International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.9 July - August – 2025
DOI:10.46647/ijetms.2025.v09i04.012 ISSN: 2581-4621

Real Time Object Detection Using You Only Look Once (Yolo) Algorithm

Mrudula Dasi[1], Deepa[2]
[1]*M.Tech Student, Department of Computer Science Engineering KLR College of Engineering & Technology, Bhadradri Kothagudem (DT), Telangana*
[2]*Assistant Professor, Department of Computer Science Engineering KLR College of Engineering & Technology, Bhadradri Kothagudem (DT), Telangana*

Abstract: Use the You Only Look Once (YOLO) principle to identify items. When compared to other methods for object detection, this one provides a number of benefits. By predicting the bounding boxes using convolutional networks and the class probabilities for these boxes, YOLO detects the image faster than other algorithms like Convolutional Neural Network and Fast Convolutional Neural Network, which do not look at the image completely.
Keyword: Digital image processing, image recognition, Convolutional Neural Network, Bounding Boxes, YOLO.

I. INTRODUCTION

Object detection is a technology that detects the semantic objects of a class in digital images and videos. One of its real-time applications is self-driving cars. In this, our task is to detect multiple objects from an image. The most common object to detect in this application is the car, motorcycle, and pedestrian. For locating the objects in the image we use Object Localization and have to locate more than one object in real-time systems. There are various techniques for object detection, they can be split up into two categories, first is the algorithms based on Classifications. CNN and RNN come under this category. In this, we have to select the interested regions from the image and have to classify them using Convolutional Neural Network. This method is very slow because we have to run a prediction for every selected region. The second category is the algorithms based on Regressions. YOLO method comes under this category. In this, we won't select the interested regions from the image. Instead, we predict the classes and bounding boxes of the whole image at a single run of the algorithm and detect multiple objects using a single neural network. YOLO algorithm is fast as compared to other classification algorithms. In real time our algorithm process 45 frames per second. YOLO algorithm makes localization errors but predicts less false positives in the background.

II. LITERATURESURVEY

You Only Look Once: Unified, Real-Time Object Detection, by Joseph Redmon. Their prior work is on detecting objects using a regression algorithm. To get high accuracy and good predictions they have proposed YOLO algorithm in this paper [1]. Understanding of Object Detection Based on CNN Family and YOLO, by Juan Du. In this paper, they generally explained about the object detection families like CNN, R-CNN and compared their efficiency and introduced YOLO algorithm to increase the efficiency [2]. Learning to Localize Objects with Structured Output Regression, by Matthew B. Blaschko. This paper is about Object Localization. In this, they used the Bounding box method for localization of the objects to overcome the drawbacks of the sliding window method [3].

III. WORKING OF YOLO ALGORITHM

First, an image is taken and YOLO algorithm is applied. In our example, the image is divided as grids of 3x3 matrixes. We can divide the image into any number grids, depending on the complexity

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.9 July - August – 2025
DOI:10.46647/ijetms.2025.v09i04.012 ISSN: 2581-4621

of the image. Once the image is divided, each grid undergoes classification and localization of the object. The objectness or the confidence score of each grid is found .If there is no proper object found in the grid, then the objectness and bounding box value of the grid will be zero or if there found an object in the grid then the objectness will be 1 and the bounding box value will be its corresponding bounding values of the found object. The bounding box prediction is explained as follows. Also, Anchor boxes are used to increase the accuracy of object detection which also explained below in detail.

Figure1: Working of YOLO

Bounding box predictions:

YOLO algorithm is used for predicting the accurate bounding boxes from the image. The image divides into S x S grids by predicting the bounding boxes for each grid and class probabilities. Both image classification and object localization techniques are applied for each grid of the image and each grid is assigned with a label.

Then the algorithm checks each grid separately and marks the label which has an object in it and also marks its bounding boxes. The labels of the gird without object are marked as zero.

Figure2: Example image with 3x3 grids

Consider the above example, an image is taken and it is divided in the form of 3 x 3 matrixes. Each grid is labeled and each grid undergoes both image classification and objects localization techniques. The label is considered as Y. Y consists of 8 values.

y =	pc
	bx
	by
	bh
	bw
	c1
	c2
	c3

Figure3: Elements of label Y

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.9 July - August – 2025
DOI:10.46647/ijetms.2025.v09i04.012 ISSN: 2581-4621

Pc– Represents whether an object is present in the grid or not. If present pc=1 else 0.
bx, by, bh, bw – are the bounding boxes of the objects (if present).
c1, c2, c3 – are the classes. If the object is a car then c1and c3 will be 0 and c2 will be 1.
In our example image, the first grid contains no proper object. So it is represented as,

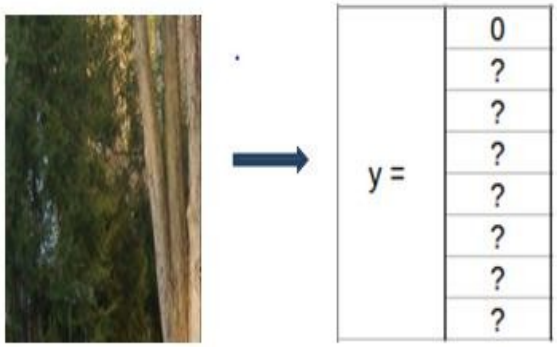

Figure4: Bounding box and Class values of grid 1.

In this grid , there exists no proper objects other value is 0.
And rest of the values are doesn't matter because there exist no object. So, it is represented as ?.
Consider a grid with the presence of an object. Both 5th and 6th grid of the image contains an object.
Let' consider the 6th grid, it is represented as.

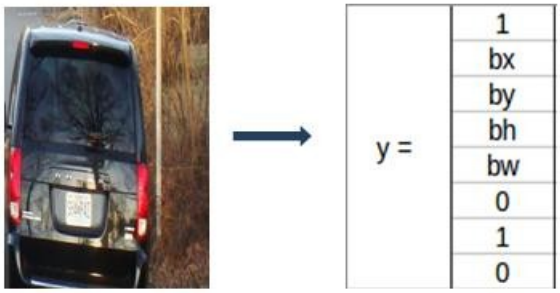

Figure5: Bounding box and Class values of grid 6.

In this table, 1 represents the presence of an object. And bx, by, bh, b ware the bounding boxes of the object in the 6th grid. And the object in that grid is a car so the classes are (0,1,0). The matrix form of Y in this is Y=3x3x8.
For the 5th grid also the matrix will be little similar with different bounding boxes by depending on the objects position in the corresponding grid.
If two or more grids contain the same object then the center point of the object is found and the grid which has that point is taken. For this, to get the accurate detection of the object we can use to methods .They are Intersection over Union and Non-Max Suppression. In IoU, it will takes the actual and predicted bounding box value and calculates the IoU of two boxes by using the formulae,
IoU =Area of Intersection/Area of Union.
If the value of IoU is more than or equal to our threshold value (0.5) the not a good prediction. The threshold value is just an assuming value. We can also take greater threshold value to increase the accuracy or for better prediction of the object.
The other method is Non-max suppression, in this, the high probability boxes are taken and the boxes with high IoU are suppressed. Repeat this until a box is selected and consider that as the bounding box for that object.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.9 July - August – 2025
DOI:10.46647/ijetms.2025.v09i04.012 ISSN: 2581-4621

ACCURACY IMPROVEMENT

ANCHOR BOX:

By using Bounding boxes for object detection, only one object can be identified by a grid. So, for detecting more than one object we go for Anchor box.

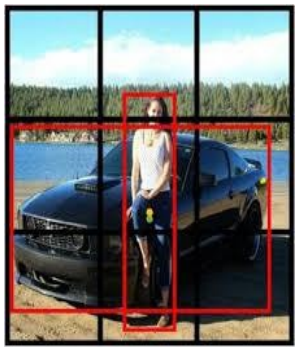

Figure6: An example image for anchor box

Consider the above picture, in that both the human and the car's midpoint come under the same grid cell. For this case, we use the anchor box method. The red color grid cells are the two anchor boxes for those objects. Any number of anchor boxes can be used for a single image to detect multiple objects. In our case, we have taken two anchor boxes.

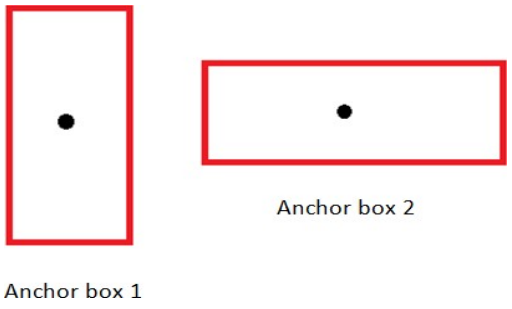

Figure7: Anchor boxes

The above figure represents the anchor box of the image we considered .The vertical anchor box is for the human and the horizontal one is the anchor box of the car.

In this type of overlapping object detection, the label Y contains16values i.e, the values of both anchor boxes.

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 1 \\ 0 \\ 0 \\ 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Anchor box 1
Human

Anchor box 2
Car

Figure8: Anchor box prediction values

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.9 July - August – 2025
DOI:10.46647/ijetms.2025.v09i04.012 ISSN: 2581-4621

Pc in both the anchor box represents the presence of the object.
bx,by,bh,b win both the anchor box represents their corresponding bounding box values.
The value of the class in anchor box 1 is (1, 0, 0) because the detected object is a human.
In the case of anchor box 2, the detected object is a car so the class value is (0, 1, 0).
In this case, the matrix form of Y will be Y= 3x3x16 or Y= 3x3x2x8. Because of two anchor box, it is 2x8.

IV. RESULTS & DISCUSSIONS

The idea of YOLO is to make a Convolutional neural network to predict a (7,7, 30) tensor. It uses a Convolutional neural network to scale back the spatial dimension to 7x7 with 1024 output channels at every location. By using two fully connected layers it performs a linear regression to create a 7x7x2 bounding box prediction. Finally, prediction is made by considering the high confidence score of a box.

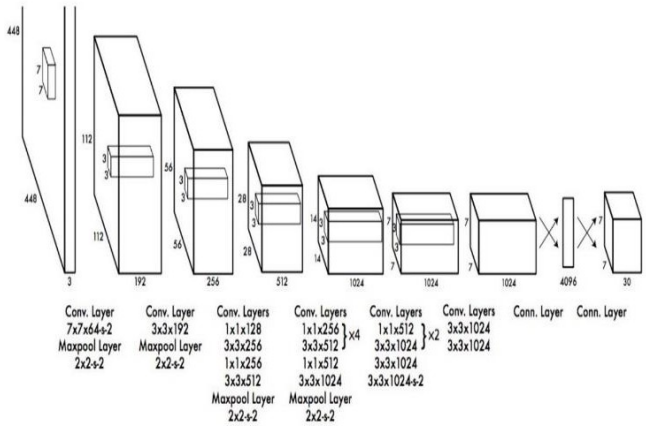

Figure9: CNN Network Design

4.1 .Loss function of YOLO algorithm:

For a single grid cell, the algorithm predicts multiple bounding boxes. To calculate the loss function we use only one bounding box for object responsibility. For selecting one among the bounding boxes we use the high IoU value. The box with high IoU will be responsible for the object. Various loss functions are:

- Classification loss function
- Localization loss function
- Confidence loss function

Localization loss means the error between the ground truth value and predicted boundary box. Confidence loss is the objectness of the box. Classification loss calculated as, the squared error of the class conditional probabilities for each class:

$$\sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

Equation1: Conditional probabilities for each class
Where,
in Equation 1, If it is 1 means the object appears in the $\mathbb{1}_i^{\text{obj}}$1, or else it is 0.
$\hat{p}_i(c)$ is the conditional class probability for class c.
The localization loss is the measure of errors in the predicted boundary box locations and the sizes. The box which is responsible for the object is only counted.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.9 July - August – 2025
DOI:10.46647/ijetms.2025.v09i04.012 ISSN: 2581-4621

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

Equation2: The localization loss

Where,
$\mathbb{1}_{ij}^{obj}$

In Equation2,Is1,if the jth bounding box of celli 4. Wei Liu, Dragomir Anguelov, Dumitru Erhan, "SSD: Single Shot Multi Box Detector", Published in Computer Vision – ECCV 2016 pp 21-37.
is responsible for detecting the object. Otherwise, it is 0.
In λ_{coord} crase the weight for the loss of bounding box coordinates.
The Confidence loss, if the object is found in a box the confidence loss is,

$$\sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

Equation3: The Confidence loss

Where,
i$\hat{C}_i$quation3,Is the confidence score of the box j in
celli.
$\mathbb{1}_{ij}^{obj}$ the jth bounding box of cell i is responsible for detecting the object. Otherwise, it is 0.
If the object is not detected then the confidence loss will be,

$$\lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

Equation4: Confidence loss if object not detected

Where,
$\mathbb{1}_{ij}^{noobj}$uation 4, Is the complement of $\mathbb{1}_{ij}^{obj}$. Is the confidence score of box j in cell i.
Are the weights down the loss when detecting the back ground.
$\hat{C}_i$

λ_{noobj}

V. CONCLUSION

In this paper, we proposed the aim of object detection using a single neural network; we presented the YOLO method in this article. When applied to new domains, our generalized approach continues to outperform competing algorithms that were trained on natural images. The method requires little development time and can be trained using a whole picture as input. Strategies that focus on proposing regions restrict the classifier to a certain area. YOLO uses the whole picture to make boundary predictions. In addition, it makes less inaccurate predictions in unlabelled regions. This approach is much quicker and more efficient than competing classification algorithms when used in real-time.

DOI:10.46647/ijetms.2025.v09i04.012 ISSN: 2581-4621

REFERENCES

1. 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.
2. YOLO Juan Du1,"Understanding of Object Detection Based on CNN Family", New Research, and Development Center of Hisense, Qingdao 266071, China.
3. Matthew B. Blaschko Christoph H. Lampert, "Learning to Localize Objects with Structured Output Regression", Published in Computer Vision – ECCV 2008 pp 2-15.
4. 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.
5. Shaoqing Ren,Kaiming He, RossGirshick, 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).
6. Joseph Redmon, Ali Farhadi,"YOLO9000: Better, Faster, Stronger", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263-7271.
7. 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).
8. Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", published in Computer Vision and Pattern Recognition (cs.CV).
9. P. Nagamani, Medikonda Asha Kiran, Mahesh Raj, Racharla Vyshnav Mani Teja, RB Pittala, Manyam Thaile, and Lakshmi Prasanna Byrapuneni, "VisuaStat: Visualizing Data and Simplifying Decisions," 2025 International Conference on Computing and Communication Technologies (ICCCT), Chennai, India, 2025, pp. 1-6, doi: 10.1109/ICCCT63501.2025.11019270.
10. R. B. Pittala, B. R. Tejopriya, and E. Pala, "Study of speech recognition using CNN," in Proc. 2nd Int. Conf. Artif. Intell. Smart Energy (ICAIS), Coimbatore, India, Feb. 23–25, 2022, pp. 150–155.
11. B. R. Krishna, M. N. Rao, and R. B. Pittala, "An algorithm to find the geo-map by multimedia communication," Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 5, pp. 1245–1250, 2019.
12. R. B. Pittala, M. Nagabhushana Rao, and M. Shiva Kumar, "Discovering the knowledge to find the affected areas of a plague for taking accurate decision," in Proc. 2016 IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Chennai, India, Dec. 15–17, 2016, Art. no. 7919723.
13. V. S. Pittala, P. Cheraku, P. Somaraju, R. B. Pittala, P. Nagababu, and K. Gopi, "High-dimensional data-driven pneumonia diagnosis using ANFIS," J. Theor. Appl. Inf. Technol., vol. 102, no. 12, pp. 5044–5051, Jun. 30, 2024.
14. B. Ramakrishna, M. Nagabhushana Rao, and R. B. Pittala, "Importance of social media in emergency communication networks: A survey," J. Adv. Res. Dyn. Control Syst., vol. 10, no. 7, pp. 1715–1720, 2018.
14. B. Ramakrishna, M. Nagabhushana Rao, and R. B. Pittala, "Low cost geo distributed data centers with big data process," J. Adv. Res. Dyn. Control Syst., vol. 10, no. 7, pp. 394–397, 2018.