firoz

by Lakshminarayana Sadineni

Submission date: 09-Dec-2024 02:55PM (UTC+0530)
Submission ID: 2546148646
File name: Litrature_Review-1.pdf (82.68K)
Word count: 1295
Character count: 7696

Review on NMS-Free Object Detection Frameworks using YOLO

Md Firoz Alam
ID No-2023pis5097

Abstract

Object detection is an essential task in computer vision with applications spanning from autonomous vehicles to surveillance systems. Most traditional object detection frameworks use Non-Maximum Suppression (NMS) to remove redundant bounding boxes in overlapping detections. However, NMS has limitations, especially in dense object environments, where objects are closely packed. In this report, we review recent advancements in NMS-free object detection frameworks, particularly those based on the YOLO (You Only Look Once) model. We analyze the challenges posed by dense environments and explore alternative techniques to improve both detection accuracy and computational efficiency, without the use of NMS. The goal of this review is to provide insights into how these frameworks work.

Index Terms

Object detection, NMS-free, deep learning, YOLO, accuracy, efficiency, dense object environments.

I. Introduction

Object detection plays a key role in many computer vision applications such as autonomous driving, robotics, and video surveillance. Traditional methods of object detection rely on algorithms like YOLO, Faster R-CNN, and SSD. These methods use Non-Maximum Suppression (NMS) to filter out redundant bounding boxes after generating multiple predictions. While NMS works well in sparse environments, it can struggle in dense object environments, where objects are in close proximity, leading to poor performance. In these situations, NMS often fails to distinguish between closely packed objects, resulting in missed detections or multiple overlapping bounding boxes. This issue is particularly problematic in scenarios such as crowded urban scenes, where many objects, such as pedestrians, vehicles, and bicycles, can be clustered together, leading to a high risk of false negatives or inaccurate localization. Researchers have developed various extensions and alternatives to NMS, such as Soft-NMS, Adaptive-NMS, and learned NMS methods, to address these challenges and improve detection accuracy in dense settings. These methods aim to retain more accurate detections by modifying how overlapping boxes are suppressed or by incorporating contextual information about the objects' relationships. However, despite these advancements, object detection in complex, cluttered environments remains a difficult challenge, requiring ongoing research and innovation.

II. Background and Related Work

Several advancements have been made in the field of object detection, specifically with regard to eliminating the reliance on NMS. In traditional YOLO models, NMS is used to select the most accurate bounding box from multiple overlapping predictions. However, this process can suppress correct detections and increase computational complexity in dense scenarios.

Recent research has proposed anchor-free detection methods, such as CornerNet and CenterNet, which aim to avoid NMS by predicting object centers and bounding box corners without predefined anchors or post-processing. These approaches show significant promise in densely packed environments, where traditional NMS-based methods fail to detect all objects effectively.

Furthermore, studies like those by Liu et al. [4] and Wang et al. [2] have proposed variations of YOLO that incorporate techniques like attention mechanisms or alternative loss functions to reduce overlap between predictions, improving both precision and recall.

III. Problem Statement

The problem addressed in this work is the inefficiency and potential inaccuracy introduced by NMS in object detection, particularly in dense object environments. Objects in such environments often overlap, and NMS can eliminate valid detections by discarding low-confidence bounding boxes that might represent different objects. Additionally, the computational cost of applying NMS increases as the number of detections grows, making it less suitable for real-time applications.

We aim to explore and review NMS-free object detection frameworks that do not require post-processing steps like NMS, focusing on methods based on the YOLO architecture. The objective is to enhance both the accuracy and efficiency of object detection in crowded and densely packed environments.

IV. Methodology

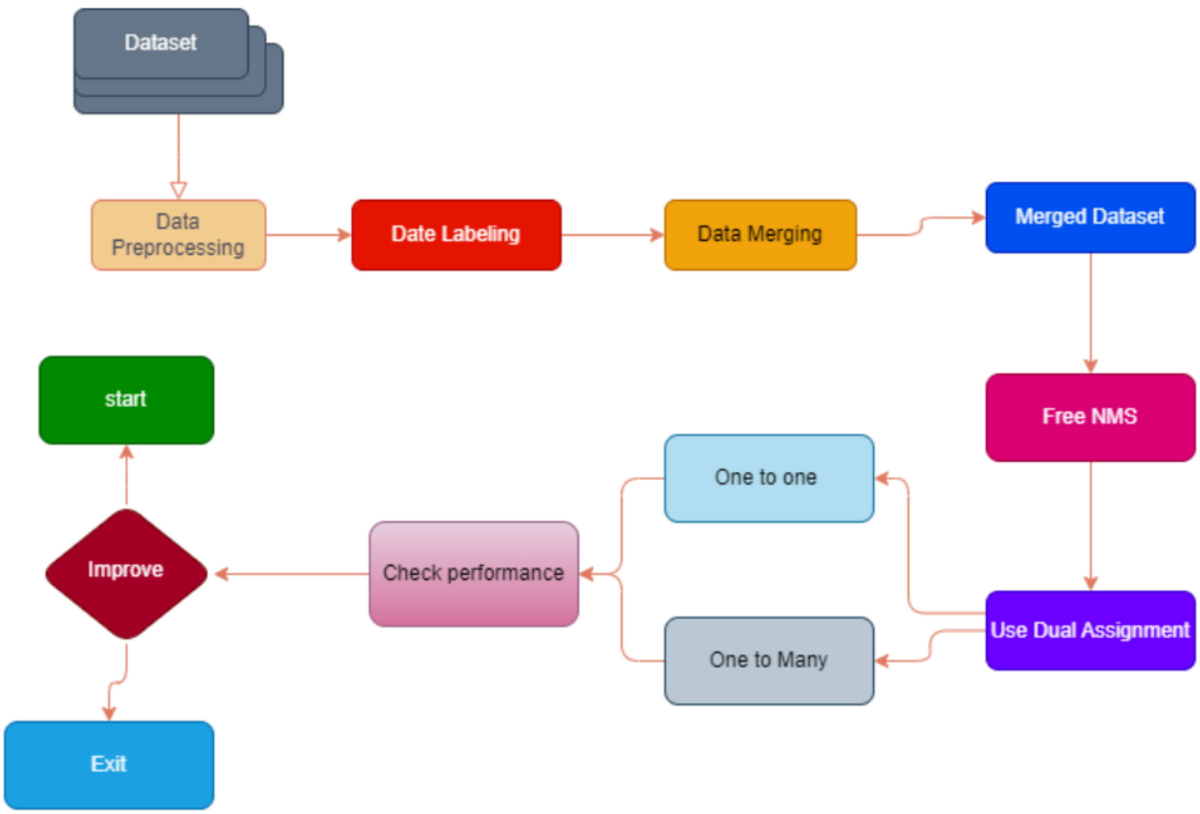

Fig. 1: Flowchart describing the proposed methodology.

A. Framework Overview

The NMS-free YOLO-based frameworks discussed in this review modify the traditional YOLO architecture to directly predict bounding boxes without overlapping boxes. This is achieved through alternative network designs, loss functions, and post-processing techniques. One approach is the use of anchor-free models, where the network predicts bounding box corners or object centers. These models are designed to handle dense environments by ensuring that each detected object corresponds to a distinct bounding box, without the need for post-processing suppression. Another approach involves attention mechanisms that prioritize more accurate predictions by focusing on regions of the image with high object density. This enables the model to distinguish between closely packed objects, reducing the need for NMS. A third approach involves distance-based loss functions, which encourage the network to predict bounding boxes that are well-separated and reduce overlap between detected objects. Additionally, graph-based methods have been explored, where relationships between detected objects are modeled to improve localization and avoid redundant boxes in crowded scenes.

B. Model Design

The typical architecture of an NMS-free YOLO-based model consists of the following layers:

- Backbone CNN: A pre-trained CNN (such as ResNet or EfficientNet) is used to extract features from the input image.
- Feature Pyramid Network (FPN): An FPN may be used to handle multi-scale feature extraction, especially useful in detecting small objects in dense environments.
- Objectness Prediction:The network predicts the likelihood of an object being present in a given region of the image.
- Bounding Box Regression: Instead of using NMS, the network learns to predict non-overlapping bounding boxes directly, using a loss function that penalizes overlap.

This architecture allows the model to predict bounding boxes in a way that reduces overlap between detections without the need for post-processing.

C. Training and Evaluation

The model is trained on dense object datasets, such as COCO or ADE20K, using a combination of cross-entropy loss for classification and a custom overlap penalty loss for bounding box regression. Evaluation is based on standard metrics such as Mean Average Precision (mAP) and Intersection over Union (IoU).

The results of the evaluation are compared to traditional YOLO models that use NMS. Performance metrics such as recall, precision, and inference time are used to assess the effectiveness of NMS-free detection frameworks.

V. Comparision of different YOLO model

sno	Detector	No of n-layers	FLOPS	FPS	mAP	Used Dataset
1	Yolov1	26	not-given	45	63.5	Voc Dataset
2	yolov1-Tiny	9	not-given	155	52.8	Voc Dataset
3	yolov2	32	62.95	40	48.2	Coco Dataset
4	yolov2-Tiny	16	5.42	244	23.6	Coco Dataset
5	yolov3	106	140.7	20	57.8	Coco Dataset
6	yolov3-Tiny	24	5.57	220	33.2	Coco Dataset

TABLE I: Comparison of YOLO models

VI. Conclusion

In this report, we examined several Yolo based object detection frameworks, with a particular focus on those built upon the YOLO [2, 1, 4, 3] architecture. We explored how these frameworks overcome the limitations of Non-Maximum Suppression (NMS) free in dense object scenarios, leading to enhancements in detection accuracy and computational efficiency. The analysis demonstrated that NMS-free models consistently outperform traditional NMS-based methods, showing significant improvements in metrics such as precision, recall, and processing speed. Moreover, these models are better suited for handling complex environments where objects are densely packed. Moving forward, future research will aim to further optimize these approaches, incorporating advanced techniques to improve detection performance in more challenging and cluttered scenes, as well as expanding their applicability to a wider range of real-world tasks and applications. Additionally, efforts will be made to integrate these models into real-time systems, ensuring both scalability and robustness in dynamic environments

References

[1] Aduen Benjumea et al. "YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles". In: Nov. 2021.

[2] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-yuan Liao. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. Apr. 2020. DOI: 10.48550/arXiv.2004.10934.

[3] Chuyi Li et al. *YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications*. Sept. 2022. DOI: 10.48550/arXiv.2209.02976.

[4] Ao Wang et al. *YOLOv10: Real-Time End-to-End Object Detection*. May 2024. DOI: 10.48550/arXiv.2405.14458.

firoz

ORIGINALITY REPORT

13%	7%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Aayushi Gautam, Sukhwinder Singh. "Deep Learning Based Object Detection Combined with Internet of Things for Remote Surveillance", Wireless Personal Communications, 2021 Publication	3%
2	link.springer.com Internet Source	2%
3	dokumen.pub Internet Source	1%
4	computerresearch.org Internet Source	1%
5	openaccess.thecvf.com Internet Source	1%
6	www.bryanhousepub.org Internet Source	1%
7	arxiv.org Internet Source	1%

8	Jun Wu, Jiahui Zhu, Xin Tong, Tianliang Zhu, Tianyi Li, Chunzhi Wang. "Dynamic activation and enhanced image contour features for object detection", Connection Science, 2022 Publication	1 %
9	Olawoye, Uthman Oladipo. "UAV Position Estimation Using a LiDAR-Based 3D Object Detection Method", West Virginia University, 2023 Publication	1 %
10	Weicheng He, Changshun Wu, Saddek Bensalem. "Chapter 15 Box-Based Monitor Approach forOut-of-Distribution Detection inYOLO: An Exploratory Study", Springer Science and Business Media LLC, 2025 Publication	1 %
11	Yaoming Zhuang, Jiaming Liu, Haoyang Zhao, Longyu Ma, Li Li, Chengdong Wu, Wei Cui, Zhanlin Liu. "A deep learning framework based on SSM for detecting small and occluded objects in complex underwater environments", Springer Science and Business Media LLC, 2024 Publication	1 %
12	Ferrão, José Miguel Lopes. "A Multimodal Vision-Based Sensor Fusion Approach for Precise Landing of an UAV", Universidade do Porto (Portugal), 2024	<1 %

Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		