolution data captured by cameras and LiDAR sensors mounted on autonomous vehicles. The dataset covers diverse and complex urban and suburban driving scenarios, supporting both 2D and 3D object detection and

DOI: 10.54254/2753-8818/86/2025.20334

tracking tasks. Waymo's dataset is significantly larger than most other datasets, offering around 12 million 3D LiDAR object annotations and 9 million 2D image annotations.

These datasets serve as valuable resources for training image recognition systems and developing robust models capable of detecting and classifying objects in various driving environments.

5. Discussion

5.1. Challenges

Despite the availability of these high-quality datasets, image recognition in autonomous driving faces several challenges, especially in handling real-time, complex driving scenarios.

(1) Recognition in Complex Environments. One of the greatest challenges for autonomous driving systems is maintaining high performance in object recognition under adverse weather conditions such as rain, fog, snow, and low visibility [9]. These environmental factors significantly impair the functioning of key sensors like cameras and LiDAR, which rely on clear line-of-sight to detect objects in the surrounding environment. Rain, for example, can distort the vision of both cameras and LiDAR. While light rain might not severely impact sensors, heavy rain causes signal attenuation and noise, reducing the accuracy of object detection. Similarly, fog leads to scattering effects that degrade both visual images and LiDAR point clouds. Dense fog can reduce detection range, resulting in missed or inaccurate object recognition, with some sensors only being able to detect objects at very close distances. Snow also poses unique challenges by creating voids in LiDAR point clouds due to snow swirl, which obstructs the sensor’s ability to properly identify objects. These environmental impacts necessitate advanced sensor fusion techniques that combine data from multiple sources, such as radar, thermal cameras, and LiDAR, to compensate for individual sensor weaknesses and improve recognition under challenging conditions.

(2) Real-Time Processing Requirements. Autonomous driving requires real-time processing and reaction to the driving environment, meaning the image recognition system must quickly detect, classify, and make decisions. For example, when a pedestrian is detected crossing the road, the system needs to react in milliseconds to avoid potential accidents. To meet this challenge, efficient deep learning models, such as YOLO and SSD, are commonly applied in autonomous driving, as they can perform object detection with minimal computational cost. In addition, hardware accelerators like GPUs and TPUs are used to speed up the inference process of deep learning models, ensuring real-time detection and decision-making even in high-speed driving scenarios.

(3) Despite the rapid advancements in autonomous driving technologies, the widespread use of image recognition systems raises several important ethical and legal concerns. One of the primary challenges is privacy. Autonomous vehicles continuously capture vast amounts of data from their surroundings, including images of pedestrians, vehicles, and other sensitive information. Ensuring that this data is anonymized and securely handled is crucial for protecting personal privacy. Constant contact between self-driving cars and the environment or cloud servers further raises the possibility of privacy violations, including identity theft, tracking, and improper use of gathered data by third parties [10].

Another major ethical concern is decision-making transparency. Deep learning models' intricate decision-making procedures sometimes operate as "black boxes," making it challenging to comprehend how choices are formed in the moment. In the case of an accident, this lack of openness may become more troublesome. Accountability requires that autonomous systems' decision-making procedures be traceable and transparent.

From a regulatory standpoint, legal frameworks for autonomous driving are still in development. As image recognition becomes more integrated into these systems, there is a growing need for stricter standards concerning transparency, accuracy, and reliability of the systems. Legal regulations must

DOI: 10.54254/2753-8818/86/2025.20334

clearly define liability in accidents, address data security issues, and ensure compliance with privacy laws. Additionally, establishing robust measures for cybersecurity is crucial to prevent attacks that could compromise the safety and functionality of autonomous vehicles.

5.2. Future Directions and Outlook

As deep learning algorithms advance, hardware capabilities improve, and more extensive datasets become available, the prospects for image recognition in autonomous driving are promising. With the development of specialized hardware like TPUs and ASICs, image recognition systems will be able to process complex driving scenarios more efficiently. In particular, real-time object detection and scene recognition will continue to improve in both speed and accuracy, especially in dynamic and high-speed driving environments. Advanced models like improved versions of YOLO and emerging architectures such as Transformers are expected to become more prevalent in the future.

Moreover, the ongoing development of multisensor fusion will enable autonomous vehicles to combine data from cameras, LiDAR, radar, and other sensors, thereby enhancing environmental perception under various conditions, including challenging weather or lighting. Integrating reinforcement learning and transfer learning techniques can further improve the system's ability to learn from limited data and adapt to new driving environments, boosting its reliability in unfamiliar settings.

On the data side, future datasets are expected to be more extensive, covering a wider range of driving scenarios, particularly rare and high-risk conditions. Techniques such as generative adversarial networks (GANs) can also generate synthetic data to help mitigate data scarcity issues and fill in gaps where real-world data is lacking.

6. Conclusion

Image recognition plays a vital role in enabling key functionalities in autonomous driving, including environmental perception, path planning, and obstacle avoidance. This paper has examined the current state of image recognition in autonomous driving, focusing on its requirements, technological overview, available datasets, challenges, and the latest advancements. While significant progress has been made, particularly in object detection, lane, and traffic sign recognition, image recognition still faces challenges such as dealing with complex environments, real-time processing demands, and data imbalance. However, with ongoing improvements in hardware capabilities, advanced deep learning models, and the integration of multisensor fusion, image recognition in autonomous driving is expected to achieve greater accuracy, robustness, and efficiency. At the same time, the ethical and regulatory challenges surrounding privacy, transparency, and accountability must not be overlooked. As these technologies evolve, ensuring that image recognition systems are not only effective but also ethically sound and legally compliant will be key to their widespread adoption. In conclusion, as the technology continues to evolve, image recognition will be a crucial driver in advancing autonomous driving to higher levels of automation, ultimately improving safety and contributing to the broader adoption of autonomous vehicles in everyday life.

References

[1] *Fujiyoshi, H., Hirakawa, T., & Yamashita, T. (2019). Deep learning-based image recognition for autonomous driving. IATSS Research, 43(4), 244-252.*

[2] *Zhao, J., Zhao, W., Deng, B., Wang, Z., Zhang, F., Zheng, W., Cao, W., Nan, J., Lian, Y., & Burke, A. F. (2024). Autonomous driving system: A comprehensive survey. Expert Systems with Applications, 242, 122836.1-122836.27.*

[3] *Lai, Y. (2019). A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition. Journal of Physics. Conference Series, 1314(1), 12148.*

DOI: 10.54254/2753-8818/86/2025.20334

[4] *Yu, W., & Ren, P. (2021). Vehicle and pedestrian target detection in auto driving scene. Journal of Physics: Conference Series, 2132(1), 12013.*

[5] *Padmavathi, B., Dhivya, S., Datchanamoorthy, K., Banu, A. K., & Karthikeyan, S. M. (2024). Lane and Traffic Sign Detection in Self-Driving Cars using Deep Learning. International Journal of Vehicle Structures and Systems, 16(1), 45-49.*

[6] *Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 3354-3361.*

[7] *Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The Cityscapes Dataset for Semantic Urban Scene Understanding. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3213–3223.*

[8] *Sun, P. , Kretzschmar, H. , Dotiwalla, X. , Chouard, A. , Patnaik, V. , & Tsui, P. , et al. (2020). Scalability in Perception for Autonomous Driving: Waymo Open Dataset. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2443-2451.*

[9] *Zhang, Y., Carballo, A., Yang, H., & Takeda, K. (2023). Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. ISPRS Journal of Photogrammetry and Remote Sensing, 196, 146-177.*

[10] *Hataba, M., Sherif, A., Mahmoud, M., Abdallah, M., & Alasmary, W. (2022). Security and Privacy Issues in Autonomous Vehicles: A Layer-Based Survey. IEEE Open Journal of the Communications Society, 3, 811–829.*