IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 2, February 2021

DOI 10.17148/IJIREEICE.2021.9210

VISUAL PERCEPTION USING NEURAL NETWORKS

Velmurugan.V[1], Harish Gowtham.S[2], Gowtham.G[3], Balaji.S[4]

Assistant Professor, Dept. of ECE, Agni College of Technology Chennai, Tamilnadu, India[1]

UG Student, Dept. of ECE, Agni College of Technology Chennai, Tamilnadu, India[2,3,4]

Abstract: The Objective is to detect of objects using You Only Look Once (YOLO) approach. This method has several advantages as compared to other object detection algorithms. In other algorithms like Convolutional Neural Network, Fast Convolutional Neural Network the algorithm will not look at the image completely but in YOLO the algorithm looks the image completely by predicting the bounding boxes using convolutional network and the class probabilities for these boxes and detects the image faster as compared to other algorithms.

Keywords: Object Detection, Neural Networks, YOLO Algorithm.

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. LITERATURE SURVEY

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. Learning to localize objects with structured output regression by M.B.Blaschko and C.H.Lampert. In Computer Vision and Pattern Recognition (CVPR) It defines Fast, accurate detection of 100,000 object classes on a single machine by . T. Dean, M.Ruzon, M.Segal, J.Shlens, S.Vijayanarasimhan, J.Yagnik .Towards unified object detection and semantic segmentation by J.Dong, Q.Chen, S.Yan, and A.Yuille.

III.WORKING OF YOLO ALGORITHM

YOLO was proposed by Joseph Redmond et al. in 2015. It was proposed to deal with the problems faced by the object recognition models at that time, Fast R-CNN is one of the state-of-the-art models at that time but it has its own challenges such as this network cannot be used in realtime, because it takes 23 seconds to predicts an image and therefore cannot be used in real-time. 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 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.

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 2, February 2021
DOI 10.17148/IJIREEICE.2021.9210

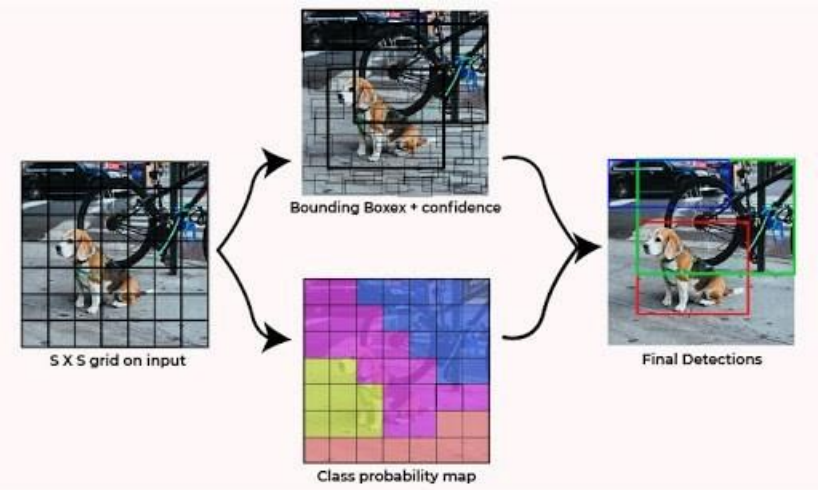

Fig 3.1 Working of YOLO

3.1. 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.

3.2. 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.

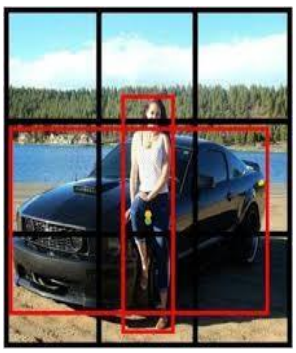

Fig 3.2 Example Image for Anchor Box

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.

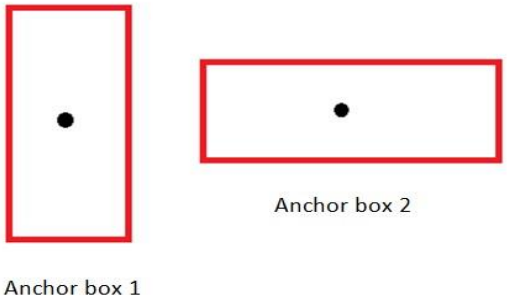

Fig 3.3 Anchor boxes

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 9, Issue 2, February 2021
DOI 10.17148/IJIREEICE.2021.9210

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 contains 16 values 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

Fig 3.4 Anchor box prediction values

Pc in both the anchor box represents the presence of the object. bx, by, bh, bw in 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 AND DISCUSSIONS

It uses a Convolutional neural network to scale back the spatial dimension to 7x7 with 1024 output channels at every location. fully connected layers it performs a linear regression to create a 7x7x2 bounding box prediction. Finally, a prediction is made by considering the high confidence score of a box.

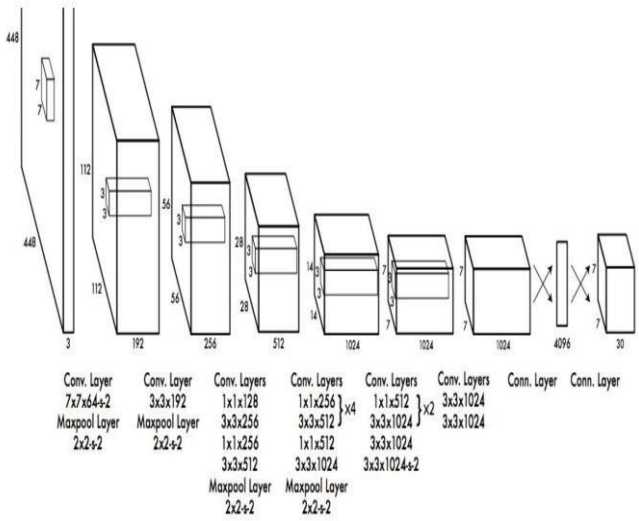

Fig 4.1 CNN Network Design

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:

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 2, February 2021

DOI 10.17148/IJIREEICE.2021.9210

- 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:

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

Equation 1: Conditional probabilities for each class

Where,

$\mathbb{1}_i^{\text{obj}}$ in Equation 1, If it is 1 means the object appears in the cell, 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.

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

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

Equation 2: The localization loss

Where,

$\mathbb{1}_{ij}^{obj}$ in Equation 2, Is 1, if the jth bounding box of cell i is responsible for detecting the object. Otherwise, it is 0.

λ_{coord} Increase the weight for the loss of bounding box coordinates.

The Confidence loss, if the object is found in a box the confidence loss is,

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

Equation 3: The Confidence loss

Where,

$\hat{C}_i$ in Equation 3, Is the confidence score of the box j in cell i.

$\mathbb{1}_{ij}^{obj}$ Is 1 if 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,

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

Equation 4: Confidence loss if object not detected

Where, in Equation 4,

$\mathbb{1}_{ij}^{noobj}$ Is the complement of $\mathbb{1}_{ij}^{obj}$.

$\hat{C}_i$ Is the confidence score of box j in cell i.

λ_{noobj} Is the weights down the loss when detecting the background.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 2, February 2021

DOI 10.17148/IJIREEICE.2021.9210

V. CONCLUSION

In this paper, we proposed about YOLO algorithm for the purpose of detecting objects using a single neural network.This algorithm is generalized, it outperforms different strategies once generalizing from natural pictures to different domains. The algorithm is simple to build and can be trained directly on a complete image.Region proposal strategies limit the classifier to a particular region. YOLO accesses to the entire image in predicting boundaries. And also it predicts fewer false positives in background areas.Comparing to other classifier algorithms this algorithm is much more efficient and fastest algorithm to use in real time.

REFERENCES

1. M.B.Blaschko and C.H.Lampert,"Learning to localize objects with structured output regression". In Computer Vision– ECCV 2008,
2. T. Dean,M.Ruzon, M.Segal, J.Shlens, S.Vijayanarasimhan,J.Yagnik, et al. Fast, accurate detection of 100,000 object classes on a single machine. In "Computer Vision and Pattern Recognition (CVPR)", 2013 IEEE Conference on IEEE 2013,
3. J.Dong, Q.Chen,S.Yan, and A.Yuille."Towards unified object detection and semantic segmentation". In Computer Vision–ECCV 2014
4. Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse,part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26, doi:10.1109/TPAMI.2004.108
5. Alexe, B., Deselaers, T., and Ferrari, V. (2010). "What is an object?," in ComputerVision and Pattern Recognition (CVPR), 2010 IEEE Conference on (San Francisco,CA: IEEE),. doi:10.1109/CVPR.2010.5540226
6. Andreopoulos, A., and Tsotsos, J. K. (2013). 50 years of object recognition: direc-tions forward. Comput. Vis. Image Underst. 117, doi:10.1016/j.cviu.2013.04.005
7. Azizpour, H., and Laptev, I. (2012). "Object detection using stronglysupervisedeformable part models," in Computer Vision-ECCV 2012 (Florence: Springer).
8. Bengio, Y. (2012). "Deep learning of representations for unsupervised and transferlearning," in ICML Unsupervised and Transfer Learning, Volume 27 of JMLRProceedings, eds I. Guyon, G. Dror, V. Lemaire, G. W. Taylor, and D. L. Silver(Bellevue: JMLR.Org).
9. Bourdev, L. D., Maji, S., Brox, T., and Malik, J. (2010). "Detecting peopleusing mutually consistent poselet activations," in Computer Vision – ECCV2010 – 11th European Conference on Computer Vision, Heraklion, Crete, Greece,September 5-11, 2010, Proceedings, Part VI, Volume 6316 of Lecture Notes inComputer Science, eds K. Daniilidis, P. Maragos, and N. Paragios (Heraklion:Springer).
10. Bourdev, L. D., and Malik, J. (2009). "Poselets: body part detectors trained using 3dhuman pose annotations," in IEEE 12th International Conference on ComputerVision, ICCV 2009, Kyoto, Japan, September 27 – October 4, 2009 (Kyoto: IEEE).