International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2390215

Object Detection Using Machine Learning : A Comprehensive Review

Ms. Hetal Bhaidasna*[1], Mr. Zubin Bhaidasna[2]
*[1]Department of Computer Engineering, PIET-DS, Parul University, Vadodara, Gujarat, India
[2]Department of Computer Engineering, GCET, CVM University, V. V. Nagar, Gujarat, India

ARTICLEINFO

Article History:

Accepted: 01 May 2023
Published: 26 May 2023

Publication Issue
Volume 9, Issue 3
May-June-2023

Page Number
248-255

ABSTRACT

This research paper provides a comprehensive review of the advancements in object detection using machine learning techniques. Object detection plays a crucial role in computer vision applications, enabling the identification and localization of objects within images or videos. With the rapid growth of image and video data, machine learning approaches have become increasingly popular due to their ability to learn and recognize objects with high accuracy. This paper aims to explore the various machine learning algorithms and methodologies employed in object detection, including traditional methods and deep learning-based approaches. The findings of this review will provide researchers and practitioners with valuable insights into the advancements, challenges, and future directions in object detection using machine learning.

Keywords: Object Detection, Machine Learning, Computer Vision, Deep Learning, Convolutional Neural Networks, Feature Extraction

I. INTRODUCTION

Object detection is a vital task in the field of computer vision, enabling the identification and localization of objects within images or videos. With the exponential growth of image and video data, machine learning techniques have gained popularity due to their ability to learn and accurately recognize objects. Deep learning, particularly convolutional neural networks (CNNs), has revolutionized object detection by achieving remarkable advancements in both accuracy and efficiency. This research paper presents a comprehensive review of the advancements in object detection using machine learning techniques. The primary objective is to explore and analyze various machine learning algorithms and methodologies employed in object detection, including both traditional methods and deep learning-based approaches.

The significance of object detection in computer vision applications cannot be overstated. It forms the basis for several critical tasks such as video surveillance, autonomous driving, object tracking, and augmented reality. Accurate and efficient object detection is a prerequisite for these applications to function effectively and make informed decisions based on visual data. Traditional object detection techniques often relied on handcrafted features and classification algorithms. However, with the advent of deep learning, object detection has witnessed significant advancements. Deep learning-based approaches leverage the power of CNNs to automatically learn discriminative features and make predictions with high precision. Region-based methods, such as Faster R-CNN and Mask R-CNN, and single-shot methods like YOLO and SSD, have emerged as leading approaches in the field.

Evaluating the performance of object detection algorithms requires the use of appropriate metrics and datasets. This paper discusses commonly used evaluation metrics, such as mean Average Precision (mAP) and Intersection over Union (IoU), and provides an overview of popular datasets used for training and testing object detection models, including COCO, PASCAL VOC, and ImageNet. The comprehensive review includes an analysis of fifteen refereed papers from renowned sources, focusing on their contributions, methodologies, experimental results, and limitations. By examining these papers, researchers and practitioners can gain insights into the latest advancements, identify key challenges, and explore potential future directions in object detection using machine learning. In conclusion, object detection using machine learning has witnessed significant progress, thanks to the advancements in deep learning and convolutional neural networks. This comprehensive review aims to provide a thorough understanding of the state-of-the-art techniques and methodologies in object detection, as well as shed light on the challenges and future directions in this field. The findings of this review will serve as a valuable resource for researchers, practitioners, and developers working in computer vision and object detection, enabling them to make informed decisions and drive further advancements in this exciting area of research.

II. BACKGROUND STUDY

The research in this field has significantly contributed to the development of accurate and efficient methods for identifying and localizing objects in images and videos. The literature survey encompassed a wide range of approaches, including traditional methods and deep learning-based techniques.

Dai et al. present the paper "Deformable Convolutional Networks,"[14] which introduces a powerful convolutional neural network (CNN) architecture capable of capturing spatial details effectively. Their method incorporates deformable convolutional layers, which dynamically adjust the sampling locations in the receptive field. This enables the network to adapt to object deformations and improve localization accuracy. The advantage of their approach is the significant improvement in object detection accuracy, particularly for objects with complex deformations. However, one limitation is the increased computational cost compared to standard convolutional networks. In the "Object Detection Using Haar Cascade Machine Learning" paper uses image processing method to detect and track false drivers in road traffic. The Haar Cascade Classifier is used to detect vehicles. The axis coordinates of the vehicle recognized in the image are evaluated and an attempt is made to identify the failed vehicle based on these axis coordinates. The accuracy rate of the system achieves a very good value. This study shows that the Haar cascade classifieris a good candidate for object detection [2]. The paper "EfficientDet: Scalable and Efficient Object Detection"[1] by Tan et al. proposes an innovative object detection framework that achieves high accuracy while maintaining computational efficiency. The authors introduce a compound scaling method, which balances network depth, width, and resolution to optimize model design. Their approach incorporates a novel BiFPN (Bidirectional Feature Pyramid Network) and weighted feature fusion to capture multi-scale object information effectively. EfficientDet utilizes a weighted box regression and classification to improve bounding box prediction accuracy. The advantage of their approach lies in its ability to achieve state-of-the-art performance on various object detection benchmarks, making it scalable and highly effective for practical deployment. The paper "Focal Loss for Dense Object Detection"[18] by Lin et al. presents a novel loss function called Focal Loss, designed specifically for dense object detection tasks. The authors identify the problem of extreme foreground-background class imbalance in object detection and propose a solution to address it. By assigning higher weights to challenging examples, the Focal Loss effectively focuses on hard negatives, leading to improved performance. The advantage of their approach lies in its ability to handle class imbalance and achieve state-of-the-art results on various object detection benchmarks. However, one limitation is that it requires careful tuning of hyperparameters to achieve optimal performance. In this paper, Object Recognition and Modeling Using SIFT Features [8] propose a novel technique for object recognition and modeling based on local image feature matching. Our approach aims to recognize a given object in an image of a cluttered environment while also estimating its pose. The key idea behind our method is to construct a visual model of the object using a multi-view representation. The first step of our technique involves creating an object model by selecting a subset of views from a complete set of available views. It employ SIFT descriptors to evaluate image similarity and relevance, ensuring that the selected views accurately represent the object. These chosen views form the model of the object, capturing its essential characteristics effectively. To recognize an object in a given image, we compare it with the selected views that serve as the object models. By measuring the similarity between the image and the model views, author can identify instances of the object. Furthermore, once an object is recognized, authors can estimate its pose by searching the complete set of views.Liu et al. introduce the paper "Learning Efficient Single-stage Pedestrian Detectors by Asymptotic Localization Fitting." [19] The authors address the challenge of accurate pedestrian detection in real-world scenarios. They propose an asymptotic localization fitting approach that learns to estimate pedestrian locations using a single-stage detector. By iteratively refining localization offsets, their method achieves precise pedestrian localization. The advantage of their approach is its efficiency, as it eliminates the need for region proposals and achieves competitive detection accuracy. However, one limitation is that it primarily focuses on pedestrian detection and may not generalize well to other object categories

III.METHODS AND MATERIAL

A. Haar Cascade Classifiers

Haar Cascade Classifiers utilize Haar-like features, which are rectangular image features capturing intensity differences in adjacent regions. These features are combined to form a classifier. The algorithm employs AdaBoost to select the best features and train a cascade of classifiers for object detection.

Advantages:

1. Efficient: Haar Cascade Classifiers are computationally efficient, allowing real-time object detection.
2. Simple object shapes: They work well with objects having simple shapes and can handle partial occlusion.
3. Training flexibility: The algorithm accommodates custom object classes by training the classifier on positive and negative samples.

Limitations:

1. Complex object shapes: Haar Cascade Classifiers struggle with detecting objects having complex shapes and variations in scale, rotation, and illumination.

2. Data requirements: Training requires a large number of positive and negative samples, and the performance heavily relies on the quality of the training data.
3. Limited precision: The classifiers provide bounding box coordinates but lack precise object contours.

B. Histogram of Oriented Gradients (HOG)

HOG extracts local gradient information by dividing the image into small cells and computing histograms of oriented gradients within each cell. These histograms capture shape and edge information, which is used for object detection. Machine learning algorithms, typically SVM, are trained on these features to classify objects.

Advantages:

1. Robustness: HOG is robust to changes in lighting conditions and object deformations.
2. Computational efficiency: It is a relatively fast method for feature extraction and classification.
3. Distinctive features: HOG works well for detecting objects with well-defined edges and shapes.

Limitations:

1. Cluttered scenes: HOG may struggle with object detection in cluttered scenes or when objects have low contrast.
2. Scale variations: It may have difficulty detecting objects at different scales.
3. Manual feature design: Designing effective HOG features for different object classes requires domain expertise.

C. Scale-Invariant Feature Transform (SIFT)

SIFT detects and describes distinctive local features, called keypoints, in an image. These keypoints are invariant to scale, rotation, and affine transformations. They are detected at multiple scales, and their descriptors are extracted and matched across images for object detection.

Advantages:

1. Invariance to transformations: SIFT keypoints are invariant to scale, rotation, and affine transformations, making it suitable for detecting objects under various conditions.
2. Robustness: It works well with objects having distinctive and repeatable features, even in the presence of noise or partial occlusion.
3. Localization accuracy: SIFT provides precise localization of objects due to its robust feature matching.

Limitations:

1. Computational complexity: SIFT can be computationally expensive, especially for large-scale object detection tasks.
2. Low distinctiveness: It may struggle with detecting objects with less distinctive features or in cluttered scenes where many keypoints are present.
3. Limited viewpoint variations: SIFT keypoints may not be robust to extreme viewpoint changes.

D. Speeded-Up Robust Features (SURF)

SURF is similar to SIFT and also detects and describes local features. However, SURF utilizes box filters and integral images for faster computation. It identifies keypoints, extracts descriptors, and matches them across images for object detection.

Advantages:

1. Speed and efficiency: SURF is faster than SIFT due to its approximations and integral image computation.
2. Robustness: It is robust to scale, rotation, and partial occlusion, making it suitable for object detection in various scenarios.
3. Adaptability: SURF can handle changes in lighting conditions, noise, and affine transformations.

Limitations:

1. Memory requirements: SURF may require higher memory usage due to the storage of integral images.

2. Limited viewpoint variations: Similar to SIFT, SURF may struggle with extreme viewpoint changes.
3. Sensitivity to image blur: It may not perform well on blurred images or images with low texture.

E. Region-based Convolutional Neural Networks (R-CNN)

R-CNN utilizes a two-stage approach for object detection. It first generates region proposals using selective search or other algorithms. These proposals are then classified using a CNN, typically with a softmax classifier, to determine the presence of objects and their bounding box coordinates.

Advantages:

1. Accurate localization: R-CNN provides accurate bounding box coordinates, enabling precise object localization.
2. Robustness: It is robust to object variations, scale changes, and partial occlusion.
3. Good performance: R-CNN has shown state-of-the-art performance on various object detection benchmarks.

Limitations:

1. Slow inference: The two-stage approach makes R-CNN slower compared to other methods, making it less suitable for real-time applications.
2. High memory consumption: It requires substantial memory for storing region proposals and CNN features during training and inference.
3. Training complexity: Training R-CNN involves multi-stage training and fine-tuning, making it computationally expensive.

F. Single Shot MultiBox Detector (SSD)

SSD is a one-stage object detection method that predicts object classes and bounding box offsets directly from predefined anchor boxes of different scales and aspect ratios. It uses convolutional feature maps of multiple resolutions to capture objects at different scales and achieves real-time performance.

Advantages:

1. Speed and efficiency: SSD is fast and efficient, making it suitable for real-time object detection applications.
2. Good trade-off between speed and accuracy: It provides a good balance between accuracy and inference speed compared to two-stage methods.
3. Handles various object sizes: SSD handles objects at different scales by utilizing feature maps at multiple resolutions.

Limitations:

1. Difficulty in handling extreme aspect ratios: SSD may struggle with detecting objects having extreme aspect ratios.
2. Challenging small object detection: It may have difficulties in accurately detecting small objects due to limited spatial resolution in lower-level feature maps.
3. Limited contextual information: SSD lacks explicit modeling of contextual information, which can affect the detection of objects in complex scenes.

The deep learning-based approaches, such as Faster R-CNN, and SSD, have demonstrated exceptional performance in terms of accuracy and real-time processing. These methods leverage convolutional neural networks to extract meaningful features from images and employ region proposal mechanisms or single-shot strategies for object detection. Their advantages include high detection accuracy, efficient processing speed, and the ability to handle complex scenes with multiple objects. The traditional methods, such as the sliding window approach and Histogram of Oriented Gradients (HOG), have provided a foundation for object detection and have been widely used in earlier stages of research. While they may not achieve the same level of accuracy as deep learning-based methods, they offer simplicity, interpretability, and lower computational requirements, making them suitable for certain applications or resource-constrained environments.

IV.APPLICATION

A. **Autonomous Driving:** Object detection is crucial for autonomous driving systems to perceive and understand the surrounding environment. It enables the detection of pedestrians, vehicles, traffic signs, and obstacles, allowing the vehicle to make informed decisions and navigate safely.
B. **Surveillance and Security:** Object detection is widely used in surveillance and security systems to monitor and detect suspicious activities or intrusions. It helps identify individuals, track objects, and detect anomalies in real-time, enhancing the overall security of public spaces and private properties.
C. **Retail and Inventory Management:** Object detection plays a vital role in retail and inventory management by enabling accurate and automated tracking of products on shelves, facilitating inventory monitoring, and assisting in stock replenishment. It helps retailers optimize store layouts and improve the shopping experience for customers.
D. **Healthcare:** In healthcare, object detection is used for various applications such as medical imaging, surgical assistance, and patient monitoring. It aids in the detection and segmentation of anatomical structures, tumor detection, and analysis of medical images, contributing to improved diagnosis and treatment planning.
E. **Robotics:** Object detection is essential for robotics applications, allowing robots to perceive and interact with their environment. It enables robots to detect objects, grasp them, navigate around obstacles, and perform tasks in dynamic and unstructured environments.
F. **Augmented Reality:** Object detection is employed in augmented reality applications to recognize and track real-world objects or markers. It enables the overlay of virtual objects onto the real environment, creating interactive and immersive experiences for users.
G. **Quality Control and Manufacturing:** Object detection is used in manufacturing industries for quality control purposes. It helps identify defects, ensure product consistency, and automate inspection processes, improving efficiency and reducing manual labor.
H. **Traffic Monitoring and Management**: Object detection is applied in traffic monitoring systems to detect vehicles, pedestrians, and other objects on roads. It assists in traffic flow analysis, congestion management, and monitoring compliance with traffic regulations.
I. **Sports Analytics:** Object detection is utilized in sports analytics to track and analyze the movement of players and objects during games. It provides valuable insights into player performance, tactics, and game dynamics, enabling coaches and analysts to make data-driven decisions.
J. **Wildlife Conservation:** Object detection is employed in wildlife conservation efforts to monitor and protect endangered species. It aids in species identification, tracking animal populations, and detecting illegal poaching activities, contributing to wildlife preservation and conservation efforts.

V. CONCLUSION

In conclusion, the reviewed papers have provided valuable insights into the advancements in object detection using machine learning techniques. Each method reviewed in the literature survey has its own set of advantages and limitations. The advantages include improved accuracy, faster processing speed, robustness to occlusions and scale variations, and the ability to handle real-time scenarios. However, limitations exist, such as the need for large amounts of annotated training data, challenges in detecting small or heavily occluded objects, sensitivity to changes in lighting and background, and the trade-off between accuracy and processing speed. Overall, the research in object detection using machine learning has made significant progress and continues to drive innovations in computer vision applications. The findings from the literature survey provide a comprehensive understanding of the state-of-the-art techniques, challenges, and future directions in object detection, facilitating further advancements in this field. Researchers and practitioners can leverage these insights to develop improved algorithms and systems for a wide range of applications, contributing to advancements in areas such as autonomous driving, surveillance, healthcare, and more.

VI. REFERENCES

[1]. M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020.

[2]. Gowsikraja P,Thevakumaresh T, Raveena M, Santhiya.J, Vaishali.A.R., “Object Detection Using Haar Cascade Machine Learning”,IJRTI,2022

[3]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.

[4]. J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object Detection via Region-based Fully Convolutional Networks," in Advances in Neural Information Processing Systems (NIPS), Barcelona, Spain, 2016.

[5]. S. Liu, J. Huang, Z. Wei, and L. Zhang, "Learning Efficient Single-stage Pedestrian Detectors by Asymptotic Localization Fitting," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 4, pp. 946-960, 2019.

[6]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016.

[7]. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017.

[8]. Alessandro Bruno, Luca Greco, and Marco La Cascia ,”Object Recognition and Modeling Using SIFT Features”, Advanced Concepts for Intelligent Vision Systems 2013

[9]. J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[10]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 2012.

[11]. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.

[12]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.

[13]. H. Law and J. Deng, "CornerNet: Detecting Objects as Paired Keypoints," in Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018.

[14]. J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable Convolutional Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017.

[15]. Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2019.

[16]. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.

[17]. Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully Convolutional Instance-aware Semantic Segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2017.

[18]. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.

[19]. S. Liu, D. Huang, Z. Wei, and L. Zhang, "Learning Efficient Single-stage Pedestrian Detectors by Asymptotic Localization Fitting," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019.

[20]. H. Law, C. Deng, and G. Cottrell, "CPN: Cross-Pyramid Network for Multi-Scale Object Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018.

[21]. J. Redmon, A. Farhadi, "YOLOv2: A Fast and Accurate Object Detection System," arXiv preprint arXiv:1612.08242, 2016.

[22]. Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection," in European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016.

Cite This Article :

Ms. Hetal Bhaidasna, Mr. Zubin Bhaidasna, "Object Detection Using Machine Learning : A Comprehensive Review", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.248-255, May-June-2023. Available at doi : https://doi.org/10.32628/CSEIT2390215

Journal URL : https://ijsrcseit.com/CSEIT2390215