Image Object Detection Algorithm for Autonomous Vehicles

Fangzhou Liu [1]

[1] University of Pittsburgh, Pittsburgh PA 15213, USA
Fal72@pitt.edu

Abstract. To improve the visual perception skills necessary for safe and effective operation, this thesis investigates the use of image object detection algorithms in autonomous vehicle systems. One essential part of the sensory framework of autonomous vehicles is object detection, which is the process of recognizing and locating different objects in the area around the vehicle. Three well-known algorithms—**:** Region-based Convolutional Neural Networks, You Only Look Once and Single Shot MultiBox Detector—that are each recognized for their distinct methods of processing and interpreting visual data are the main focus of this study's evaluation. An overview of the history of autonomous driving technologies is given at the outset of the study, with a focus on the importance of object detection for visual perception systems. The thesis compares the benefits and drawbacks of R-CNN, YOLO and SSD, focusing on detection accuracy, processing speed and adaptability to environmental changes. The performance of these algorithms in various driving scenarios is highlighted by the experimental results, which provide a solid assessment of their usefulness in autonomous driving. Aim to further enhance autonomous vehicle technologies by improving object detection capabilities, the conclusion reviews the research findings and makes recommendations for future developments and research directions.

Keywords: Machine Learning, Autonomous Vehicle, Object Detection

1 Introduction

One of the biggest developments in contemporary transportation technology is autonomous vehicles (AVs), which have the potential to completely transform daily commutes by lowering the risk of human error and raising traffic safety. The ability of visual perception systems to precisely detect and interpret the surrounding environment is essential to the operation of autonomous vehicles. One of the most important parts of these systems is object detection, which helps cars find and identify other cars, pedestrians, road signs, and other obstacles. It is impossible to overestimate the significance of strong object detection algorithms because they have a direct impact on how an AV's control system makes decisions. The foundation of vehicle autonomy is the algorithms' ability to provide accurate and dependable detections in a variety of unpredictable environmental conditions. In order to improve

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_26

object detection algorithms' accuracy and dependability in the context of autonomous driving, this research looks into how well they work. Techniques for object detection have significantly improved as a result of recent developments in deep learning. The three main object detection algorithms covered in this research are SSD (Single Shot MultiBox Detector), YOLO (You Only Look Once) and R-CNN (Region-based Convolutional Neural Networks). Each of these techniques has benefits, but they are all subject to certain restrictions regarding computational efficiency, detection precision, and processing speed. By means of comparative analysis and experimental validation, this research aims to improve autonomous vehicles' perception skills and open the door to safer, more effective roads where machine-driven and human vehicles coexist peacefully.

2 Theoretical Foundations of Object Detection

Object detection is a critical component of computer vision, enabling the recognition and localization of objects in digital images and videos. This technology is essential for applications like medical diagnostics, security surveillance, and autonomous vehicles (AVs). In AVs, object detection systems allow vehicles to perceive and interpret their surroundings, ensuring safe navigation. The main goal is to predict the bounding boxes of objects such as cars, pedestrians and traffic signs. Metrics like precision and recall measure the effectiveness of these systems. Efficiency is crucial, particularly in dynamic driving environments where quick decisions are vital. Challenges include partial occlusions, varying lighting conditions and different viewing angles, requiring robust handling. Deep learning has significantly advanced object detection, enhancing efficiency and reliability. Key algorithms include Single Shot MultiBox Detectors (SSD), You Only Look Once (YOLO) and Region-Based Convolutional Neural Networks (R-CNNs). R-CNNs generate region proposals and classify each region with deep networks, but they can be slow and computationally demanding. YOLO speeds up processing by dividing the image into a grid and predicting bounding boxes and class probabilities in a single pass, though it may struggle with small or obscured objects. SSD improves YOLO by predicting multiple bounding boxes and class probabilities at various scales directly from feature maps, effectively handling different object sizes and types. Each method has trade-offs between speed, computational load, and accuracy, which must be understood to develop robust object detection systems for AVs.

2.1 Analysis of R-CNNs

Region-based Convolutional Neural Networks (R-CNNs) create a basic framework for object detection by combining convolutional neural networks with region proposals. The fundamental R-CNN model uses a selective search algorithm to make initial approximations of object locations in an image. Each candidate region is processed independently by CNN to extract high-dimensional feature vectors. Linear regressors adjust the bounding box coordinates for each region, and support vector machines (SVMs) classify the feature vectors for each object class. Although R-CNN achieves high accuracy, its computational processing is too slow for real-time applications like autonomous driving. "Simulations results indicate that the SVM poorly performs, and its speed cannot assure real-time response, while the YOLO model and SSD can reach higher accuracy with a notable ability to detect objects in real-time when rapid driving decisions need to be made" [1]. Fast R-CNN improves R-CNN by introducing a simplified architecture that reduces computation time. Instead of processing each region proposal independently, Fast R-CNN processes the entire image once with a CNN to produce a feature map. Regions of interest (RoIs) are then proposed on this map. RoI pooling extracts a fixed-length feature vector from each RoI's feature map, which is fed into fully connected layers to generate class probabilities and bounding box coordinates. This method enhances efficiency and allows end-to-end training with a multi-task loss that maximizes classification accuracy and bounding box localization simultaneously. However, it still falls short of real-time processing rates due to the need for precomputed region proposals. Faster R-CNN further improves Fast R-CNN by integrating a Region Proposal Network (RPN) that shares the CNN-generated feature map, predicting object bounds and objectness scores simultaneously at each position. This integration creates a unified, efficient detection system with near real-time object detection. The RPN proposes regions using anchor boxes at each feature map position, which higher-level network layers refine and classify. This enhancement significantly speeds up the process and increases accuracy, making it suitable for complex, dynamic environments like autonomous vehicles. The Faster R-CNN process starts with high-resolution optical images for both training and detection. During training, images are cropped, and positive and negative sample images are selected and labeled to create datasets. These datasets are fed into the Faster R-CNN model for network training. For detection, the trained model processes high-resolution images to detect targets, such as ship targets. Network testing, evaluation, and further training refine the model's accuracy and efficiency. This comprehensive workflow, illustrated in Figure 1, highlights R-CNN's application in real-world scenarios like autonomous driving, where accurate object detection and classification are crucial for safety and operational efficiency.

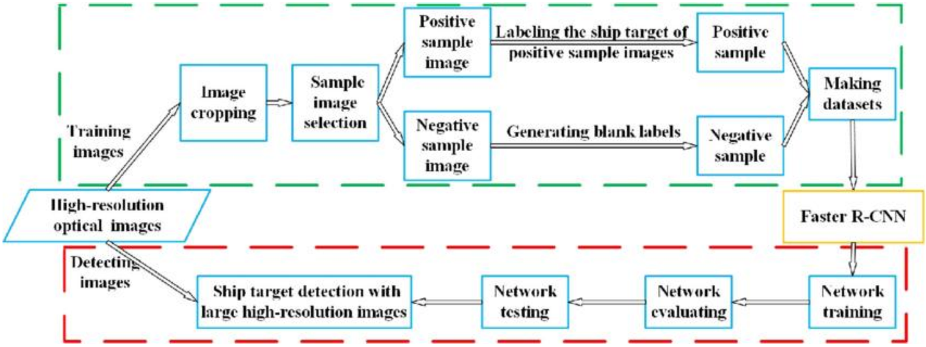

Fig. 1. Workflow of the Faster R-CNN algorithm.[2]

2.2 Analysis of YOLO

You Only Look Once (YOLO) transforms the notion of region-proposal-driven detection by reducing the problem to a single regression from image pixels to bounding box coordinates and class probabilities. The first step in YOLO's workflow, as illustrated in Figure 2, is processing an input image through a sequence of convolutional layers, each of which oversees extracting hierarchical features that get more complex over time. Each grid cell in the divided image directly predicts bounding boxes and related class probabilities. Multiple bounding boxes are predicted by this unified model for each grid cell, and each bounding box has a confidence score that indicates how likely it is that an object will be present as well as how accurate the bounding box is. Convolutional layers and pooling layers help to reduce the feature maps' spatial dimensions while maintaining the critical features required for precise detection. These features are combined in the last fully connected layer to generate the final output, which contains the class probabilities and bounding box coordinates for every grid cell. Because of its ease of use and effectiveness, YOLO can process images fast and efficiently, which is an essential feature for autonomous driving, where real-time processing is critical. Nevertheless, the grid-based method presents difficulties when trying to identify objects whose sizes don't precisely fit into the designated grid cells. This restriction may make it more difficult for the algorithm to recognize small or far-off obstacles, which are frequent in driving situations. Despite this flaw, YOLO is still a useful tool in the field of autonomous driving, especially in situations where speed is crucial, due to its quick and reasonably accurate detections. "The use of sensor fusion and Deep Neural Network (DNN) have played a predominant role in overcoming these limitations" [3].

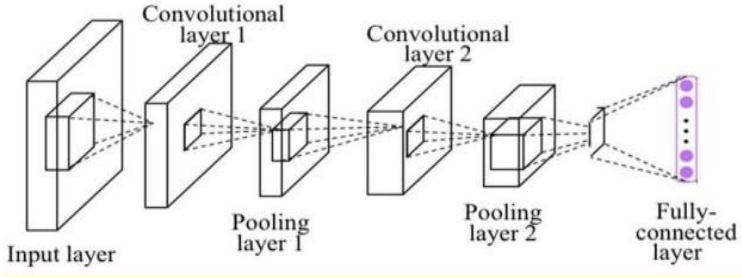

Fig. 2. Structure of a convolutional neural network [4].

2.3 Analysis of SSD

Some of the drawbacks of YOLO are mitigated in part by the Single Shot MultiBox Detector (SSD), which uses multiple feature maps at different scales to predict bounding boxes and class scores. The SSD architecture begins with a base network, such as VGG16, as shown in Figure 3, which processes the input image to produce a number of feature maps. The SSD can then detect objects at different scales by adding additional feature layers that gradually get smaller. The use of a fixed set of default bounding boxes over various aspect ratios and scales at each feature map location eliminates the need for a separate proposal generation step. The network learns to classify the contents of each box by modifying these default boxes during training so that they more closely resemble the shapes of the objects in the training data. SSD is especially helpful in autonomous vehicle environments where it is essential to precisely identify both large, nearby objects and smaller, distant objects. This multiscale approach allows SSD to detect objects of different sizes. The model performs better at detection across a broad range of object sizes because of the feature maps' varied resolutions, which improve its capacity to capture minute details and comprehensive contextual information. As seen in Figure 3, the architecture employs a series of convolutional layers followed by pooling layers in the VGG16 base network, which extracts high-level features from the input image. These features are then passed through additional convolutional layers that create feature maps of different scales, facilitating the detection of objects at various resolutions. Each of these layers is responsible for predicting class scores and refining bounding box coordinates. Detecting objects in dynamic and complex environments is crucial for autonomous vehicles, and SSD's ability to integrate multiple scales and directly predict bounding boxes makes this possible. This in-depth analysis of the model demonstrates the important developments and unmet obstacles still present in the object detection space. Autonomous vehicles need fast, accurate, and dependable object detection systems in order to safely navigate complex environments. Each model offers a unique strategy for striking a balance between these demands, supporting continued technological progress in this crucial area. "By adding shallow high-resolution features and changing the size of the output feature map, the detection ability of the algorithm for small objects is significantly improved" [5].

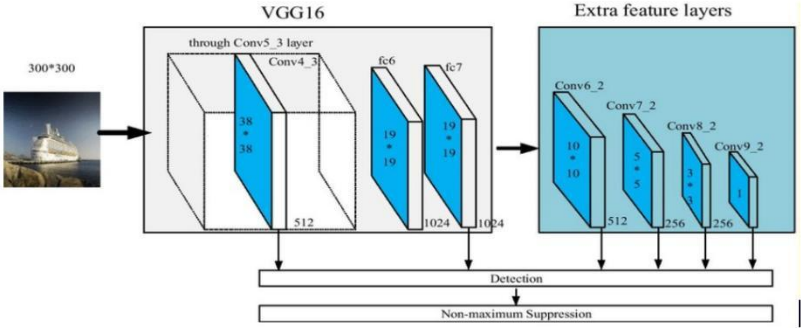

Fig. 3. Architecture of the VGG16-based object detection system [6].

3 Comparative analysis

R-CNN combines convolutional neural networks with region proposals to create a fundamental framework for object detection. It uses a selective search algorithm to produce initial guesses regarding object locations in an image, processes each candidate region with a CNN to extract high-dimensional feature vectors, and classifies these vectors using support vector machines (SVMs) while fine-tuning the bounding box coordinates with linear regressors. Despite achieving high accuracy, R-CNN is too slow for real-time applications like autonomous driving due to its computational intensity."The mean Average Precision of the improved YOLOv5s algorithm on BDD100K dataset increased by 3.2 percentage points, and the average detection speed is 74.6 FPS" [7]. Fast R-CNN improves upon R-CNN by reducing computation time; it processes the entire image once with a CNN, producing a feature map with suggested regions of interest (RoIs), which are then subjected to RoI pooling to extract fixed-length feature vectors. These vectors are fed into fully connected layers to generate class probabilities and bound box coordinates. However, the need for precomputed region proposals still limits its speed. "In order to evaluate the inference speed of MFDS, each function in MFDS was timed" [8]. Faster R-CNN incorporates a Region Proposal Network (RPN) that shares the CNN-generated feature map, predicting object bounds and objectness scores simultaneously, creating a unified, faster detection system suitable for near real-time applications. "The anchor-free FCOS detector is a slightly faster alternative to RetinaNet, with similar precision and lower memory usage" [9]. YOLO changes the object detection paradigm by treating it as a single, integrated regression problem, mapping image pixels to bounding box coordinates and class probabilities. It divides the image into a grid, with each cell responsible for predicting bounding boxes and confidence scores. This approach offers extremely fast processing but can struggle with small or overlapping objects. "Despite the rising popularity of one-stage detectors, our findings show that two-stage detectors still provide the most robust performance" [9]. Later iterations, YOLOv2 and YOLOv3, introduce improvements like anchor boxes, higher resolution inputs, batch normalization, multi-scale predictions, and a deeper network architecture to enhance accuracy while maintaining speed. "There is no general guideline for network architecture design, and questions of 'what to fuse', 'when to fuse', and 'how to fuse' remain open" [10]. The Single Shot MultiBox Detector (SSD) improves YOLO by using multiple feature maps at varying scales to predict bounding boxes and class scores, recognizing objects of different sizes more effectively. SSD uses a fixed set of default bounding boxes over different aspect ratios and scales at each feature map location, eliminating the need for a separate proposal generation step. It processes input images using a base network like VGG-16, applying smaller convolutional layers to predict default bounding boxes and class scores across various scales and aspect ratios, making it suitable for real-time applications in autonomous driving. "In recent years, deep learning has become the de-facto approach for object detection, and many probabilistic object detectors have been proposed" [11]. The advancements in these methodologies have significantly enhanced the performance and applicability of object detection systems in various real-world scenarios. "We compared the performance of object detection using FL to the traditional deep learning approach and noticed a significant difference between the two models" [12].

4 Conclusion and future expectations

Expectations for object detection algorithms rise as autonomous vehicle technology develops. Taking advantage of new technologies and resolving existing issues will be key to the future of autonomous driving object detection. Improved small- and occluded object detection, real-time processing efficiency, robustness to environmental variability, integration with advanced sensor fusion, deep learning and AI developments, scalability, and generalization are important areas for future development. Future algorithms must accurately recognize small and partially occluded objects under various conditions because urban environments are dense and dynamic. Autonomous vehicles require effective algorithms that balance speed and accuracy to make decisions quickly. Sustaining high performance requires robustness to a variety of environmental conditions, including changing lighting, weather, and complicated terrain. Combining data from several sensors, such as cameras, radar, and LiDAR, can improve the accuracy of object detection, so advanced sensor fusion techniques should be the focus of future research. More potent and effective object detection models may result from utilizing cutting-edge deep learning and artificial intelligence architectures and training techniques, such as transfer learning, reinforcement learning, and unsupervised learning. For widespread adoption, it is imperative to ensure scalability and generalization across diverse autonomous vehicle types and driving environments globally. "The Edge YOLO system can effectively avoid excessive dependence on computing power and uneven distribution of cloud computing resources" [13]. This thesis examined the theoretical underpinnings, important algorithms, and comparative analysis of object detection techniques for autonomous cars. It also closely looked at how R-CNN, YOLO, and SSD functioned and evaluated their advantages and disadvantages in relation to autonomous driving. This analysis demonstrated the vital role these algorithms play in allowing self-driving cars to recognize and safely navigate their environment. The field of object detection is still dynamic and fast developing, with new developments constantly pushing the envelope of what is feasible. Realizing the full potential of self-driving cars will depend heavily on the development of increasingly complex, precise, and effective object detection systems as autonomous driving technology advances. The insights gathered from this study advance the understanding of object detection and open the door to new developments that will eventually result in more dependable and safer autonomous cars. Future research and development efforts will continue to improve the capabilities of autonomous vehicles by concentrating on improving detection accuracy, real-time processing, environmental robustness, sensor fusion, and leveraging AI advancements. The road toward completely autonomous vehicles is still long, but the tremendous strides being made in this revolutionary field are demonstrated by the developments in object detection algorithms.

References

1. Xiao, B., Guo, J., He, Z.: Real-Time Object Detection Algorithm of Autonomous Vehicles Based on Improved YOLOv5s. In: 2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI), pp. 1–6. Tianjin, China (2021). doi: 10.1109/CVCI54083.2021.9661149
2. Gao, L., He, Y., Sun, X., Jia, X.: Incorporating Negative Sample Training for Ship Detection Based on Deep Learning. Sensors 19(684), 684 (2019). doi: 10.3390/s19030684
3. Ravindran, R., Santora, M. J., Jamali, M. M.: Multi-Object Detection and Tracking, Based on DNN, for Autonomous Vehicles: A Review. IEEE Sensors Journal 21(5), 5668–5677 (2021)
4. Pan, W., Duan, Y., Zhang, Q., Tang, J., Zhou, J.: Deep Learning for Aircraft Wake Vortex Identification. IOP Conference Series: Materials Science and Engineering 685, 012015 (2019). doi: 10.1088/1757-899X/685/1/012015
5. Masmoudi, M., Ghazzai, H., Frikha, M., Massound, Y.: Object Detection Learning Techniques for Autonomous Vehicle Applications. In: 2019 IEEE International Conferences on Vehicular Electronics and Safety (ICVES), pp. 1–5. Cairo, Egypt (2019). doi: 10.1109/ICVES.2019.8906437
6. Li, A., Zhu, X., He, S., Xia, J.: Water surface object detection using panoramic vision based on improved single-shot multibox detector. EURASIP Journal on Advances in Signal Processing 2021(1), 10.1186/s13634-021-00831-6 (2021)
7. Hoffmann, J. E., Tosso, H. G., Santos, M. D., Justo, J. F., Malik, A. W., Rahman, A. U.: Real-Time Adaptive Objection Detection and Tracking for Autonomous Vehicles. IEEE Transactions on Intelligent Vehicles 6(3), 450–459 (2021). doi: 10.1109/TIV.2020.3037928
8. Person, M., et al.: Multimodal Fusion Object Detection System for Autonomous Vehicles. ASME Digital Collection (2019)
9. Carranza-Garcia, M., et al.: On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data. MDPI (2020)
10. Feng, D., et al.: 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 (2021)
11. Feng, D., Harakeh, A., Waslander, S. L., Dietmayer, K.: A Review and Comparative Study on Probabilistic Object Detection in Autonomous Driving. IEEE Transactions on Intelligent Transportation Systems 23(8), 9961–9980 (2022)
12. Jallepalli, D., Ravikumar, N. C., Badrinath, P. V., Uchil, S., Suresh, M. A.: Federated Learning for Object Detection in Autonomous Vehicles. In: 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), pp. 107–114. Oxford, United Kingdom (2021). doi: 10.1109/BigDataService52369.2021.00018
13. Liang, S., et al.: Edge YOLO: Real-Time Intelligent Object Detection System Based on Edge-Cloud Cooperation in Autonomous Vehicles. IEEE Transactions on Intelligent Transportation Systems 23(12), 25345–25360 (2022)