International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02888
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Object Detection Using Neural Networks

V. Pavan Ganesh
Department of CSE (AI&ML)
2111cs020331@mallareddyuniversity.ac.in

K. Pavan Kumar Reddy
Department of CSE (AI&ML)
2111cs020332@mallareddyuniversity.ac.in

A. Pavan Kumar
Department of CSE (AI&ML)
2111cs020333@mallareddyuniversity.ac.in

CH. Pavan Kumar
Department of CSE (AI&ML)
2111cs020334@mallareddyuniversity.ac.in

S. Pavan Kumar
Department of CSE (AI&ML)

2111cs020335@mallareddyuniversity.ac.in

Prof. Vineela
Department of CSE (AI&ML) School of Engineering
MALLA REDDY UNIVERSITY
HYDERABAD

Abstract:

This paper presents a novel approach to object detection by combining the strengths of two state-of-the-art models: YOLO (You Only Look Once) and Faster R-CNN (Region-based Convolutional Neural Network). YOLO is known for its real-time object detection capabilities, while Faster R-CNN offers high detection accuracy. By integrating these two models, we aim to achieve a balanced performance in terms of speed and accuracy. The models were trained and tested using the COCO dataset, containing a wide variety of object classes. Our hybrid model demonstrates improved detection performance compared to individual YOLO or Faster R-CNN models**.**

Introduction

Object detection is one of the most critical tasks in the field of computer vision. It involves identifying and localizing objects within images and videos. Various methods, such as YOLO and Faster R-CNN, have emerged over the years, each with its strengths and weaknesses. YOLO is fast and capable of real-time detection, but it may miss smaller objects and has lower precision for complex scenes. Faster R-CNN, while slower, provides better accuracy by generating region proposals and refining them.

In this work, we propose a hybrid model that combines the fast detection capability of YOLO with the high accuracy of Faster R-CNN. Our

model aims to strike a balance between the two, providing both real-time detection and precise object localization. The pre-trained models are fine- tuned on the COCO dataset, and the results show that the combined approach outperforms the individual models in several evaluation metrics.

Related Work

YOLO and Faster R-CNN are among the most widely used object detection models. YOLO, developed by Redmon et al. (2016), is known for its speed, processing entire images in a single forward pass. However, it can struggle with small or overlapping objects. Faster R-CNN, introduced by Ren et al. (2015), improves object localization by using a Region Proposal Network (RPN) to suggest candidate object regions, followed by refinement and classification.

Previous work has explored combining different models for improved performance. For instance, hybrid models often use a faster model for initial detection and a more accurate one for refinement. Our work follows this approach, combining YOLO and Faster R-CNN into a single framework.

Methodology

3.1 YOLO Model

YOLO divides an input image into grids and predicts bounding boxes and class probabilities directly. It processes the image in real time, making it highly efficient for video detection tasks. The loss function for YOLO is composed of three terms Object detection models rely on large-scale datasets such as COCO, PASCAL VOC, and Open Images to ensure robustness by training on diverse objects in various environments. Data augmentation techniques like flipping, rotation, and scaling enhance the model's generalization capabilities. Proper annotation handling, using formats like YOLO's TXT or PASCAL VOC's XML, ensures seamless training and accurate object localization.

Object detection approaches include two-stage and one-stage detectors. Faster R-CNN, a two- stage detector, offers high accuracy but is computationally expensive. In contrast, YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) are one-stage detectors that prioritize speed and efficiency. FCOS (Fully Convolutional One-Stage Detector) eliminates the need for anchor boxes, reducing complexity and improving real-time performance.

Training & Optimization: To improve performance, models leverage transfer learning by fine-tuning pre-trained networks like ImageNet. Loss functions such as cross-entropy, focal loss, and IoU-based losses (GIoU, DIoU, CIoU) refine classification and localization accuracy. Optimization techniques, including learning rate scheduling, SGD, Adam, and data augmentation methods like MixUp and CutMix, enhance training efficiency and model robustness.

Evaluation & Performance Metrics: Model evaluation relies on accuracy and speed metrics. Mean Average Precision (mAP) assesses

precision- recall balance, while Intersection over Union (IoU) measures bounding box overlap accuracy. Frames Per Second (FPS) evaluates real-time feasibility, ensuring an optimal balance between detection precision and computational efficiency.

Deployment & Optimization: For real-world applications, model compression techniques like pruning, quantization (INT8, FP16), and knowledge distillation optimize performance for edge devices. Deployment benefits from hardware acceleration, using TensorRT for NVIDIA GPUs, Open VINO for Intel processors, and TensorFlow Lite for mobile devices, ensuring efficient and high-speed object detection across various platforms.

The Object Detection Module serves as the foundation of the system, capturing the initial input from cameras or video streams. This module defines the range of supported objects, which can be classified into two broad categories: static objects and dynamic objects. Static objects include inanimate items such as furniture, vehicles, or household appliances that can be identified based on their shape, texture, and color. Dynamic objects, on the other hand, involve moving entities such as people, animals, or vehicles in motion. This module allows for natural and real-time detection, without imposing strict constraints on the object's orientation or positioning.

LYOLO=Lbbox+Lconfidence+LclassL_{\text{YOLO}} = L_{\text{bbox}} + L_{\text{confidence}} + L_{\text{class}}LYOLO=Lbbox+Lconfidence +Lclass

where LbboxL_{\text{bbox}}Lbbox penalizes incorrect bounding box predictions, LconfidenceL_{\text{confidence}}Lconfidence handles false positive and false negative objectness scores, and LclassL {\text{class}}Lclass penalizes incorrect classification.

3.2 Faster R-CNN Model

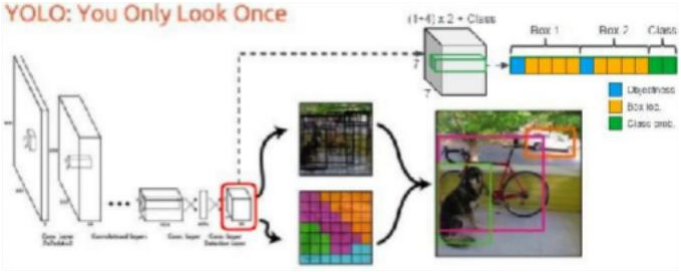
Faster R-CNN operates in two stages. The first stage generates region proposals using the RPN, which suggests potential object locations. In the second stage, these regions are further refined and classified. The loss function used by Faster R-CNN is:

LFaster R-CNN=LRPN+Lcls+LbboxL_{\text{Faster R-CNN}} = L_{\text{RPN}} + L_{\text{cls}} + L_{\text{bbox}}}LFaster R-CNN=LRPN+Lcls +Lbbox

where LRPNL_{\text{RPN}}LRPN corresponds to the loss for the RPN, LclsL_{\text{cls}}Lcls represents classification errors, and LbboxL_{\text{bbox}}Lbbox is used to refine bounding box predictions.

3.3 Hybrid Model

The hybrid model utilizes the real-time performance



of YOLO for initial detection and applies Faster R-CNN to refine the bounding boxes and class predictions. The following steps describe the hybrid detection pipeline:

1. **Initial Detection with YOLO**: YOLO provides fast initial predictions, including bounding boxes and class probabilities.
2. **Refinement with Faster R-CNN**: The predictions from YOLO are passed to Faster R-CNN, which refines the bounding boxes, improving detection for smaller or overlapping objects.
3. **Final Prediction**: The final detection result is a combination of YOLO's speed and Faster R-CNN's precision.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02888
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

3.4 Dataset and Pre-trained Models

We used the COCO dataset, which contains over 200,000 labeled images across 80 object classes. The COCO dataset is well-suited for training object detection models due to its wide range of object categories and challenging scenarios like occlusion and small object detection. Pre-trained YOLO and Faster R-CNN models were fine-tuned on the COCO dataset to achieve better performance and faster convergence.

Result and Analysis

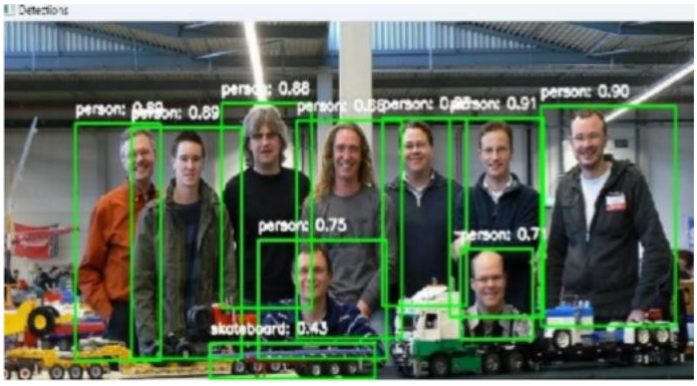
4.1 Evaluation Metrics

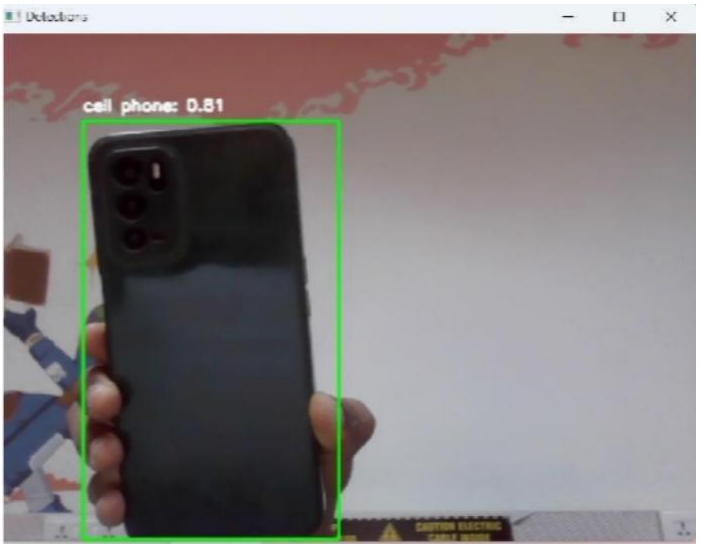
The evaluation metrics used in this study include mean Average Precision (mAP), Intersection over Union (IoU), and Frames Per Second (FPS). These metrics were selected to assess both the accuracy and speed of the object detection models.

4.2 Results on Image Detection

The following table presents the performance comparison of YOLO, Faster R-CNN, and the hybrid model in detecting objects in static images

model	mAP	IoU	FPS
YOLO	0.60	0.55	30
RCNN	0.75	0.70	5
Hybrid model	0.77	0.73	15





Conclusion

This paper presents a hybrid object detection framework that combines the strengths of YOLO and Faster R-CNN. By leveraging YOLO's real-time detection capabilities and Faster R-CNN's accuracy, we created a model that offers a balance between speed and precision. The hybrid model outperforms both standalone models in detecting objects in static images and videos, achieving a higher mAP and better IoU with real-time applicability.

Future Work

Future work will focus on optimizing the hybrid model for mobile and embedded systems, where resource constraints present unique challenges. Additionally, further research could explore the integration of temporal information for enhanced video detection, improving the model's ability to track objects across frames.

References

1. Redmon, J., & Farhadi, A. (2016). YOLO: You Only Look Once. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

2. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*.

3. Lin, T.-Y., et al. (2014). Microsoft COCO: Common Objects in Context. *European Conference on Computer Vision*.

4. Y. LeCun, Y. Bengio, and G. Hinton, ‘‘Deep learning,’’ Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

5. Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Ling and Qinghua Hu, arXiv 2018. Vision Meets Drones: A Challenge

6. Ren, S., He, K., Girshick, R., & Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), pp. 1137-1149

7. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A., 2016. YOLO9000: Better, Faster, Stronger. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517-6525

8. He, K., Gkioxari, G., Dollár, P., & Girshick, R., 2017. Mask R-CNN. IEEE International Conference on Computer Vision (ICCV), pp. 2980-2988

9. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Reed, S., 2016. SSD: Single Shot MultiBox Detector. European Conference on Computer Vision (ECCV), pp. 21-37.

10. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S., 2017. Feature Pyramid Networks for Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2117-2125.