A review of the application of CNN-based computer vision in auto-driving

Tongwei Zhang

Department of Computer Science and Software Engineering (CSSE), Concordia University, Montreal, QC, H3H 2L9, Canada

330aihebe@gmail.com

Abstract. Beginning with Tesla, self-driving technology has become commercially available in recent decades. Target recognition and semantic segmentation remain significant obstacles for autonomous driving systems. Given that these two tasks are also part of the primary tasks of computer vision and that deep learning techniques based on convolutional neural networks have made advancements in the field of computer vision, a great deal of research has begun to apply convolutional neural networks to autonomous driving in the past few years. In this paper, we examine recent publications on CNN-based techniques for autonomous driving, classify them, and offer insights into future research directions.

Keywords: Convolutional Neural Networks; Computer Vision; Autonomous Driving; Image Recognization; Object Detection; Semantic Segmentation.

1. Introduction

People have envisioned autonomous cars for decades. Due to rapid technological advancement, this idea is now commercially available from Tesla. Target recognition and semantic segmentation are persistent autonomous driving challenges [1]. Without accurate detection of artificially determined elements, such as lane markers, guardrails, and other cars, autonomous vehicles could cause property damage and endanger lives. It's important to quickly and accurately recognize images from the front center camera video and generate vehicle control signals. Pattern recognition is used to require time-consuming, inaccurate manual feature extraction and image classifiers.

Deep learning methods, especially convolutional neural networks (CNN), have outperformed state-of-the-art machine learning techniques in several fields, including computer vision [2]. They are usually better than earlier systems that relied on manual feature extraction and develop more complex techniques for tasks like image classification [3], object recognition [4], and semantic segmentation [5]. These three responsibilities are interconnected and mutually supportive. That's because they all have their roots in CNN's original ideas. Down this link, each task gets harder. Object recognition and segmentation in picture classification require basic network models. CNN-based image classification algorithms include cutting-edge object detection and segmentation methods. CNN can automatically extract complex features and use this training data in large-scale picture recognition applications, an improvement over human-based systems. Using convolutional kernels to scan a global image requires fewer parameters and reduces loading time.

This paper reviews the publications of the three main computer vision roles and deep learning based on CNN. These technologies have been used in business and academic fields since their inception. Their use in driverless car research is promising. CNN-based deep learning applied to car front camera images may help to solve the problems of pattern recognition in the Autonomous Driving System.

The remainder of the paper is structured as follows: In the second part, we detail the research methods that were applied to this overview. The papers gathered using the methodology are discussed in the third section. The conclusion is presented in Section 4.

2. Methodology

This review summarizes CNN-based deep learning research on autonomous driving. This systematic review analyzed existing studies using predetermined criteria. These reviews helped determine research knowledge. All primary source information was analyzed. The systematic literature review answers the study's main question with knowledge, logic, and strength [6].

2.1. Research Framework

This systematic review begins with structure. It includes the general plan followed in the systematic review of the literature. The plan has three layers: planning, data selection, and evaluation.

2.1.1. Research Problem. To conduct a systematic literature review of a topic effectively, one must first define the research questions. This paper examines which aspects of autonomous driving technology CNN-based deep learning techniques have enhanced.

2.1.2. Database and Searching Strategy. To extract useful information from domain data, use a methodical search strategy. In this step, meaningful and relevant data were extracted from the vast amount of data. Creating an automated search mechanism to exclude data from domain-related sources. Reference lists of research papers, case studies, Tesla reports, and relevant publications were examined. IEEE Explore, Science Direct, Springer Link, and ISI Web of Knowledge were searched for relevant articles. Only academic publications were searched, with phrases like "autonomous driving; CNN-based deep learning; image classification" in the title, keywords, or abstracts.

2.1.3. Initial Filtering Criteria. Initially, papers were chosen based on specific criteria such as the language of the paper, the year of publication, and the topic's relevance within the desired field. This study only includes English-language research papers. Our review papers are based on research published between 2014 and 2022. The papers selected must be relevant to the search terms specified in the strategy.

2.2. Literature Evaluation

This study's goal is to identify gaps in the scientific literature, not to find all the literature on autonomous driving, which would produce a large number of results. Instead, an overview and classification are needed to identify field gaps.

2438 papers and presentations were extracted using the initial search criteria. To analyze the evolution of these technologies, researchers study CNN-based deep learning methods to perceive and control autonomous driving. This study eliminates duplicate publications and accurately reflects the search depth by carefully analyzing abstracts, titles, and journals. This work combines prior knowledge of CNN-based deep learning and computer vision to evaluate publications to better understand the literature's status, emphasis, and future. 20 papers were chosen.

3. Result

Deep neural networks are crucial for achieving total autonomy in vehicle operation. They consist of a network of nodes that are interconnected. Their structure is analogous to the connections between neurons in the human brain. The network nodes collaborate to address specific issues. Neural networks that have been programmed to complete a set of tasks are capable of providing expert-level knowledge in their trained field. Convolutional neural networks are a prevalent type of deep neural network employed in computer vision. CNN is an excellent tool for gathering and learning both global and local data because they use simple features (like curves and edges) to generate more complex ones (like shapes and corners) [7]. In my research, CNN-based deep neural networks were utilized for image classification, object detection, and semantic segmentation for autonomous vehicle control.

3.1. Image classification and Object Detection

CNN's image categorization and target identification relate to autonomous driving perception. The perception subsystem of an autonomous driving system is responsible for recognizing objects in the driving environment. ADS [8] uses multi-label image classification instead of single-label, which identifies one class of objects per image. Multilevel image classification is called object classification throughout the study. This section summarizes ADS object classification and detection work published in the last three years using autonomous vehicle sensing technologies.

Camera-based image data comes first. Weather and lighting can affect an autonomous vehicle's ability to analyze and recognize scenes. The hardest part of ADS design research is managing driving scenarios, especially in natural settings. Algorithms that improve picture classification may solve many ADS design issues.

Li et al. presented ML-ANet, a deep adaptive neural network for multi-label image classification [9]. The proposed strategy uses transfer learning to go from fully labeled to limited or unlabeled domains to achieve successful knowledge transfer. MK-MMD loss is used to evenly distribute features from the source and target domains, reducing disparities. Domain changes from clear to overcast skies are the focus of the experiments. The experiment uses the KITTI, Cityscapes, and Foggy Cityscapes datasets, which are widely used in their respective domains, and compares the accuracy of the new algorithm with state-of-the-art algorithms, demonstrating that the new algorithm can adapt to all-weather illumination under different weather conditions, achieving greater image classification accuracy than other algorithms. Additionally, the development period is shortened due to a lack of properly labeled training data.

Fog, rain, or snow can make it hard to see while driving. [10] presents three YOLO CNN-based models [11] for spotting people in foggy weather. By using convolutional and linear bottleneck talents separately, these models reduce computing costs and parameter counts, making the proposed networks more effective. The proposed model outperforms state-of-the-art methods in accuracy and speed using a custom foggy weather pedestrian dataset.

Rain and snow also inspire writers. Hnewa et al. [12] study object detection in bad weather. This paper examines techniques for mitigating rain's impact on performance, including image alteration [13], domain adaptation [14], and de-drainage [15]. Experiments used the BDD100K dataset of weather-tagged images. Every picture was labeled with weather information, including fog, rain, etc. The mitigation method improves CNN rain detection, according to evaluations.

Nighttime driving is dangerous because lighting can obscure important details. Due to security concerns, many have tried to improve low-light photo ID accuracy. [16] proposes LE-net for low-light image enhancement. A pipeline converts daylight shots to low-light images for model training. Training and validating LE-net use low-light images. In real-world low-light evening conditions, LE-net outperforms other models.

Most of the reviewed materials used camera-based data. However, time and weather affect camera images' lighting and shading. Even small changes in lighting can alter perception, causing algorithm failure. Meanwhile, Pure camera perceptrons have trouble responding to changes in illumination due to the finiteness of the visual space. Multi-view or stereo systems are more robust [17], but a single camera can capture scale in dynamic activities [18].

Combined camera, LIDAR, and radar sensors improve precision and eliminate single points of failure [19]. Several recent studies have focused on improving the perceptual quality of autonomous driving scenes and real-time object categorization and detection. [20] categorized objects using LIDAR and visual data. Initial pixel-level point cloud data is upsampled and turned into depth feature maps. CNN was fed RGB and depth data. The KITTI dataset showed greater object categorization accuracy than depth or RGB data. LiDAR data accelerates element learning and convergence.

The expense of LIDAR is addressed by a second fusion-based system [15] that employs a 4-beam LIDAR instead of a 64-beam LIDAR and a stereo camera for 3D object identification. Improved depth estimation improves 3D target identification on the KITTI dataset.

3.2. Semantic Segmentation

Segmenting each pixel of an input picture into one of a series of predefined object classes is one of the most essential scene comprehension challenges in contemporary computer vision. Semantic segmentation is more concerned with the characteristics of individual objects, such as people, cyclists, traffic signs and lights, road markings, etc. than with the categorization of images in distinct scenarios, which was the emphasis of the preceding section. Segmentation will segment every pixel and divide the whole picture into distinct semantic components. CNN has dominated comparable perceptual challenges for decades.

Semantic image segmentation improves object categorization and scene-specific properties. People and traffic lights must be identified for autonomous driving.

Boyuan and Muqing proposed a pedestrian ID model based on YOLOv4 [21]. This detection model combines an SPP (Spatial Pyramid Pool) network, K-mean clustering, and YOLOv4 to extract features. A Mish activation function is added to the detection model's neck to replace the leaky ReLU function and improve detection. Titan XP achieved 84.7% AP at 36.4 FPS on the Caltech pedestrian dataset.

Traffic signal detection involves recognizing traffic signs, lights, and ground symbols to determine if autonomous driving can follow traffic rules. Using HOG features and SVM to detect and classify the shapes of circular or triangular signs, [22] proposed a two-stage traffic sign detection and recognition method based on SVM and CNN. In [23], a deep learning strategy for adaptive single-shot detection (SSD) traffic signal recognition is described. SSD had trouble recognizing small objects for traffic signal detection. By updating to an Inception-v3 CNN, objects smaller than 10 pixels can be detected without changing the input image size. On the DriveU traffic light dataset, the model performs well.

Road markings, a traffic signal component, have also been studied. Ye et al. [24] dealt with distorted and worn road markings using YOLO-v. The first stage uses a YOLO-v2 CNN to identify initial road markers. RM-Net, a novel, lightweight, transformation-invariant classification network, is used to identify road markers in the second step. [30] provides a public dataset for the road sign identification challenge based on daytime and nighttime images in varying weather. On this dataset, the model scores 86.5% on mAP, outperforming all others.

After detection and classification, divide the image. Scene/object geometry can inform computer vision applications. Geometry can be depicted using depth maps, which include shape, texture, and distance from the picture plane. Producing a depth map for application-specific datasets can be time-consuming (e.g., via depth cameras, LIDAR sensors, etc.). Unsupervised/self-supervised CNN for depth estimation has recently been presented [25].

RGB material with a single view was once considered an appropriate input for CNN semantic segmentation. [26] compared two CNN networks, DeepLabv3 and FCN, built on ResNet-50 and ResNet-101, for RGB data. Using data augmentation methods, studies on the MiniCity urban sceneries dataset illustrated the neural networks' efficacy in various trials.

Sometimes RGB photos do not provide enough category clues. A recent study [27] introduced a unique regularizer for punishing the semantic-depth gap during segmenter training. Neither ground truth depth maps nor special data (e.g., stereo 3D) are needed during the training phase. Experiments were done to objectively assess autopilot's public scene parsing. Using the depth map during training without multitasking (which may require a more sophisticated backbone CNN to prevent underfitting) is sufficient to improve accuracy during deployment/inference without increasing the processing burden.

4. Conclusion

This review discusses CNN-based computer vision in autonomous driving systems. Image categorization and object recognition are core computer vision foci for autonomous driving systems. This research uses CNN-based deep learning techniques to improve image recognition accuracy and speed. Review of recent studies.

Autonomous driving technology focuses on image recognition. Existing approaches in natural environments have sensor and cost limitations. Fluctuating weather and lighting conditions make it difficult for autonomous driving systems to switch control inputs in a timely manner. Future research should improve target detection in challenging driving conditions to improve autonomous driving safety.

The automatic organization is a medically-pioneered concept in deep learning. Automatic organizing detects traits and establishes relationships or patterns in a sample of images. Automatic organization strategies paired with convolutional neural networks (CNNs) have improved medical expert system feature representation [28]. The automatic organization is a developing paradigm. The research could improve future image processing systems' precision. Future image recognition studies for autonomous driving can improve the ability to perceive traffic signals in harsh environments.

References

[1] W. Xu, B. Li, S. Liu, and W. Qiu, “Real-time object detection and semantic segmentation for autonomous driving,” Feb. 2018, p. 44. doi: 10.1117/12.2288713.

[2] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision: A Brief Review,” Computational Intelligence and Neuroscience, vol. 2018, pp. 1–13, 2018, doi: 10.1155/2018/7068349.

[3] S. Zhou and W. Song, “Deep learning-based roadway crack classification using laser-scanned range images: A comparative study on hyperparameter selection,” Automation in Construction, vol. 114, p. 103171, Jun. 2020, doi: 10.1016/j.autcon.2020.103171.

[4] J. C. P. Cheng and M. Wang, “Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques,” Autom. Constr., vol. 95, pp. 155–171, Nov. 2018, doi: 10.1016/j.autcon.2018.08.006.

[5] M. Wang and J. C. P. Cheng, “A unified convolutional neural network integrated with conditional random field for pipe defect segmentation,” Computer-Aided Civil and Infrastructure Engineering, Jul. 2019, doi: 10.1111/mice.12481.

[6] M. Dildar et al., “Skin Cancer Detection: A Review Using Deep Learning Techniques,” Int. J. Environ. Res. Public. Health, vol. 18, no. 10, p. 5479, May 2021, doi: 10.3390/ijerph18105479.

[7] M. ur Rehman, S. H. Khan, S. M. Danish Rizvi, Z. Abbas, and A. Zafar, “Classification of Skin Lesion by Interference of Segmentation and Convolotion Neural Network,” in 2018 2nd International Conference on Engineering Innovation (ICEI), Jul. 2018, pp. 81–85. doi: 10.1109/ICEI18.2018.8448814.

[8] L. Chen, W. Zhan, W. Tian, Y. He, and Q. Zou, “Deep Integration: A Multi-Label Architecture for Road Scene Recognition,” IEEE Trans. Image Process., vol. 28, no. 10, pp. 4883–4898, 2019, doi: 10.1109/TIP.2019.2913079.

[9] G. Li et al., “ML-ANet: A Transfer Learning Approach Using Adaptation Network for Multi-label Image Classification in Autonomous Driving,” Chin. J. Mech. Eng. Ji Xie Gong Cheng Xue Bao Engl. Ed, vol. 34, no. 1, Dec. 2021, doi: 10.1186/s10033-021-00598-9.

[10] G. Li, Y. Yang, and X. Qu, “Deep Learning Approaches on Pedestrian Detection in Hazy Weather,” IEEE Trans. Ind. Electron., vol. 67, no. 10, pp. 8889–8899, 2020, doi: 10.1109/TIE.2019.2945295.

[11] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, pp. 6517–6525. doi: 10.1109/CVPR.2017.690.

[12] M. Hnewa and H. Radha, “Object Detection Under Rainy Conditions for Autonomous Vehicles: A Review of State-of-the-Art and Emerging Techniques,” IEEE Signal Process. Mag., vol. 38,

no. 1, pp. 53–67, 2021, doi: 10.1109/MSP.2020.2984801.

[13] M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised Image-to-Image Translation Networks.” arXiv, Jul. 22, 2018. doi: 10.48550/arXiv.1703.00848.

[14] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, “Domain Adaptive Faster R-CNN for Object Detection in the Wild,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 2018, pp. 3339–3348. doi: 10.1109/CVPR.2018.00352.

[15] N. A. M. Mai, P. Duthon, L. Khoudour, A. Crouzil, and S. A. Velastin, “Sparse LiDAR and Stereo Fusion (SLS-Fusion) for Depth Estimationand 3D Object Detection.” arXiv, May 28, 2021. doi: 10.48550/arXiv.2103.03977.

[16] G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, “A deep learning based image enhancement approach for autonomous driving at night,” Knowledge-Based Systems, vol. 213, p. 106617, Feb. 2021, doi: 10.1016/j.knosys.2020.106617.

[17] X. Cheng, P. Wang, and R. Yang, “Learning Depth with Convolutional Spatial Propagation Network,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 10, pp. 2361–2379, 2020, doi: 10.1109/TPAMI.2019.2947374.

[18] X. Ma, Z. Wang, H. Li, P. Zhang, W. Ouyang, and X. Fan, “Accurate Monocular 3D Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving,” in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6850–6859. doi: 10.1109/ICCV.2019.00695.

[19] P. Radecki, M. Campbell, and K. Matzen, “All Weather Perception: Joint Data Association, Tracking, and Classification for Autonomous Ground Vehicles.” arXiv, May 07, 2016. doi: 10.48550/arXiv.1605.02196.

[20] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, “Object Classification Using CNN-Based Fusion of Vision and LIDAR in Autonomous Vehicle Environment,” IEEE Trans. Ind. Inform., vol. 14, no. 9, pp. 4224–4231, Sep. 2018, doi: 10.1109/TII.2018.2822828.

[21] W. Boyuan and W. Muqing, “Study on Pedestrian Detection Based on an Improved YOLOv4 Algorithm,” in 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 1198–1202. doi: 10.1109/ICCC51575.2020.9344983.

[22] A. Hechri and A. Mtibba, “Two-Stage Traffic Sign Detection and Recognition Based on SVM and Convolutional Neural Networks,” IET Image Processing, Dec. 2019, doi: 10.1049/iet-ipr.2019.0634.

[23] J. Müller and K. Dietmayer, "Detecting Traffic Lights by Single Shot Detection," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 266-273, doi: 10.1109/ITSC.2018.8569683.

[24] X.-Y. Ye, D.-S. Hong, H.-H. Chen, P.-Y. Hsiao, and L.-C. Fu, “A two-stage real-time YOLOv2-based road marking detector with lightweight spatial transformation-invariant classification,” Image Vis. Comput., vol. 102, p. 103978, Oct. 2020, doi: 10.1016/j.imavis.2020.103978.

[25] S. Papadopoulos, I. Mademlis and I. Pitas, "Neural vision-based semantic 3D world modeling," 2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW), 2021, pp. 181-190, doi: 10.1109/WACVW52041.2021.00024.

[26] A. Kherraki, M. Maqbool, and R. El Ouazzani, “Traffic Scene Semantic Segmentation by Using Several Deep Convolutional Neural Networks,” in 2021 3rd IEEE Middle East and North Africa COMMunications Conference (MENACOMM), 2021, pp. 1–6. doi: 10.1109/MENACOMM50742.2021.9678270.

[27] S. Papadopoulos, I. Mademlis and I. Pitas, "Semantic Image Segmentation Guided By Scene Geometry," 2021 IEEE International Conference on Autonomous Systems (ICAS), 2021, pp. 1-5, doi: 10.1109/ICAS49788.2021.9551117.

[28] M. S. S. Mahecha, O. J. S. Parra, and J. B. Velandia, “Design of a System for Melanoma Detection Through the Processing of Clinical Images Using Artificial Neural Networks,” Lecture Notes in Computer Science, pp. 605–616, 2018, doi: 10.1007/978-3-030-02131-3_53.