OBJECT DETECTION USING YOLO V5

L. Manikandan[1], Rapelli Malavika[2], Borlakunta Anjali[3], Sathuri Chandana[4]

Article History: Received: 07.05.2023 **Revised:** 19.06.2023 **Accepted:** 14.07.2023

Abstract:

Detecting objects is an important computer vision task that involves recognizing and locating objects in pictures or videos. YOLOv5 is a cutting-edge object detection algorithm that uses deep ConvNet to predict object classes and locations in instantaneously. In this article, we present the object detection application of YOLOv5, including the detection of people, vehicles, and animals in various environments. Test the performance of the method on the COCO dataset and make it clearer and faster compared to YOLO's previous model. Our outcome shows that YOLOv5 achieves good modern performance for a variety of sensing applications, including surveillance, robotics, and autonomous driving. Additionally, we provide a comprehensive review of the algorithm's strengths and weaknesses and discuss future directions for improving its performance. Overall, this article demonstrates the potential of YOLOv5 to be a powerful tool for object detection in real-world applications.

Keywords: Object detection, COCO dataset. YOLOv5.

---

[1]Assistant Professor, Department of Computer Science & Engineering, Sridevi Womens Engineering College, Hyderabad, Telangana, India.
[2,3,4]Undergraduate Student, Department of Computer Science & Engineering, Sridevi Womens Engineering College, Hyderabad, Telangana, India.

Email: [1]swecmanicse90@gmail.com, [2]malavikarapelli@gmail.com, [3]anjaliborlakunta38@gmail.com, [4]chandanasathuri06@gmail.com

DOI: 10.31838/ecb/2023.12.s3.698

1. INTRODUCTION

Detecting of objects is a challenging project in computer vision that includes figuring out and localizing objects of interest within a photo or video. With the expanding availability regarding high-resolution cameras and the proliferation of video data, object detection has become an essential tool for a huge variety of utilization. However, item detection remains a rigorous task as a consequences of difficulty and variability of actual environments, the presence of occlusions and clutter, and the large number of objects that must be detected simultaneously.

In recent times, deep learning strategies have revolutionized object detection, enabling significant improvements in accuracy and speed. YOLO is a desired family of object spotting algorithms that use deep CNN to carry out real-time detection. The You Only Look Once algorithm detects objects by using splitting the input photo into a grid and detecting bounding boxes and sophistication probabilities for each grid square. YOLO also utilizes novel loss function that unite localization and classification errors, allowing it to handle overlapping objects and improve the accuracy of detection.

Detecting objects is a rapidly evolving field and there are many ongoing research efforts aimed at improving the accuracy and efficiency of object detection algorithms. Future developments in object detection may involve the use of more advanced deep learning architectures, the incorporation of additional sensor modalities, and the integration of object detection with additional computer vision tasks, like semantic segmentation and instance segmentation.

Recently, a new version of YOLO, called YOLOv5, has been released, which improves the accuracy and speed of object detection even further. YOLOv5 uses a novel architecture based on a hybrid backbone network that combines features from different levels of abstraction, allowing it to detect objects with high accuracy and efficiency. YOLOv5 also includes several optimizations, including a new anchor box design, feature pyramid network, and improved training techniques, which further improve its performance.

Here, we present an implementation of YOLOv5 in object detection tasks and evaluate its performance on the COCO dataset. We examine that YOLOv5 achieves trailblazing performance on these datasets and can be used for a wide range of the algorithm’s strengths and weaknesses and discuss future directions for improving its performance. Overall, this paper demonstrates the potential of YOLOv5 as a powerful tool for object detection in real-world applications.

2. RELATED WORK

Existing object recognition methods frequently depend on manually labelled data, which places severe constraints on the development of completely independent machine vision systems. In this paper, we show a machine vision system with intelligence that can recognize and autonomously learn about specific objects in a real environment. Salient object detection is used by this technology. We took cues from the early phases of the human visual system when designing it. We offer a novel, quick method for detecting visually striking objects in this situation that is resistant to poor lighting. Afterward, we apply it to extract salient items that occur effectively to utilize for training suggested system's ML based detection and recognition feature. We evaluate the performance of our important object detection set of rules against other cutting-edge methods using data from the MSRA Salient Object Database benchmark. A humanoid robot with the suggested system installed has increased autonomy in studying and social interactivity. We present and debate the findings, confirming the suggested ideas.

A recent development in the manufacturing of food or agriculturally related materials is traceability using video. However, these apps have constrained bandwidth and processing power. For these applications, it is essential to enhance conventional object detection techniques. As part of this, we present an algorithm for traceability video analysis that combines the non-parametric method and frame difference. The suggested method outperformed the conventional frame difference and GMM, per the experimental findings.
One of the most crucial steps in video surveillance, background subtraction is used in

a variety of real-world uses including surveillance, interaction of human-machine, capturing of optical motion, and intelligent visual observation of micro and macro organisms. One of the initial steps used to distinguish the foreground objects from the largely stationary backdrop is background subtraction. A pixel is typically regarded as foreground if its value exceeds that of the corresponding pixel in the reference picture. Therefore, to distinguish between the foreground and background pixels, each pixel must be contrasted. By first classifying the blocks in the frame as background and then classifying the remaining blocks using the correlation coefficient, the technique described in this article enhances the frame difference method. Blocks that aren't regarded as backgrounds are classified at the pixel level for additional precision. Studies are carried out using common data sets, and the performance measures work well under some challenging circumstances.

With the uncommon evolution of Convolutional Neural Network (ConvNet) and its variations from 2012, detecting objects has expanded as a vital usage of image processing. The mean average precision (mAP) of the ConvNet series when it develops to faster regions with CNN has hit 76.four, but the frame per second (FPS) of faster R-ConvNet remains at range of five and eighteen, which is a great deal slower than the real-rime impact. Therefore, increasing speed is crucial in detecting objects process improvement. Presenting YOLO among CNN, which breaks through the CNN circle of relatives' subculture and innovates a completely unique way of fixing object detection in the straightforward and highly effective manner. This depends on information related to history and ConvNet main solution. With 155 frames per second at its fastest pace and a map that can reach up to 78.6 pixels in size, it has significantly outperformed Faster R-CNN in terms of performance. Additionally, YOLOv5 outperforms the most recent, most sophisticated solution in terms of both streak and perfection, as well as owning detector with excellent hypothesis capabilities to mean the entire picture.

Between most effective and frequently used methods for object location are sliding window classifiers. However, the method that training is typically carried out is unsuitable for localizing. A binary algorithm is first skilled on a selected favorable and unfavorable instances before being applied to various areas of test images. Instead, we suggest treating object localization as a complication of detecting ordered data, modeling the issue as the projection of the bordering of items found in input photo rather than as difficulty of binary classification. We can structure the training process as an SVM generalization that can be solved quickly by using a joint-kernel framework. By employing a branch-and-bound approach for determination through both coaching and examining, we further increase computational effectiveness. The results of development on the PASCAL VOC and TU Darmstadt datasets demonstrate that the ordered skilling method outperforms binary training and the highest prior scores.

Existing system:

The Region-based convolutional neural network demonstrated by Ross Girshick's team, is a DL algorithm for detection of objects. It is region-based approach that combines region proposals and CNNs to predict objects in an image. R-CNN model has effective results on datasets like ImageNet and was a significant breakthrough in object detection. However, it has some disadvantages such as slow training and inference times, high memory usage, and a complex pipe.

Disadvantages of the Existing system:

- R-CNN is a computationally expensive approach, with slow training and inference times due to the need to generate region proposals and extract features for each proposal separately. It requires a large amount of memory to store the extracted features for each region proposal.
- R-CNN has a complex pipeline that involves multiple stages, including region proposal generation, feature extraction, and classification. It has difficulty in handling objects that vary in scale or aspect ratio, which can lead to missed detections or false positives.

Proposing system:

Proposing object detection using YOLOv5, as it is a famous object detection algorithm that detects from images, video, and live camera feeds. This algorithm is pre-skilled with all pictures and assigns a completely unique elegance call to each unique image and then generates a version, this set of rules converts each picture into layers, and then for every layer it extracts capabilities and adds weight to the model. Every time we provide an image, it will be carried out to a pre-trained weight version to get the first-class accuracy matching the picture label. Here we train the model by using a pre-defined COCO dataset.

Advantages of the proposed system:

- YOLOv5 is a lightweight and fast object detection algorithm. It achieves high accuracy in standard object detection of the COCO dataset.
- YOLOv5 is versatile and can be used for various object detection tasks. It is easy to use and implement, with a straightforward training process and simple configuration files. It is an open-source algorithm.
- COCO dataset, contains a large number of images with detailed object annotations. Using the COCO dataset in combination with YOLOv5 can lead to highly accurate object detection results.

Structure of the project

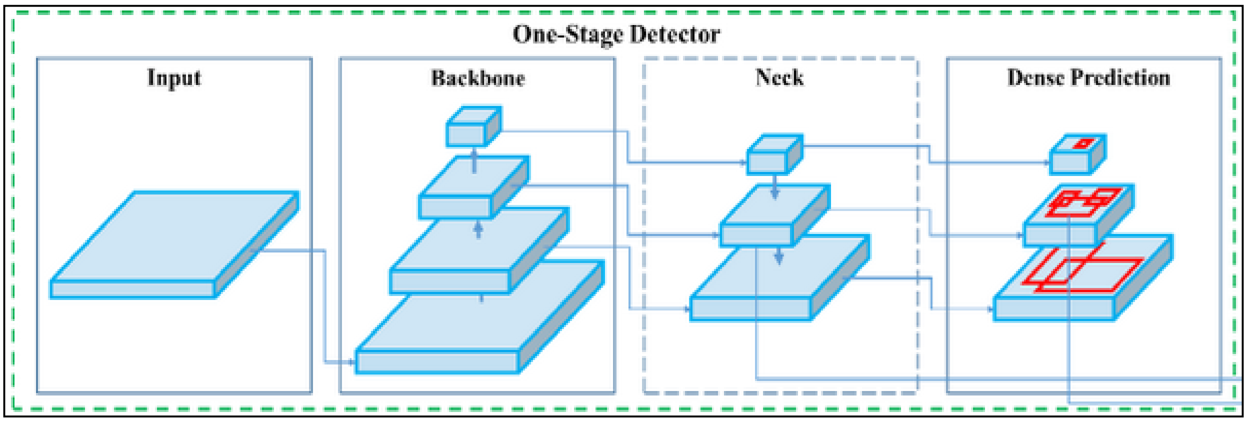


Fig-1: Algorithm version

The System architecture describes a one-stage detector as the system uses the YOLO V5 algorithm. The admin develops the application where the input is trained by a machine to detect the objects. From the input, the essential features of different resolutions will be exploited which is known as the backbone model. From this model, the fusing of features of different resolutions is done and represented as a neck model. Finally, the dense prediction of objects along with dimensions will be predicted. Here, it only represents the object detection process internally, whereas externally the admin generates the model and the user signs up and uses this application to detect objects along with the dimensions. As it is one staged detector the detecting of objects will be fast and multiple objects can be detected within no time. The user details and credentials will be stored by using the database. The detection of objects can be chosen in two ways. One is detecting objects by live data. Alternatively, the user can upload an image that can be uploaded in .jpg format and the application detects the learned objects. All these processes will be done by using the web server.

Modules

1. Data exploration:

Data exploration is the process of analyzing and understanding the data available for a particular problem or project. It involves examining the data to identify patterns, relationships, and insights that can inform decision-making or guide further analysis. Here, it includes identifying the dataset (COCO dataset) that will be used for training and testing the YOLOv5 model.

2. Data Processing:

Data processing is the step in the object detection pipeline where raw data is transformed and prepared for use in training and testing an object detection model. In the context of YOLOv5, data processing typically involves data loading, data cleaning, data augmentation, data normalization, data batching, and data splitting.

3. Spitting data into train and test:
Splitting the data into training and testing sets is a critical step in the object detection pipeline using YOLOv5. The motive of this step is to evaluate the overall performance of the skilled version on unseen statistics and prevent overfitting. The typical approach for splitting data into train and test sets is to use a random split. This involves randomly selecting a portion of the data, typically around 20-30%, to be utilized as the testing set, while the remaining data is considered into the training set.

4. Building a model:
Building a model for object detection using YOLOv5 involves several steps, including data preparation, model configuration, and training.

5. User signup & login:
User signup and login are crucial components of any web application or mobile app that requires user authentication. Here, the user should signup and log in to use the application.

6. User input:
User input refers to any data, command, or requests that a user enters into a computer program or system using a keyboard, mouse, touchscreen, or another input device. User input is essential for programs and systems to function properly, as it allows users to interact with the software and provide the necessary information to achieve their desired outcome. Here, the user uploads the image or live camera data to detect objects.

7. Prediction:
Prediction is the process of using data and statistical algorithms to make informed estimates about future outcomes or events. Based on the artificial intelligence and machine learning, prediction involves using algorithms to analyze patterns in data and create models that can be used in a wide range of applications. The first-class predictions depends on the quality and quantity of data used to create the model, as well as the algorithm used to analyze the data and make predictions. Here, by using YOLOv5 and COCO datasets the prediction of objects and their detection will be done.

Algorithm
YOLOv5:
It is an object detection algorithm that was developed by Ultralytics, a computer vision, and AI research company. It is an updated model of the popular YOLO (You Only Look Once) algorithm, which is known for its fast and accurate object detection capabilities. YOLOv5 builds on the strengths of the original YOLO algorithm, while also introducing a number of improvements and new features. The YOLOv5 is primarily depends on deep CNN structure which skilled utmost for locating objects in images. The set of rules uses a single neural network in order to expect item bordering and class probabilities. This differs from traditional object detection algorithms that require more than one level to detect objects, making YOLOv5 quicker and extra correct than conventional algorithms. Overall, YOLOv5 is an accurate object detection algorithm that offers high accuracy and pace, making it properly appropriate to more than a few applications in computer imaginative and prescient. Its architecture and training techniques have been optimized to improve its overall performance on a variety of datasets, making it a famous preference for item detection tasks.

COCO dataset:
The Common Objects in Context dataset is a huge-scale photo recognition, segmentation, and captioning dataset that is extensively utilized in computer vision and machine cutting-edge research. The dataset consists of greater than 330000 images with over 2.five million object instances categorized across 80 unique object classes. The snap shots have been gathered from a huge variety modern-day sources and are annotated with object bounding boxes, segmentation masks, and photo captions.
The COCO dataset is widely used for training and evaluating object detection and segmentation algorithms, and also for other works such as photo captioning and visual question answering. Particularly useful for object detection as the dataset contains a large number of objects in a variety of contexts and poses, making it challenging for algorithms to accurately detect and classify objects. The COCO dataset has become a benchmark dataset in laptop imaginative and prescient studies,

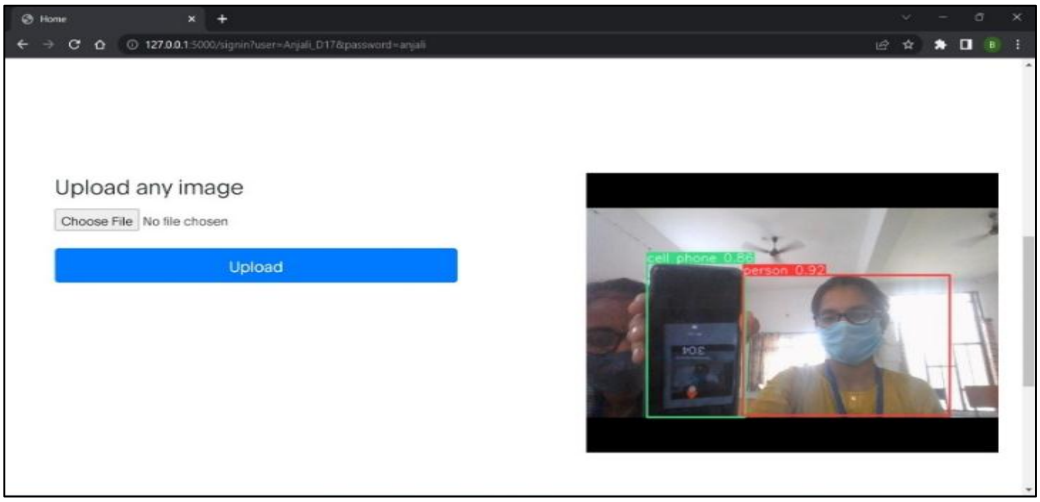
with many algorithms and strategies evaluated and tested on the dataset.

Thus, YOLOv5 along with the COCO dataset the object detection application is developed.

Test instances:

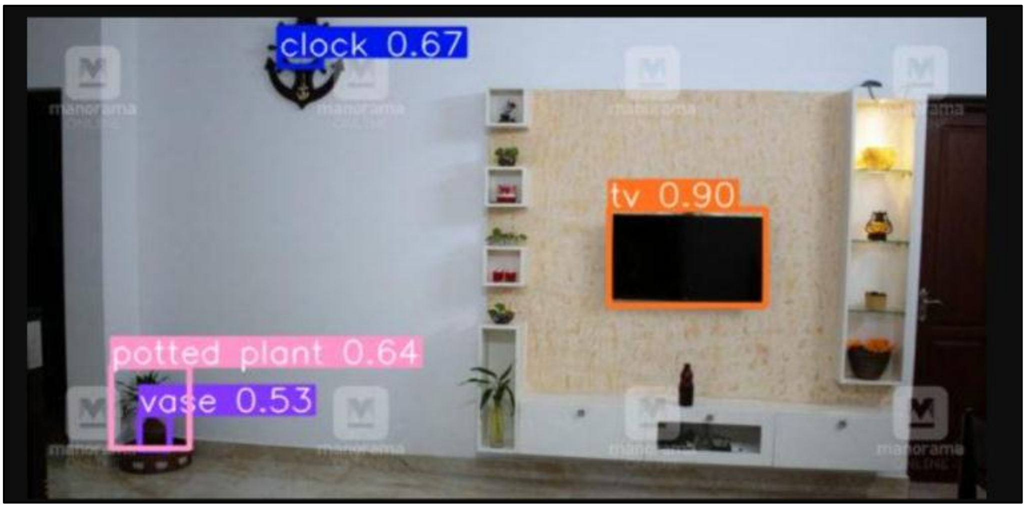
Instances	**Expected end result**	**Actual end result**	**Test result**
User register by providing the required credentials and logins.	Viewing the user's page.	Viewing the user's page.	Successful
User logins and camera will be active to detect.	Detection of live feed data objects.	Detection of live feed data objects.	Successful
The user uploads an image.	Object detection along with the location.	Object detection along with the location.	Successful
The user uploads the file with other than the .jpg extension.	Kindly upload the file with .jpg.	Kindly upload the file with .jpg.	Successful
The user gives incorrect credentials.	No change of page.	No change of page.	Successful

3. **Results**:



Above screen represents live object detection.

Above screen represents object detection of uploaded image.



Above image represents object detection of the uploaded image.

4. CONCLUSION & FUTURE SCOPE

YOLOv5 is a tremendously efficient and powerful item detection algorithm that offers significant improvements in phrases of velocity and accuracy in comparison to its predecessors. It has shown great promise in a diversity of applications, autonomous driving, surveillance, and robotics. However, like any technology, it is important to use YOLOv5 ethically and responsibly to avoid potential negative consequences. This includes ensuring that the data used for training and testing is diverse and representative as well as taking steps to address any biases that may be present in the algorithm. Overall, YOLOv5 represents a significant step forward in object detection technology. The COCO dataset, which is predefined of 80 objects is very helpful as the objects can be detected within no time.
Yolov5 is an effective set of rules for detecting the items that has shown promising effects in real time detection of objects and tracking. In the future, there will be several potential areas in which YOLOv5 might be in addition implemented:

Object recognition in complex environments: YOLOv5 could be enhanced to improve the accuracy of object detection in complex environments, such as crowd scenes or low-light conditions.

Real-time object tracking: YOLOv5 has already demonstrated impressive real-time object detection capabilities, but further research could focus on improving its object tracking capabilities to allow for more advanced applications.

Multi-object detection and tracking: YOLOv5 could be extended to detect and track multiple objects simultaneously, allowing for more sophisticated surveillance and monitoring systems.

3D object detection: YOLOv5 could be developed to detect objects in 3D space, which would have significant applications in robotics, autonomous vehicles, and segmented reality.

Integration with AI technologies: YOLOv5 could be integrated with other AI technologies, such as natural language processing or facial recognition, to create more advanced and intelligent systems.

COCO dataset: Extend the number of objects to the existing ones so that the COCO dataset can be used even further.
Thus, further research and development could lead to many exciting and innovative applications in a wide range of industries.

5. REFERENCE

1. Dumitru Erhan, Christian Szegedy, Alexander Toshev, “Scalable Object Detection using Deep Neural Networks”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147-2154.
2. Jifeng Dai, Yi Li, Kaiming He, Jian Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks”,

published in: Advances in Neural Information Processing Systems 29 (NIPS 2016).
3. Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time Object Detection",The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
4. Joseph Redmon, Ali Farhadi,"YOLO9000: Better, Faster, Stronger", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263-7271.
5. Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", published in Computer Vision and Pattern Recognition (cs.CV).
6. Lichao Huang, Yi Yang, Yafeng Deng, Yinan Yu DenseBox, "Unifying Landmark Localization with End to End Object Detection", Published in Computer Vision and Pattern Recognition (cs.CV).
7. Matthew B. Blaschko Christoph H. Lampert, "Learning to Localize Objects with Structured Output Regression", Published in Computer Vision – ECCV 2008 pp 2-15.
8. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Published in Advances in Neural Information Processing Systems 28 (NIPS 2015).
9. Wei Liu, Dragomir Anguelov, Dumitru Erhan, "SSD: Single Shot MultiBox Detector", Published in Computer Vision – ECCV 2016 pp 21-37.
10. YOLO Juan Du1,"Understanding of Object Detection Based on CNN Family",New Research, and Development Center of Hisense, Qingdao 266071, China.