Advanced Two Stage AI Technique for Object Detection

Shashank Tiwari*[1], Balla Sahithi*[2], M. Phani Krishna*[3], M. Hari Charan Reddy*[4]

*[1] Assistant Professor Of Department Of CSE (AI & ML) Of ACE Engineering College, India.

*[2,3,4] Students Of Department Of CSE (AI & ML) Of ACE Engineering College, India.

ABSTRACT

Object detection in computer vision uses AI, mainly deep learning, to identify and locate objects in images or videos. It involves an AI system that spots various objects, determines their type, and marks their positions with bounding boxes. Built on advanced deep learning models like Convolutional Neural Networks (CNNs), YOLO, or Faster R-CNN, it excels in real-time detection. Trained on large datasets like COCO, it recognizes diverse objects across different scenes. It tackles challenges investigates challenges like varying object sizes, scales, lighting, and occlusion using techniques such as transfer learning, data augmentation, and model optimization for fast, accurate results.

Keywords:

Deep learning, YOLO / Faster R-CNN, bounding boxes, and COCO dataset.

INTRODUCTION

Object detection constitutes a critical subdomain of computer vision, entailing the precise localization and classification of semantic entities within visual data streams. Classical methodologies such as Histogram of Oriented Gradients (HOG), Deformable Part Models (DPM), Viola-Jones (VJ) detectors, and Local Binary Patterns (LBP) have demonstrated appreciable performance; however, their efficacy diminishes in scenarios demanding high computational efficiency and complex feature abstraction. The paradigm shift to deep learning, particularly Convolutional Neural Networks (CNNs), has engendered transformative advancements—exemplified by architectures like AlexNet, Fast R-CNN, and Faster R-CNN—which have redefined the landscape of feature extraction and object proposal generation. The YOLO (You Only Look Once) framework introduced an end-to-end, single-stage detection mechanism, enhancing real-time applicability without sacrificing accuracy. Among its iterations, YOLOv5 has emerged as a state-of-the-art model due to its architectural efficiency, reduced inference latency, and superior generalization.

LITERATURE REVIEW

1. Liu, Ouyang, Wang, Fieguth, Chen, Liu, Pietikäinen[2019][1]:

This paper offers an in-depth survey of over 300 deep learning approaches for object detection, systematically grouping them by framework designs, feature extraction methods, and proposal generation techniques. It assessed models ranging from two-stage systems like Faster R-CNN to single-stage detectors like YOLO, leveraging datasets such as MS COCO and PASCAL VOC. The methodology centered on comparing architectural components and performance metrics like mean Average Precision. The findings underscored the field's evolution, highlighting the balance between detection speed and accuracy, and the pivotal role of advanced feature extraction in boosting performance across diverse scenarios.

2. Ayesha, Iqbal, Ahmad, Alassafi, Alfakeeh, Alhomoud [2023][2]:

This study reviews object detection techniques, organizing them into anchor-based, anchor-free, and transformer-based categories, and comparing convolutional neural networks like Faster R-CNN with vision transformers like DETR. The methodology involved evaluating speed and accuracy metrics on standard datasets, emphasizing the strengths of each approach. The findings revealed that transformers provide superior accuracy in complex scenes due to their global context awareness, while CNNs offer faster processing for real-time applications, illuminating key trade-offs in computational demands and detection quality.

3. Lin, Goyal, Girshick, He, Dollár[2017][3]:

This work proposed RetinaNet, a single-stage detector, and introduced focal loss to address class imbalance in dense object detection tasks. Unlike two-stage models relying on region proposals, RetinaNet processes images in one pass, optimized on the COCO dataset. The methodology compared focal loss with traditional cross-entropy loss, focusing on its ability to prioritize challenging examples. The findings showed that RetinaNet matches or exceeds two-stage model accuracy by mitigating foreground-background imbalance, making it highly effective for dense detection scenarios.

4. V.D. Soni[2017][4]:

This paper investigates traditional and early deep learning object detection methods, emphasizing feature detection and hypothesis verification components. It proposed a hybrid methodology for object localization and classification, evaluated on datasets like PASCAL VOC. The approach integrated handcrafted features with early convolutional neural networks to improve localization. The findings highlighted the shortcomings of sliding window methods and demonstrated that the proposed framework enhances detection accuracy, particularly for smaller objects, offering a practical solution for early deep learning applications.

5. Shang, Zhao, Li, Wu, Cao[2025][5]:

This research developed an optimized YOLOv5-based algorithm for ship detection in unmanned surface vessels, designed for real-time performance with minimal computational resources. The methodology involved enhancing YOLOv5 with lightweight convolutional layers and testing on maritime-specific datasets. The findings indicated improved detection accuracy and lower latency compared to standard YOLOv5, proving the algorithm's suitability for resource-constrained maritime environments and its potential for autonomous navigation systems.

COMPARISON TABLE

S. No	Author Name	Title	Methodology	Findings
1	L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen	A comprehensive check of deep knowledge ways for general object discovery	Conducted a systematic review of over 300 deep learning-based object detection methods, categorizing them by frameworks, feature extraction, and proposal generation. Analyzed datasets like COCO and PASCAL VOC, and evaluated metrics like mAP.	Identified key trends in deep learning for object detection, highlighting the shift from region-based to single-stage detectors and the importance of robust feature representations for improved accuracy.
2	Ayesha, M. J. Iqbal, I. Ahmad, M. O. Alassafi, A. S.	A review of object discovery with deep knowledge, convolutional	Reviewed object detection techniques, classifying them	Found that vision transformers offer superior accuracy in complex scenes but are computationally

	Alfakeeh, A. Alhomoud	neural networks, and vision manufactories	into anchor-based, anchor-free, and transformer-based methods. Compared CNN-based models (e.g., Faster R-CNN) with vision transformers (e.g., DETR) using metrics like speed and accuracy.	intensive, while CNN-based methods remain faster for real-time applications.
3	T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár	Addressing class imbalance in thick object discovery with focal loss	Introduced RetinaNet, a single-stage detector, and proposed focal loss to mitigate class imbalance in dense object detection. Trained and tested on the COCO dataset, comparing with two-stage detectors like Faster R-CNN.	Demonstrated that focal loss significantly improves single-stage detector performance, achieving state-of-the-art accuracy on COCO by reducing the impact of background class dominance.
4	V. D. Soni	An analysis and methodology for detecting objects in images	Analyzed traditional and early deep learning object detection techniques, focusing on components like feature detectors and hypothesis verifiers. Proposed a methodology for object localization and categorization, tested on standard datasets.	Highlighted limitations of sliding window approaches and early CNNs, proposing a hybrid methodology that improves localization accuracy for small objects.
5	Y. Shang, J. Zhao, S. Li, T. Wu, J. Cao	Effective single-stage boat discovery algorithm for unmanned face vessels	Developed an optimized YOLOv5-based algorithm for ship detection in unmanned surface vessels, incorporating	Achieved high detection accuracy and low latency, making the algorithm suitable for resource-constrained maritime environments, with

			lightweight convolutional layers and real-time processing. Evaluated on maritime-specific datasets.	improved performance over standard YOLOv5.

Table 1: Research Study Comparison

RESEARCH GAPS IN EXISTING SYSTEMS:

Based on the literature review, several research gaps have been identified:

1. **Intensive Computational Demands:**

Many algorithms, such as YOLOv5, require significant GPU resources and memory for training and real-time detection, posing challenges for deployment on devices with limited capabilities.

2. **Data Annotation Challenges:**

Despite advancements in tools like Roboflow, annotating data remains labor-intensive and prone to errors, with inaccurate labels hindering model accuracy and slowing progress.

3. **Limited Efficiency on Edge Devices:**

Even with optimized versions, YOLOv5 struggles to balance real-time performance and accuracy on edge or mobile devices due to hardware constraints.

4. **Sensitivity to Environmental Variations:**

Detection models often falter under changing lighting, occlusions, or cluttered backgrounds, undermining reliability in dynamic outdoor settings.

5**. Lack of Interpretability:**

Deep learning models function as opaque systems, providing little insight into their decision-making processes, which reduces trust in high-stakes applications like healthcare or autonomous navigation.

PROPOSED SYSTEM

The emergence of deep learning in the 2010s transformed object detection, with Convolutional Neural Networks (CNNs) serving as the foundation for feature extraction, significantly enhancing accuracy and adaptability across diverse environments. Two key approaches developed: region-based methods and single-stage detectors, the latter simplifying detection by directly predicting object locations and classes, offering speed ideal for real-time applications. YOLO (You Only Look Once) (Redmon et al., 2016) redefined detection as a regression task, dividing images into grids to predict bounding boxes and class probabilities simultaneously, achieving remarkable speed but initially struggling with small objects and precise localization. SSD (Single Shot MultiBox Detector) (Liu et al., 2016) improved on this by using multi-resolution feature maps to detect objects of varying sizes, balancing speed and accuracy effectively for diverse scenarios.

CONCLUSION AND FUTURE SCOPE

The object detection project successfully implemented advanced computer vision techniques using deep learning models like YOLO, SSD, or Faster R-CNN to achieve efficient and accurate object identification in images and video streams. Key accomplishments include selecting and training pre-trained models on datasets like COCO or Pascal VOC, testing the system on diverse scenarios for robustness, and evaluating performance with metrics like mean Average Precision (mAP), precision, and recall. The project utilized frameworks like YOLO and tools like OpenCV, highlighting the potential of object detection in industries such as autonomous driving, security, healthcare, and retail. Future advancements include improving real-time detection speed and accuracy for edge devices, optimizing models for applications like autonomous vehicles and surveillance, and integrating multi-modal data from LiDAR, radar, or thermal cameras for challenging conditions. Further scope involves enhancing small or occluded object detection for satellite imagery and medical imaging, advancing 3D object detection for navigation and AR/VR, applying detection agriculture and environmental monitoring, and addressing ethical AI concerns to reduce bias in diverse scenarios.

REFERENCES:

1. L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu and M. Pietikäinen, “A comprehensive check of deep knowledge ways for general object discovery,” International Journal of Computer Vision, vol. 128, no. 1, 2019, pp. 300 – 322.
2. Ayesha, M. J. Iqbal, I. Ahmad, M. O. Alassafi, A. S. Alfakeeh and A. Alhomoud, “A review of object discovery with deep knowledge, convolutional neural networks, and vision manufactories,” IEEE Access, vol. 11, no. 1, 2023, pp. 46581 – 46607.
3. T.- Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, “Addressing class imbalance in thick object discovery with focal loss,” IEEE International Conference on Computer Vision, vol. 1, no. 1, 2017, pp. 2980 – 2988.
4. V. D. Soni, “An analysis and methodology for detecting objects in images,” International Journal of Computer Vision and Image Processing, vol. 7, no. 2, 2017, pp. 1 – 22.
5. Y. Shang, J. Zhao, S. Li, T. Wu and J. Cao, “Effective single- stage boat discovery algorithm for unmanned face vessels,” Sensors, vol. 25, no. 9, 2025, pp. 3054 – 3070.