Ahmedabad University

CSE 641 Computer Vision

Weekly Report

Project Title: Evaluate Performance of YOLO family models in case of small object detection (HBB)

Section - 1

Group - 02

Name	Roll No
Romir Bedekar	AU2240254
Hetav Shah	AU2240253
Namra Maniar	AU2240080

Introduction

This study aims to evaluate the performance of various YOLO models in small object detection. We explore different YOLO variants, their architectures, evaluation methodologies, and key differences.

List of Yolo Family:

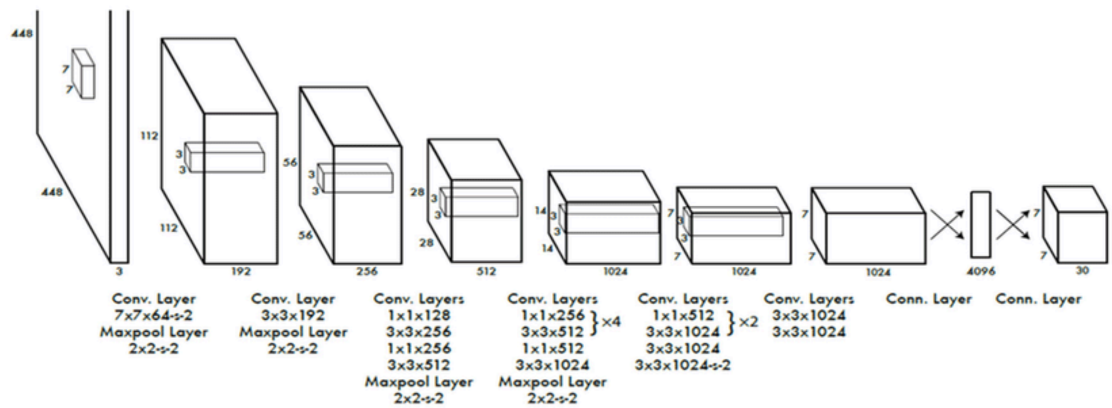
- YOLO (June 8, 2015)
- YOLOv2 aka YOLO9000 (December 25, 2016)
- YOLOv3 (April 8, 2018)
- YOLOv4 (April 23, 2020)
- Scaled-YOLOv4 (June 9, 2020)
- PP-YOLO (July 23, 2020)
- YOLO-R (November 16, 2020)
- YOLOX (June 1, 2021)
- YOLOS (May 10, 2021)
- YOLOv5 (April 23, 2020)
- YOLOv6 (July 18, 2021)
- YOLOv7 (June 2022)
- YOLOv8 (January 6, 2022)
- YOLOv6 3.0 (January 10, 2023)
- YOLOv9 (January 30, 2024)
- YOLOv10 (February 21, 2024)
- YOLOv11 (May 23, 2024)
- YOLO-NAS (May 2, 2023)
- YOLO-World (September 30, 2024)

From all these models, we are considering YOLO -World, YOLO-NAS , YOLOX ,Scaled YOLOv4, PP-YOLO for evaluating performance in case of small object detection.

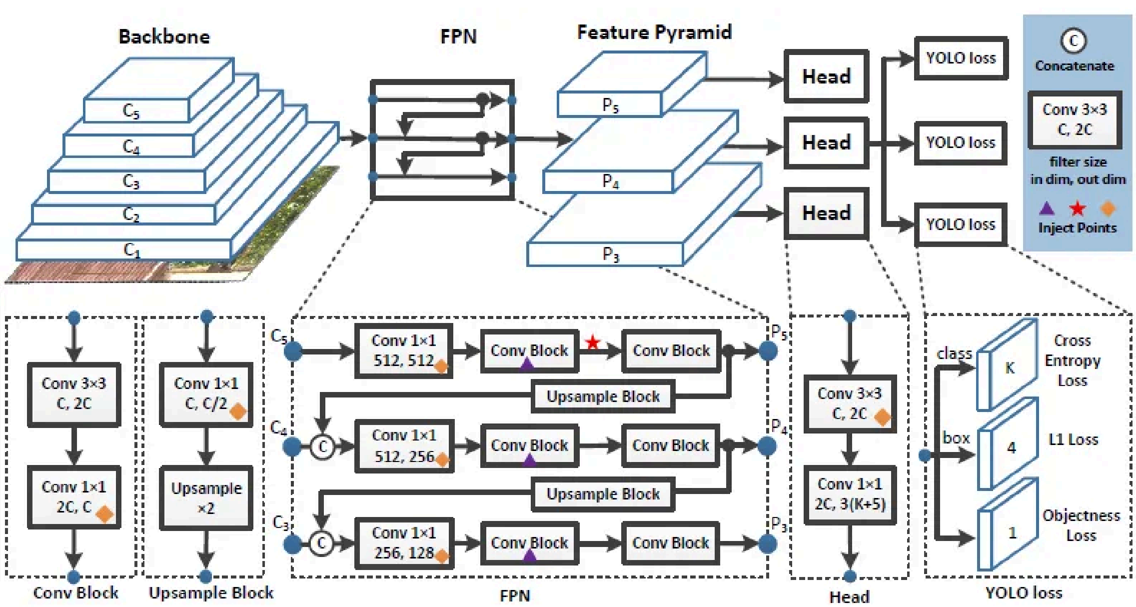
1. YOLO Family Models Overview

1.1 YOLO (You Only Look Once)

- Introduced by Joseph Redmon, YOLO is a one-stage object detection model.
- Compared to two-stage detectors like Faster R-CNN, YOLO processes images in a single pass that is quick but sometimes inaccurate.
- Later improved in the accuracy, efficiency, and also in adaptability from YOLOv2 to YOLOv8.



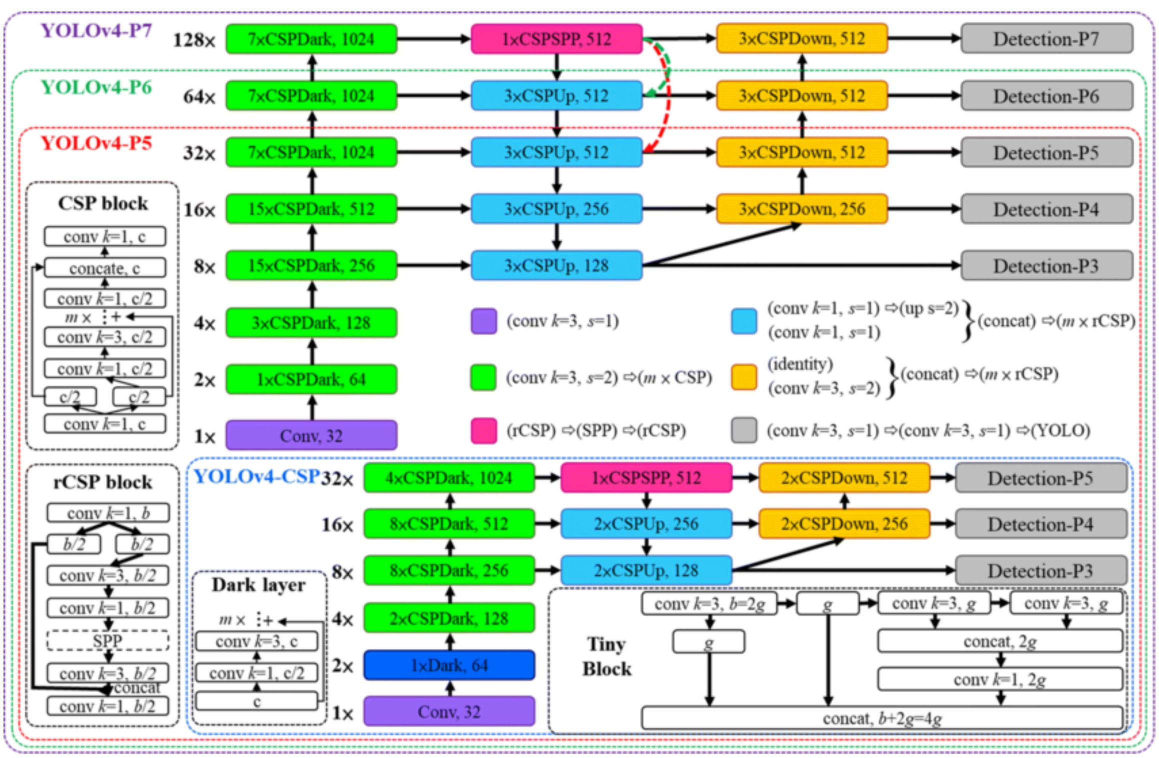
1.2 PP-YOLO



- PP-YOLO was built upon YOLOv4, embedded with the following improvements:
- Feature fusion via Path Aggregation Network (PANet).
- Use of ResNet50-vd as a backbone to enhance feature extraction.
- Synchronized batch normalization to stabilize the training.
- IoU-aware prediction to improve the accuracy of localization.
- Performance: Faster and more accurate than YOLOv4 while also being computationally efficient.

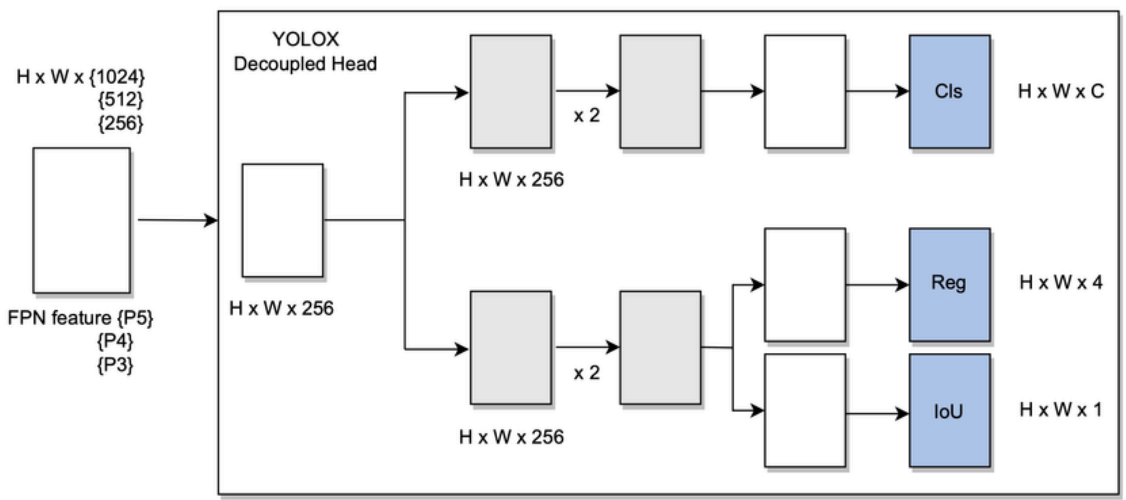
1.3 Scaled YOLOv4

- Introduced scaling of YOLOv4 for various hardware settings.
- It uses CSPDarknet53 as a backbone and improves feature extraction.
- It features Mish activation, CSPNet, and multi-input spatial attention for improved small object detection.
- Performance: Scaled to accommodate different size runs (tiny, small, large), making it adaptable to various deployment environments.



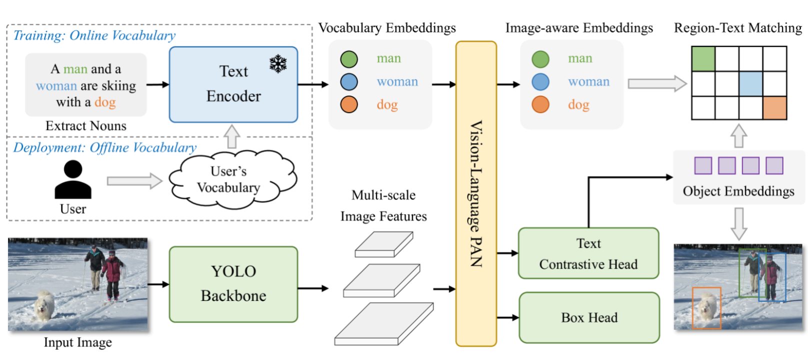
1.4 YOLOX

- It re-implements YOLOv4 with an anchor-free paradigm.
- It uses decoupled heads to classify and regress separately.
- It brings SimOTA (Optimal Transport Assignment) for better assignment of training labels during training.
- Performance: Results in higher accuracy at a bit of speed loss.



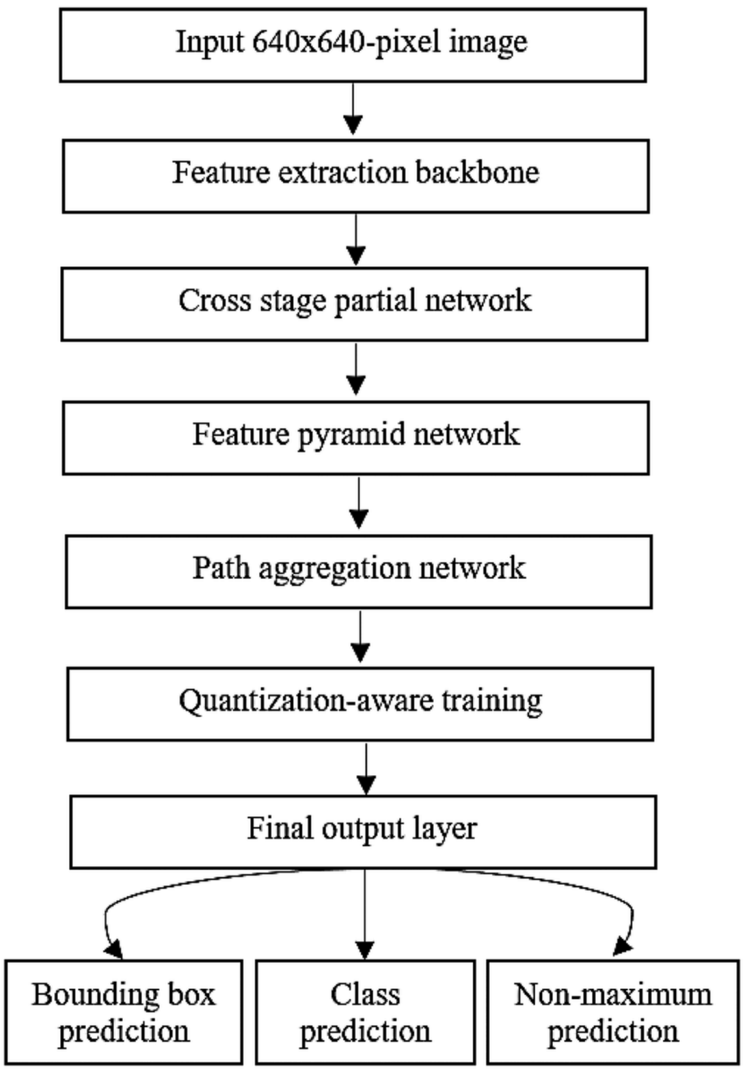
1.5 YOLO-World

- It was developed by Meta AI with a focus on open vocabulary object detection.
- Several datasets were used in training for the on-set of generalization.
- Can recognize previously unseen objects by their textual description.
- Performance: Good for applications needing varied ranges of objects.



1.6 YOLO-NAS

- This is a very recent YOLO model that uses neural architecture search (NAS) for its optimization.
- Use of RepVGG-based backbone allows for better feature extraction.
- Features quantization-aware training for optimized performance on edge devices.
- Performance: Optimized for real-time inference at low latency.



2. Evaluation Criteria

To compare these models for small object detection, we focus on:

- **mAP (Mean Average Precision):** Evaluates detection accuracy.
- **IoU (Intersection over Union):** Measures bounding box precision.
- **FPS (Frames Per Second):** Determines inference speed.
- **Small Object Recall:** Measures detection capability on small objects.