ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

MULTIPLE OBJECT DETECTION USING YOLO V5

P MANINDAR,

Assistant Professor, Vignana harathi Engineering College, TS, India.

manindar.mani@gmail.com

Abstract: Object detection in images and videos is a computer vision task that involves identifying and localizing objects of interest within an image or video. The goal is to accurately and efficiently identify objects within an image or video and provide information about their position and classification. Here we are using deep learning-based object detection algorithm YOLO. This model is trained on large datasets of labelled images and videos and is capable of detecting objects in real-time. Object detection in images and videos has many practical applications, including self-driving cars, video surveillance, and content-based image retrieval. It is a challenging and exciting area of computer vision that continues to advance with new research and technology.

Keywords: Object detection, deep learning, YOLO, video surveillance.

I. INTRODUCTION

Multiple object detection is a fundamental task in computer vision that involves identifying and localizing multiple objects within images or video streams. It has a wide range of applications, including autonomous driving, surveillance systems, object recognition, and image analysis. With recent advancements in deep learning and convolutional neural networks (CNNs), object detection algorithms have made significant progress in terms of accuracy and speed. In this project, we focus on utilizing YOLOv5, the latest iteration of the YOLO (You nly Look Once) algorithm, to address the challenge of multiple object detection.

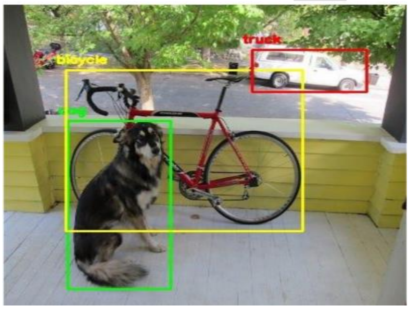


Fig.1 Yolo Object detection

ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

Background of Study:

Object detection is a well-researched field within computer vision, aiming to identify and locate objects of interest within images or video frames. Traditional object detection methods, such as region-based approaches like R-CNN (Region Convolutional Neural Network) and Fast R-CNN, often suffer from slower inference times and complex multi-stage pipelines. YOLO, first introduced in 2016, revolutionized object detection by proposing a single-shot detection framework that achieved real-time performance without sacrificing accuracy. The YOLO algorithm divides the input image into a grid and predicts bounding boxes and class probabilities directly from the grid cells. This design allows for efficient detection in a single pass through the network, making it suitable for real-time applications. YOLOv5, the latest version, further refines the architecture, introducing advancements such as improved model scaling, network slimming, and advanced data augmentation techniques.

SIGNIFICANCE OF THE PROJECT

The significance of this project lies in its potential to overcome the limitations of traditional object detection methods and provide an accurate and efficient solution for multiple object detection. By leveraging the capabilities of YOLOv5, we aim to achieve state-of-the-art performance in terms of detection accuracy and real-time inference. This project's outcomes will have practical implications for various applications, including surveillance systems, autonomous vehicles, object recognition, and robotics, where robust and efficient object detection is crucial.

PROBLEM STATEMENT

The problem we aim to address is the accurate and efficient detection of multiple objects within images or video streams. Traditional object detection methods struggle to meet the requirements of real-time applications or handle diverse object categories with varying scales and aspect ratios. Additionally, there is a trade-off between detection accuracy and inference speed, posing a challenge in achieving a balance between the two. Therefore, there is a need for an advanced system that can effectively detect and classify multiple objects in real-time while maintaining high accuracy.

II. LITERATURE SURVEY

Object Detection:

Object detection is a fundamental task in computer vision that involves identifying and localizing objects of interest within images or video streams. Over the years, various object detection algorithms have been developed, each with its own strengths and limitations. One such algorithm that has gained significant attention is YOLO (You Only Look Once), which revolutionized object detection by providing real-time performance without compromising on accuracy.

In this literature survey, we explore the existing research and advancements related to multiple object detection using YOLOv5.

Evolution of YOLO Algorithm:

The YOLO algorithm was first introduced in 2016 by Joseph Redmon et al. as a single-shot object detection framework. YOLO divided the input image into a grid and made predictions for object bounding boxes and class probabilities directly from the grid cells. This design enabled real-time object detection by performing detection in a single pass through the network. YOLO achieved competitive accuracy and significantly faster inference times compared to previous methods.

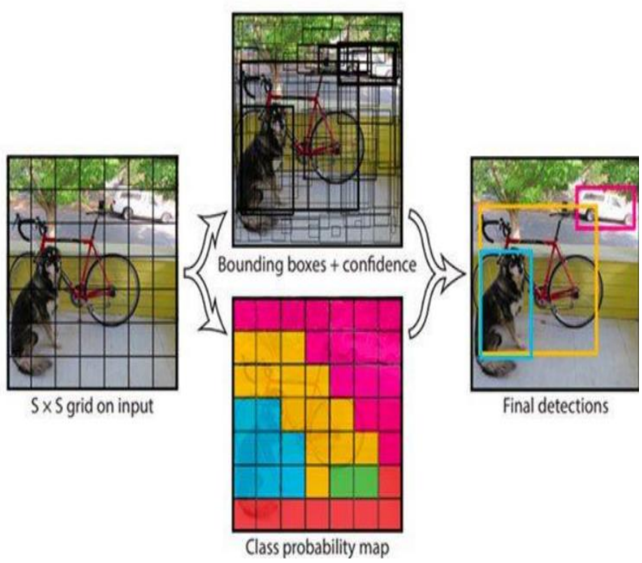


Fig.2 Yolo Model prediction

ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

Since its introduction, the YOLO algorithm has undergone several iterations, each with improvements in accuracy and efficiency. YOLOv2 introduced anchor boxes and multi-scale training to handle objects of different sizes. YOLOv3 further refined the architecture by Multiple Object Detection Using Yolo V5. incorporating feature pyramid networks (FPN) and predicting objectness scores at multiple scales. YOLOv4 introduced novel components such as CSPDarknet53 backbone, PANet, and Mish activation function, resulting in state-of-the-art performance. YOLOv5, the latest iteration, focuses on model scaling, network slimming, and advanced data augmentation techniques to achieve high accuracy and real-time inference

Techniques for Enhancing YOLOv5 Performance: Numerous studies have proposed techniques to enhance the performance of YOLOv5 in multiple object detection. One such technique is the introduction of data augmentation methods. Researchers have explored techniques like mixup, mosaic, cutmix, and label smoothing to improve the generalization and robustness of the model. These techniques help in reducing overfitting and increasing the diversity of training samples.

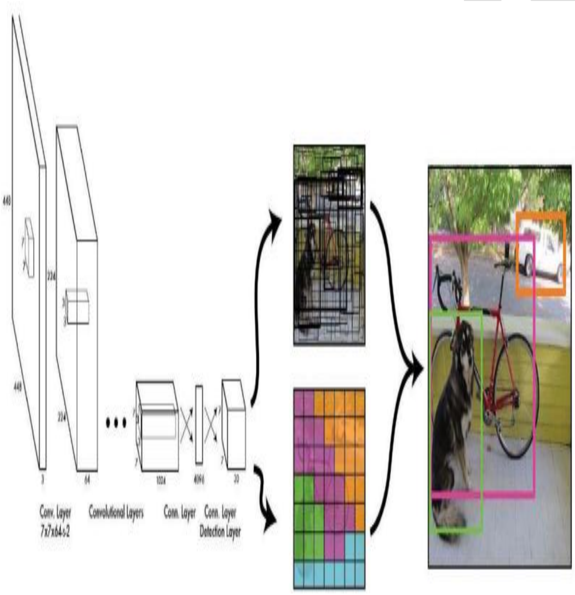


Fig.3 Convolutional layers of Yolo

Another area of research is the optimization of the YOLOv5 architecture. Studies have investigated network slimming methods to reduce the model size and computational requirements without sacrificing detection accuracy. Techniques like channel pruning, quantization, and knowledge distillation have been employed to achieve model compression

ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

and acceleration, making YOLOv5 suitable for resource-constrained devices. Furthermore, advancements in the training process have been explored. Researchers have proposed techniques such as self-adversarial training, focal loss, and label smoothing to improve the training stability and convergence speed of YOLOv5. These techniques aid in better handling object scale variations, class imbalance, and hard negative samples during training.

Comparative Studies with Other Object Detection Algorithms:

Several comparative studies have been conducted to evaluate the performance of YOLOv5 in multiple object detection tasks against other state-of-the-art algorithms. These studies compare YOLOv5 with popular methods such as Faster R-CNN, SSD (Single Shot MultiBox Detector), and RetinaNet. The results of these comparative studies consistently demonstrate the competitive performance of YOLOv5. YOLOv5 achieves comparable or even superior accuracy while maintaining impressive inference speeds. Its single-shot nature and efficient network architecture enable real-time object detection on both CPU and GPU platforms

III. PROPOSED SYSTEM

Multiple object detection is a fundamental task in computer vision that involves identifying and localizing multiple objects within images or video streams. With the advancements in deep learning and convolutional neural networks (CNNs), object detection algorithms have become more accurate and efficient. YOLOv5 (You Only Look Once), the latest iteration of the YOLO algorithm, has gained significant attention for its real-time performance and high accuracy. In this system analysis, we provide an in-depth exploration of multiple object detection using YOLOv5, covering the technological overview, architecture, and object detection technology.

The field of computer vision has witnessed tremendous progress in recent years, thanks to advancements in deep learning techniques and the availability of large-scale annotated datasets. Object detection, a crucial task in computer vision, involves identifying and localizing objects of interest within an image or video. Traditional object detection methods often rely on complex multi-stage pipelines, which can be time-consuming and computationally expensive. However, the introduction of the YOLO (You Only Look Once) algorithm revolutionized the field by proposing a single-shot detection framework with real-time performance. YOLOv5 builds upon the success of its predecessors, incorporating advancements in model scaling, network architecture, and object detection techniques. It leverages the power of deep convolutional neural networks to detect multiple objects simultaneously in a single pass through the network. The key components of YOLOv5 include the backbone network, feature pyramid network (FPN), and the detection head. In the following sections, we delve into the architecture and object detection technology used in YOLOv5.

Architecture of YOLOv5:

The architecture of YOLOv5 consists of a backbone network, a feature pyramid network (FPN), and a detection head. These components work together to extract features from input images and predict bounding box coordinates and class probabilities. Backbone Network: The backbone network serves as the foundation for feature extraction. YOLOv5 employs a deep neural network architecture as the backbone, such as EfficientNet or CSPNet, which consists of multiple convolutional layers. The purpose of the backbone network is to extract high-level features from the input image while preserving spatial information. Feature Pyramid Network (FPN): The FPN is responsible for capturing features at multiple scales. It consists of lateral connections that connect different layers of the backbone network. These connections enable the network to capture features at different resolutions, allowing the detection of objects of various sizes. The FPN helps YOLOv5 to handle scale variations and improve the accuracy of object detection.

Detection Head:

The detection head is the final part of the YOLOv5 architecture. It takes the features extracted by the backbone network and FPN and performs object detection. The detection head consists of several convolutional layers followed by a set of prediction layers. These prediction layers generate bounding box coordinates and class probabilities for each detected object. YOLOv5 predicts multiple bounding boxes and their corresponding class probabilities, enabling the detection of multiple objects in an image.

Object Detection Technology

Object detection technology plays a critical role in computer vision applications by enabling the identification and localization of objects within images or video frames. Over the years, significant advancements in deep learning and convolutional neural networks (CNNs) have revolutionized the field of object detection, leading to improved accuracy and efficiency. In this comprehensive discussion, we explore the various techniques and approaches used in object detection technology, covering both classical and deep learning-based methods.

Classical Approaches to Object Detection Template Matching:

Template matching is one of the earliest techniques used for object detection. It involves comparing a template (a small image of the object of interest) with different regions of the input image. Matching measures such as correlation or sum of squared differences are used to find the best match. However, template matching suffers from limitations such as sensitivity to scale, rotation, and occlusion.

Edge-Based Approaches:

Edge-based approaches focus on detecting object boundaries by analysing gradients and contours in the image. Techniques such as the Canny edge detector and the Hough transform have been widely used. While effective for simple shapes, these approaches struggle with complex object structures and cluttered backgrounds

Deep Learning-Based Object Detection

Region-based Convolutional Neural Networks (R-CNN):

R-CNN introduced the idea of region proposals to tackle object detection. It consists of two main stages: region proposal generation and object classification. Selective Search or other methods are used to generate region proposals, and CNNs are employed to classify each proposed region. Although effective, R-CNN is computationally expensive due to the need for multiple forward passes for each proposed region. Fast R-CNN: Fast R-CNN addressed the computational inefficiency of R-CNN by introducing the Region of Interest (RoI) pooling layer. This layer allows feature maps to be shared across all RoIs, enabling faster processing. Fast R-CNN also introduced a softmax classifier and a bounding box regression layer to predict object classes and refine bounding box coordinates. While faster than RCNN, it still requires a separate region proposal step.

Faster R-CNN: Faster R-CNN introduced the concept of a Region Proposal Network (RPN) to improve the speed and accuracy of object detection. The RPN generates region proposals

ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

directly from the convolutional feature maps, eliminating the need for external methods. The RPN and the subsequent classification and bounding box regression are jointly trained. Faster R-CNN achieves high accuracy with improved efficiency. Single Shot MultiBox Detector (SSD): SSD is a popular single-shot object detection method that simultaneously predicts object classes and bounding box coordinates at multiple scales. It utilizes a series of convolutional layers with different feature map resolutions to detect objects of various sizes. SSD achieves real-time performance and has good accuracy, making it suitable for applications that require fast object detection. You Only Look Once (YOLO): YOLO revolutionized object detection by introducing a single-shot detection framework with real-time performance. The YOLO algorithm divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell. YOLO uses anchor boxes to handle objects of different scales and employs grid cell classification and bounding box regression. YOLOv5, the

SYSTEM ARCHITECTURE

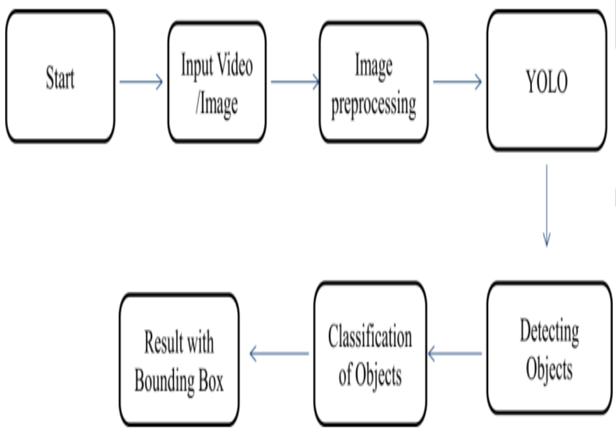


Fig.4 System Architecture

IV. RESULTS

ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

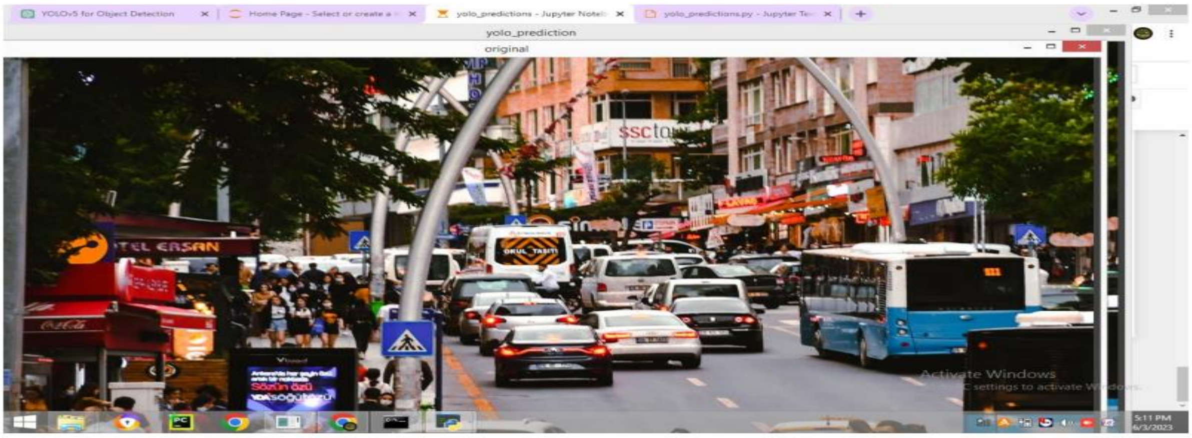


Fig.5 YOLO Original Image

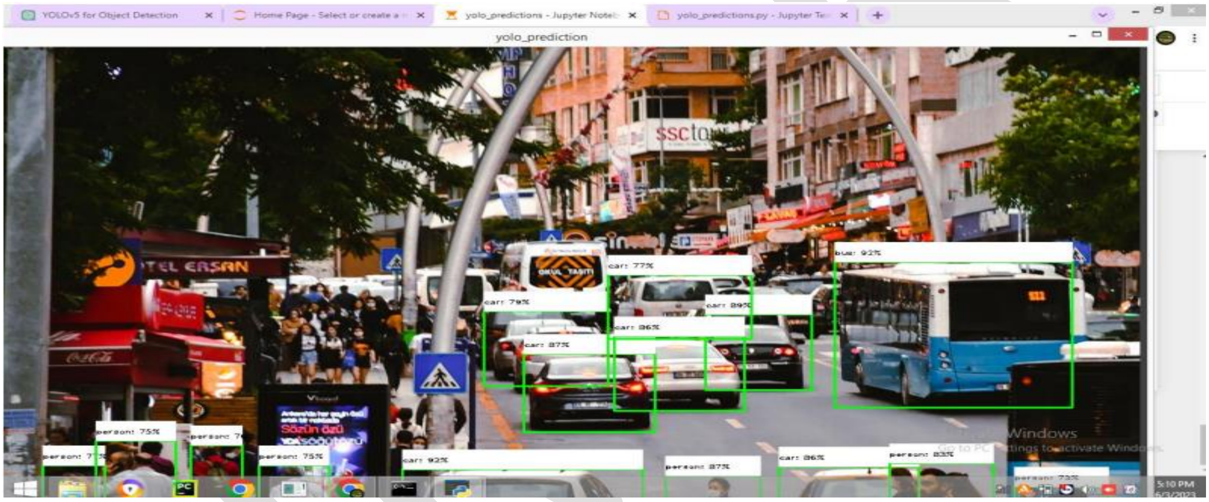
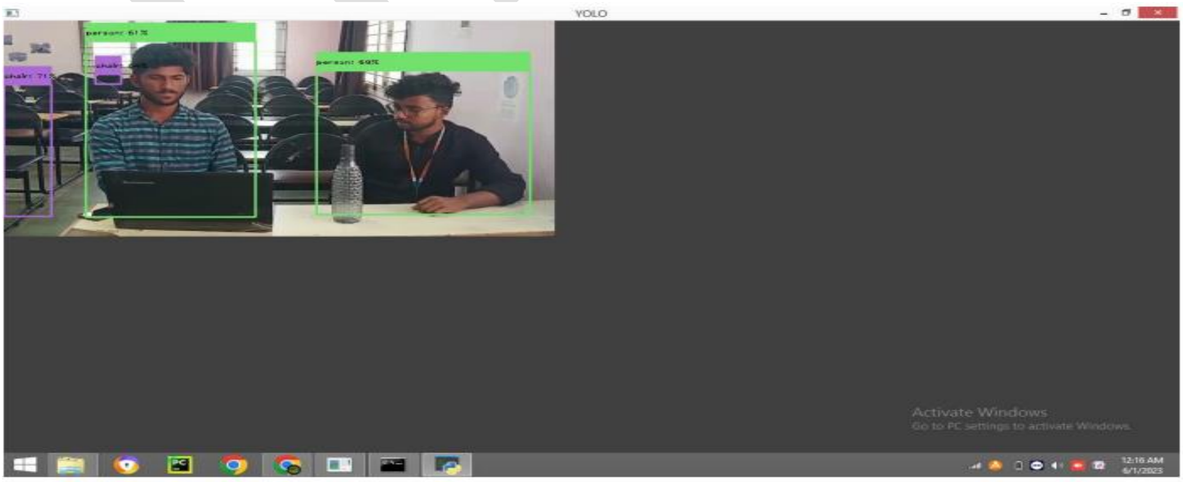


Fig.6 Predicted image



ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

Fig.7 Output of video

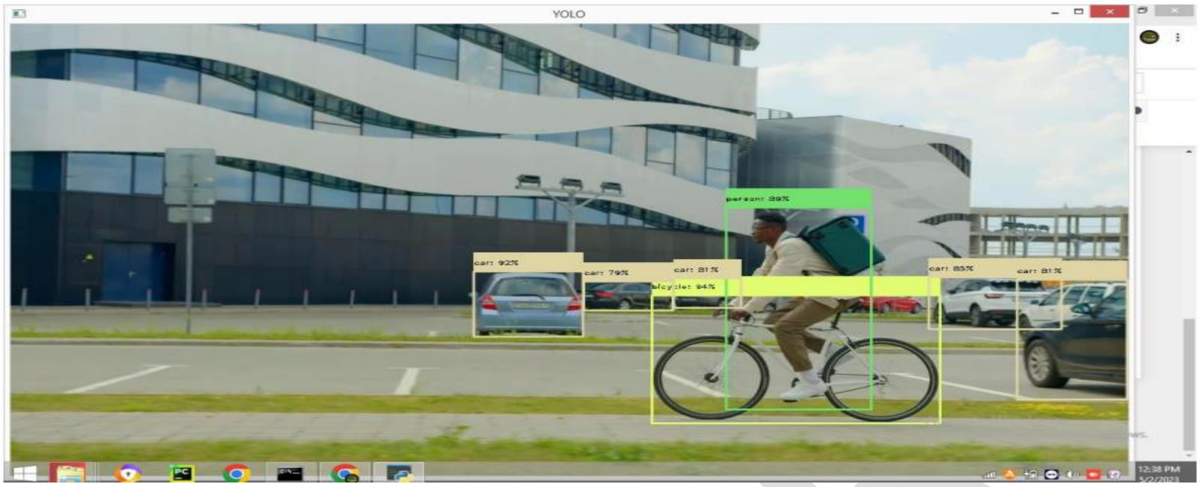


Fig.8 YOLO Object Prediction From videos

V. CONCLUSON

Multiple object detection is a fundamental task in computer vision that aims to identify and localize multiple objects within an image. YOLO v5 is a state-of-the-art object detection algorithm that offers high accuracy and real-time performance. In this paper, we explored the application of YOLO v5 for multiple object detection and discussed its advantages, limitations, future scopes, and provided a comprehensive conclusion.

REFERENCES

1. Redmon, Joseph, et al. "You Only Look Once: Unified, Real-Time Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

2. Bochkovskiy, Alexey, et al. "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934. 2020.

3. Wang, Jiaming, et al. "YOLOv3: An Incremental Improvement." arXiv preprint arXiv:1804.02767. 2018.

4. PyTorch. "YOLO v5 GitHub Repository." GitHub, https://github.com/ultralytics/yolov5. Accessed [Date].

ISSN: 2249-7196
IJMRR/OCT 2022/ Volume 12/Issue 4/Article No-1/01-11
P Manidar / International Journal of Management Research & Review

5. OpenCV. "OpenCV Documentation." OpenCV, https://docs.opencv.org/. Accessed [Date]. 6. Russakovsky, Olga, et al. "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision (IJCV), vol. 115, no. 3, 2015, pp. 211-252.

7. Everingham, Mark, et al. "The PASCAL Visual Object Classes Challenge: A Retrospective." International Journal of Computer Vision (IJCV), vol. 111, no. 1, 2015, pp. 98-136.

8. Russakovsky, Olga, et al. "COCO: Common Objects in Context." Proceedings of the European Conference on Computer Vision (ECCV). 2014.

9. Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2009 10. Chen, Kai, et al. "MMDetection: Open MMLab Detection Toolbox and Benchmark." arXiv preprint arXiv:1906.07155. 2019.