ICSSR Sponsored - International Conference on Need of Digital Transition and CSR For Innovative Sustenance in the 2023 Economy
© 2023 IJSRST | Volume 10 | Issue 11 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
International Journal of Scientific Research in Science and Technology

Deep Learning Methods for Generic Object Detection: A Review

Shereena V B[1], Greeshma P V[2], Hafisa Beegum M C[2]
Prabal Sharma[1], Rinkey Sahu[2]
[1]Assistant Professor, [2]M.Sc. Student
Department of Computer Science, MES College Marampally, Kerala, India

ABSTRACT

Object detection is the most essential and challenging task in computer vision and image processing. The most important application in object detection are medical diagnosis, civil military, security, video surveillance etc. Deep learning techniques have come up as a power full strategy for object detection. The field of generic object detection has experienced significant advancement and the emergence of deep learning techniques, which offer a potent approach to learning feature representation from data. In recent years, deep learning approaches have significantly advanced the state-of-the-art in object detection. The main aim of this paper is to provide a comprehensive review of the rapid evolution in object detection using deep learning techniques. The paper also emphasizes the challenges and opportunities in object detection using deep learning and the future directions in this exciting field.

KEYWORDS: Generic Object Detection, Deep Learning, Convolutional Neural Network, Deep convolutional Neural Network

I. INTRODUCTION

Object detection is the one of the major and challenging problem in computer vision. The major goal of object detection is identifying images which are belongs to such as human, dogs, cats, bicycles etc [1]. It is a computer vision technique or task. Which works to determine or locate object within an image or video. object detection is completely interlinked with other similar computer vision techniques such as image segmentation and image recognition that assist as to understand and analyse the scenes in videos and image. [2] Object detection is used for solving such high level or complex tasks such as scene understanding, segmentation, object tracking, image captioning, activity recognition and event detection [3].
The first deep neural network for object detection was overfeat. Here introduced a multi-scale sliding window approach using CNN's and showed that object detection also improved image classification. They were shortly followed by R-CNN region with CNN features. Later proposed spatial pyramid pooling networks. Spatial pyramid pooling layer, which enables a CNN to generate a fixed-length representation regardless of the size of image/region of interest without rescaling it. [1] In 2017 T-Y Linetal proposed feature pyramid Network on basis of faster R-CNN. EPN has now become a basic building block of many latest detectors. And then YOLO was proposed by R. Joseph at al., in 2015. YOLO is extremely fast. In 2015 W. Liu at al proposed SSD. The main

contribution of SSD is the introduction of the multi-reference and multi-resolution detection techniques, after that in 2017 T-Y-Lin et al has discovered RetinaNet. They claimed that the extreme foreground-background class imbalance encountered during training of dense detectors is the central cause [1].

Object detection involves two key tasks: object localization, which entails identifying the location of objects within an image, and object classification, which involves determining the category or type of each object.[2] To achieve these objectives, traditional object detection models follow a three-stage pipeline. This pipeline involves selecting informative regions within the image, extracting relevant features from these regions, and finally classifying the object based on the extracted features. This approach has proven to be highly effective in accurately detecting and classifying object in various contexts.[2]

The rest of this paper is structured in the following manner. Literature review is summarized in section 2. Scope of object detection is discussed in section 3. Disadvantages in object detection is summarized in section 4. Section 5 explains the dataset for implementation. And section 6 concludes the work.

II. LITERATURE REVIEW

Li Liu et al., [1] presented a survey of recent achievements in object detection by deep learning techniques. This survey includes more than 300 research contributions. This survey focuses on major progress of the last 5 years, and restricting attention to still pictures, leaving the important subject of video object detection as a topic for separate consideration in the future. The demerit of this method is high quality detection must accurately localize and recognize objects in image or video frames.

Zhong-Qiu Zhao et al., [2] reviews deep learning-based object detection frameworks, covering generic architectures and modifications to improve performance, as well as specific tasks like salient object, face, and pedestrian detection. Experimental analyses are included, and future directions are suggested.

Xiongwei Wu et al., [3] provided a thorough analysis of the recent developments in deep learning-based visual object detection through a comprehensive survey. The survey covers a vast range of literature on the topic, divided into three major parts detection components, learning strategies, and application and benchmark. The factors that influence detection performance, such as detector architecture, proposal generation, feature learning, and sampling strategies. Furthermore , the article suggests future directions for research to further advance visual object detection using deep learning.

Syed Sahil Abbas Zaidi et al., [4] presented a survey of the latest advancement in deep learning-based object detection. It provides a brief summary of the benchmark datasets and evaluation metrics used for detection, as well as highlighting the key backbone architectures used for recognition tasks. Additionally, the article explores contemporary lightweight classification models that are suitable for edge devices. Finally, the performances of these architectures are compared across multiple metrics to determine their effectiveness in object detection.

The survey commenced by Enoch Arulprakash and Martin Aruldoss [5] highlight the key aspects of deep learning applied to object detection. It presents a thorough examination of object representation, focusing on the application of Convolutional Neural Network and various Deep Convolutional Neural Network architecture. The review provides a brief overview of well-known dataset and standard measurement metrics, which serve as the fundamental benchmarks to assess the efficiency of the detection framework. The survey then delves into an in-depth analysis of one-stage and two-stage detection framework and evaluates each frameworks significance with standard datasets. The study further explores various challenges faced in object detection, including multi-scale, intra-class variations, generalization, and security. The survey outlines the essential steps involved in creating an object detector for different conditions based on the reviewed literature.

A study by Nipun D Nath,Amir H Behzadan [6] evaluates YOLO-based CNN models for fast detection of construction objects using a large-scale dataset. Results show that YOLO-v3 outperforms YOLO-v2, and a combination of crowdsourced and web-mined images improves performance. The proposed methodology can also predict the relative distance of detected objects. The study lays the groundwork for technology-assistive systems to aid human interpretation of visual data in complex environment.

Mouna Afif, et al., [7] proposed Deep Convolutional Neural Network for indoor object detection for visually impaired persons and a new indoor dataset are proposed. The novel DCNN design is based on a pre-trained DCNN called YOLO-v3 to test the proposed DCNN. Experimental results prove the high performance of the proposed indoor object detection as its recognition rate is 73.19%. Dataset provides image that are high relevant for VIP mobility. The challenge faced is the need of huge amount of data.

Fereshteh S. Bashiri et al., [8] performed a study whose objective was to create computer vision algorithm using deep neural networks that could aid visually impaired individuals in navigation clinical settings by detecting important landmarks such as stairs, doors, and signs with high accuracy. The results of quantitative experiments indicated that the network was capable of accurately recognizing these objects of interest with over 98% precision within a very short time frame, provided that a sufficient number of training samples were available.

Mouna Afif at al., [9] introduced a novel classification system for recognizing indoor objects, utilizing a Deep convolutional Neural Network model that can be implemented on mobile embedded platforms. The proposed approach was evaluated on natural images from the MCIndoor 20000 dataset, captured under natural illumination conditions, and the experimental results demonstrate that the system achieves nearly perfect accuracy in classifying indoor objects, with an accuracy rate of almost 100%.

Mouna Afif at al., [10] developed a novel indoor object detector using a deep Convolutional Neural Network framework. The framework was constructed using the RetinaNet Neural Network architecture and evaluated using ResNet, DenseNet, and VGGNet, to enhance detection accuracy and processing speed. The results of the evaluation were quite promising, with a detection precision of up to 84.61% mAP. These findings suggest that the proposed approach could be effective for indoor object detection.

Bernardo Calabrese, et al., [11], presented a solar-powered wearable device that uses a camera, a computing unit, and an ultrasonic sensor to recognize objects for visually impaired people. The device provides auditory feedback about the objects location and identity, achieving an 86% correct recognition rate. The device can recognize 91 objects and custom objects and human faces. The researcher introduced a scalable methodology for training the device to recognize new objects using image datasets. Comprehensive training with 100 images achieved 89% recognition, while fast training with 12 images achieved 55%.

Fahad Ashiq et al., [12] presented a system to assist visually impaired individuals with navigation and safety. It uses automated voice and a web-based app to share location with family. The device combines multiple features and outperforms existing devices with a score of 9.1/10. The system employs Mobile Net architecture and a deep CNN model with an 83.3% accuracy rate. Six pilot studies show satisfactory results.

Xintao Ding et al., [13] proposed a Convolutional Neural Network based pipeline for indoor object detection. The pipeline involves pre-training an offline CNN model using both public indoor dataset and private frames of videos dataset. A selective search process is then employed to extract regions of interest from the input video frames, which are classified into candidates using the pre-trained deep model. The candidates are refined using detection fusion between the nearest frame images, and the annotated frames are merged to produce the final output video. Experimental results demonstrate the high efficiency of our pipeline for indoor object detection.

Y.C. Wong et al., [14] proposed a smart objects detection system for visually impaired people using a Convolutional Neural Network and audio –based detector. Edge box algorithm was used to reduce complexity, and fine-tuned CaffeNet model was employed . Results were evaluated using mean average precision and frame-per-second, showing that the single Shot Multibox Detector achieved higher accuracy and faster speed compared to fast R-CNN.

The study by Mohamed L. Mekhalfi et al., [15] aimed to enhance object recognition for the visually impaired by including multiple objects and sacrificing detail. Two labelling strategies, using Euclidean distance or Gaussian process estimation, are used with compressed image representation for faster processing. Promising results were obtained on two indoor datasets.

Xintao Ding et al., [16] proposed a prior knowledge-based deep learning method for robot object recognition. A CNN is trained using a combined dataset. Mean images and scene knowledge are used as prior knowledge to improve detection accuracy. The method is applicable to video object recognition and has potential for robot vision.

Jaskirat Kaur and Williamjeet Singh [17] provided an in-depth literature review of object detection techniques, following a systematic review approach to summarize current research findings and address seven research questions. Contribution to the research includes analysing traditional, two-stage, and one-stage detection techniques, dataset preparation and standard dataset, annotation tools, and performance evaluation metrics. Here also conducted a comparative analysis of the proposed techniques, which differ in architecture, optimization function, and training strategies. The success of deep neural networks in object detection has significantly improved detector performance. This study also discusses research challenges and potential future detection for object detection.

Mouna Afif, et al., [18] presented a new deep learning-based indoor wayfinding assistance system that detects landmark indoor signs to support daily activities and improve social inclusion for both blind and sighted individuals. We conducted training and testing experiments using our proposed indoor signage dataset and demonstrated the system's effectiveness. The results of the experiments showed that the proposed indoor wayfinding and signage detection system achieved mean average precision of 93.45%.

The research article by Ajeet Ram Pathaka et al., [19] clarifies the application of convolutional neural network-based deep learning techniques for object detection. It also discusses available deep learning frameworks and services for object detection. The study evaluates deep learning techniques used in state-of –the-art object detection systems.

The review article by Abinash Bhandari et al., [20] presented the use of deep learning systems in navigational tools for visually impaired individuals and proposes a framework to guide future research. Here comparing current deep learning systems used in navigational tools and create a taxonomy of essential features for such systems. The study also addresses the challenges faced in detection. Taxonomy of improved navigational systems demonstrates sufficient robustness for general application.

III. SCOPE

As technology continues to advance, the significance of biometrics for security purposes has increased. Biometric authentication is a highly reliable method of identifying individual identity, as it relies on unique biological characteristic such as fingerprints, DNA, retina, ears, etc. Various object detection techniques have utilized in previous research studies to analyse biometric data.

Autonomous robots have emerged as a compelling field of research in recent years. One of the critical tasks these robots undertake is object detection, enabling them to identify objects in their vicinity and execute various operation, including offering information, opening, and closing doors, sounding alarms, and more. Object detection is therefore a crucial function that autonomous robots carry out.[17]

Various technical video surveillance systems are used to monitor numerous public areas, including subways, parks, schools, shopping centres, and more. As it is impossible for humans to continuously monitor video clips, object detection has become a crucial aspects of video surveillance. This technology plays a vital role in identifying and tracking specific objects in a scene, such as a suspected individual or vehicle, thereby enabling effective surveillance and security measures.

Detecting humans using computer vision is a complex task due to the diverse appearances and poses that individuals can adapt. To address this challenge, various object detection architecture has been developed for identifying humans in images or videos, including pedestrian detection. Object detection has also enabled rapid crowd counting in densely populated areas such as park and malls. This technology has proven to be highly effective in enhancing surveillance and security measures in such areas [17].

IV. GENERIC OBJECT DETECTION: PROBLEMS AND CHALLENGES

Generic object detection is the task of identifying and localizing objects of various classes in an image or video. While recent advances in deep learning have led to significant improvements in object detection accuracy, there are still several challenges and problems that need to be addressed in order to achieve more robust and accurate object detection systems. Here are some of the main problems and challenges in generic object detection

- Scale variation: Objects can appear at different scales in images and videos, making it challenging to detect them reliably
- Occlusion: Objects can be partially or fully occluded by other objects or background clutter, making them difficult to detect.
- Object shape variation: Objects can come in different shapes, making it challenging to develop a single model that can detect all objects accurately
- Limited training data: Object detection models require large amounts of labelled training data to generalize well to new images and scenes.
- Real-time processing: Object detection systems must often operate in real-time or near real-time, requiring efficient algorithms and hardware implementations to meet speed and latency requirements.
- Handling small objects: Small objects can be difficult to detect, especially when they are surrounded by clutter or appear at low resolution.
- Domain shift: Object detection models may not generalize well to new or unseen domains, leading to a drop in detection accuracy

V. DATASET

A dataset refers to a relevant collection of data such as text, image, and video that is formatted according to the requirement of an algorithm. When machines are trained artificially, they require well-prepared model that is both mathematically and logically sound. To create such models, the training process is divided into two main components: the technique and the dataset. The technique outlines the learning steps to be taken by the machine, while the dataset is used to train the machine using input samples. The quality and volume of the dataset play a crucial role in the performance of the recognition model, which is why datasets are also used to assess the accuracy of proposed techniques. The quality and quantity of the dataset are vital in developing an accurate and effective recognition model, and evaluating the proposed techniques. [1,17]

VI. CONCLUSION

In conclusion, object detection is a critical computer vision task that aims to identify and locate objects within an image or video. Deep learning techniques, such as convolutional neural networks (CNNs) and its variations have proven to be highly effective for object detection. Object detection in deep learning continues to be an active area of research, with ongoing efforts to improve speed, accuracy, and efficiency, and to adapt to new and challenging environments. This review is also meaningful for the developments in neural networks and related learning systems, which provides valuable insights and guidelines for future progress.

VII. REFERENCES

[1]. Li Liu , Wanli Ouyang , Xiaogang Wang, Paul Fieguth,Jie Chen, Xinwang Liu, Matti Pietikainen. Deep Learning for Generic Object Detection: A Survey (2019) springer International Journal of Computer Vision.
[2]. Zhong-Qiu Zhao, Peng Zheng, Shou-Tao Xu, and Xindong Wu. Object Detection With Deep Learning: A Review (2019).
[3]. Xiongwei Wu, Doyen Sahoo, Steven C.H. Hoi. Recent Advances in Deep Learning for Object Detection (2019).
[4]. Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoona Asghar, Brian Lee A survey of modern deep learning based object detection models (2022)
[5]. Enoch Arulprakash, Martin Aruldoss. A study on generic object detection with emphasis on future research directions (2022)
[6]. Nipun D. Nath,and Amir H. Behzadan. Deep Convolutional Networks for Construction Object Detection Under Different Visual Conditions (2020)
[7]. Mouna Afif, Riadh Ayachi, Edwige Pissaloux, Yahia Said, Mohamed Atri. Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people (2020)
[8]. Fereshteh S. Bashiri, Eric LaRose, Jonathan C. Badger, Roshan M. D'Souza, Zeyun Yu, Peggy Peissig. Object Detection to Assist Visually Impaired People: A Deep Neural Network Adventure (2018)
[9]. Mouna Afif, Riadh Ayachi, Edwige Pissaloux, Yahia Said, Mohamed Atri. Indoor Object Classification for Autonomous Navigation Assistance Based on Deep CNN Model (2019)
[10]. Mouna Afif, Riadh Ayachi, Edwige Pissaloux, Yahia Said, Mohamed Atri.An evalution of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation (2020)
[11]. Bernardo Calabrese, Ramiro Velazquez, Carolina Del-Valle-Soto, Roberto de Fazio, Nicola Ivan Giannoccaro, Paolo Visconti. Solar-Powered Deep Learning-Based Recognition System of Daily Used Objects and Human Faces for Assistance of the Visually Impaired (2020)

[12]. Fahad Ashiq, Muhammed Asif, Maaz Bin Ahemad, Sadia Zafar, Khalid Masood, Toqeer Mahmood, Muhammad Tariq Mahmood, Ik Hyun Lee. CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People (2022)

[13]. Xintao Ding, Yonglong Luo, Qingying Yu, Qingde Li, Yongqiang Cheng, Guorong Cai. Indoor Object Recognition Using Pre-trained Convolutional Neural Network (2019)

[14]. Y.C. Wong, J.A. Lai, S.S.S. Ranjit, A.R. Syafeeza, N. A. Hamid. Convolutional Neural Network for Object Detection System for Blind People (2019)

[15]. Mohamed L. Mekhalfi, Farid, Melgani, Yakoub Bazi, Naif Alajlan. A Compressive Sensing Approach to Describe Indoor Scenes for Blind People (2014) , IEEE Transactions on Circuits and Systems for Video Technology.

[16]. Xintao Ding, Yonglong Luo,Qingde Li, Yongqiang Cheng, Guorong Cai, Robert Munnoch, Dongfei Xue, Qingying Yu, Xiaoyao Zheng,Bing Wang. Prior knowledge-based deep learning method for indoor object recognition and application (2018)

[17]. Jaskirat Kaur, Williamjeet Singh .Tools, techniques, datasets and application areas for object detection in an image: a review (2022)

[18]. Mouna Afif, Riadh Ayachi, Yahia Said, Mohamed Atri. Deep learning-based application for indoor wayfinding assistance navigation (2021)

[19]. Ajeet Ram Pathaka, Manjusha Pandeya, Siddharth Rautaray. Application of Deep Learning for Object Detection (2018)

[20]. Abinash Bhandari, P. W. C. Prasad, Abeer Alsadoon and Angelika Maag. Object detection and recognition: using deep learning to assist the visually impaired (2019)