Application of Small Object Detection in Autonomous Driving Abstract

Zichun Tang
Central South University, Changsha, Hunan, China

Abstract: This paper aims to provide a comprehensive review of the current application status and development trends of small object detection technology in the field of autonomous driving. With the rapid development of autonomous driving technology, small object detection, as a crucial part of environmental perception, is of great significance for improving the safety and reliability of autonomous driving systems. The article first introduces the definition, characteristics, and challenges of small object detection. Then, it elaborates on the deep-learning-based small object detection methods, mainly including data augmentation, multi-scale feature fusion, super-resolution, and context information utilization. Subsequently, the specific application scenarios of small object detection in autonomous driving, such as pedestrian, vehicle, and traffic sign detection, are explored. Finally, the existing challenges in current research are analyzed, and the future development directions are prospected, providing references for related research.

Keywords: **Small Object Detection; Autonomous Driving; Deep Learning; Environmental Perception; Computer Vision**

1. Introduction

With the continuous progress of autonomous driving technology, its requirements for the accuracy and comprehensiveness of environmental perception are becoming increasingly higher. Autonomous driving vehicles need to obtain detailed information about the surrounding environment in real - time, including the position, size, and category of various target objects, in order to make safe and reasonable driving decisions. Small - sized target objects, such as distant pedestrians, small obstacles, and traffic signs, are common in actual driving scenarios and are crucial for driving safety.

In recent years, with the rapid development of deep-learning technology, significant progress has been made in the accuracy and efficiency of small object detection. However, due to the characteristics of small objects, such as fewer pixels occupied in the image and indistinct features, their detection still faces many challenges. Hoping to offer references and inspiration for related research, this paper intends to explore the future development directions, review the application status of small object detection technology in the autonomous driving field, and analyze the characteristics and limitations of existing research methods.

In recent years, autonomous driving technology has witnessed remarkable progress in many aspects [1], attributed to the rapid development of artificial intelligence, sensor technology, and computing power. Autonomous driving vehicles are now capable of achieving partial or full - autonomous driving in complex environments. However, with the continuous expansion of autonomous driving application scenarios, the demand for small object detection has become increasingly prominent. Small objects, such as distant pedestrians, small obstacles, and traffic signs, are extremely common in the driving environment. Nevertheless, their low resolution, blurred features, and susceptibility to environmental interference make detection challenging. Accurately detecting these small objects is of utmost importance for ensuring the safety and reliability of autonomous driving systems. Particularly in urban roads and complex traffic scenarios, small object detection technology enables vehicles to promptly identify potential hazards and make rational decisions, thus effectively reducing the risk of accidents. Therefore, small object detection has emerged as one of the key technologies in autonomous driving environmental perception, and its development will directly impact the popularization and application of autonomous driving technology.

In autonomous driving, small - target detection is of utmost importance. It can accurately identify pedestrians, vehicles, and traffic signs at long distances or those with small sizes. Pedestrians move variably on the road, and small - target detection can promptly detect their movements to avoid collisions. Small vehicles are easily overlooked in complex traffic conditions, and small - target detection ensures their precise positioning to guarantee driving safety. Small - sized traffic signs carry crucial traffic information, and small - target detection can quickly and accurately recognize them, providing a basis for decision - making in autonomous driving and facilitating the safe and efficient operation of the autonomous driving system.

Accurately detect small targets to timely discover potential hazards on the road, such as distant small obstacles, suddenly intruding pedestrians or animals. Autonomous driving vehicles can then react in advance, take measures such as braking and avoiding, and thus avoid collision accidents, significantly improving the safety of autonomous driving.

Enhance the reliability and stability of autonomous driving. By optimizing small object detection algorithms, reduce the occurrence of false positives and false negatives, enabling autonomous driving vehicles to operate stably in various complex environments and enhancing users' trust in autonomous driving technology.

Expand the application scenarios of autonomous driving. Good small object detection capabilities can enable autonomous driving vehicles to better adapt to different road conditions and environments, thus expanding their application scenarios (such as logistics transportation and public transportation).

Promote the integration and development of multiple disciplinary fields. The application of small object detection in autonomous driving involves multiple disciplinary fields such as computer vision, machine learning, sensor technology, and automotive engineering. Research on small object detection technology can not only promote the development of these disciplines themselves but also facilitate their cross - integration, generating new technologies and methods.

Accurately detecting small targets can enable autonomous driving vehicles to better cooperate with other traffic participants, achieving more precise traffic flow control and management.

2. Basic Concepts and Challenges of Small Object Detection

2.1 Definition of Small Objects

The definitions of small objects generally consist of two categories. First, based on relative scale, if the size of an object is less than 0.1 times the size of the original image, it is considered a relatively small object [2]. Second, based on absolute scale, in the MS - COCO dataset, if the size of an object is less than 32×32 pixels, it is considered an absolute small object [3].

2.2 Technical Difficulties Faced by Small Object Detection

Low resolution and few available features. Compared with large/medium - sized objects, small objects have a lower resolution. After pooling operations, some characteristics of small objects may be removed, resulting in less information, making it difficult to extract discriminative features.

High - precision positioning requirements. Small objects are too small in the image and are extremely vulnerable to environmental interference. A one - pixel shift in network prediction can have a significant impact on small objects.

Small objects account for a small proportion in existing datasets. In the field of object detection, most existing datasets focus on large/medium - scale objects and pay less attention to the special type of small objects.

Small objects are difficult to annotate, with high labor costs and greater sensitivity to errors.

Sample imbalance. When training, a threshold is set to determine whether an anchor box belongs to a positive sample, which may lead to sample imbalance problems for objects of different sizes. Therefore, when the artificially set anchor box differs greatly from the true bounding box, the model may ignore the detection of small objects.

Small object aggregation. Small objects are more likely to aggregate. In this case, the predicted bounding boxes of the network model may filter out a large number of correct bounding boxes due to non - maximum suppression, resulting in missed small objects, or the bounding boxes are too close together, making it difficult for the model to converge.

Network structure. Current networks (e.g., Faster R-CNN) are optimized for medium/large objects; small object detection requires enhanced functionality, such as adding high-resolution feature layers in shallow modules. Coupled with the difficulties brought by the characteristics of small objects themselves, existing algorithms generally perform poorly in small object detection.

3. Small Object Detection Methods

Currently, small object detection methods are primarily improved based on mainstream object detection network models. To address the challenges of insufficient and uneven distribution of small object data, researchers have proposed various data augmentation techniques. These methods expand the data scale and improve distribution by performing operations such as transformation and synthesis on existing data. To overcome the weak representation ability of a single feature layer for small objects, multi-scale fusion methods are widely used. These methods integrate feature information at different scales, enabling the model to capture small-sized objects more effectively. Considering the weak visual features of small objects, super-resolution methods have emerged. By enhancing image resolution, these methods strengthen the visual features of small objects, making them easier to recognize. Given the limited feature information carried by small objects, context information learning methods aim to explore the context clues around small objects to enrich their feature expressions. To address the weak adaptability of prior anchor box sizes to small objects, the anchor box mechanism strategy is continuously optimized by adjusting the size and proportion of anchor boxes. Additionally, attention mechanism methods have been introduced to enable the model to focus more on the key features of small objects. Furthermore, small object detection methods designed for specific scenarios are constantly evolving, with detection models customized to improve accuracy and adaptability. Among these, data augmentation, multi-scale feature fusion, super-resolution, and context information utilization are the four most commonly used methods.

3.1 Based on Data Augmentation

Due to the small amount and uneven distribution of small object data, models may suffer from insufficient training and poor generalization ability. Data augmentation expands the data scale and enriches diversity by performing operations such as rotation, scaling, cropping, adding noise, and changing colors on the original data. This allows the model to learn more small object features and improves its recognition ability. At the same time, data augmentation can improve data distribution, avoid overfitting, and enhance the model's performance in small object detection tasks.

Kisantal[4] et al. increased the recognition and segmentation accuracy of small objects through oversampling and augmentation methods. They proposed three copy-paste strategies, demonstrating that copying and pasting each small object yields the best results [3]. In YOLOv4 [5], the Mosaic data augmentation method was introduced, which enriches the dataset by randomly splicing four different images. However, this method can reduce the size of small objects in the image, potentially decreasing the performance of small object detection.

Other notable data augmentation methods include MixUp, CutOut, and CutMix. MixUp linearly combines two training images and their labels, increasing data diversity and improving generalization. CutMix cuts out a part of one image and pastes it onto another, enabling the model to learn target features in different backgrounds. CutOut randomly selects rectangular areas in the image and sets their pixel values to zero, enhancing the model's robustness to missing parts of the target.

3.2 Based on Multi - Scale Feature Fusion

Multi-scale feature fusion leverages convolutional layers with different receptive fields or feature pyramid structures to extract features at various scales. Small-scale feature maps contain detailed information, while large-scale feature maps have stronger semantic information. By fusing these features, the model can better represent small objects and improve detection performance.

Lin [6] et al. proposed the Feature Pyramid Network (FPN), which combines high-resolution detail information with high-semantic information through a bottom-up and top-down path. FPN has been successfully applied to Faster R-CNN, significantly improving small object detection performance. However, FPN has limitations [7], such as the loss of high-level feature information and the semantic gap between features. To address these issues, Guo

[8] et al. proposed AugFPN, which further enhances detection accuracy.
Liu [9] et al. introduced the Path Aggregation Network (PANet), which improves upon FPN by adding a bottom-up path and adaptive feature pooling. Another notable method is the Adaptively Spatial Feature Fusion (ASFF)[10], which adaptively learns fusion spatial weights for feature maps at different scales, improving the speed and accuracy of single-stage detectors.

3.3 Based on Super - Resolution

Super-resolution techniques restore or enhance small objects in low-resolution images to high-resolution, improving their clarity and detail. This enhances the feature expression ability of small objects, making them easier to detect.
SRCNN [11], a pioneering algorithm, introduced deep learning into super-resolution by dividing the process into patch extraction, non-linear mapping, and reconstruction. ESPCN [12] uses sub-pixel convolution to increase image resolution while reducing computational cost. SRGAN [13], based on generative adversarial networks (GANs), generates high-resolution images that are as close as possible to real high-resolution images.

3.4 Based on Context Information Utilization

Context-based algorithms use the environment and scene around the target to assist in detecting small objects. These methods consider global information, semantic information, and spatial relationships to infer the existence and location of small objects, improving detection accuracy and recall.
The F-SSD [14] algorithm obtains context information through feature fusion, enhancing small object detection accuracy. The FA-SSD [15] algorithm combines feature fusion with an attention module, further improving performance by focusing on both the target and its context.

4. Application Scenarios of Small Objects in Autonomous Driving

4.1 Pedestrian Detection

Pedestrian detection mainly includes traditional pedestrian detection and deep - learning - based pedestrian detection. Traditional pedestrian detection mainly relies on hand - designed features for target representation, while deep - learning - based pedestrian detection uses deep - learning technology to identify and locate pedestrian targets from images or image sequences. Its algorithms can be roughly divided into two categories: anchor - box - based pedestrian detection and anchor - free pedestrian detection. Anchor - box - based pedestrian [16] detection uses the prior information of the dataset to set anchor boxes and obtains detection results through the classification and regression of convolutional neural networks. It specifically includes methods based on pedestrian parts (methods based on part detectors and attention mechanisms), methods based on the weighted combination of pedestrian wholes and parts, cascade - based methods (cascades of two - stage and single - stage detectors), and scale - based methods (methods based on feature fusion and scale adaptation). Anchor - free pedestrian detection [16] does not need to set prior box parameters and directly uses feature maps for classification prediction and position regression. The mainstream methods include point - based pedestrian detection methods (such as CSP, which detects pedestrians by predicting the heatmap of the center points of pedestrian targets and the height of the detection boxes) and line - based pedestrian detection methods (such as TLL, which divides pedestrian detection into three sub - tasks: predicting the upper vertices, lower vertices, and central axes of pedestrian targets).

4.2 Vehicle Detection

With the increase in vehicle demand, the number of small vehicles in urban areas has increased. Small vehicle detection faces challenges in complex background interference, and the blurry appearance of small vehicles increases the detection difficulty. Anchor - based detectors rely on specific regression anchors to identify targets when detecting small vehicles, and are prone to misclassification when encountering small objects with blurry appearances. To solve this problem, anchor - free detectors have emerged. They directly detect the key points of targets and generate bounding boxes, eliminating the complex pre - processing of anchor initialization. Although they are widely used in optical remote - sensing target detection, their detection performance for small vehicles still needs to be improved. Therefore, Liu [17] et al. designed an adaptive local context (ALC) embedding method and integrated it into the anchor - free detection method to form ALC - Net. This method can adaptively set effective local context features. Through multiple experiments on datasets such as UCAS - AOD [18] and VEDAI [19], it has been verified that ALC - Net has more advantages in small vehicle detection performance than other detectors.

4.3 Traffic Sign Detection

Currently, deep - learning algorithms are widely used in traffic sign detection technology. Traditional traffic sign recognition technology has poor performance in dealing with the influence of natural environmental factors, while deep - learning algorithms represented by convolutional neural networks have obvious advantages. They can extract features while training and have better performance and accuracy in recognizing traffic signs in complex situations. Convolutional neural network detection models are mainly divided into single - stage and two - stage models. The former can directly recognize the position and classify the target through a deep convolutional network, while the latter needs to complete the target candidate region extraction, coordinate correction, and target classification through two detectors successively. The two - stage model has high detection accuracy but slow speed. Deep - learning - based object detection algorithms are divided into two categories: candidate - region - extraction - based and end - to - end. The former was widely used in the early stage and has been continuously improved to increase detection efficiency and accuracy, but it still has problems with detection speed. The latter has excellent real - time detection capabilities, such as the continuously optimized YOLO series of algorithms. In addition, when traffic signs are small in the image, the detection ability of mainstream convolutional neural networks is weak. Therefore, improved algorithms such as combining feature pyramid networks are used to enhance the detection accuracy of small targets. In terms of the influence of complex environments, methods such as combining image enhancement with algorithms are used to improve the detection and recognition accuracy.

5. Conclusion

Small object detection is crucial for autonomous driving's environmental perception, enhancing system safety and reliability. Despite research progress, it still faces challenges in accuracy, efficiency, and robustness. These include issues like data alignment in multi - modal fusion, high computational costs in high - resolution processing, difficulty in designing efficient deep - learning methods, meeting real - time requirements while ensuring accuracy, constructing high - quality datasets, and integrating information in real - time. Ethical and safety concerns also pose challenges.

References

[1] Yin Fanqing, Li Chuanyou, Jiang Liangchao, et al. Development Status and Prospect of Autonomous Driving Vehicles [J]. Motorcycle Technology, 2018(11): 33 - 36.
[2] Liu Ying, Liu Hongyan, Fan Jiulun, et al. Review of Research and Application of Small Object Detection Based on Deep Learning [J]. Acta Electronica Sinica, 2020, 48(03): 590 - 601.
[3] Pan Xiaoying, Jia Ningxin, Mu Yuanzhen, et al. Review of Small Object Detection Research [J]. Journal of Image and Graphics, 2023, 28(09): 2587 - 2615.
[4] Kisantal M ,Wojna Z ,Murawski J , et al.Augmentation for small object detection.[J].CoRR,2019,abs/1902.07296
[5] Bochkovskiy A， Wang C Y and Liao H Y M. 2020. Yolov4： Optimal speed and accuracy of object detection. Computer Vision and Pat-tern Recognition [EB/OL]. [2020-04-23]
[6] LIN T Y, DOLLÁR P, GIRSHICK R B, et al.Featurepyramid networks for object detection[C]//2017 IEEEConference on Computer Vision and Pattern Recognition （CVPR），2017：936-944.
[7] Dong Gang, Xie Weicheng, Huang Xiaolong, et al. Review of Deep Learning - based Small Object Detection Algorithms [J]. Computer Engineering and Applications, 2023, 59(11): 16 - 27.
[8] GUO C，FAN B，ZHANG Q，et al.AugFPN：improvingmulti- scale feature learning for object detection[C]//2020 IEEE/CVF Conference on Computer Vision andPattern Recognition（CVPR），2020：12592-12601.
[9] LIU S，QI L，QIN H，et al.Path aggregation network forinstance segmentation[C]//2018 IEEE/CVF Conferenceon Computer Vision and Pattern Recognition，2018：8759-8768.
[10] Liu S T， Huang D and Wang Y H. 2019. Learning spatial fusion for single-shot object

detection. Computer Vision and Pattern Recognition [EB/OL]. [2019-11-25]. http://arxiv.org/pdf/1911.09516.pdf
[11] Dong C ,Loy C C ,He K , et al.Image Super-Resolution Using Deep Convolutional Networks.[J].CoRR,2015,abs/1501.00092
[12] WENZHE SHI, JOSE CABALLERO, FERENC HUSZÁR, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]//29th IEEE Conference on Computer Vision and Pattern Recognition: 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 26 June – 1 July 2016, Las Vegas, Nevada.:Institute of Electrical and Electronics Engineers, 2016:1874-1883.
[13] CHRISTIAN LEDIG, LUCAS THEIS, FERENC HUSZÁR, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition: CVPR 2017, Honolulu, Hawaii, USA, 21-26 July 2017, pages 1-751, [v.1].:Institute of Electrical and Electronics Engineers, 2017:105-114.
[14] Li Z , Zhou F .FSSD: Feature Fusion Single Shot Multibox Detector[J]. 2017.
[15] Liu Kuan, Lang Lei. Research on Lightweight SSD Object Detection Method [J]. Journal of Hubei Minzu University (Natural Science Edition), 2021, 39(4): 418 - 424.
[16] Luo Yan, Zhang Chongyang, Tian Yonghong, et al. Review of Deep - Learning - based Pedestrian Detection Methods [J]. Journal of Image and Graphics, 2022, 27(07): 2094 - 2111.
[17] Liu S, Zhuang Y, Dong H, et al. Adaptive Local Context Embedding for Small Vehicle Detection from Aerial Optical Remote Sensing Images[C]//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, Hangzhou: IEEE, 2022: 1712-1715.
[18] Zhu H, Chen X, Dai W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]//2015 IEEE International Conference on Image Processing (ICIP), Quebec City: IEEE, 2015: 3735-3739.
[19] Razakarivony S, Jurie F. Vehicle detection in aerial imagery: A small target detection benchmark[J]. Journal of Visual Communication and Image Representation, 2016, 34: 187-203.