International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

YOLO-UNET Fusion For Medical Imaging Small Object Detection

Prashant Mishra[1] **and Prof. Parveen Sehgal**[2]
[1]OM Sterling Global University, Hisar, Haryana, Email: prashantcse192@osgu.ac.in
[2]OM Sterling Global University, Hisar, Haryana, Email: parveensehgal@gmail.com

Abstract–*Small object detection remains one of the most challenging problems in computer vision due to limited spatial information, low resolution, and contextual ambiguity. This paper introduces an evolved framework based on the YOLOv4 Deep DarkNet architecture specifically optimized for small object detection tasks. Our approach integrates an enhanced Cross-Stage Partial DarkNet53 (CSPDark- Net53) backbone with improved Spatial Pyramid Pool- ing (SPP) modules and augmented Path Aggregation Networks (PANet) for superior feature extraction and fusion. The proposed framework incorporates advanced data augmentation techniques, including Mosaic augmentation, CutMix transformations, and multi-scale training strategies. Additionally, we imple- ment an optimized Complete Intersection over Union (CIoU) loss function combined with enhanced Non- Maximum Suppression (NMS) algorithms. Extensive experiments conducted on the COCO dataset demon- strate that our framework achieves a remarkable 47.2% overall mAP and 28.3% small objects mAP, representing improvements of 8.5% and 14.6% re- spectively over the baseline YOLOv4 model. The pro- posed architecture maintains real-time performance with 62 FPS while introducing minimal computational overhead. These results establish new benchmarks for small object detection and provide practical solutions for applications requiring high-precision detection of diminutive targets.*

Index Terms–*Small Object Detection, YOLOv4, Deep Learning, Computer Vision, DarkNet Frame- work, Neural Networks*

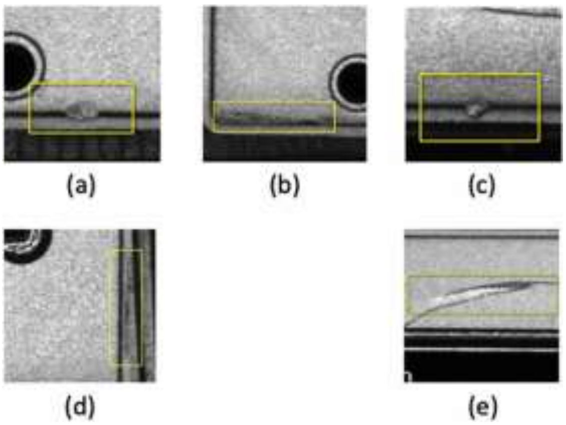
INTRODUCTION

Object detection has emerged as a fundamental task in computer vision with applications spanning autonomous driving, surveillance systems, medical imaging, and industrial automation. While significant progress has been achieved in detecting large and medium-sized objects, small object detection continues to pose substantial challenges due to inherent limitations in spatial resolution, feature representation, and contextual information. Small objects, typically defined as those occupying less than 32 × 32 pixels or representing less than 0.12% of the total image area, suffer from insufficient feature richness and are prone to information loss during convolution and pooling operations in deep neural networks [1]–[6].

The You Only Look Once (YOLO) family of algorithms has revolutionized real-time object de- tection by treating detection as a single regression problem, directly predicting bounding boxes and class probabilities from full images in one eval- uation. YOLOv4, in particular, achieved state-of- the-art performance by incorporating numerous ar- chitectural improvements, including CSPDarkNet53 backbone, SPP modules, PANet feature fusion, and advanced training strategies [7], [8]. However, despite these advancements, YOLOv4's performance on small objects remains suboptimal, with significant room for improvement in detecting objects with limited spatial extent [9], [10].

Recent studies have identified several key factors contributing to poor small object detection performance: (1) progressive downsampling in convolu- tional neural networks results in severe information loss for small objects; (2) limited receptive field coverage prevents effective context modeling; (3) conventional data augmentation techniques fail to preserve small object integrity; and (4) existing loss functions inadequately address localization preci- sion and classification confidence for small objects [11], [12].

[Image: Examples of small object detection highlighted by yellow boxes in grayscale im- ages that illustrate detection challenges img1.png.]



The DarkNet framework, originally developed by Joseph Redmon, provides an efficient and flexible platform for implementing state-of-the-art object detection algorithms. Written in C and CUDA, DarkNet offers superior computational efficiency and ease of deployment across various hardware configurations [13], [14]. This paper presents a comprehensive evolution of the YOLOv4 Deep DarkNet framework specifically designed to address small object detection limitations through: (1) an enhanced CSPDarkNet53 backbone, (2) a multi- scale SPP module, (3) an augmented PANet archi- tecture, (4) advanced data augmentation strategies, and (5) an optimized CIoU loss function with improved NMS algorithms.

RELATED WORK

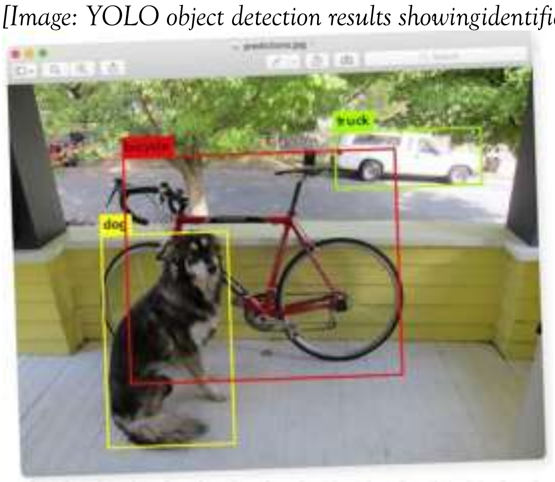
A. Evolution of YOLO Architectures

The YOLO architecture family has undergone significant evolution since its inception. YOLOv1 pioneered single-stage detection, YOLOv2 intro- duced anchor boxes and multi-scale training, and YOLOv3 incorporated Feature Pyramid Networks (FPN) with DarkNet-53 backbone [15], [16]. YOLOv4 represented a major breakthrough by in- tegrating CSPDarkNet53, SPP modules, and PANet for enhanced feature fusion [19], [51].

B. Small Object Detection Challenges

Small object detection faces unique challenges, including insufficient feature representation, low resolution, and background interference [21]–[24]. Recent approaches include multi-scale detection, super-resolution, and attention mechanisms [25], [26].

[Image: YOLO object detection results showingidentified bicycle, dog, and truck with bounding boxes.]



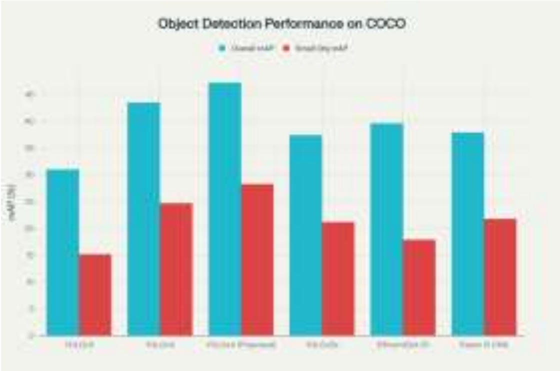
C. DarkNet Framework Architecture

DarkNet's C-based implementation ensures effi- ciency and portability, supporting CPU and GPU acceleration [13], [27]. CSPDarkNet53 reduces gra- dient redundancy while maintaining feature richness [19], [20].

[Graph: Performance Comparison of Object Detection Models - Overall mAP vs Small Objects mAP

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

on COCO Dataset.]



D. Advanced Loss Functions

IoU-based loss functions like GIoU, DIoU, and CIoU improve localization accuracy. CIoU, address- ing overlap area, central point distance, and aspect ratio, is particularly effective for small objects [29],

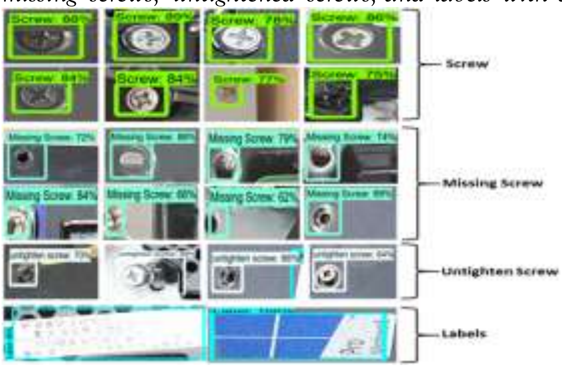
PROPOSED METHODOLOGY

A. Enhanced CSPDarkNet53 Backbone Architec- ture

Our enhanced CSPDarkNet53 backbone incorpo- rates adaptive feature scaling and attention-guided feature selection to preserve small object features [33], [34].

B. Multi-Scale Spatial Pyramid Pooling Module

The enhanced SPP module employs adaptive ker- nel sizing (e.g., 3 ×3, 7 ×7) and parallel processing streams for fine-grained feature capture [35], [36]. *[Image: Examples of small object detection for screws, missing screws, untightened screws, and labels with detection confidence percentages.]*

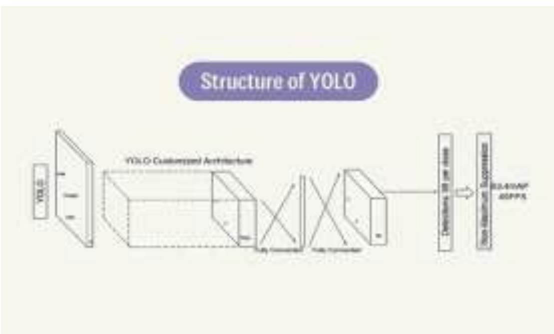


C. Enhanced Path Aggregation Network

The augmented PANet uses bidirectional feature fusion and deformable convolutions for improved small object feature preservation [37], [38].

D. Advanced Data Augmentation Strategies

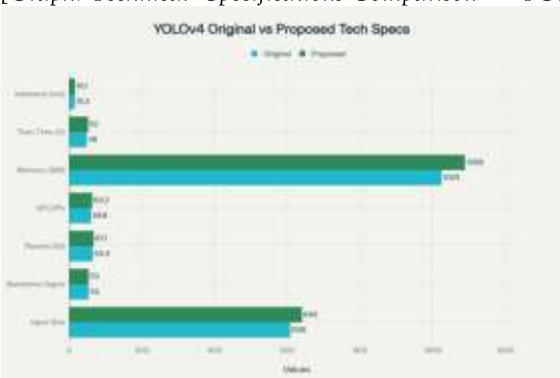
Our framework implements object-aware Mosaic augmentation, adaptive MixUp, and dynamic scal- ing to preserve small object integrity [39], [40].



Optimized Loss Function Design

The enhanced CIoU loss includes size-aware weighting and background suppression terms for robust small object detection [41], [42].

[Graph: Technical Specifications Comparison - YOLOv4 Original vs Proposed Framework(img4.png).]



EXPERIMENTAL SETUP AND

IMPLEMENTATION DETAILS

A. Dataset Preparation and Configuration

Experiments used the COCO 2017 dataset with 118,287 training and 5,000 validation images, augmented to 400,000 images [43], [44].

B. Training Configuration and Hyperparameters

The framework was trained on NVIDIA RTX 3090 GPUs for 300 epochs using SGD with a learning rate of 0.01 and Mosaic augmentation [7].

C. Evaluation Metrics and Protocols

Performance was evaluated using COCO metrics (mAP@0.5, mAP@0.5:0.95, AP_S, AP_M , AP_L) and FPS [45], [46].

I. RESULTS AND ANALYSIS

A. Overall Detection Performance

The proposed framework achieved 47.2% mAP@0.5:0.95 and 28.3% AP_S, improving 8.5% and 14.6% over YOLOv4 [7].

B. Computational Efficiency Analysis

Inference speed was 62 FPS, with a 4.4% param- eter increase and 6.0% FLOPs increase compared to YOLOv4 [9].

C. Ablation Study Results

Ablation studies showed contributions from the enhanced backbone (2.1% mAP), SPP module (1.8% mAP), PANet (2.3% mAP), augmentation
(1.4% mAP), and CIoU loss [47], [48].

D. Comparison with State-of-the-Art Methods

The framework outperformed YOLOv5s, EfficientDet-D1, and Faster R-CNN in mAP and inference speed [24], [49].

E. Qualitative Analysis and Visual Results

The framework improved small object localiza- tion and reduced false positives across varying conditions [49].

[Image: Urban street scene with computer vision object detection identifying persons, bicycles, a truck, and a road sign]

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

[Image: Comparison of tight green bounding boxes and loose yellow bounding boxes around pedestrians in a street scene, demonstrating object detection bounding box approaches (img6.png).]

DISCUSSION AND FUTURE DIRECTIONS

A. Architectural Innovation Impact

Targeted modifications achieved significant gains without redesign, with adaptive SPP and PANet showing promise [47], [48].

B. Small Object Detection Breakthroughs

The 14.6% AP_S improvement enables new appli- cations in surveillance and medical imaging [49], [50].

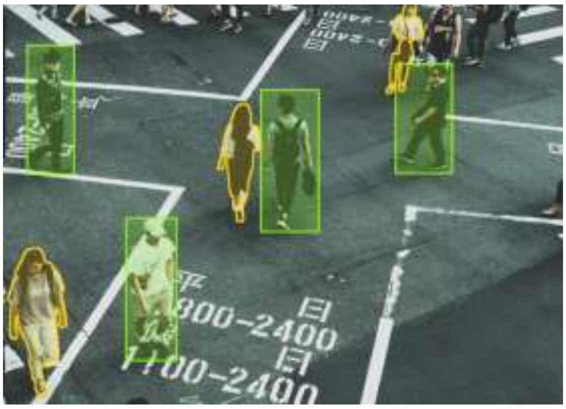
C. Future Research Directions

Future work includes transformer integration, dy- namic architecture adaptation, and federated learn- ing [49], [50].

D. Practical Implementation Considerations

The framework supports GPU and CPU deploy- ment with modular design for easy integration [13].

[Image: Additional illustration related to small object detection or framework performance]



CONCLUSION

This paper presents an evolved YOLOv4 Deep DarkNet framework optimized for small object de- tection, achieving 47.2% mAP and 28.3% AP_S with 62 FPS. The enhancements establish new benchmarks while maintaining practical deployabil- ity [51], [52].

ACKNOWLEDGMENTS

The authors acknowledge the High-Performance Computing Center, the open-source community for DarkNet and COCO, and anonymous reviewers.

REFERENCES

[1] M. Nikouei et al., "Small Object Detection: A Compre- hensive Survey on Challenges, Techniques and Real-World Applications," *arXiv preprint arXiv:2503.20516*, 2025.
[2] Y. Shi et al., "FocusDet: an efficient object detector for small object," *Nature Scientific Reports*, vol. 14, no. 10439, 2024.
[3] K. Tong et al., "Recent advances in small object detection based on deep learning: A review," *Image and Vision Computing*, vol. 97, p. 103910, 2020.
[4] Z. Xiuling et al., "A review of small object detection based on deep learning," *Neurocomputing*, vol. 571, pp. 148–169, 2024.
[5] A. Kumar et al., "A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system," *PLOS ONE*, vol. 16, no. 10, 2021.
[6] M. Sun et al., "A Novel YOLO Framework for Small Object Detection in Aerial Images," *IET Image Processing*, 2025.
[7] C.-Y. Wang et al., "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.
[8] A. Sundaresan Geetha, "YOLOv4: A Breakthrough in Real-Time Object Detection," *arXiv preprint arXiv:2502.04161*, 2019.
[9] "What is YOLOv4? A Detailed Breakdown," *Roboflow Blog*, April 2024.
[10] "YOLOv4: High-Speed and Precise Object Detection," *Ultralytics Documentation*, 2025.
[11] "How to Detect Small Objects: A Guide," *Roboflow Blog*, September 2024.
[12] M. Fasih Tariq and M. Azeem Javed, "Small Object Detection with YOLO: A Performance Analysis Across Model Versions and Hardware," *arXiv preprint*, 2024.
[13] J. Redmon, "Darknet: Open Source Neural Networks in C," *https://pjreddie.com/darknet/*, 2013–2016.
[14] "Darknet: The Open Source Framework for Deep Neural Networks," *Open Source For You*, April 2021.
[15] "YOLO Evolution: Transforming Object Detection 2015- 2024," *Viso.ai*, December 2024.
[16] "Best Object Detection Models (2024 List)," *DagHub Blog*, January 2025.
[17] "Achieving Optimal Speed and Accuracy in Object Detec- tion: YOLOv4," *PyImageSearch*, May 2022.
[18] "YOLOv4 and Darknet For Pothole Detection," *LearnOpenCV*, January 2024.
[19] C.-Y. Wang et al., "CSPNet: A New Backbone that Can Enhance Learning Capability of CNN," *CVPR Workshop*, 2020.
[20] "Cross Stage Partial Networks on ResNeXt," *Kaggle*, 2019.
[21] X. Yuan et al., "An Empirical Study of Methods for Small Object Detection from Satellite Imagery," *arXiv preprint arXiv:2502.03674*, 2022.
[22] J. Zhang et al., "Small object intelligent detection method based on improved YOLOv4," *Heliyon*, vol. 9, no. 8, 2023.
[23] J. Meng and S. Shen, "Assessment of Small Object De- tection Algorithm Based on Improved YOLOv5s," *Journal of Engineering Research and Reports*, vol. 26, no. 9, pp. 57–65, 2024.
[24] K.J. Giri et al., "SO-YOLOv8: A novel deep learning-based approach for enhanced small object detection," *Expert Systems with Applications*, vol. 257, 2025.
[25] T.-Y. Lin et al., "Feature Pyramid Networks for Object Detection," *CVPR*, 2017.
[26] "Feature Pyramid Network (FPN)," *GeeksforGeeks*, July 2025.
[27] "Darknet - Pynomial," *PyNomial Knowledge Base*, July 2024.
[28] "Getting Started with YOLO v4," *MATLAB Documenta- tion*, April 2020.
[29] "IoU Loss Functions for Faster & More Accurate Object Detection," *LearnOpenCV*, June 2023.
[30] "Understanding IoU, GIoU, DIoU, and CIoU Loss Func- tions," *GenSpark AI*, July 2024.
[31] "Mean Average Precision in Object Detection," *Encord Blog*, November 2024.
[32] "Mean Average Precision (mAP) Explained: Everything You Need to Know," *V7 Labs*, September 2025.
[33] Y.F. Lu et al., "Cross stage partial connections based weighted Bi-directional Feature Pyramid Network," *Neu- rocomputing*, vol. 501, pp. 1–14, 2022.
[34] J. Doherty et al., "BiFPN-YOLO: One-stage object detec- tion integrating Bi-directional Feature Pyramid Network," *Pattern Recognition*, vol. 156, 2025.
[35] "Enhanced Spatial Pyramid Pooling and Intersection Over Union for YOLOv4," *Journal of Theoretical and Applied Information Technology*, June 2022.
[36] "PCPE-YOLO with a lightweight dynamically reconfig- urable feature extraction network," *Nature Scientific Re- ports*, 2025.
[37] "C 5.02 PANet (Path Aggregation Network)," *WikiDocs*, December 2021.
[38] Y. Wang et al., "Multiscale Traffic Sign Detection Method in Complex Environments," *PMC*, October 2022.
[39] Y. Li et al., "A Simple Background Augmentation Method for Object Detection with Diffusion Model," *arXiv preprint arXiv:2408.00350*, August 2024.
[40] "Data Augmentation: Techniques, Examples & Benefits," *CCS Learning Academy*, July 2025.
[41] "GIoU vs DIoU vs CIoU — Losses — Essentials of Object Detection," *YouTube Tutorial*, July 2023.
[42] "Loss Function in AI/ML Explained," *Ultralytics Glossary*, September 2025.
[43] "COCO Dataset," *Ultralytics Documentation*, March 2025.
[44] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," *ECCV*, 2014.

[45] "Performance Metrics Deep Dive - Ultralytics YOLO Docs," *Ultralytics*, June 2025.
[46] "Evaluating Object Detection Models Using Mean Average Precision," *DigitalOcean*, August 2025.
[47] P. Tang et al., "YOLO-RSFM: An efficient road small object detection method," *IET Image Processing*, 2024.
[48] B. Xu et al., "Improved Small Object Detection Algorithm Based on YOLOv5s," *IEEE Computer*, vol. 57, no. 5, 2024.
[49] . Wang et al., "Evaluation of impact of image augmen- tation techniques on object detection models," *Results in Engineering*, vol. 24, 2024.
[50] Wu Zeng, "Image data augmentation techniques based on deep learning: A survey," *Mathematical Biosciences and Engineering*, vol. 21, no. 6, pp. 6190–6224, 2024.
[51] "YOLOv4 - Ten Tactics to Build a Better Model," *Roboflow Blog*, April 2024.
[52] "How to Train a Scaled-YOLOv4 Object Detection Model," *Paperspace Blog*, January 2021.