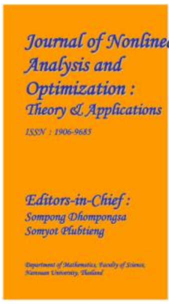
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.3 : 2024
ISSN : **1906-9685**



AN ANALOGY ANALYSIS OF THE OBJECT DETECTION ALGORITHMS USING YOLOV5, YOLOV7, AND YOLOV8

Mrs. S. Lakshmi devi, Research Scholar, Department of Computer Science Sri Krishna Adithya College of Arts and Science Kovaipudur, Coimbatore, Tamil Nadu, India
Dr. D. Sridhar Assistant Professor Department of Computer Science Sri Krishna Adithya College of Arts and Science Kovaipudur, Coimbatore, Tamil Nadu, India

Abstract—
The term "object recognition" refers to a branch of computer vision that deals with finding explicit instances of semantic items in digital photos and videos. Object detection's primary mission is to locate and identify one or more effective targets from still images or video footage. It broadly integrates a critical range of techniques, such as image processing, pattern recognition, artificial intelligence, and machine learning. Several additional technologies, like face recognition, self-driving automobiles, vehicle detection, and others, utilize object recognition. Since real-time object recognition is a dynamic and challenging area of computer vision, it requires more processing power to quickly identify the item at that precise moment. In this work, the most recent versions of YOLOv7 and YOLOv8, as well as the widely used YOLOv5, are compared. This article compares the widely used YOLOv5 with the most recent YOLO variants, YOLOv7 and YOLOv8. Experiments were carried out by training a bespoke model using YOLOv5, YOLOv7, and YOLOv8 individually in order to ascertain which of the three models performs best in terms of precision, recall, and mAP@0.5:0.95. to determine which performs better in terms of mAP@0.5:0.95, accuracy, and recall. A bespoke dataset consisting of 2,216 images was used in the experiment to detect objects. The information has been taught to identify characteristics and distinguish across states. Using training a custom version with YOLOv5, YOLOv7, and YOLOv8 individually, studies were carried out to determine which performs better in terms of accuracy, recall, and mAP@0.5:0.95. of 32.5%, YOLOv5 had mAP@0.5:0.95 of 37.5, recall values of 56.4, and precision scores of 52.8%. The trial's findings showed that YOLOv8 performed better.

Keywords—
Fatigue, drowsiness, object detection, IOU, darknet53, YOLO.

I Introduction

Finding and detecting objects in images and videos is an obstacle for object detection in computer vision. The launch of You Only Look Once (YOLO) in 2016 signalled a significant advancement in the field of object identification. Due to its extraordinary speed and accuracy, it constituted a substantial development, outperforming the performance of the most effective algorithms at the time (Redmon et al., 2016). Because of YOLO's remarkable efficiency and accuracy in identifying and establishing object coordinates, it established a new standard for object identification algorithms. YOLO is an object detection technique that views object detection as a regression challenge in order to find many items within one picture.

Real-time picture analysis is accomplished by the original YOLO model at a frame rate of 45 FPS. Fast A more compact version called YOLO processes photographs at an astounding 155 frames per second and achieves twice the mean Average Precision (mAP) compared to previous real-time detectors. The YOLOv8 'n' version works well for embedded devices like the Jetson Nano. Numerous variations of YOLO have been freed, including YOLOv1 to YOLOv7, and YOLOv8. The point of analysis is to contrast the effectiveness of various YOLO algorithms in order to identify which one performs the best overall. The YOLO technique is used in a variety of fields, such as computer vision

(CV) tasks, drones, military operations, autonomous vehicles, healthcare settings, and wildlife monitoring (Górriz et al., 2020). YOLO has seen different versions over the years, including YOLOv1 to YOLOv7. The intent of this study is to gauge the effectiveness of various YOLO algorithms in order to identify which one performs the best overall. According to earlier research, In terms of accuracy and speed, YOLOv5 outperforms YOLOv3 and YOLOv4 (Sahal, 2021; Ramya et al., 2021). YOLOv8, which was just released, has also shown promising results when used in a cohesive framework. This leads to the primary goal of the study, which is to compare how well YOLOv5, YOLOv7, and YOLOv8 perform in object detection tasks.

II Framework of YOLO

The YOLO (You Only Look Once) phenomenon has its roots in 2015 when University of Washington researcher At the Computer Vision and Pattern Recognition (CVPR) conference, Joseph Redmon delivered a talk entitled "You Only Look Once: Unified, Real-Time Object Detection". End-to-end training is possible with the YOLO architecture, which also provides real-time processing speeds without sacrificing high average precision. The technique requires creating an SS grid from the input image. When an object's centre lies inside a grid cell, that specific cell is responsible for detecting the object. Bounding boxes are forecasted in each grid cell, along with associated confidence scores for those boxes. Confidence is measured using the intersection over union (IOU) between the predicted box and the ground truth box. If an object is absent from a given cell, the confidence ratings are reset to zero. However, if an object is present, the estimated IOU value equals the confidence score. Each bounding box is described by five predictions: (w, h) that represent the box's width and height in reference to the entire image; and (x, y) coordinates that show the box's centroid in relation to the borders of the grid cell. The fifth prediction is the confidence value, which shows the IOU between the expected box and the actual box.

III ARCHITECTURE – YOLOv8

The backbone and the head are the two main parts of the convolutional neural network used by YOLOv8, an improvement over earlier YOLO algorithms. The basis of the system's structure is the CSPDarknet53 architecture, which is made up of 53 convolutional layers with cross-stage partial connections to enhance information flow between layers. Several convolutional layers followed by fully connected layers comprise the head of YOLOv8, which forecasts boundaries, objectness scores, and class probabilities for recognized objects in an image. The inclusion of a self-attention mechanism in the network's brain is one of YOLOv8's standout features. Using this approach, the model may vary its focus to different areas of the image and change how important different features are depending on the job at hand. The capacity of YOLOv8 to carry out multi-scaled object identification is another crucial feature. This is accomplished using a feature pyramid network, which consists of layers with different specializations for finding objects in the image that are different sizes and scales. As a result, YOLOv8 performs better in object detection tasks since it can recognize both big and tiny things with equal effectiveness.

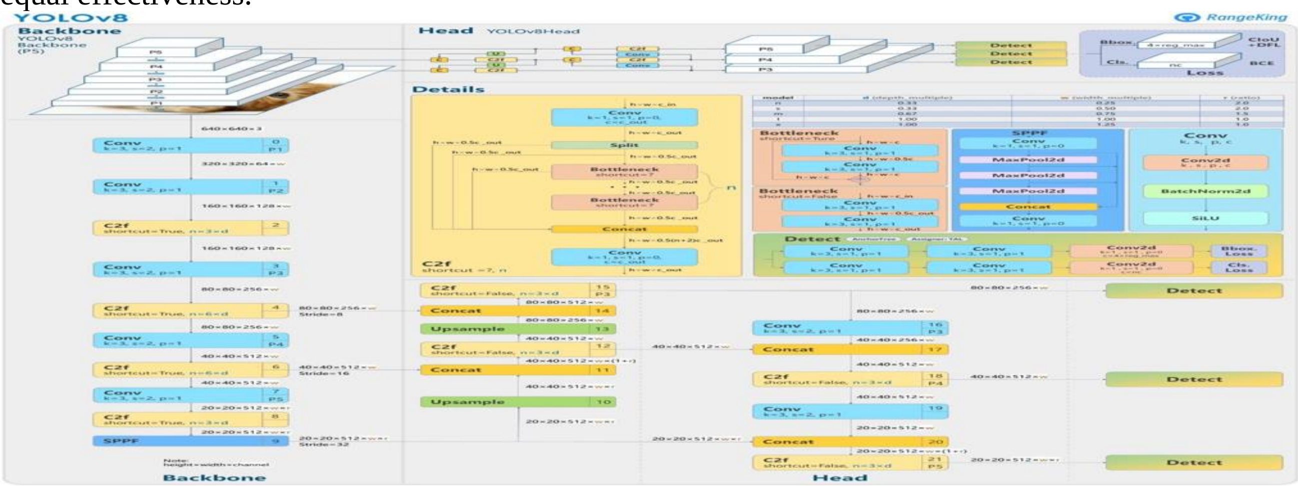


Fig 1. YOLOv8 architecture

Fig 2. YOLO releases

Ameliorations of YOLO –Versions

S.NO	YOLO RECENSION	Growth	Result
1	YOLO (Joseph Redmon, 2016)	• YOLO – single shot detection approach • It predicts object bounding boxes and class identification in one pass.	Real-time detection with high accuracy compare to RCNN.
2	YOLOv2 and YOLO 9000 (J. Redmon and A. Farhadi, 2017)	• Multi-scale detection, Darknet-19 backbone architecture -19 conventional layers. • Evolutionary enhancements in Batch Normalization, object detection at higher resolutions, and the integration of anchor boxes have been progressively implemented.	Exhibiting increased speed and higher precision compared to original YOLO
3.	YOLOv3(Joseph Redmon & and A. Farhadi,2018)	• Detecting objects at multiple scales Darknet-53, 53 conventional layers. • Sophisticated anchor box clustering algorithm to determine the default bounding box shapes.	Improved accuracy with three different scales and detect smaller objects.
4	YOLOv4 (Alexey Bochkovskiy, April 2020)	• CSPDarknet-53 backbone • PaNet for feature aggregation • CSPResNeXt as an alternative backbone. • CutMix and Mixup regularization techniques used during training. • SAM module focus on spatial regions.	State-of-art-of performance
5.	YOLOv5 (Glen Jocher, June 2020)	• G Simplified architecture with high accuracy. • Utilizes PyTorch. • Grid-based prediction mechanism to detect objects. • Improved augmentations. • Multiple model size(YOLO%s, YOLO5m, YOLOv5l, YOLO5x)	Competitive results with reduced complexity

6	YOLOv6 (Chuyi & September 2022)	• Anchor-free paradigm • SIoU bounding box regression loss; to further increase the detection accuracy, dynamically allocate positive samples.	YOLOv6 improves both speed and accuracy compared to its predecessors,
7	YOLOv7(Alexey Bochkovskiy & July 2022)	• E-ELAN layer aggregation, trainable bag of freebies, and a reduction of 35% in network parameters. Scaling a model based on concatenations • Ylo7 basic models YOLOv7-tiny, YOLOv7-W6, YOLOv7-E6, YOLOv7-D6, YOLOv7-6E6.	Quickness and precision enhancement, ease of training, and inference
8	YOLOv8 (Glenn Jocher January 2023)	• The most recent iteration of Ultralytics' YOLO. • state-of-the-art, cutting-edge (SOTA) model,	Key strength – performance and versatility supports AI tasks – detection, segmentation, tracking, pose estimation, classification.

Comparison between structures of YOLOv5, YOLOv7, YOLOv8

	YOLOV5	**YOLOV7**	**YOLOV8**
Neural Network	Fully convolution - Efficient Net network	Fully convolution - ResNeXt.	Fully convolution
Backbone Network	CSP Darknet53	CBS, E-ELAN, MP, and SPPCSPC modules	CSP Darknet53, C2F modules
Loss Function	Binary cross entropy & Logits loss function	Focal Loss	DFL – cross entropy optimization, IoU – between predicted and bonding boxes
Neck	PANet	Features Pyramids	PAN - Path Aggregation Network
Detection Head	YOLO layer	Lead Head & Auxiliary Head	
Training Techniques	Dynamic anchor box	Re-parameterized Convolution (RepConvN)	Swish Activation

IV Methodology

Individual YOLOv5, YOLOv7, and YOLOv8 models were trained on custom datasets as part of the tests. The primary goal was to determine which of these models performed best in terms of mAP@0.5, mAP@0.5:0.95, recall, and precision. These metrics evaluate the detecting system's overall performance. The quantitative examination of the models employed the following metrics.

Precision

This statistic compares the number of correctly and mistakenly classified positive samples (True Positive + False Positive) to the proportion of correctly identified positive samples (True Positive). The precision of the model refers to how well it detects favourable events.

$$Precision = \frac{TP}{TP + FP}$$

Recall

The recall value is calculated by dividing the total number of actual positive samples (True Positive + False Negative) by the count of True Positives. It assesses the model's ability to identify positive instances among all of the actual positives.

$$Recall = \frac{TP}{TP + FN}$$

mAP 0.5 to 0.95

The average mAP is determined by this metric with respect to a range of IOU criteria, namely from 0.5 to 0.95 in steps of 0.05. It provides a thorough assessment of the model's performance in relation to several IoU criteria.

$$\boldsymbol{mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k}$$
$$\boldsymbol{AP_k = \text{ the AP of class } k}$$
$$\boldsymbol{n = \text{ the number of classes}}$$

The experimentation aimed to determine which model excels in terms of these keys metrics, thus offering insights into their respective detection capabilities.

V Data Description

The study employed a heterogeneous dataset that was obtained from multiple sources, such as locally taken photos, the Roboflow Public Dataset, and the Google Open Photos Dataset. The primary contribution from the Google Open Images Dataset consisted of 3006 images, encompassing distinct classes like human, vehicle, animals and birds. Furthermore, locally acquired images were integrated into the dataset, depicting individuals, vehicles, face from different environment. Captured using a high-definition camera, these images adhered to a resolution of 1200 x 720 pixels, consistent with the specifications of YOLOv8, YOLOv7 and YOLOv5 models. To facilitate model training, meticulous annotation was conducted using the VGG Image Annotator tool, categorizing instances into human, vehicle, animals and birds classes.

VI Data Pre-processing

The images were downsized to 420 × 420 pixels (width x height) as part of the data pre-processing steps to comply with the input requirements of YOLOv5, YOLOv7, and YOLOv8. Auto-Orient adjustments were also made. The curated dataset comprised a total of 3639 images, featuring 4,561 annotations across the four classes. For model training, a split ratio of 60:20:20 was applied to segregate the dataset into training, testing, and validation subsets.

VII Result and Discussion

Table 2 shows the performance results, along with the corresponding output values, from testing the YOLOv8, YOLOv7, and YOLOv5 models**.**

Class	**PERCISION**			**RECALL**			**mAP@0.5to 0.95**		
	YOLO v8	YOLO v7	YOLO v5	YOLO v8	YOLO v7	YOLO v5	YOLO v8	YOLO v7	YOLO v5
Human	0.912	0.756	0.816	0.898	0,756	0.786	0.774	0.581	0.596
Vehicl es	0.832	0.587	0.715	0.867	0.746	0.693	0.546	0.426	0.487
Anima ls	0.686	0.378	0.509	0.427	0.514	0.407	0.189	0.173	0.181
Birds	0.542	0.353	0.448	0.312	0.217	0.254	0.192	0.0865	0.105

Table 1: Performance result of YOLOv8, YOLOv7 and YOLOv5

7. 1 Precision

For more accurate analysis, Table 2's comparison of the YOLOv8, YOLOv7, and YOLOv5 findings shows that YOLOv8 outperforms YOLOv5 & YOLOv7 in every case. In every class, YOLOv8 performs better than YOLOv5, which scored 62.6% and 81.9% overall and 71.6%, 51.1%, and 45.8% for the human, vehicle, animal, and avian classes, respectively. Conversely, YOLOv7 attains 52.8% across all classes and 77.8%, 58.8%, 38.2%, and 36.3% for the animal, bird, vehicle, and human classes, in that order. This study shows that, in comparison to YOLov5 and YOLOv7, YOLOv8 and YOLOv5 show a 9.8% difference in overall class detection, with a greater ratio of true positives to the total number of detected items.

7. 2 Recall

The recall analysis highlights humans as the class with the highest recall rate, with YOLOv8 achieving a recall of 79.6% in contrast to YOLOv7's 77.8%, resulting in a minor difference of 0.7%. YOLOv7 proves more efficient at differentiating between the human and Person classes, which added to a 3% total class recall advantage over YOLOv5. Interestingly, compared to YOLOv7, YOLOv5 also shows better recall rates in the identification of vehicle and animal classes. Comparing the outcomes presented in Table 2 for both mAP@0.5 and mAP@0.5:0.95, it is evident that YOLOv8 achieved higher accuracy than YOLOv5 and YOLOv7 across all instances. The comprehensive class-specific outcomes for mAP@0.5 and mAP@0.5:0.95 were 55.3% and 34.2%, respectively, for YOLOv5, surpassing YOLOv7's 51.2% and 31.5%. These mAP values, evaluated at an Intersection over Union (IOU) of 0.5, demonstrate the model's precise object detection within frames. YOLOv5's 4% variance from YOLOv7 in mAP@0.5 underscores its adeptness at accurately identifying objects compared to ground truth objects.

7.3 Accuracy in terms of mAP@0.5:0.95

Contrasting the outcomes presented in Table 2 for mAP@0.5:0.95, it is obvious that YOLOv8 achieved higher accuracy than YOLOv5 and YOLOv7 across all instances. The comprehensive class-specific outcomes for mAP@0.5:0.95 were for YOLOv8, 6.7% and 62.3% respectively for 55.6% and 34.4% respectively for YOLOv5, surpassing YOLOv7's 51.5% and 31.7%. These mAP values, evaluated at an Intersection over Union (IOU) of 0.5, demonstrate the model's precise object detection within frames. YOLOv5's 7% variance from YOLOv5 and YOLOv7 in mAP@0.5 underscores its adeptness at accurately identifying objects compared to ground truth objects.

VIII Conclusion

When mAP@0.5:0.95 findings are examined, it is clear that YOLOv8 outperformed YOLOv5 and YOLOv7 in terms of accuracy in every one of the cases. Regarding particular class outcomes, YOLOv5 surpassed YOLOv7 with values of 51.2% and 31.5%, exhibiting comprehensive results of 55.3% and 34.2% for mAP@0.5 and mAP@0.5:0.95, respectively. These mAP values demonstrate the model's accurate object detection inside individual frames, and they are computed with an Intersection over Union (IOU) of 0.5. The minuscule 4(deviation seen in mAP@0.5 between YOLOv5 and YOLOV7, YOLOv8 highlights YOLOv8's capacity to distinguish items correctly in comparison to the real ground truth objects.

IX References

1. Alexey B., Chien-Yao W., Hong-Yuan M. L. (2020) Yolov4: Optimal speed and accuracy of object detectionarXiv:2004.10934.
2. Oluwaseyi Ezekiel Olorunshola, Martins Ekata Irhebhude : A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms, December 2022.
3. Sahla Muhammed Ali, Comparative Analysis of YOLOv3, YOLOv4 and YOLOv5 for Sign Language Detection, IJARIIE-ISSN(O)-2395-439, Vol-7 Issue-4 2021
4. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
5. Hekai Zhang1 , Sichao Cheng ,YOLOv5-based fatigue state detection method, ISSN 2616-5775 Vol. 5, Issue 3: 28-34

6. REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advance in neural information processing systems, 2015, 28: 91-99.
7. Chanda Manideep, D.L.N.N.Manikant, Dr.V.Vaissnav, Real-Time Driver Drowsiness Detection Using YOLO and Machine Learning Technique, ISSN 2349-9249, April 2023 Volume 10, Issue 4.
8. Muhammad Hussain, YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection, June 2023, 11(7), 677
9. Jialin Zhang, Jiucai Jin, Lightweight object detection algorithm based YOLOv5 for unmanned surface vehicles, vol 9, January 2023.
10. Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark LiaoYOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, July 2022