



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 10    Issue: VI    Month of publication: June 2022**

**DOI: <https://doi.org/10.22214/ijraset.2022.44530>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Object Tracking in Crowd Environment Using Deep Learning

Lingala Hrithika Goud<sup>1</sup>, Ramidi Pooja Reddy<sup>2</sup>, Sai Surya Teja Gedam<sup>3</sup>, Dr. K. Sateesh Kumar<sup>4</sup>

<sup>4</sup>Assistant Professor, <sup>1,2,3</sup> Department of Electronics and Computer Engineering (ECM)

Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

**Abstract:** Object tracking is fetching a primary set of object identification, allocating a unique ID to each one of them, and then ensuing each of the objects as they move around frames in a movie while safe keeping the ID assignment. Person tracking is a difficult task in video surveillance. In recent years, many computer vision, deep and machine learning have been developed. Convolutional Neural Networks (CNN) are transforming target tracking. The project's purpose is to recognize and track images utilizing object identification techniques such as Region based Convolutional Neural Networks (RCNN), Faster RCNN, Single Shot Detector (SSD), and You Only Look Once (YOLO). Faster-RCNN and SSD have superior accuracy among them, whereas YOLO performs better when speed is prioritized above accuracy. Deep learning combines SSD and Mobile Nets for efficient identification and tracking implementation. This technique detects objects quickly and efficiently without sacrificing performance. Convolutionally default boxex are passed over several feature maps. If a detected object matches one of the object classifiers during prediction, a score is produced..

**Keywords:** Computer vision, Convolutional-Neural-Networks (CNN), Deep Learning, Mobile Net, You-Only-Look-Once (YOLO).

## I. INTRODUCTION

Real-time object recognition and tracking for Object Detection has been the focus of much investigation in current years due to its close alliance with video analysis and image elucidation [1]. Fabricated characteristic and superficial formable construction are the foundation of conventional object detection systems [2]. Combining a variety of low-level visual features with high-level context provided by object detectors and location categorization, they can be effortlessly put out of action. With swift advancements in deep learning, more influential tools capable of semantic learning, higher level and in depth features are being advance to solve problems that infestation current systems [3]. As far as of network architecture, training techniques, and enhancement functions, these models act in a unexpected way. In this review, we study deep learning-based object identification architectures. Our investigation begins with a short history of deep learning and its most delegate instrument, the Convolutional Neural Network (CNN).

## II. LITERATURE SURVEY

- 1) A deep neural community is used to comprehend things in photographs. SSD discretizes the result area of bounding box containers into a hard and fast of default packing containers in keeping with characteristic map role during numerous thing ratios and scales. The community generates ratings for the presence of every object kind in every default field at prediction time, after which adjusts the field to higher fit the item shape.
- 2) Introducing a class of proficient models called MobileNets for mobile and inserted image handling applications. MobileNets is in light on a simplified design that builds a lightweight deep neural network with depth-separable convolutions. We present two simple global hyperparameters to optimize the balance between latency and accuracy. Based on the limitations of the problem, these hyperparameters help the modeler choose the right model size for the application.
- 3) The YOLO object sensor isolates the input image into SxS grids, and each grid cell predicts only one object. If a cell contains multiple small objects, YOLO will not be able to detect them and will miss the object detection.
- 4) In this project objects are followed in view of variety, development of single and different items (vehicles) are perceived and included in various edges. Further single calculation might be intended for object distinguishing proof by considering shape, variety, surface, object of interest, development of article in multi heading.
- 5) Discovers objects in a video sequence using object detection algorithms such as Gaussian mixed models, Haar algorithms, directed gradient histograms, and local binary patterns.

### III. EXISTING METHODOLOGY

#### A. ResNet (Residual Neural Network)

We utilize a similar strategy concerning DSSD to train the organization model all the more really in this paper (the exhibition of the remainder of the organization is superior to that the presentation of the VGG organization). The aim is to increase precision.

#### B. RCNN (Region based Convolutional Neural Network)

To get around the difficulty of selecting so many places, Ross Girshick et al. intended a strategy where we can use indiscriminate chase to get only 2000 regions from an image, which is called region suggestion. Consequently, rather than attempting to rank countless positions, you can focus in just on 2,000.

#### C. Faster R-CNN

To find regional recommendations, the two algorithms above (RCNN and Fast RCNN) use selective search. Selective search is a sluggish and tedious activity that corrupts network execution.

#### D. MANet

Target detection has been a hotbed in computer vision for many years because it is a fundamentally difficult problem. The motive and goal of target identification is to see if an image contains examples of a particular type of object. Target detection returns the spatial coordinates and spatial degree of component cases on the off chance that there is an item to be recognized in a given picture (for example in view of jumping box utilization).

### IV. PROPOSED METHODOLOGY

#### A. YOLO

"You See Only Once" is referred to as YOLO. It's an algorithm to distinguish and recognize various components of an image. The object recognition in this algorithm is performed as a reverting issue and the class chances of the identified pictures are provided. Convolutional Neural Networks are used in the YOLO technique to detect objects in instantaneous time. For object detection, this method only performs one onward go through the neural network, as the name recommend. Yolo architecture is like FCNN (full accumulation neural network), passing the image (nxn) once over FCNN and generating a prediction (mxm), in contrast to other region recommender classifier networks (Fast) RCNN), do the detection on different suggested regions and thus end up doing the predictions multiple times for different regions of the image. This architecture splits the input image into mxm grids, with two bounding boxes and a class probability generated for each grid generation. It should be noted that the bounding box is more likely to be larger than the grid. This shows that a single run of the algorithm is utilized to foresee the whole image. CNN is utilized to foresee numerous bounding boxes and class probabilities simultaneously. There are few other forms of the YOLO algorithm. Tiny-YOLO and YOLO-v3 are two famous instances.

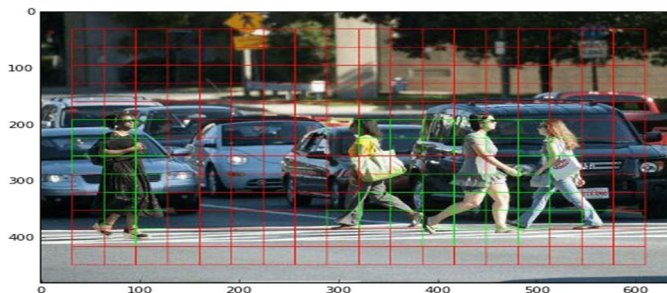
#### B. How The Yolo Algorithm Works

YOLO algorithm works of the succeeding 3 ways:

##### 1) Residual blocks.

In the image is gridded up into different grids. The dimension of each raster is SxS. The picture shows how the feed in image is split into raster.

There are many grid cells of the same size. The objects displayed in each grid cell are recognized. For example, if the centre of an object appears in a particular grid cell, that cell is accountable for discovering it.





### 2) Bounding box Regression

A bounding box is a border that emphasizes objects in an picture. Each bounding box in the image includes the succeeding attributes:

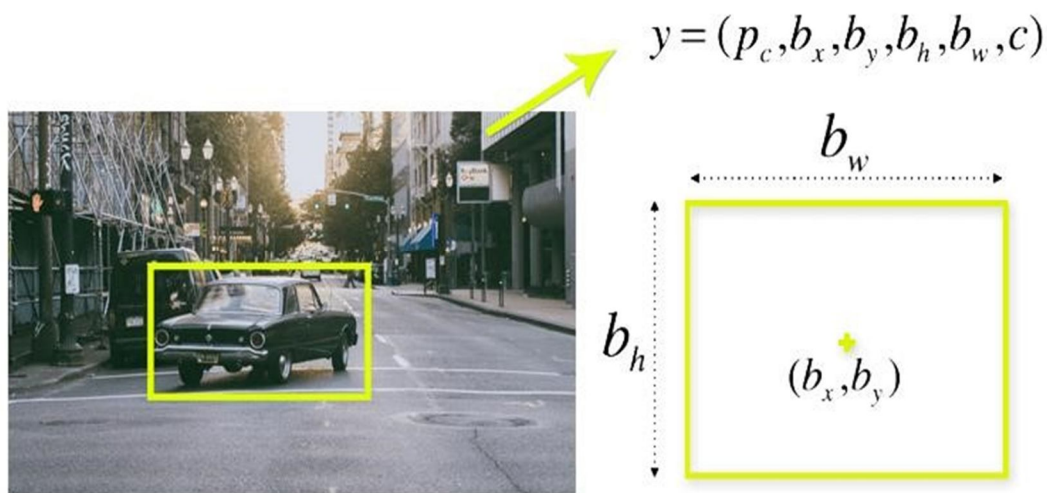
Width ( $b_w$ )

Height ( $b_h$ )

Class (for example, person, car, traffic light, etc.)- This is represented by the letter  $c$ .

Centre of bounding box ( $b_x, b_y$ )

The picture shows an instance of bounding box. The bounding box is signified by a yellow colour outline.

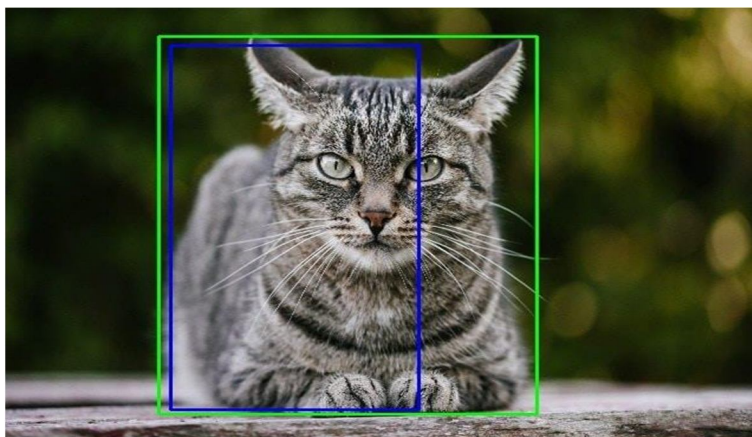


To predict the height, width, centre and class of objects, YOLO uses a single bounding box regression. The probability that an object appears in the bounding box is shown in the graph above.

### 3) Intersection Over Union (IOU)

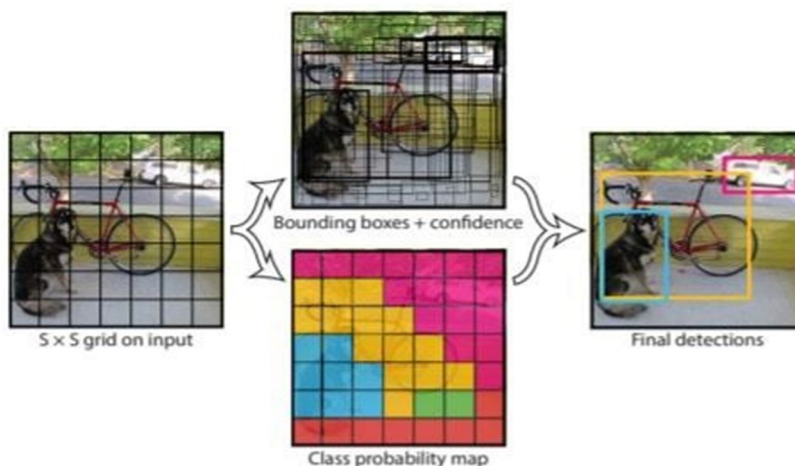
The proposal of Intersection Over Union (IOU) represents how boxes cross-over in object detection. IOU is used by YOLO algorithm to make a outcome field that pleasantly surrounds the items.

The Bounding boxes and it's self-assurance rankings are anticipated through every grid cell.



Initially, the image is distributed into matrix cells. Every framework cell predicts a B jumping box and gives their confidence opinion. The cell predicts the class likelihood and decides the class for each object.

For example, you can perceive objects of not less than three categories: Car, Canine, and Bicycle. All expectations are made at the same time utilizing a solitary convolutional brain organization.



Intersections on the merger ensure that the forecasted bounding box is even to the actual box of the object. This action eliminates extraneous bounding boxes that don't correspond to the object's attributes (like level and width). The last recognition consists of a unique bounding box that exactly matches the object.

For instance, a pink bounding box encompasses a vehicle and a yellow bounding box encompasses a bike. The blue bounding box was utilized to underline the canine.

## V. RESULTS

In this chapter we will discuss about the results

### 1) Commands to Execute

```

Anaconda Prompt (Anaconda3)

(base) C:\Users\Ramsha Akhter>cd C:\Users\Ramsha Akhter\Desktop\yolo-object-detection\yolo-object-detection

(base) C:\Users\Ramsha Akhter\Desktop\yolo-object-detection\yolo-object-detection>conda activate project

(project) C:\Users\Ramsha Akhter\Desktop\yolo-object-detection\yolo-object-detection>python yolo.py --image images/baggage_claim.jpg --yolo yolo-coco
[INFO] loading YOLO from disk...
[INFO] YOLO took 0.494382 seconds

(project) C:\Users\Ramsha Akhter\Desktop\yolo-object-detection\yolo-object-detection>python yolo.py --image images/dining_table.jpg --yolo yolo-coco
[INFO] loading YOLO from disk...
[INFO] YOLO took 0.540004 seconds

(project) C:\Users\Ramsha Akhter\Desktop\yolo-object-detection\yolo-object-detection>python yolo.py --image images/baggage_claim.jpg --yolo yolo-coco
[INFO] loading YOLO from disk...
[INFO] YOLO took 0.495597 seconds

(project) C:\Users\Ramsha Akhter\Desktop\yolo-object-detection\yolo-object-detection>python yolo.py --image images/baggage_claim.jpg --yolo yolo-coco
[INFO] loading YOLO from disk...

```

Fig 6.1 – commands to execute

In the Fig 6.1, these are the commands which are being executed. Python YOLO.py script is being run . After executing this command, —image Specifying the path of an input image is sent into the script Coco data is being passed following – yolo





Fig 6.1 Car Chase Input

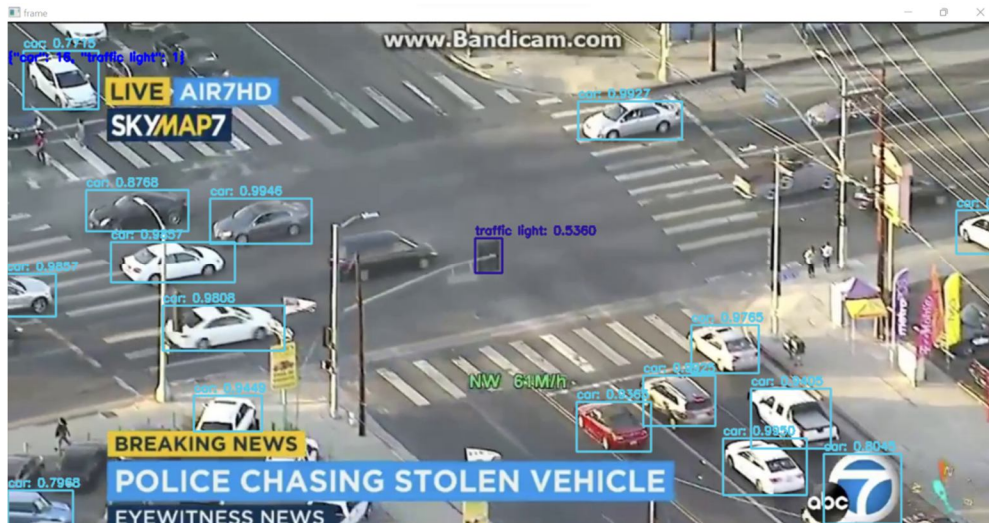


Fig 6.2 Car Chase Output

In the Fig 6.2, that the model successfully identified the objects in a video displaying news, thereby it is evident that this model can be integrated with CCTV software and can be used in complex applications to improve the security of people's lives. Irrespective of the video quality available, the model is able to identify the objects and returns the list at the top left corner and also highlights the objects with a rectangular box with the confidence intervals being displayed



Fig 6.3 Airport Input

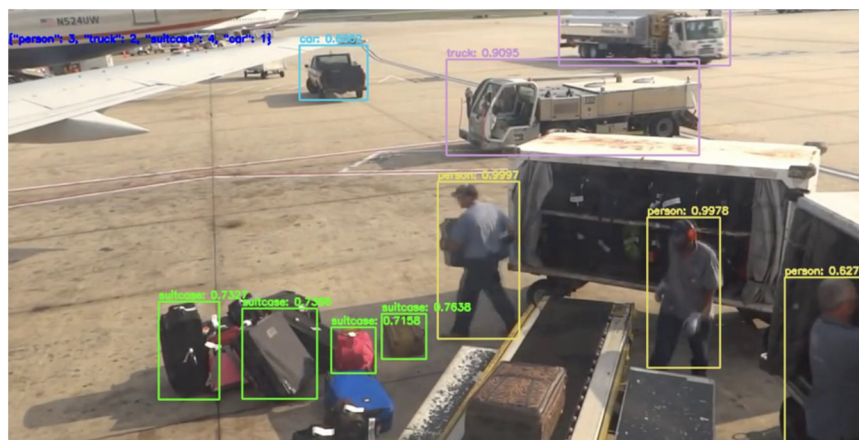


Fig 6.4 Airport Output

In the Fig 6.4, the objects are clumsy, moreover the suitcases were kept together indistinguishably. yet, the model was able to identify each suitcase separately and also the trucks at the corner of the image  
The end detection consists of distinctive set of bounding boxes which are displayed around the detected objects with confidence value for the given inputs.

## VI. CONCLUSION AND FUTURE SCOPE

In this project, we suggested about YOLO algorithm for the motive of objects identification using a single neural network. This algorithm is stereotype, it exceeds different strategies once stereotype from natural pictures to different domains. The algorithm is simple to construct and can be instructed directly on a complete image. Region proposal strategies limit the classifier to a particular region. YOLO accesses to the entire image in forecasting boundaries. And it forecasts fewer false positives in background areas. Comparing to other classifier algorithms this algorithm is much more well organized and fastest algorithm to use in real time. In the proposed system, we have identified several objects in a frame of a video with some accuracy. This work can be extended in a possible way to get the description of objects along with the caption. For example, car number plate detection, model description, owner name, model name for a detected car. Similarly, Bio-graphic details about a person.  
Example: Height, Skin tone, Texture details (Colours of outfit, age, gender)

## REFERENCES

- [1] J. Kang, S. Tariq, H. Oh and S. S. Woo, "A Survey of Deep Learning-Based Object Detection Methods and Datasets for Overhead Imagery," in IEEE Access, vol. 10, pp. 20118-20134, 2022, doi: 10.1109/ACCESS.2022.3149052.
- [2] Asra Aslam, Edward Curry, A Survey on Object Detection for the Internet of Multimedia Things (IoMT) using Deep Learning and Event-based Middleware: Approaches, Challenges, and Future Directions, Image and Vision Computing, Volume 106, 2021, 104095, ISSN 0262-8856, <https://doi.org/10.1016/j.imavis.2020.104095>
- [3] N. Kumaran and U. S. Reddy, "Object detection and tracking in crowd environment — A review," 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017, pp. 777-782, doi: 10.1109/ICICI.2017.8365242.
- [4] C. P. Papageorgiou, M. Oren and T. Poggio, "A general framework for object detection," Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), 1998, pp. 555-562, doi 10.1109/ICCV.1998.710772.
- [5] C. Tang, Y. Feng, X. Yang, C. Zheng and Y. Zhou, "The Object Detection Based on Deep Learning," 2017 4th International Conference on Information Science and Control Engineering (ICISCE), 2017, pp. 723-728, doi: 10.1109/ICISCE.2017.156.
- [6] X. Zhou, W. Gong, W. Fu and F. Du, "Application of deep learning in object detection," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 2017, pp. 631-634, doi: 10.1109/ICIS.2017.7960069.





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)