International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Object Detection Using YOLO

Mr. Shreyash K. Patil[1], Mr. Atharva N. Kharade[2],
Mr. Abhishek M. Kesarkar[3], Mr. Udayraj D. Bankarmali[4],
Mr. Prof. N.K. Magdum[5]

[1, 2, 3, 4]Sanjeevan Group of Institutions, Panhala, Maharashtra, India
[5]Department of Computer Science and Engineering

Abstract

Object detection is a fundamental task in computer vision with wide-ranging applications such as surveillance, autonomous driving, and robotics. This project explores the implementation of object detection using the YOLO (You Only Look Once) algorithm, a real-time, deep learning-based approach known for its speed and accuracy. Unlike traditional methods that apply a classifier to various regions of an image, YOLO frames object detection as a single regression problem, directly predicting bounding boxes and class probabilities from full images in one evaluation. This makes YOLO highly efficient for real-time systems. In this work, we train and evaluate YOLO on a standard dataset, analyze its performance, and compare it with other state-of-the-art object detection models. The results demonstrate that YOLO provides a strong balance between accuracy and computational efficiency, making it a powerful tool for real-time object detection tasks.

Object detection using YOLO (You Only Look Once) is a real-time deep learning technique that identifies and classifies multiple objects within an image in a single pass. It offers high speed and accuracy by framing detection as a regression problem, making it ideal for applications like surveillance and autonomous driving.

INTRODUCTION

Object detection is a crucial area in the field of computer vision that combines techniques for both classification and localization of objects within an image or video. It aims not only to identify what objects are present but also to determine their positions using bounding boxes. This technology has widespread applications in areas such as autonomous driving, video surveillance, medical imaging, facial recognition, and industrial automation.

Traditional object detection methods, such as R-CNN (Region-based Convolutional Neural Network), Fast R-CNN, and Faster R-CNN, have shown promising results but are often computationally expensive and not suitable for real-time processing. These approaches typically involve multiple stages, such as generating region proposals, feature extraction, and classification, which increase both complexity and inference time.

To address these limitations, the YOLO (You Only Look Once) algorithm was introduced as a unified and real-time object detection system. YOLO frames object detection as a single regression problem, directly predicting bounding box coordinates and class probabilities from an input image in one evaluation. This makes YOLO significantly faster compared to previous approaches, enabling it to process images at high frame rates while maintaining competitive accuracy.

The strength of YOLO lies in its ability to generalize well to new domains and its global understanding of the image, which reduces false positives in background regions. Over the years, several versions of YOLO have been developed, each improving in terms of speed, accuracy, and efficiency, including YOLOv2, YOLOv3, YOLOv4, YOLOv5, and the most recent YOLOv8 and YOLO-NAS.

This project/paper focuses on implementing object detection using the YOLO algorithm, evaluating its performance on benchmark datasets, and exploring its real-time capabilities. Through this work, we aim to highlight the advantages of YOLO in practical scenarios and demonstrate its effectiveness as a state-of-the-art solution for fast and accurate object detection.

PROBLEM STATEMENT

Many real-world applications, accurately identifying and localizing multiple objects in images or videos is critical. Traditional object detection methods are often slow, complex, and not suitable for real-time tasks due to their multi-stage processing pipelines

1. **Real-time Detection Challenge**: In many real-world applications such as autonomous driving, smart surveillance, and robotics, the ability to detect and identify objects in real-time is essential. A delay of even a fraction of a second can lead to incorrect decisions.For example, in autonomous vehicles, object detection must be performed on live video streams to avoid collisions. Similarly, in real-time surveillance, quick detection of threats is necessary for prompt action.
2. **Complex and Multi-stage Pipelines**: Traditional object detection frameworks like R-CNN, Fast R-CNN, and Faster R-CNN rely on multi-stage processing pipelines. These methods first generate region proposals, extract features, classify regions.
3. **Need for a Fast and Accurate Solution:**To overcome the limitations of traditional methods, there is a pressing need for a model that can detect multiple objects accurately and in real time. YOLO (You Only Look Once) addresses this need by treating object detection as a single regression problem.
4. **Overlapping Objects**: Object detection models, including YOLO, often struggle with accurately detecting small or closely packed objects in an image. In real-world scenarios like traffic scenes or crowded areas, multiple objects may overlap or appear very small.
5. **Dataset and Environment Variability:**YOLO’s performance can vary significantly depending on the dataset and environment in which it is used. Lighting conditions, background complexity, and object diversity all impact detection accuracy.

OBJECTIVES

- Develop a real-time object detection system using the YOLO algorithm.
- Understand YOLO’s architecture and how it improves over traditional methods.
- Train and evaluate YOLO on a standard dataset for accuracy and speed.
- Compare YOLO with other detection models like Faster R-CNN and SSD.

SYSTEM ARCHITECTURE

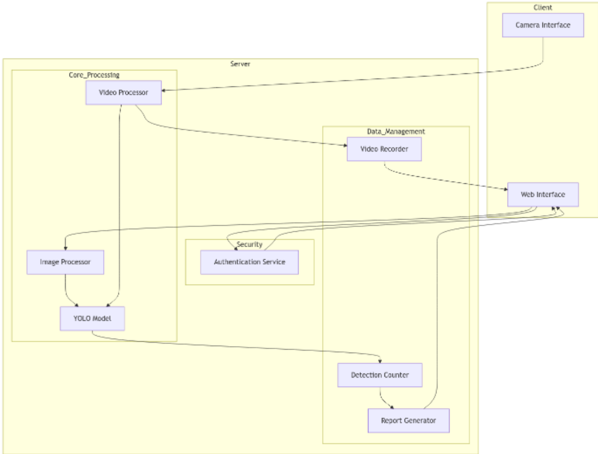

Fig.1 System Architecture

YOLO (You Only Look Once) is a single-stage, end-to-end deep learning model for object detection. Unlike traditional object detection methods that use multiple stages YOLO treats detection as a single regression problem, predicting both class probabilities and bounding box coordinates directly from the full image in one pass. The architecture is composed of the following key components:.

MODULES

There are Five Modules:

1.Authentication Module

2.Image Detection

3.Live Video Detection

4.Analytics and Reporting

5.Video Export Module

The system architecture of YOLO (You Only Look Once) is built as a unified, single-stage convolutional neural network for real-time object detection.

1. **Authentication Module**:

An authentication module is a system that verifies a user's identity before granting access. When combined with object detection using YOLO (You Only Look Once), it authenticates users by detecting specific objects (like a face, ID card, badge, or unique item) in real-time through a camera.Instead of entering a password, the user shows a registered object. YOLO detects and identifies the object, and if it matches the pre-approved data, access is granted.

2. **Image Detection:**

Image detection in authentication systems refers to identifying specific objects in an image or video feed to verify a user's identity. Instead of using passwords or PINs, users present a physical object—like a face, ID card, or badge—that the system recognizes. This is achieved using object detection algorithms like YOLO is a deep learning-based algorithm that can detect and classify multiple objects in a single pass with high speed and accuracy.This method is fast, contactless, and reduces the risk of password theft or spoofing. It's commonly used in smart security systems, offices, and IoT-based authentication.

3. **Live Video Detection**:

Live video detection is the process of analyzing real-time video streams to detect and track objects as they appear in each frame. Using advanced algorithms like YOLO, the system processes video from a webcam or CCTV camera to identify objects instantly.In authentication systems, live video detection is used to recognize faces, ID cards, or other objects continuously rather than from a single image..

4. Detection & Reporting:

Analytics and reporting in an authentication system involve collecting, processing, and visualizing data related to user access and system performance. It tracks metrics such as successful logins, failed attempts, object detection logs, and time of access.This data helps administrators understand usage patterns, detect suspicious activities, and improve system security.

5. Video Export Module:

The Video Export Module allows the system to record and save live video footage during authentication eventsWhen a user attempts to authenticate (whether successful or not), the relevant video clip is captured and stored.This is useful for auditing, security reviews, or evidence, especially in high-security environments.

REFRENCES

[1] **Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016).Title:***You Only Look Once: Unified, Real-Time Object Detection*
Publisher: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[2] **Redmon, J., & Farhadi, A. (2017).**
Title:*YOLO9000: Better, Faster, Stronger*
Publisher:IEEECVPR
Info: YOLO9000 improved the accuracy of the original model and introduced the ability to detect over 9,000 object categories by combining classification and detection datasets.
[3]**Redmon, J., & Farhadi, A. (2018).**
Title:*YOLOv3: An Incremental Improvement*
Publisher: arXiv preprint
Info: YOLOv3 made enhancements in backbone architecture (Darknet-53), used residual connections, and introduced multi-scale predictions. It significantly improved performance on small object detection.
[4]**Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020).Title:***YOLOv4: Optimal Speed and Accuracy of Object Detection*
Publisher: arXiv preprint
Info: YOLOv4 was developed to run efficiently on conventional GPUs. It introduced new techniques such as Weighted Residual Connections, Cross-Stage-Partial connections, and Self-Adversarial Training.