Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

YOLOv11: A Next-Generation Approach to Real-Time Object Detection

Mrs. Ch. Naga Lakshmi[1], Puvvula Vamsi Pravallika[2], Ponnuri Sandhya[3], Paritala Venkata Lavanya[4], Padamata Prasanth[5]
Assistant Professor[1], UG Scholar[2,3,4,5]
Department of Computer Science and Engineering[1, 2,3,4,5],
Seshadri Rao Gudlavalleru Engineering College (Affiliated to JNTUK) [1, 2,3,4,5],
nagalakshmi@gmail.com[1], vamsipravallikapuvvula2108@gmail.com[2], sandhyaponnuri07@gmail.com[3],
lavanyaparitala2004@gmail.com[4], padamataprasanth6@gmail.com[5]

Abstract: Vision impairment is one of the top ten disabilities that can be affected by humans, with India having the highest visually impaired population. This study introduces a novel framework utilizing YOLOv11 for real-time object detection and recognition for the objects and to assist the visually impaired individuals in navigating their surroundings independently. The YOLOv11 detector, trained on the COCO dataset with an additional class for enhanced detection, enables precise classification of various objects. The system integrates Python and OpenCV to detect objects from videos and webcams, ultimately providing audio feedback to the user. The proposed system ensures fast and accurate detection, improving accessibility and reducing dependency on external assistance.

***Index terms** -YOLOv11, Object Detection, Deep Learning, Computer Vision, Assistive Technology, Visual Impairment, Real-Time Detection, OpenCV, COCO Dataset, AI for Accessibility*

1. INTRODUCTION

Vision impairment and blindness are among the most prevalent disabilities affecting millions worldwide, significantly impacting individuals' independence and mobility. India has the highest visually impaired population, making it crucial to develop assistive technologies that enhance their ability to perceive and navigate their surroundings safely. Traditional mobility aids, such as canes and guide dogs, provide limited assistance, often requiring external support. With advancements in artificial intelligence and computer vision, deep learning-based object detection systems offer a promising solution for improving accessibility.

This study focuses on developing a real-time object detection and recognition framework using YOLOv11 to assist visually impaired individuals. YOLO (You Only Look Once) is a well-known deep learning-based object detection technique known for its speed and accuracy. In this work, YOLOv11 is trained on the COCO dataset, with additional classes incorporated to enhance detection capabilities. The system processes video frames in real-time using OpenCV and Python, identifying objects and providing audio feedback to users, thereby enabling them to recognize and avoid obstacles efficiently.

Before the emergence of deep learning techniques, traditional computer vision approaches struggled to achieve high accuracy in object detection tasks. Conventional methods such as edge detection and feature matching had significant limitations in varying lighting conditions and complex environments. The advent of deep learning, particularly convolutional neural networks (CNNs), has led to significant improvements in image classification and object recognition. Modern detectors like YOLO and SSD (Single Shot Detector) have revolutionized real-time object detection, making them ideal for assistive technologies.

The proposed system integrates a laptop camera to continuously scan the surroundings, detect objects, and inform the user about their presence. This technology eliminates the need for close physical interaction, reduces reliance on external assistance, and enhances mobility for visually impaired individuals. By leveraging YOLOv11, this framework ensures high-speed detection with accurate classification, making it a valuable tool for improving accessibility and independence.

2. LITERATURE SURVEY

2.1 MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

https://arxiv.org/abs/1704.04861

ABSTRACT: This paper employs on yolov11 mechanism for object detection and classification, followed by recognition of objects .yolov11 detector is trained on coco dataset, in which a new class is added, to enable the detection of different We provide a pair of straightforward global hyper-parameters that strike a good compromise between accuracy and latency. Based on the constraints of the problem at hand, model builders can select the ideal model size using these hyper-parameters. We surpass other well-known ImageNet classification methods after doing thorough investigations of the accuracy vs. resource tradeoff. Following this, we demonstrate that MobileNets are effective at identifying items, classifying fine-grained features, identifying faces, and performing large-scale geo-localization.

2.2 Real Time Object Detection and Tracking Using Deep Learning and OpenCV:

http://researchgate.net/publication/331421347_Real_Time_Object_Detection_and_Tracking_Using_Deep_Learning_and_OpenCV

ABSTRACT: In the past several years, deep learning has had a significant influence on how society adjusts to AI. Some well-known methods for object recognition are RCNN, Faster-RCNN, SSD, and YOLO. If speed is your first priority, YOLO will outperform Faster-RCNN and SSD in terms of accuracy. For effective tracking and identification, deep learning employs SSD and Mobile Nets. Without compromising performance, this approach efficiently detects objects.

2.3 Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles:

https://doi.org/10.36893/JNAO.2025.V16I01.088

https://www.sciencedirect.com/science/article/pii/S1877050915030136

ABSTRACT: Unmanned aerial vehicles (UAVs) are used extensively in this study to identify and monitor moving items that represent a major security risk to the southern border of the United States. Immigrant border encroachment and illegal trespassing pose a serious threat to the Department of Homeland Security and the US border security force. The vast volume of data involved makes it impossible to guarantee suspicious conduct, which is monitored for extended periods of time by human operators. The primary goal of this study is to support human operators by putting in place intelligent visual surveillance technologies that aid in identifying and monitoring odd or suspicious activity in the video sequence. The visual surveillance system needs reliable and quick ways to monitor and identify moving objects. In this study, we looked at ways to monitor and identify things from unmanned aerial vehicles. Adaptive background subtraction was successfully used to detect moving objects, while Lucas-Kanade optical flow tracking and Continuously Adaptive Mean-Shift tracking were used to track these items. The simulation results demonstrate how well these methods work to identify and follow moving objects in the UAV-acquired video sequences.

2.4 Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixed camera:

https://ieeexplore.ieee.org/abstract/document/8053292

ABSTRACT: For security applications, video surveillance is a crucial and vital responsibility. In the past, surveillance involved taking a video with a camera, saving the data in a database, and then manually retrieving the necessary information from the database. Sensitive data might be lost instantly as a result. Automated video monitoring is crucial in these situations. Automated video surveillance allows for real-time object tracking and detection, discovers the necessary data, and instantly notifies the administrator. This document explains how to count the number of items and detect them in real time. It also explains how things are categorised; using the characteristics extraction and comparison methods, they are divided into five predetermined classes: people, vehicles, motorcycles, buses, and horses.

2.5 Object Tracking Algorithms for Video Surveillance Applications

https://ieeexplore.ieee.org/document/8524260

ABSTRACT: Because objects move in different ways, object tracking is one of the most important study topics. In particular, feature selection is essential to object tracking. Object identification, traffic management, gesture recognition, human-computer interaction, and video surveillance are just a few of its many uses. To get over tracking problems relating to object movement and appearance, a lot of methods concentrate on the tracking algorithm to smoothen the video sequence. These techniques employ multidirectional tracking for video surveillance applications by using object form, colour, texture, object of interest, and motion. This study outlines a thorough analysis of several object tracking algorithms in a range of environmental settings and identifies the most effective algorithms for each type of tracking. The motion of both single and many objects (vehicles) is recognised and tallied in several frames in this work, which tracks objects based on colour. Additionally, a single algorithm for object tracking might be created by taking into account the item's form, colour, texture, object of interest, and multidirectional motion.

3. METHODOLOGY

i) Proposed Work:

The proposed system aims to provide real-time object detection and recognition to assist visually impaired individuals in navigating their surroundings independently. It utilizes YOLOv11, a deep learning-based object detection model, trained on the COCO dataset with an additional class to improve detection accuracy. The system is designed to recognize various objects, including indoor items like chairs and mobile phones, as well as outdoor elements such as vehicles and pedestrians. Implemented using Python and OpenCV, the system processes video frames captured through a laptop camera, identifies objects, and delivers audio feedback to the user, enabling them to recognize and avoid obstacles effectively.

One of the key advantages of this system is its ability to function in real-time, offering quick and accurate object detection regardless of environmental conditions. Unlike traditional mobility aids, this system ensures continuous object tracking and provides immediate alerts without requiring close interaction. It enhances user independence by reducing reliance on external assistance while improving safety and mobility. Additionally, the system can be used anytime and anywhere, making it a practical and accessible solution for visually impaired individuals seeking greater autonomy in their daily lives.

ii) System Architecture:

The YOLOv11 architecture is structured into three key components: Backbone, Neck, and Head, each playing a crucial role in enhancing object detection efficiency. The Backbone is responsible for feature extraction, utilizing multiple convolutional layers (Conv) and C3K2 modules to refine spatial and contextual information from input images. This deep feature extraction process ensures that important object details are retained while reducing unnecessary noise. The extracted features are then passed to the Neck, which further processes and fuses the information to improve detection accuracy.

The Neck employs operations such as Concat, Upsample, SPFF, and C2PSA to enhance multi-scale feature representation, making the model capable of detecting objects of different sizes effectively. The processed features are then sent to the Head, where multiple detection layers generate bounding boxes and class predictions. This enables real-time object recognition with high precision. Implemented using Python and OpenCV, this architecture provides quick and reliable detection, making it particularly useful for assisting visually impaired individuals by converting detected object information into audio feedback for safer and more independent navigation.

https://doi.org/10.36893/JNAO.2025.V16I01.088

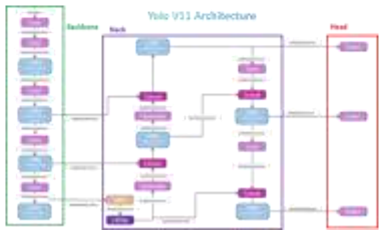

Fig 1 Proposed architecture

iii) Modules:

a) Browse System Videos

This module allows users to upload a pre-recorded video from their system for object detection. Once the video is loaded, the application plays it while analyzing each frame to detect objects in real time. Any identified objects are highlighted with bounding boxes for easy recognition. If the user wants to stop the tracking process at any point, they can simply press the 'q' key, which will immediately stop video playback and detection. This module provides a flexible way to analyze objects in existing videos without requiring a live camera feed.

b) Start Webcam Video Tracking

This module enables real-time object detection using a system's built-in webcam. When activated, the application starts live video streaming and continuously scans the surroundings for objects. If an object is detected, the system marks it with a bounding box, helping users identify objects dynamically. This feature is particularly useful for visually impaired individuals, as it allows them to recognize objects in their environment in real time. The tracking process can be stopped anytime by pressing the 'q' key, making the system user-friendly and efficient for practical applications.

iv) Algorithms:

a. Video-Based Object Detection Algorithm

The video-based object detection algorithm is designed to analyze pre-recorded videos uploaded by the user. When the application starts, the user selects a video file, which is then loaded into the system. The YOLOv11 model processes each frame of the video, identifying objects and marking them with bounding boxes. These detected objects are displayed on the screen in real time. If the user wants to stop the detection process at any point, they can press the 'q' key, which will immediately halt the video playback. The algorithm continues processing frames until the video ends or the user manually stops it, ensuring a smooth and efficient detection process.

b. Real-Time Webcam Object Detection Algorithm

For real-time webcam object detection, the system connects to the built-in webcam and continuously captures live video frames. The YOLOv11 model analyzes each frame to detect and classify objects, highlighting them with bounding boxes. This enables real-time recognition of objects in the user's surroundings, making it particularly beneficial for visually impaired individuals. The processed video stream is displayed on the screen, providing instant feedback. Similar to the video-based detection, pressing the 'q' key stops the webcam stream, allowing the user to exit detection mode easily. This approach ensures real-time object identification with high accuracy and efficiency.

4. EXPERIMENTAL RESULTS

https://doi.org/10.36893/JNAO.2025.V16I01.088

Precision: Precision calculates the fraction of True Positives to the Sum of True Positives and False Positives. The formula is as follows

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: Recall is a metric in that is present in machine learning that measures fraction of TP to the Sum of TP and FN. The ability of a model is used to identify all patinent instances of a class can be done by machine learning recall. Looking at the ratio of expected positives to actual positives, it reveals how accurately the model is used to classify the data. The Formula for the recall is as follows:-

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

mAP: Mean Average Precision (MAP) is a one of the ranking quality metric in machine learning. It can be calculated as the average of Arithmetic mean of all the classes of k to the number of the classes n. In the formula Apk stands of Arithmetic Mean of classes k and n stands for number of classes n. The formula is asfollows:-

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

Fig 2 object tracking

Fig 3object detect

5. CONCLUSION

The proposed system utilizing YOLOv11 for object detection provides an efficient and accurate solution for recognizing objects in both pre-recorded videos and real-time webcam streams. By integrating deep learning techniques, the system ensures quick and reliable detection, making it particularly useful for visually impaired individuals. The ability to process frames dynamically and present results in real time enhances navigation and awareness of surroundings. Overall, this framework significantly improves object recognition capabilities and can be effectively deployed in various applications.

6. FUTURE SCOPE

https://doi.org/10.36893/JNAO.2025.V16I01.088

Future enhancements can focus on integrating audio feedback systems to provide real-time verbal descriptions of detected objects, further assisting visually impaired users. Additionally, improving the system's adaptability by training the model on a wider range of datasets can enhance detection accuracy across different environments. Implementing edge computing for real-time processing on portable devices can also improve efficiency and accessibility.

REFERENCES

[1] Wei Liu and Alexander C. Berg, "SSD: Single Shot MultiBox Detector",Google Inc., Dec 2016.

[2] Andrew G. Howard, and Hartwig Adam, "MobileNets: EfficientConvolutional Neural Networks for Mobile Vision Applications", GoogleInc., 17 Apr 2017.

[3] Justin Lai, Sydney Maples, "Ammunition Detection: Developing a Real-Time Gun Detection Classifier", Stanford University, Feb 2017

[4] ShreyamshKamate, "UAV: Application of Object Detection andTracking Techniques for Unmanned Aerial Vehicles", Texas A&MUniversity, 2015.

[5] Adrian Rosebrock, "Object detection with deep learning and OpenCV",pyimagesearch.

[6] Mohana and H. V. R. Aradhya, "Elegant and efficient algorithms for realtime object detection, counting and classification for video surveillanceapplications from single fixed camera," 2016 International Conferenceon Circuits, Controls, Communications and Computing (I4C),Bangalore, 2016, pp. 1-7.

[7] AkshayMangawati, Mohana, Mohammed Leesan, H. V. RavishAradhya, "Object Tracking Algorithms for video surveillanceapplications" International conference on communication and signalprocessing (ICCSP), India, 2018, pp. 0676-0680.

[8] ApoorvaRaghunandan, Mohana, PakalaRaghav and H. V. RavishAradhya, "Object Detection Algorithms for video surveillanceapplications" International conference on communication and signalprocessing (ICCSP), India, 2018, pp. 0570-0575.

[9] ManjunathJogin, Mohana, "Feature extraction using ConvolutionNeural Networks (CNN) and Deep Learning" 2018 IEEE InternationalConference On Recent Trends In Electronics InformationCommunication Technology,(RTEICT) 2018, India.

[10] ArkaPrava Jana, AbhirajBiswas, Mohana, "YOLO based Detection and Classification of Objects in video records" 2018 IEEE InternationalConferenceOn Recent Trends In Electronics InformationCommunication Technology

https://doi.org/10.36893/JNAO.2025.V16I01.088