International Journal of Research Publication and Reviews, Vol 4, no 12, pp 3953-3956 December 2023

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Deep Learning Based Object Detection of Big and Tiny Objects

Gautam Dematti [a], Ahona Adhya [b], Divya Hariwal [c], Heena Ansari [d], Priyanka Mudhol [e]

[a,b,c,d,e] Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India

A B S T R A C T

Object detection is a computer-based technology that relates to image processing and computer vision that are basically used for detecting instances of objects of assured class task classification and localization in digital images, videos, action recognition, environment surveillance, sports video analysis, and scene understanding, etc. Small Object Detection is an important and interesting concept in computer vision. It has several real-world applications like autonomous driving, traffic monitoring, personal protective equipment (PPE), healthcare, retail. There are many different object detection algorithms, such as R-CNN, Fast R-CNN, YOLO, and Single Shot MultiBox Detector (SSD). These algorithms use a variety of techniques to accurately identify and locate objects inside an image, video, etc. We also evaluate the literature on crucial SOD tasks, including, small pedestrian detection, small face detection and aerial image object detection. Incisive run-through of benchmark datasets and evaluation techniques used in detection is also provided along with some of the notable backbone architectures used in identification and recognition tasks. We provide an acute assessment of the advantages and limitations of the models. This project also includes backbone networks like CNN, loss functions and training strategies, classical object detection architectures, datasets and evaluation metrics, future development instructions, as well as a review and analysis of deep-learning based object detection techniques conducted in priorly by researchers. This article surveys and evaluates recent developments in deep learning-based object detection techniques.

Keywords: Deep Learning, Object Detection, Convolutional Neural Network, Datasets

Introduction

Object detection (OD) stands as a fundamental task that underpins various other computer vision assignments, including object tracking, instance segmentation, action recognition, environment surveillance, and sports video analysis, among others. The robust feature-learning capabilities of deep convolutional neural networks (CNNs) have fuelled substantial growth in object detection research over the past decade. Deep learning-based approaches for object detection can be categorized into two main types: one-stage models and two-stage models. Noteworthy two-stage models comprise R-CNN, Fast R-CNN, Faster R-CNN, SPP-Net, and feature pyramid network (FPN) models. These models initiate by generating a region of interest (ROI) in the first stage and subsequently fine-tune the ROI to classify objects and determine bounding box localization in the second stage. On the other hand, one-stage models such as YOLO, SSD, and various anchor-free models like feature selection anchor-free module (FSAF), CornerNet, FCOS, and CenterNet, directly classify and localize objects from the feature map without the need for an intermediate (region of interest) ROI stage.

Early models for object detection relied on assembling handcrafted feature extractors like the Viola-Jones detector and Histogram of Oriented Gradients (HOG). Unfortunately, these models were characterized by sluggish performance, low accuracy, and struggled when faced with unfamiliar datasets. The paradigm shift occurred with the resurgence of convolutional neural networks (CNNs) and deep learning for image classification, catalysed by the success of AlexNet in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012. This breakthrough inspired extensive research in applying CNNs to computer vision. Today, object detection plays a pivotal role in diverse applications ranging from self-driving cars and identity recognition to security and medical domains. In recent years, the field has experienced exponential growth with the rapid development of new tools and techniques.

Detecting small objects poses a considerable challenge in computer vision due to the diminutive representations of objects and the inherent diversity in input images. The task becomes even more daunting when images vary in resolution, potentially hindering the detection of small objects, particularly in cases of low resolution where visual information is significantly limited. Complicating matters, small objects may be deformable or obscured by other objects. While various detection methods, particularly those rooted in deep learning, have been proposed over the years, their focus has primarily been on normal-sized objects rather than small ones. Consequently, evaluating small object detection approaches becomes indispensable for advancing object detection studies. Recent cutting-edge approaches, trained initially on ImageNet and then applied to detection tasks, have shown promise in addressing challenges posed by datasets like PASCAL VOC and COCO. Notably, our prior work extensively evaluates three state-of-the-art models—You Only Look Once (YOLO), Single Shot MultiBox Detector (SSD), and Faster R-CNN—considering factors such as accuracy, execution time, and resource constraints. In contrast to previous efforts, our evaluation extends beyond real-time models, encompassing both one-stage (e.g., YOLOv3, RetinaNet) and two-stage approaches (e.g., Fast RCNN, Faster RCNN) to provide a comprehensive understanding of their strengths, weaknesses, and design considerations.

Challenges

In addition to the common challenges in object detection, such as continual object detection, imbalance problems, etc. There are typical challenges when it comes to SOD, including feature representation with noise, small object information loss, the effect of the receptive field, location variation sensitivity and the scarcity of small object datasets. Feature representation with noise. The features of small objects are often contaminated by noise in the background after CNN implementation, making it difficult for the network to capture the discriminative information that is pivotal for the localization and classification tasks. Besides, small objects are often occluded and clustered, so it is particularly difficult to distinguish small objects from noisy clutter and precisely locate their boundaries. Small object information loss.

The features of a small object are virtually eliminated after the down-sampling operations in deeper neural networks due to the small number of pixels occupying each small object. The weak information wipeout of small objects is fatal to SOD because it is hard for the detection head to give accurate predictions in the presence of highly structural representations. Effect of the receptive field. Large receptive fields are typically chosen by deep neural networks to prevent the loss of information. However, the receptive field for the prediction low-resolution feature map may not match the size of small objects. If the receptive field is larger than the small object, it will cause the object to be detected to become the background, and no features will be extracted by backbone networks, resulting in poor SOD performance. Location variation sensitivity.

Small location deviation of the bounding box in the IoU-based metric produces a more significant disturbance for small objects than for larger objects, which makes it difficult to find a suitable IoU threshold and deliver high-quality positive and negative samples to train the networks. Scarcity of small object datasets. There are still not enough large-scale general small object datasets to match the cost of annotating small objects. Although MS COCO has a reasonably large number of small objects (31.62%), each image has too many instances, which leads to the uneven distribution of small objects.

Overall, there are several problems relating to challenges that need to be solved with object detection. Object detection itself draws much attention from researchers, but after a period, challenges just tackle a part; particularly, COCO challenges provide a standard regarding small and medium detection, and accuracy in most of detectors is still low with this standard. (Therefore, in terms of small object detection, it is harder to researchers because apart from normal challenges alike object detection, it owns challenges for small objects. Besides, the definition of small objects is not obviously clear. (e following presentation make it more obvious.

Future Directions

According to the challenges of SOD and the analysis of performance results, we discuss several potential directions for future research in SOD:

1) Weakly supervised, unsupervised, and self-supervised SOD. Existing deep learning-based SOD techniques use a fully supervised model. For model training, a sizable number of images with bounding-box annotations (fully supervised information) are required. However, the annotation work is labor-intensive and time-consuming. Weakly supervised object detection can use image-level labels (such as image categories) as supervised signals to train object localization models without the need for pixel-level annotation, which lessens the workload associated with the annotation.

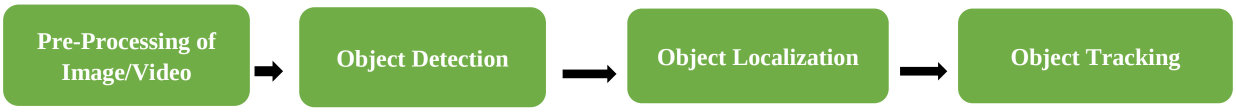
2) Suitable metric for SOD. IoU-based metrics, including the original IoU and its extensions (DIoU, GIou, etc.), are extremely sensitive to the position deviation of small objects and significantly reduce the detection performance when utilized in anchor-based detectors. The authors of use a new Wasserstein distance-based SOD metric, which outperformed the standard fine-tuning baseline by an AP value of 6.7 AP, as well as the state-of-the-art SOTA model by an AP value of 6.0. Therefore, designing a suitable metric for small objects will be crucial to further research.

3) Multi-task joint optimization. Even though techniques like scale-aware training strategies, incorporating contextual information, data augmentation and increasing the resolution of input features help to improve SOD performance, they are still far from adequate, and the combined use of these methods may be able to further improve SOD performance.

4) Open world or few-shot SOD. Few-shot object detection has produced prominent achievements, and SOD in the few-shot scenario is also in urgent need of solutions. Open world SOD seeks to overcome the SOD conundrum while enabling incremental learning in the model, and this type of issue will be a significant research topic in the future.

Stages of an Object Detecting Framework

Object detection and object classification are two phases that preface the tracking process and serve a critical role in enhancing tracking accuracy. Although this distinction between the three is minor and might be overlooked upon the first glimpse, it is critical to recognize them since each is distinct and requires its own research. The initial step determines which objects are visible in the video frame. Next, contingent on what we want to track, we need to categorize these objects. Afterwards, the actual tracking commences. The three phases are explained as follows, along with the various strategies utilized for each category.



- **Pre-Processing of Image/Video:** An image is the general impression of representing a thing, person, etc. A video is a collection of frames. Each frame represents a unique state of the object's state. Detect an object from a frame and track a specific object through a video series. A video is a collection of frames. Each frame represents a unique state of the object's state. Detect an object from a frame and track a specific object through a video or image series.
- **Object Detection:** Object detection refers to computer technology that detects the presence of semantic objects related to a certain category (for example, cars, buildings, people, etc. in videos and digital images. Several basic object detection strategies include background subtraction, optical flow, and frame extraction.
- **Object Localization:** Object localization is a procedure in which objects detected in a frame are classified according to which objects of interest they are. It is just a matter of figuring out what type of object it is. Several factors such as texture, color, motion, and shape can be used to identify an object. Depending on the parameter used, we can perform texture-based classification, color-based classification, motion-based classification, or shape-based classification.
- **Object Tracking:** Object tracking is a technique for determining how an object moves relative to other things by tracking it through successive image frames. The most common way is to measure the displacement of the center of gravity of the object at points (x, y) in successive frames. Kernel-based tracking, point-based tracking and silhouette-based tracking are three types of object tracking.

Illustrations

All figures should be numbered with Arabic numerals (1,2, 3,). Every figure should have a caption. All photographs, schemas, graphs, and diagrams are to be referred to as figures. Line drawings should be good quality scans or true electronic output. Low-quality scans are not acceptable. Figures must be embedded into the text and not supplied separately. In MS word input the figures must be properly coded. Lettering and symbols should be clearly defined either in the caption or in a legend provided as part of the figure. Figures should be placed at the top or bottom of a page wherever possible, as close as possible to the first reference to them in the paper.

The figure number and caption should be typed below the illustration in 8 pt and left justified [***Note:*** one-line captions of length less than column width (or full typesetting width or oblong) centered]. For more guidelines and information to help you submit high quality artwork please visit: http. Artwork has no text along the side of it in the main body of the text. However, if two images fit next to each other, these may be placed next to each other to save space. For example, see Fig. 1.

Fig. 1 - (a) Object Detection;(b) Detection with accuracy.

Comparison Table

Sl. No	Title	Authors	Advantages	Disadvantages	Accuracy
1	A Survey of Deep Learning-based Object Detection.	Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi and Rong Qu, Senior Member	Offers an extensive implementation technique for detection of tiny objects.	Neglects basics of object detection as it focuses only on small object detection.	78.4%
2	An Evaluation of Deep Learning Methods for Small Object Detection.	Nhat-Duy Nguyen, Tien Do, Thanh Duc Ngo, and Duy-Dinh Le	Greater Accuracy and time processing along with proper resource consumption and output of results.	After a period, challenges just come through a part; particularly, COCO challenges provide a standard regarding small	73.6%

				and medium detection, and accuracy in most of detectors is still low with this standard.	
3	Deep – learning based detection from the perspective of small and tiny objects.	Kang Tong, Yiquan Wu	Does not include or explain about small object detection.	Includes various algorithms that can be used for detection of small objects.	88.7%
4	A survey of modern deep learning-based object detection models.	Syed Sahil Abbas Zaidi, Mohammad Samar Ansari b, Asra Aslam, Nadia Kanwal, Mamoona Asghar, Brian Lee	Talks about lightweight object detection models that is necessary for small and tiny objects.	Talks about lightweight object detection models that is necessary for small and tiny objects. Does not explain the datasets or computer models used.	93.3%
5	A survey of deep learning-based small object detection	Qihan Feng, Xinzheng Xu, and Zhixiao Wang	Tell about several things related to small object detection.	Complex description of methods and techniques used for object detection.	77.3%

Table 1.1 Comparison Table

4. Conclusion

A comprehensive description of all proposed object tracking algorithms for deep learning (DL) techniques is presented focusing on four main steps to characterize a general object tracking pipeline: data processing, object detection, target location and tracking. The use of DL in object tracking has been investigated. Although most of the curriculum focused on various DL applications of learning and application of flow characteristics, there were some exceptions. However, DL is only used in several approaches to derive the correlation algorithm Object detection is currently being researched and the entire field of artificial intelligence (AI) is largely being built in the data. While the demand for object tracking applications includes all the real problems, it is a reality. The performance of the application is not sufficient and this is due to the lack of data collection capabilities for construction and development models The information presented in this study can be considered to help researchers provide an overview of the current situation advances in human object tracking, autonomous driving, and medical detection.

References

[1] I. Ahmed and G. Jeon, A real-time person tracking system based on SiamMask network for intelligent video surveillance, J. Real-Time Image Process. 18 (2021), no. 5, 1803–1814.

[2] A.F. Al-Battal, Y. Gong, L. Xu, T. Morton, C. Du, Y. Bu, I.R. Lerman, R. Madhavan and T.Q. Nguyen, A CNN segmentation-based approach to object detection and tracking in ultrasound scans with application to the vagus nerve detection, Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, 2021, pp. 3322–3327.

[3] D. Anitta, Human head pose estimation based on HF method, Microprocess. Microsyst. 82 (2021), 103802.

[4] M.R. Bachute and J.M. Subhedar, Autonomous driving architectures: insights of machine learning and deep learning algorithms, Mach. Learn. Appl. 6 (2021), 100164.

[5] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi and P.H.S. Torr, Fully-convolutional siamese networks for object tracking, Eur. Conf. Comput. Vis. Springer, Cham., 2016, pp. 850–865.

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, Yolov4: optimal speed and accuracy of object detection, arXiv Prepr. arXiv2004.10934, (2020).

[7] J. Chen, C. Zhang, J. Luo, J. Xie and Y. Wan, driving manoeuvres prediction based autonomous driving control by deep monte Carlo tree search, IEEE Trans. Veh. Technol. 69 (2020), no. 7, 7146–7158.

[8] H.-K. Chiu, J. Li, R. Ambrus, and J. Bohg, Probabilistic 3d multi-modal, multi-object tracking for autonomous driving, IEEE Int. Conf. Robotics and Automation (ICRA), 2021, pp. 14227–14233. A survey of deep learning-based object detection: Application and open issues 1503

[9] G. Ciaparrone, F. Luque S´anchez, S. Tabik, L. Troiano, R. Tagliaferri and F. Herrera, Deep learning in video multi-object tracking: a survey, Neurocomput. 381 (2020), 61–88.

[10] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, Y. Li and D. Cao, Deep learning for image and point cloud fusion in autonomous driving: a review, IEEE Trans. Intell. Transp. Syst. 23 (2022), no. 2, 722–739.