IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Object Detection Using YOLOv8

[1]Pragati Satpute, [2] Prof. Yuvraj Nikam
[1]Student [2]Professor
[12]MIT School of Engineering, MIT-ADT University, Pune, India

Abstract: This study presents a groundbreaking approach to enhance the accuracy of the YOLOv8 model in object detection, focusing mainly on addressing the limitations of detecting objects in varied image types, particularly for small objects. The proposed strategy of this work incorporates the Context Attention Block (CAB) to effectively locate and identify small objects in images. Furthermore, the proposed work improves the feature extraction capability without increasing model complexity by increasing the thickness of the Coarse-to-Fine(C2F) block. In addition, Spatial Attention (SA) has been modified to accelerate detection performance. The enhanced YOLOv8 model (Namely YOLOv8-CAB) strongly emphasizes the performance of detecting smaller objects by leveraging the CAB block to exploit multiscale feature maps and iterative feedback, thereby optimizing object detection mechanisms. As a result, the innovative design facilitates superior feature extraction, "especially the weak features," contextual information preservation, and efficient feature fusion. Rigorous testing on the Common Objects in Context (COCO) dataset was performed to demonstrate the efficacy of the proposed technique. It is resulting in a remarkable improvement over standard YOLO models. The YOLOv8-CAB model achieved a mean average precision of 97% of detecting rate, indicating a 1% increase compared to conventional models. This study highlights the capabilities of our improved YOLOv8 method in detecting objects, representing a breakthrough that sets the stage for advancements in real-time object detection techniques.

Index Terms **-** Artificial Intelligence, Deep Learning, Computer Vision, Object Detection.

I. INTRODUCTION

Object recognition for small objects in images is a critical and indispensable task in the field of computer vision, finding applications in various domains, such as identifying pathological cells, crime prediction , plant classification , Epidemic prevention, human age recognition , and navigation assistance. Despite the advancements in object detection models, accurately detecting small and irregularly shaped objects remains challenging. This difficulty arises because most models primarily concentrate on medium or large objects, often ignoring the intricacies associated with smaller objects. Recent efforts have focused on creating network structures that are both efficient and accurate for real-time applications. One of the notable examples of these structures include MobileNet, ShuffleNet, ResNet, and DarkNet, specifically designed for performance on hardware platforms, such as CPUs and GPUs, as proposed by researchers. During training, convolutional neural network (CNN) models learn directly from the original pixel data. That allows them to discover data features and express complex contextual information effectively. Certain CNNs have exhibited substantial enhancements in accuracy and generalizability, successfully addressing various image analysis tasks, such as image categorization , image region segmentation, and image quality enhancement. Object detection algorithms can be broadly classified into two categories: Two-stage algorithms, including Fast ReCNN, Faster ReCNN, and Mask ReCNN , and one-stage algorithms, such as the well-known You Only Look Once (YOLO) series algorithms and single shot multi-box detector (SSD) algorithms, among others. The YOLO algorithms have undergone substantial development and are widely recognized as some of the most effective algorithms in the field. Notably, the YOLOv8 algorithm, introduced in 2023, has achieved exceptional accuracy, surpassing previous iterations. The YOLO algorithm is primarily designed to identify and categorize objects that occupy the entire image. However, its performance for detecting smaller-scale objects may be comparatively less than certain contemporary algorithms when configured to operate in a unique environment

with specific dimensions. CNN robustness becomes evident when evaluating its models' performance on visible images. The ability to capture profound input characteristics has led to intensive research in the challenging area of discovering frail and tiny objects in videos using CNNs.
Recent advancements in object detection algorithms have provided robust solutions for identifying medium to large objects in various contexts. However, detecting small and geometric objects still represents an essential challenge, especially for objects whose detection is critical for applications like micro-organism classification or precision agriculture. While the YOLOv8 algorithm introduced an innovative approach, it falls short in environments where object scale and clarity are compromised. The work presents YOLOv8-CAB, an evolution of YOLOv8, specifically engineered to enhance small object detection. Integrating the Context Attention Block (CAB) within the model's architecture addresses the intricacies associated with detecting fine-scale objects without compromising the real-time processing capabilities. This is a significant step forward from the conventional Coarse-to-Fine models. Empirical evaluations on the COCO dataset demonstrate a 2.1 % increase in mAP for objects under a certain size threshold compared to the baseline YOLOv8, outperforming contemporary models like the NanoDet model in speed and precision. These improvements are not just incremental; they enable the application of YOLOv8-CAB in scenarios where rapid and precise detection of small objects can be life-saving, such as in medical diagnostics or disaster response scenarios. These advantages substantially contribute to enhancing object detection precision in images. Subsequently, the presented work enhances the existing methodology and proposes an innovative detection framework, YOLOv8-CAB, explicitly designed for identifying diminutive and feeble entities within visual representations. The model demonstrates a high level of reliability and efficiency in the task of object detection within images. The network prioritizes shallow information and optimizes feature extraction by replacing the Coarse-to-Fine (C2F) module with CAB in its backbone.

II. THE PRIMARY CONTRIBUTION OF THIS STUDY CAN BE BRIEFLY OUTLINED AS FOLLOWS:

a) This study presents the YOLOv8-CAB approach for detecting small and geometric objects by examining their distinctive characteristics. The method builds upon the YOLOv8 framework and involves an analysis of the network structure, channel compression, parameter optimization, and other relevant factors. Substantial advancements have been implemented to enhance the design of this novel network, YOLOv8- CAB. Specifically, the feature extraction network has been meticulously engineered to fully exploit shallow characteristics while adding four layers to the detection head network to prioritize detecting small and fragile objects. The proposed models exhibit enhanced speed and accuracy compared with current image object detection algorithms.
b) The developed feature extraction network expands upon and iterates the shallow C2F module, replacing it with the Context Attention Block (CAB), explicitly designed to capture local and global context effectively and efficiently, allowing the network to detect small objects with better performance.
c) In the head, the C2F has been modified by increasing the thickness of the C2F module, which allocates more layers and filters in the convolution operations. This enhancement may improve the model's performance on certain tasks due to the added capacity for feature extraction.
d) Improving spatial attention by adding Selective Kernel (SK) attention to spatial attention, the proposed contribution enhances the spatial attention module by incorporating modifications inspired by the SK attention mechanism. These modifications include a split operation for multiscale spatial modelling , separate fuse and scale steps for flexible feature weighting, and integrating local and global context. These improvements enhance small object detection and selectively highlight important spatial regions and channels. The modified spatial attention module shows potential for surpassing the performance of the original module on various computer vision tasks.

III. KEY ISSUES AND FUTURE DIRECTIONS

a) **Dynamic Scene Understanding:** Current object detection algorithms often struggle to accurately localize and classify objects in dynamic scenes with multiple moving objects and complex interactions. There is a need for algorithms capable of modelling temporal dependencies and incorporating motion cues to improve object detection performance in live video streams.
b) **Robustness to Environmental Variations:** Variations in lighting conditions, weather, and scene clutter can significantly impact the performance of object detection algorithms. Research efforts should focus on developing algorithms that are robust to environmental variations and can adapt

dynamically to changing conditions without compromising detection accuracy.

c) **Real-time Processing Efficiency:** Achieving real-time performance in object detection requires algorithms that are computationally efficient and can process video streams at high frame rates. Investigating techniques for reducing computational complexity, optimizing algorithmic implementations, and leveraging hardware acceleration is essential to meet the stringent processing speed requirements of real-time applications.

d) **Scalability and Generalization:** Object detection algorithms should be scalable and capable of generalizing across different environments, object categories, and camera viewpoints. Research efforts should aim to develop algorithms that can adapt seamlessly to diverse scenarios and perform reliably across various real- world applications

IV. METHODOLOGIES:

Object detection is a fundamental task in computer vision with numerous applications spanning across various domains such as surveillance, autonomous vehicles, robotics, healthcare, and more. The utilization of advanced deep learning techniques has significantly enhanced the accuracy and efficiency of object detection systems. Among these techniques, YOLOv8 (You Only Look Once version 8) has emerged as a prominent algorithm due to its real-time processing capabilities and robust performance.

4.1.Significance of Object Detection using YOLOv8: Object detection using YOLOv8 holds significant importance across various applications for several reasons:

4.1.1. Real-Time Processing: YOLOv8's ability to perform object detection in real-time enables applications where timely detection and response are critical, such as surveillance, autonomous vehicles, and security systems.

4.1.2. Efficiency and Accuracy: YOLOv8 achieves a balance between speed and accuracy, making it suitable for applications requiring rapid and precise object detection, including robotics, industrial automation, and retail analytics.

4.1.3. Versatility: YOLOv8 can detect a wide range of object classes in diverse environments, making it adaptable to various use cases, including object tracking, counting, 19 recognition, and classification.

4.1.4. Scalability: YOLOv8's architecture is scalable and can be optimized for deployment on resource-constrained devices, opening up opportunities for edge computing and Internet of Things (IoT) applications.

4.1.5. Ease of Deployment: YOLOv8 models can be deployed across different platforms and integrated into existing software systems with relative ease, facilitating rapid prototyping and development of object detection solutions.

ARCHITECTURE :

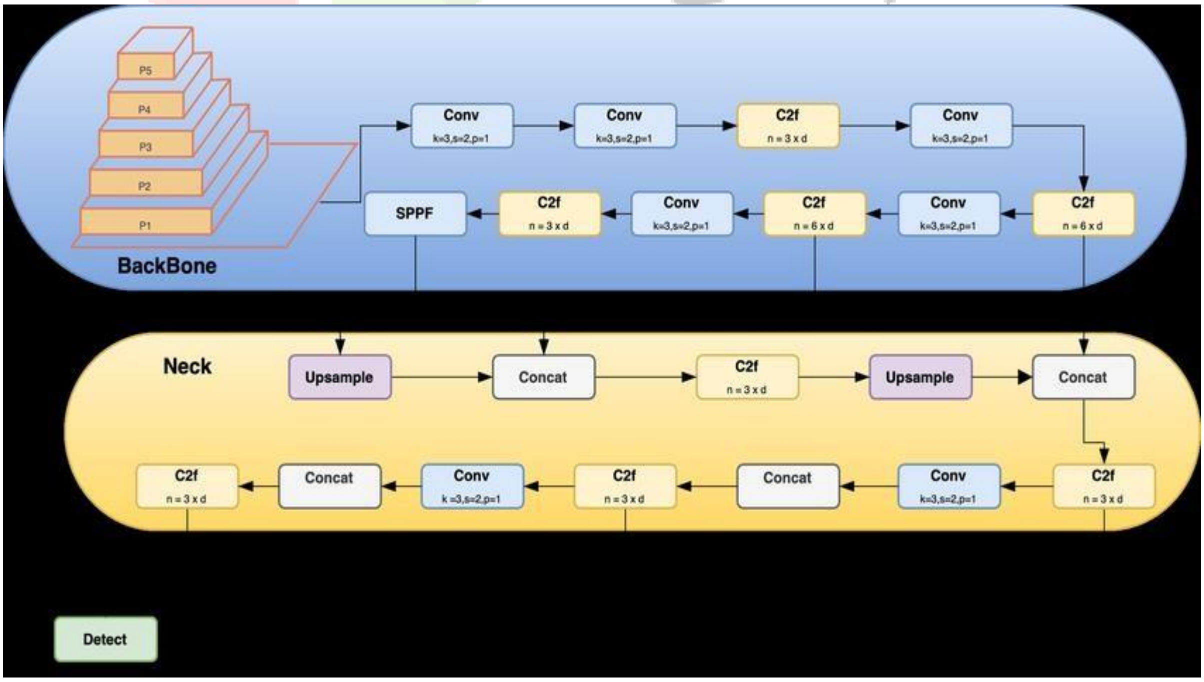

Fig.2. YOLOv8 architecture

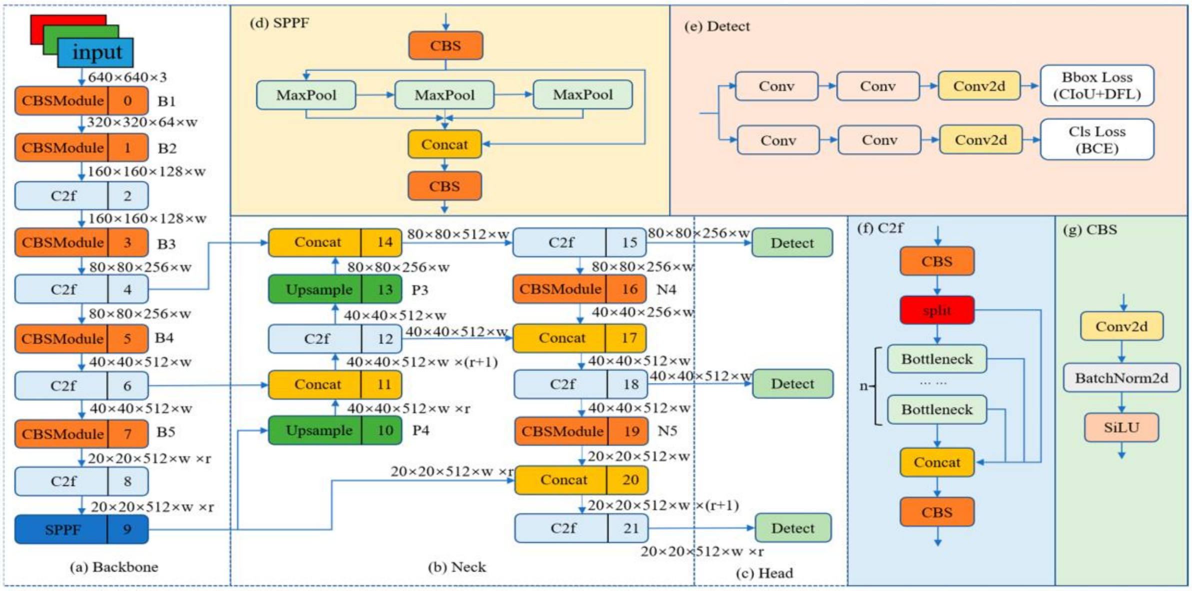

Fig.2. Mathematical YoloV8

The backbone part of YOLOv8 is basically the same as that of YOLOv5, and the C3 module is replaced by the C2f module based on the CSP idea. The C2f module learned from the ELAN idea in YOLOv7 and combined C3 and ELAN to form the C2f module , so that YOLOv8 could obtain more abundant gradient flow information while ensuring its light weight. At the end of backbone, the most popular SPPF module was still used, and three Maxpools of size 5 × 5 were passed serially, and then, each layer was concatenation, so as to guarantee the accuracy of objects in various scales while ensuring a light weight simultaneously.
In the neck part, the feature fusion method used by YOLOv8 is still PAN-FPN, which strengthens the fusion and utilization of feature layer information at different scales. The authors of YOLOv8 used two up sampling and multiple C2f modules together with the final decoupled head structure to compose the neck module. The idea of decoupling the head in YOLOx, was used by YOLOv8 for the last part of the neck. It combined confidence and regression boxes to achieve a new level of accuracy.
YOLOv8 can support all versions of YOLO and can switch between different versions at will. It can also run on various hardware platforms (CPU-GPU), giving it strong flexibility.

V. IMPLEMENTATION

Tools and Technologies:

- **YOLOv8:** A state-of-the-art object detection algorithm.
- **Python:** Programming language for implementation.
- **OpenCV:** Library for computer vision tasks.
- **GPU (optional):** Accelerates processing speed for real-time detection.

Implementation Steps:

Setup Environment:

- **Install necessary libraries:** OpenCV, NumPy, etc. Download pre-trained YOLOv8 weights and configuration files.
- **Load YOLOv8 Model:** Load YOLOv8 model architecture and pre-trained weights. Configure parameters such as confidence threshold and non-maximum suppression threshold.
- **Read Input Data:** Capture frames from a video stream or read images from a directory. Preprocess input data (resize, normalization, etc.).
- **Object Detection:** Apply YOLOv8 model to detect objects in input frames/images. Extract bounding boxes, class labels, and confidence scores for detected objects.
- **Post-processing:** Apply non-maximum suppression to remove redundant bounding boxes. Filter out detections below a certain confidence threshold.
- **Visualization:** Draw bounding boxes and labels on input frames/images to visualize detected objects. Display or save the annotated frames/images.
- **Performance Evaluation (Optional):** Measure the detection speed (frames per second) using a GPU if available. Evaluate the accuracy of object detection against ground truth annotations.

Simulation:
Run the implemented object detection system on sample images or videos. Analyze the results to ensure accurate detection of objects and assess the system's performance in real-time scenarios. Experiment with different input data, including images/videos with varying lighting conditions, backgrounds, and object scales, to evaluate the robustness of the system.

Performance Measurement:
Performance Measurement for object detection using YOLOv8 involves assessing speed (FPS) and accuracy metrics such as mean Average Precision (mAP) and Intersection over Union (IoU). These metrics evaluate the system's real-time detection capabilities and accuracy in detecting objects in images or videos.

VI. RESULTS

```
0: 384x640 2 persons, 1 bottle, 1 sandwich, 5 chairs, 2 dining tables, 62.2ms
Speed: 3.4ms preprocess, 62.2ms inference, 1.9ms postprocess per image at shape (1, 3, 384, 640)
```

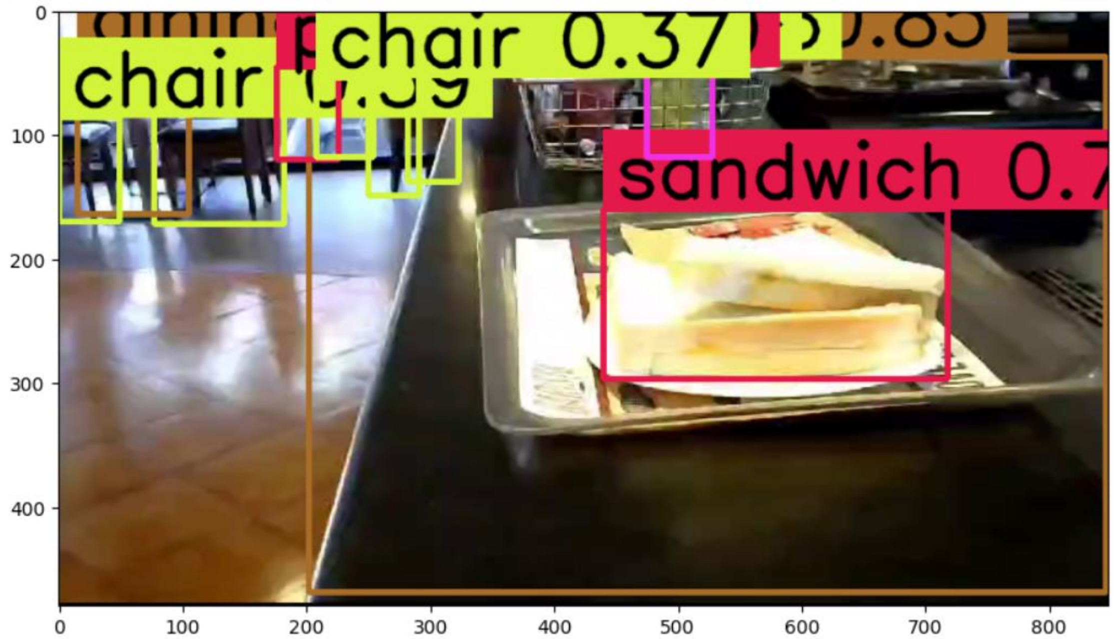

Fig.4. Object detection using Yolov8

```
100% 1015/1015 [01:11<00:00, 17.53it/s]

0: 384x640 2 persons, 1 bottle, 1 sandwich, 5 chairs, 2 dining tables, 62.5ms
Speed: 5.0ms preprocess, 62.5ms inference, 2.6ms postprocess per image at shape (1, 3, 384, 640)
None --- <class 'NoneType'>

0: 384x640 2 persons, 1 bottle, 1 sandwich, 5 chairs, 2 dining tables, 63.7ms
Speed: 4.2ms preprocess, 63.7ms inference, 9.5ms postprocess per image at shape (1, 3, 384, 640)
None --- <class 'NoneType'>

0: 384x640 2 persons, 1 bottle, 1 sandwich, 6 chairs, 2 dining tables, 55.3ms
Speed: 3.8ms preprocess, 55.3ms inference, 2.8ms postprocess per image at shape (1, 3, 384, 640)
None --- <class 'NoneType'>

0: 384x640 2 persons, 1 bottle, 1 sandwich, 5 chairs, 2 dining tables, 45.9ms
Speed: 6.4ms preprocess, 45.9ms inference, 2.6ms postprocess per image at shape (1, 3, 384, 640)
None --- <class 'NoneType'>
```

Fig.5. Each object count from frame

VII. CONCLUSION

In conclusion, this review paper has provided a thorough examination of object detection using YOLOv8, highlighting its pivotal role in advancing computer vision applications. YOLOv8 stands out for its real-time processing capabilities, achieving impressive accuracy and efficiency in detecting objects across diverse environments. By synthesizing existing literature, we have elucidated the strengths and limitations of YOLOv8, showcasing its superiority compared to previous iterations and alternative methods.
Moreover, our analysis has identified key research gaps, including the need for improved interpretability, robustness in challenging scenarios, and scalability for large-scale deployments. Addressing these gaps will be crucial for further enhancing the applicability and reliability of YOLOv8 in real-world settings.
Overall, YOLOv8 represents a milestone in object detection technology, paving the way for innovative applications in surveillance, autonomous driving, healthcare, and beyond. As research and development in this field continue to evolve, YOLOv8 remains at the forefront, promising continued advancements and transformative impacts in the future.

REFERENCES:

1. A Convolutional Neural Network based Live Object Recognition System as Blind Aid (2018) Kedar Potdar ,Chinmay D. Pai, Sukrut Akolkar.
2. Object detection and tracking using OpenCV (2022) B Sangamitra, MS Teja ,PK Kongoti, J Dileep, K Bheema.
3. Object Detection System Based on Convolution Neural Networks Using Single Shot Multi-Box Detector (2020) A Kumar, S Srivastava.
4. Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network (2021) H Aung, AV Bobkov, NL Tun.
5. Real Time Live Fish Object Detection and Tracking In Under Water Stereo Videos(2014)M Girija, M Rajasekar, S Nithiya.
6. Car Vehicle Image Object Detection Using You Only Live Once (YOLO)(2023)N Azizah, Y sharia, S Sahwari, M Iskandar.
7. Object detection using OpenCV with python (2023)P Nikitha, T Nikitha, M Nikshitha, T Nikhil, T Nithin, KN Reddy, SJ Hamilpure.
8. An novel approach object detection of video surveillance system using OpenCV(2024)V Kakulapati, S Potla, M Sista, AK Chelukali, R Dhabale, SR Katta.
9. Real-time object detection and robotic manipulation for agriculture using a YOLO-based learning approach(2024)H Zhao, Z Tang, Z Li, Y Dong, Y Si, M Lu, G Panoutsos.
10. Real Time Object Detection and Tracking Using Deep Learning and OpenCV(2018)G Chandan, A Jain, H Jain.
11. DC-YOLOv8:Small-Size Object Detection Algorithm Based on Camera Sensor(2023)Haitong Lou ,Xuehu Duan, Junmei Guo, Haiying Liu,Jason Gu, Lingyun Bi and Haonan Chen.
12. YOLOv8-CAB: Impr Ov8-CAB: Improved YOLOv8 for Real-time object detection(2024) Talib, Moahaimen;Al-Noori, Ahmed H. Y.; and Suad, Jameelah.
13. Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8(2023)Debasis Kumar, Naveed Muhammad.
14. Efficient Small-Object Detection in Underwater Images Using the Enhanced YOLOv8 Network(2024) Minghua Zhang, Zhihua Wang, Wei Song, Danfeng Zhao, Huijuan Zhao.
15. Real-Time Flying Object Detection with YOLOv8(2023)Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi.