International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

A Review paper on object detector based on changeable size light weight convolution and context argumentation module or images captured by UAVs

Priya S. Dhongade[1], Prof. Sunil Kuntawar[2]

Department of Electronics and Telecommunication Engineering,
Ballarpur Institute of Technology Ballarpur, Maharashtra, India

Abstract:

In recent years, object detection has become a crucial component in various applications such as surveillance, autonomous vehicles, medical imaging, and smart cities. This project focuses on object detection using state-of-the-art algorithms YOLOv8 and YOLOv9. Leveraging the power of deep learning, YOLO (You Only Look Once) models are renowned for their speed and accuracy in identifying multiple objects within images and video streams. The proposed system allows users to upload images or videos and utilizes the webcam for live object detection. The frontend is built using Streamlit, ensuring an intuitive and interactive user interface, while the backend is implemented in Python and executed on Google Colab to leverage GPU acceleration. The system processes the input through pre-trained YOLO models, identifies objects, and displays them with bounding boxes and confidence scores. This project demonstrates how advanced object detection algorithms can be integrated into user-friendly applications, promoting accessibility and real-world deployment. Additionally, it serves as a foundation for further enhancements such as object tracking, anomaly detection, and integration with IoT systems for smarter decision-making in dynamic environments.

Keywords: YOLOv8, YOLOv9, Object Detection, Streamlit, Analysis

1. Introduction

Object detection, a subset of computer vision, plays a pivotal role in recognizing and localizing multiple objects within images or video streams. As the demand for intelligent systems grows across various sectors—ranging from surveillance and traffic monitoring to healthcare and smart manufacturing—the ability to detect and interpret visual data in has become indispensable. Deep learning has revolutionized object detection by introducing advanced neural network architectures capable of learning complex patterns and features from large datasets. Among these, YOLO (You Only Look Once) has emerged as one of the most efficient and widely adopted object detection frameworks. The latest versions, YOLOv8 and YOLOv9, offer substantial improvements in speed, accuracy, and computational efficiency over their predecessors.

This project is designed to implement a object detection system using YOLOv8 and YOLOv9. The solution supports three modes of operation: image upload, video file analysis, and live webcam streaming. The frontend is built using Streamlit, which simplifies deployment and enhances user interactivity. The backend utilizes Python and runs on Google Colab to benefit from GPU acceleration. This combination allows the system to process and analyze frames quickly, drawing bounding boxes around detected objects and displaying confidence scores. Overall, the project highlights the potential of deep learning in building scalable, responsive, and accessible object detection solutions for real-world applications.

2. Objective:

The primary objective of this project is to design and implement a object detection system that is accurate, fast, and user-friendly. By leveraging the latest advancements in deep learning, specifically the YOLOv8 and YOLOv9 models, the system aims to deliver high performance in diverse scenarios such as video surveillance, smart traffic management, and public safety applications.

Key objectives include:

- To build an intuitive frontend interface using Streamlit that supports image, video, and webcam input.
- To integrate pre-trained YOLOv8 and YOLOv9 models capable of detecting multiple object classes with high accuracy and low latency.
- To utilize Google Colab for backend processing, ensuring GPU-accelerated inference for faster results.
- To display results dynamically, including bounding boxes and confidence scores, for each detected object.
- To maintain modularity and scalability so that the system can be extended in the future for tasks like object tracking, face recognition, or suspicious activity detection.

Ultimately, the system is intended to bridge the gap between complex AI models and end-users by providing a lightweight, accessible platform that can be utilized in both academic research and real-world deployments.

3. Literature Survey

1. CNN-Based Live Object Recognition for Blind Aid (2018):

This system employs a Convolutional Neural Network (CNN) for real-time object recognition to assist visually impaired individuals. CNNs automatically extract hierarchical features from input images and classify objects with high accuracy. The live feed from a camera is processed by the CNN model, which identifies everyday objects and generates voice feedback to help blind users navigate their surroundings.

The model is trained on a diverse dataset to generalize well in dynamic environments. This real-time application emphasizes low latency, accuracy, and a lightweight architecture to suit embedded devices used by the visually impaired.

2. Object Detection and Tracking Using OpenCV (2022):

This system uses traditional computer vision techniques available in OpenCV for object detection and tracking in video streams. Algorithms like Haar cascades, background subtraction, and contour detection are combined with tracking methods such as Kalman filters or MeanShift to identify and follow objects. While not deep learning-based, this approach is computationally efficient and suitable for real-time applications where lightweight processing is essential. It serves well in surveillance and motion analysis but may lack the robustness of deep learning models under varying lighting, occlusion, or object deformation conditions.

3. SSD-Based Object Detection Using CNNs (2020):

The Single Shot MultiBox Detector (SSD) is a CNN-based model designed for fast and accurate object detection. Unlike two-stage detectors (e.g., R-CNN), SSD performs object localization and classification in a single pass through the network, making it suitable for real-time use. The model uses multiple feature maps at different scales to detect objects of various sizes. This system benefits from speed and efficiency without compromising much on accuracy, making it effective for tasks such as surveillance, autonomous vehicles, and smart devices. It also reduces computational overhead compared to other deep learning detection models.

4. YOLO with VGG16 for Real-Time Face Detection (2021):

This model integrates the YOLO (You Only Look Once) detection framework with the VGG16 architecture to perform real-time face detection. YOLO divides the image into a grid and simultaneously predicts bounding boxes and class probabilities, offering high-speed detection. VGG16 contributes robust feature extraction due to its deep architecture with 16 layers. Together, they enhance accuracy while maintaining speed, making this combination suitable for surveillance and biometric applications. The system can detect multiple faces in live video with reduced false positives and is optimized for performance in cluttered or low-light environments.

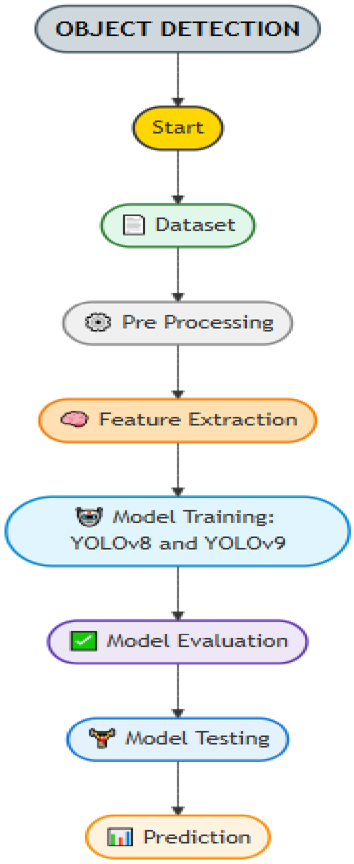
5. Real-Time Fish Detection in Underwater Videos (2014):

This system focuses on detecting and tracking fish in stereo underwater videos using classical image processing and motion detection techniques. Techniques such as optical flow, background subtraction, and contour-based segmentation are used to locate fish amidst noisy and dynamic aquatic environments. Although deep learning was not commonly used at the time, the system managed to deliver effective tracking by leveraging stereo vision for depth estimation and robust movement tracking algorithms. The application is valuable in marine biology, fishery monitoring, and aquatic research for behavioral analysis of underwater species.

6. Vehicle Detection Using YOLO (2023):

This application utilizes the YOLO (You Only Look Once) deep learning algorithm for detecting cars and vehicles in real-time images. YOLO processes the entire image with a single neural network and predicts bounding boxes and class probabilities simultaneously. It is highly efficient and accurate, making it ideal for real-time traffic monitoring, autonomous driving, and smart city infrastructure. The YOLO model is trained on annotated vehicle datasets, allowing it to detect different vehicle types under various lighting and weather conditions. The system prioritizes speed, making it suitable for edge devices and real-time deployment in urban areas.

4. WORK FLOW:



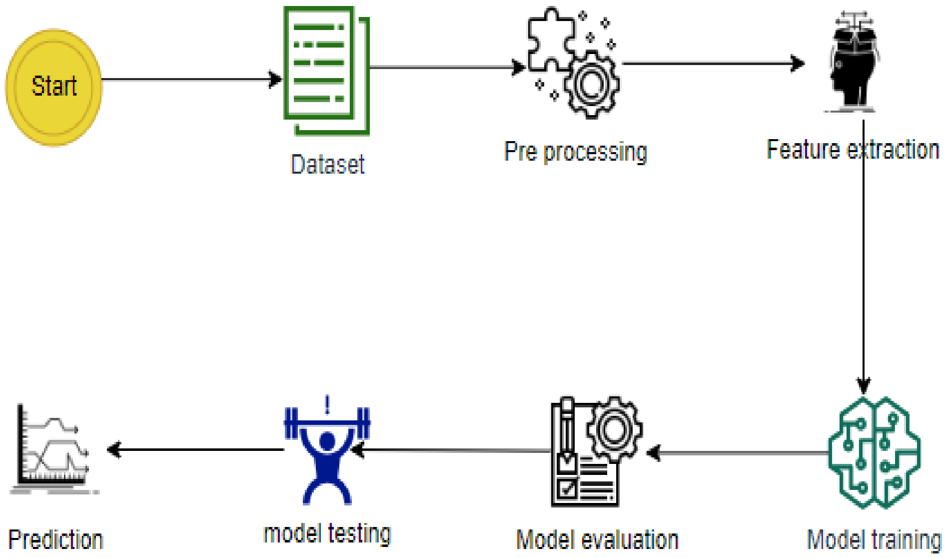
H/W Configuration:

- Processor - I3/Intel Processor
- Hard Disk -160 GB
- RAM - 8 GB

S/W Configuration:

- Operating System : Windows 7/8/10 .
- Server side Script : HTML, CSS & JS.
- IDE : Vscode
- Libraries Used : Numpy, Pandas,Sklearn,Tensorflow
- Technology : Python 3.6+.

Architecture



5. Conclusion

The project effectively demonstrates the capabilities of modern object detection algorithms, particularly YOLOv8 and YOLOv9, in delivering real-time performance with high accuracy. By integrating these models into a Streamlit-based interface, the system achieves both technical robustness and user-friendliness, making complex deep learning solutions more accessible to end-users. The use of Google Colab for backend execution ensures cost-effective GPU support, enabling efficient processing of image and video data for live detection tasks. Whether the input is a static image, pre-recorded video, or live webcam feed, the system consistently identifies and labels multiple objects with bounding boxes and confidence scores. This implementation bridges the gap between theoretical advancements in computer vision and practical deployment in real-world scenarios. The project showcases how cutting-edge models like YOLOv9 can be harnessed in flexible, scalable applications with minimal infrastructure, making them suitable for domains such as surveillance, healthcare, smart cities, and autonomous systems. Overall, the system highlights the evolution and effectiveness of single-shot detectors in object recognition tasks and sets a strong foundation for more complex, intelligent systems capable of contextual decision-making.

6. Future Enhancement

While the current system offers robust object detection capabilities, several future enhancements can significantly expand its functionality and application scope. One promising direction is the integration of **object tracking**, which would allow the system to maintain consistent object identities across frames, essential for video analytics and surveillance. Incorporating **anomaly detection** could elevate the system's utility in critical sectors such as security and healthcare by flagging irregular patterns or behaviors. Another enhancement involves **training on custom datasets**, enabling industry-specific object recognition, such as detecting defective items in manufacturing or specific tools in construction. Integrating the solution with **IoT ecosystems** can lead to smarter automation, like alert systems or adaptive controls based on detected objects. Edge computing support could be explored for **offline, low-latency deployments** in remote or bandwidth-constrained environments. Additionally, incorporating **voice or gesture controls** can make the system more interactive and inclusive. As AI hardware becomes more accessible, the system can evolve into a fully embedded solution, such as a mobile app or standalone camera device. These advancements would not only broaden its usability but also contribute to the development of smarter, real-time, and context-aware environments across industries.

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. 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. www.ijcrt.org © 2024 IJCRT | Volume 12, Issue 12 December 2024 | ISSN: 2320-2882 IJCRT2412497 International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org e583

10. Real Time Object Detection and Tracking Using Deep Learning and OpenCV(2018)G Chandan, A Jain, H Jain.
11. https://www.researchgate.net/figure/YOLOv8-architecture_fig5_371888779
12. 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.
13. YOLOv8-CAB: Impr Ov8-CAB: Improved YOLOv8 for object detection(2024) Talib, Moahaimen;Al-Noori, Ahmed H. Y.; and Suad, Jameelah.
14. Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8(2023)Debasis Kumar, Naveed Muhammad.
15. Efficient Small-Object Detection in Underwater Images Using the Enhanced YOLOv8 Network(2024) Minghua Zhang, Zhihua Wang, Wei Song, Danfeng Zhao, Huijuan Zhao.
16. Flying Object Detection with YOLOv8(2023)Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi.