International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 315-319 www.ijaem.net ISSN: 2395-5252

A Survey on Small Object Detection Based on Deep Learning

Avarnita Chauhan, Prof. Sapna Choudhary
[1]*Student, Shri Ram Group of Institutions, Jabalpur, M.P.*
[2] *Prof, Shri Ram Group of Institutions, Jabalpur, M.P.*

Submitted: 25-01-2022 Revised: 05-02-2022 Accepted: 08-02-2022

ABSTRACT:
With the improving of the intelligent driving awareness, object detection as an important part of intelligent driving, has now become a research hotspot in the world. In recent years, convolutional neural network (CNN) has attracted more and more attention in the field of computer vision. CNN has made a series of important breakthroughs in the field of object detection.
This paper introduces the object detection method based on deep learning. This paper mainly introduces the detection algorithm based on regional suggestion and regression, and analyses the advantages and disadvantages of the detection algorithm. Then, the disadvantages of these detection methods in detecting small objects and the difficulties in detecting small objects are analysed. On this basis, the public data sets and evaluation criteria related to small object detection are introduced.
Keywords: Detection of moving objects; tracking of moving objects; behavior understanding, Neural Network, Caffe model, CNN.

I. INTRODUCTION

Modern Small object detection is a fundamental computer technology related to image understanding and computer vision that deals with detecting instances of small objects of a certain class in digital images and videos. As an indispensable and challenging problem in computer vision, small object detection forms the basis of many other computer vision tasks, such as object tracking [1], instance segmentation [2, 3], image captioning [4], action recognition [5], scene understanding [6], etc. In recent years, the compelling success of deep learning techniques has brought new blood into small object detection, pushing it forward to a research highlight. Small object detection has been widely used in academia and real world applications, such as robot vision, autonomous driving, intelligent transportation, drone scene analysis, military reconnaissance and surveillance. There are mainly two definitions of small objects. One refers to objects with smaller physical sizes in the real world.

Another definition of small objects is mentioned in MS-COCO [7] metric evaluation. Objects occupying areas less than and equal to 32 × 32 pixels come under "small objects" category and this size threshold is generally accepted within the community for datasets related to common objects. Some instances of small objects (e.g. "baseball", "tennis" and traffic sign "pg") are shown in Figure 1.1.

Figure 1.1: Some instances of small objects.

Although many object detectors perform well on medium and large objects, they perform poorly on the task of detecting small objects. This is because that there are three difficulties in small object detection. First, small objects lack appearance information needed to distinguish them from background or similar categories. Then the locations of small objects have much more possibilities. That is to say, the precision requirement for accurate localization is higher. Furthermore, the experiences and knowledge of small object detection are very limited because the majority of prior efforts are tuned for the large object detection problem. In this paper we provide a comprehensive and in-depth survey on small object detection in the deep learning era. Our survey aims to cover thoroughly five respects of small object detection algorithms, including multi-scale feature learning, data augmentation, training

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 315-319 www.ijaem.net ISSN: 2395-5252

strategy, context-based detection and GAN-based detection. Aside from taxonomically reviewing the existing small object detection methods, we investigate datasets and evaluation metrics of small object detection. Meanwhile, we thoroughly analyse the performance of small object detection methods and present several promising directions for future work.

1.1. History and scope compared with other computer vision tasks, the history of small object detection is relatively short. Earlier work on small object detection is mostly about detecting vehicles utilizing hand-engineered features and shallow classifiers in aerial images [8, 9]. Before the prevalent of deep learning, color and shape-based features are also used to address traffic sign detection problems [10]. With the rapid advancement of convolutional neural networks (CNNs) in deep learning, some deep learning-based small object detection methods have sprung up. However, there are relatively few surveys and researches focusing only on small object detection. Most of the state-of-the-art methods are based on existing object detection algorithms with some modifications so as to improve the detection performance of small objects. To the best of our knowledge, Chen et al. [11] are perhaps the first to introduce a small object detection (SOD) dataset, an evaluation metric, and provide a baseline score in order to explore small object detection. Later, Krishna and Jawahar [12] build upon their ideas and suggest an effective upsampling-based technique that performs better results on small object detection. Different from the R-CNN (regions with CNN features) used in [11, 12], Zhang et al. [13] use deconvolution R-CNN [14] for small object detection on remote sensing images. Faster R-CNN [15] and single shot detector (SSD) [16] are two major approaches in object detection. Based on Faster R-CNN or SSD, some small object detection methods [17–18] are proposed.

1.2 Challenges in small object detection

Accurately detecting and tracking object from a video sequence is a challenging task because of the fact the object can have complicated structure and can change shape, size or orientation over subsequent video frames. Designing efficient and accurate system is always a big challenge. Some of the major challenges that arise while detecting small objects because of occlusions, short comings of capturing devices, variations in scenes or appearance, shadows of the objects appearing in the frame. These tend to degrade the performance of the developed algorithm and results of detections may be poor.

To overcome these challenges, the developed algorithm must take care of these issues while proposing solutions for specific applications. The major design issues are summarized in this section that can act as various challenges for researchers to focus upon.

1. Low quality of Image capturing tool results in generation of noisy or blurry image that in turn leads to false detection. The image could also be noisy because of weather conditions like rain, fog etc. The system shall be able to work in noisy images and able to detect the objects in videos with precise boundaries. The quality of camera deployed for capturing images need to be considered together with the weather conditions.

2. Camera jitter It makes the captured object look blurred with prolonged boundaries. The proposed system shall provide effective ways to handle **camera jitter** that occurs due to high velocity winds blowing at time of image capturing. The detection method shall be able to overcome this limitation.

3. Video captured on moving cameras like cameras installed on top of vehicles add another dimension to this already challenging area. The movement of camera needs to be simulated by the algorithm for effective and accurate object detections. The problem of moving object detection within moving camera is one of the most happening areas being explored by researchers.

4. Changes in illumination can occur because of presence or disappearance of a light source in background for eg bulb, tube light, sun etc Rapid Illumination changes in the scene of interest lead to false detections in consecutive frames or over multiple frames. The developed solution shall be able to work on different levels of illuminations.

II. LITERATURE REVIEW

2.1 Object detection

Object detection can be defined as the task of predicting the location and class of instances belonging to a specific class from an input image. This task has been actively researched in the domain of computer vision for some time. Deep learning has attracted much attention from various fields since a method using deep learning [19] won overwhelmingly in the classification task at ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) e 2012 compared with its counterparts in the same year and in previous competitions. This trend has continued, and many object detection algorithms using deep learning have been presented since Regions with CNN

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 315-319 www.ijaem.net ISSN: 2395-5252

features (RCNN) [20] was presented in 2014. Recent object detection algorithms can be classified into two main groups: ‘‘two-stage detection’’ such as RCNN and Faster-RCNN [21], in which object proposal and classification are done step by step, and ‘‘onestage detection’’ such as YOLO [22] and SSD [23], in which they are done simultaneously. A characteristic of detectors classified as ‘‘two-stage detection’’ is the high accuracy of the location of the resulting bounding boxes. In contrast, detectors classified in the ‘‘one-stage detection’’ group run much faster than those in the two-stage detection group. Some consumer-oriented drones can fly at a maximum speed of 72km=h (M 42.0S, 2019). For surveillance, a long processing time per frame leads to a delay in the detection of approaching drones. This means that countermeasures cannot be implemented promptly. Hence, object detection should preferably be performed quickly, such as at 10 fps which is the camera frame rate of the proposed system. For these reasons, detection processing speed is important for drone surveillance.

Deep learning is a transformative technology in machine learning, especially in computer vision. With the advantages of Convolutional Neural Networks in extracting high-level features of images, deep learning has achieved great success in the field of target detection. In 2014, Girshick et al. first proposed the deep learning model R-CNN based on convolutional neural network on CVPR 2014. In 2015, Girshick and Ren et al. proposed Fast-RCNN and Faster R-CNN algorithms. In 2016, Redmon et al. proposed the YOLO3algorithm. The detection speed is 10times faster than Faster R-CNN4. The extracted candidate regions are directly classified and regressed in the main network structure, and it will classify and locate the integrated thoughts provide new ideas for subsequent research. Based on the YOLO algorithm, in 2016, LIU and Redmon proposed the SSD algorithm. The SSD algorithm simplifies the entire process of object detection, integrates target determination and recognition, and greatly improves the operating speed. In 2020, YOLO v4 and YOLO v5 have been released one after another; introducing new data enhancement methods, and the detection speed and accuracy are greatly enhanced.

2.2 GAN-based detection

The generative adversarial networks (GAN) introduced by Goodfellow et al. [24] in 2014, has received great attention in recent years. GAN is structurally inspired by the two-person zero-sum game in game theory. A typical GAN consists of a generator network and a discriminator network, contesting with each other in a minimax optimization framework. The generator learns to capture the potential distribution of true data samples and generates new data samples, while the discriminator aims to discriminate between instances from the true data distribution and those produced by the generator. To the best of our knowledge, Li et al. [25] put forward a novel perceptual GAN model that made the first attempt to accommodate GAN on object detection task to boost small object detection performance via generating super-resolved representations for small objects to narrow representation difference of small objects from the large ones. Its generator learns to enhance the poor representations of the small objects to super-resolved ones that are similar enough to real large objects to fool its discriminator, while its discriminator competes with its generator to recognize the generated representation. Meanwhile, on the generator, its discriminator imposes an additional perceptual requirement that the generated representations must be beneficial for detecting small objects. Bai et al. [26] presented a multi-task generative adversarial network, namely MTGAN, in order to handle the detection problem of small objects. The generator in the MTGAN is a super-resolution network which up samples the small blurred images into fine-scale clear images. Unlike the generator, the discriminator in the MTGAN is a multi-task network. In the discriminator, each super-resolved image patch is described by a real or fake score, object category scores and regression offsets. Moreover, the bounding box regression and classification losses in the discriminator are back-propagated to the generator during training in order to make the generator obtain more details for more accurate detection. Extensive experiments show the superiority of the above two GAN-based detection methods in detecting small objects such as traffic signs, over state-of-the-art algorithms.

III. CONCLUSION

The main motivation of this paper is to provide an insight about detecting and curing small objects using data mining technique. For this paper we have surveyed various detection techniques with their advantages and limitations.

Thus, in an environment similar to that of the used this survey is a good selection to obtain a robust prediction model for small object detection.

REFERENCES

[1] K. Kang, H. Li, J. Yan, X. Zeng, B. Yang, T. Xiao, C. Zhang, Z. Wang, R. Wang, X. Wang, W. Ouyang, T-CNN: tubelets with convolutional neural networks for object detection from videos, IEEE Trans. Circ. Syst. Video Tech. 28 (10) (2018) 2896–2907.

[2] J. Dai, K. He, J. Sun, Instance-aware semantic segmentation via multi-task network cascades, Computer Vision and Pattern Recognition 2016, pp. 3150–3158.

[3] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 386–397.

[4] Qi Wu, Chunhua Shen, Peng Wang, Anthony R. Dick, Anton van den Hengel, Image captioning and visual question answering based on attributes and external knowledge, IEEE Trans. Pattern Anal. Mach. Intell. 40 (6) (2018) 1367–1381.

[5] S. Herath, M. Harandi, F. Porikli, Going deeper into action recognition: a survey, Image Vis. Comput. 60 (2017) 4–21.

[6] B. Zhou, Z. Hang, X. Puig, T. Xiao, S. Fidler, A. Barriuso, A. Torralba, Semantic understanding of scenes through the ADE20K dataset, Int. J. Comput. Vis. 127 (3) (2016) 302–321.

[7] T.Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C.L. Zitnick, P. Dollar, Microsoft COCO: common objects in context, European Conference on Computer Vision 2014, pp. 740–755.

[8] A. Kembhavi, D. Harwood, L.S. Davis, Vehicle detection using partial least squares, IEEE Trans. Pattern Anal. Mach. Intell. 33 (6) (2011) 1250–1265.

[9] V.I. Morariu, E. Ahmed, V. Santhanam, D. Harwood, L.S. Davis, Composite discriminant factor analysis, IEEE Winter Conference on Applications of Computer Vision 2014, pp. 564–571.

[10] T.T. Le, S.T. Tran, S. Mita, T.D. Nguyen, Real time traffic sign detection using color and shape-based features, Asian Conference on Intelligent Information and Database Systems 2010, pp. 268–278.

[11] C. Chen, M.-Y. Liu, O. Tuzel, J. Xiao, R-CNN for small object detection, Asian Conference on Computer Vision 2016, pp. 214–230.

[12] H. Krishna, C.V. Jawahar, Improving small object detection, Asian Conference on Pattern Recognition 2017, pp. 340–345.

[13] W. Zhang, S. Wang, S. Thachan, J. Chen, Y. Qian, Deconv R-CNN for small object detection on remote sensing images, IEEE International Geosciences and Remote Sensing Symposium 2018, pp. 2483–2486.

[14] R.B. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, IEEE Conference on Computer Vision and Pattern Recognition 2014, pp. 580–587.

[15] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, Adv. Neural Inf. Proces. Syst. (2015) 91–99.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S.E. Reed, C.-Y. Fu, A.C. Berg, SSD: single shot multibox detector, European Conference on Computer Vision 2016, pp. 21–37.

[17] C. Eggert, S. Brehm, A.Winschel, D. Zecha, A closer look: small object detection in faster R-CNN, International Conference on Multimedia and Expo 2017, pp. 421–426.

[18] C. Cao, B. Wang, W. Zhang, X. Zeng, X. Yan, Z. Feng, Y. Liu, Z. Wu, An improved faster R-CNN for small object detection, IEEE Access 7 (2019) 106838–106846.

[19] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097e1105.

[20] Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proc. Of the IEEE Conf. On Computer Vision and Pattern Recognition, pp. 508e587.

[21] Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91e99.

[22] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: Proc. Of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 779e788.

[23] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. Ssd: single shot multibox detector. In: European Conf. On Computer Vision. Springer, pp. 21e37.

[24] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative adversarial nets, Neural Information Processing Systems 2014, pp. 2672–2680.

[25] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, S. Yan, Perceptual generative adversarial networks for

small object detection, IEEE Conference on Computer Vision and Pattern Recognition 2017, pp. 1951–1959.

[26] Y. Bai, Y. Zhang, M. Ding, B. Ghanem, SOD-MTGAN: small object detection via multitask generative adversarial network, European Conference on Computer Vision 2018, pp. 210–226.