IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 5, April 2024

YOLO-V2
(You Only Look Once)

Ms. N P Mohod[1], Mr. Nikhil Patil[2], Mr. Ashutosh Shrungare[3], Mr. Om Labhasetwar[4], Mr. Abhijeet Ghadge[5], Mr. Sumit Bade[6]

Assistant Professor, Department of Computer Science of Engineering[1]
Students, Department of Computer Science of Engineering[2,3,4,5,6]
SIPNA College of Engineering, Amravati, India

Abstract*: The you-only-look-once (YOLO) v2 object detector uses a single stage object detection network. YOLO v2 is faster than other two-stage deep learning object detectors, such as regions with convolutional neural networks (Faster R-CNNs).The YOLO v2 model runs a deep learning CNN on an input image to produce network predictions. The object detector decodes the predictions and generates bounding boxes YOLO v2 uses anchor boxes to detect classes of objects in an image. For more details, see Anchor Boxes for Object Detection. The YOLO v2 predicts these three attributes for each anchor box:* ***Intersection over union (IoU)*** *— Predicts the objectness score of each anchor box.* ***Anchor box offsets*** *— Refine the anchor box position.* ***Class probability*** *— Predicts the class label assigned to each anchor box. The figure shows predefined anchor boxes (the dotted lines) at each location in a feature map and the refined location after offsets are applied. Matched boxes with a class are in color. You can design a custom YOLO v2 model layer by layer. The model starts with a feature extractor network, which can be initialized from a pretrained CNN or trained from scratch. The detection subnetwork contains a series of Conv, Batch norm, and ReLu layers, followed by the transform and output layers, yolov2TransformLayer and yolov2OutputLayer objects, respectively.yolov2TransformLayertransforms the raw CNN output into a form required to produce object detections.yolov2OutputLayerdefines the anchor box parameters and implements the loss function used to train the detect.*

Keywords: R-CNN, YOLOv2, Object classification, Object detection, F-CNN

I. INTRODUCTION

Yolo v2: You Only Look Once

YOLOv2 is the second version in the YOLO family, significantly improving accuracy and making it even faster. The improved YOLOv2 model used various novel techniques to outperform state-of-the-art methods like Faster-RCNN and SSD in both speed and accuracy. One such technique was multi-scale training that allowed the network to predict at varying input sizes, thus allowing a trade-off between speed and accuracy.

At 416×416 input resolution, YOLOv2 achieved 76.8 mAP on VOC 2007 dataset and 67 FPS on Titan X GPU. On the same dataset with 544×544 input, YOLOv2 attained 78.6 mAP and 40 FPS.

They also proposed the YOLO9000 model trained on the COCO detection dataset with the ImageNet classification dataset. The idea behind this type of training was to detect object classes that did not have ground truth for object detection but use supervision from object classes with ground truth. The domain for such training is referred to as weakly supervised learning. This approach helped them achieve 16 mAP on 156 classes that did not have detection ground truth.

In terms of speed, YOLO is one of the best models in object recognition, able to recognize objects and process frames at the rate up to 150 FPS for small networks. However, in terms of accuracy mAP, YOLO was not the state-of-the-art model but has fairly good Mean average Precision (mAP) of 63% when trained on PASCAL VOC2007 and PASCAL VOC 2012. However, Fast R-CNN which was the state of the art at that time has anmAP of 71%. YOLO v2 and YOLO 9000 was proposed by J. Redmon and A. Farhadi in 2016 in the paper titled YOLO 9000: Better, Faster, Stronger. At 67 FPS, YOLOv2 gives mAP of 76.8% and at 67 FPS it gives anmAP of 78.6% on VOC 2007 dataset bettered the models like Faster R-CNN and SSD. YOLO 9000 used YOLO v2 architecture but was able to detect more than 9000 classes. YOLO 9000, however, has anmAP of 19.7%.

II. LITERATURE REVIEW

Real-Time Object Detection with YOLO, by Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam. In this paper, their work is to detect multiple objects from an image using YOLO approach [1]. You Only Look Once: Unified, RealTime Object Detection, by Joseph Redmon. This paper explains object detection as regression problem and repurposes classifier using YOLO approach [2]. Object Detection and Recognition in Images, by Sandeep Kumar, Aman Balyan, Manvi Chawla. This paper used Easynet model to recognize images and detection of objects for instances of real objects like bicycles, fruits, animals and buildings in images [3]. Object Detection and Classification Algorithms using Deep Learning for video Surveillance Applications, by Mohana and H. V. Ravish Aradhya. This paper prior work is the classification of objects in images and video, have use YOLOv2 approach [4].

III. WORKING OF YOLOV2 ALGORITHM

Step 1- An image is taken and divide it into a grid cell. Here, example has taken where the image splits into grids of 7x7 matrices. It will divide the image into any number of grids, looking on the complexity of the image.

Step 2- Once the image is split, classification and localization of the image is performed in each grid cell. If an object is detected then it represents the probability of each grid vector. The output of this is the dimension of bounding box and class.

Step 3- Now, thresholding is performed and based on the value grid cells with the highest probabilities are picked. This step produces the removal of bounding boxes which doesn't have object or the confidence score less than a threshold of 0.35.

STEP 4-The yolo v2 algorithm uses anchor boxes which detect the objects in single grid cell and give the location of object. finally, non-max suppression uses intersection over union for final detection.

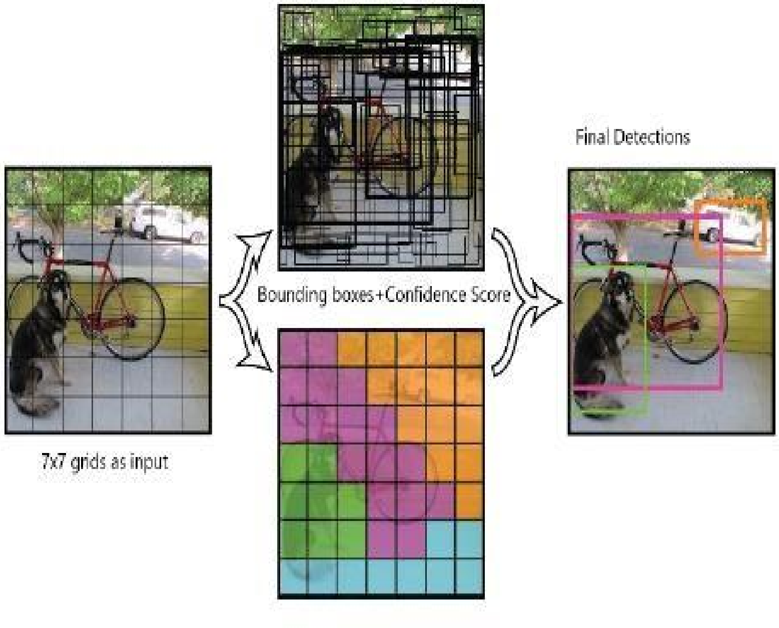
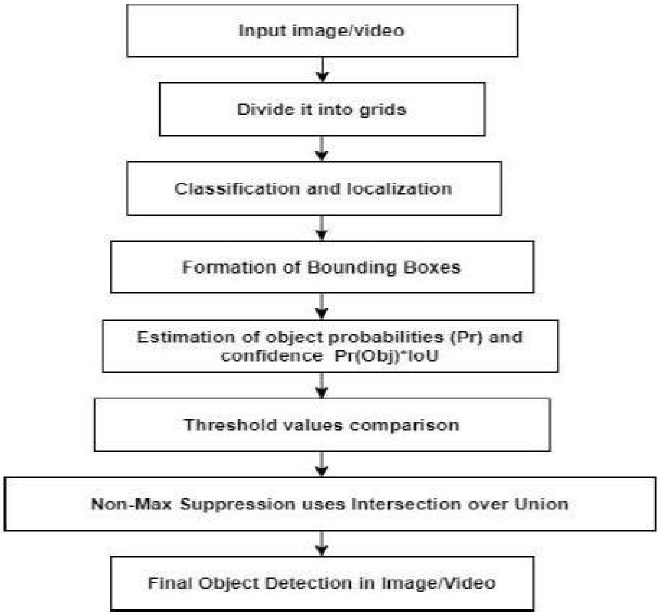


Fig. 2. Working of YOLOv2

In fig 2. Shows the working of the algorithm which divides the image into SxS grid cell and every grid cell predicts bounding boxes and class probabilities which is mapped in different colors [2].

DOI: 10.48175/IJARSCT-17515

. Flow Diagram of YOLOv2 Model


Input image/video
Divide it into grids
Classification and localization
Formation of Bounding Boxes
Estimation of object probabilities (Pr) and confidence Pr(Obj)*IoU
Threshold values comparison
Non-Max Suppression uses Intersection over Union
Final Object Detection in Image/Video

IV. ANALYSIS OF PROBLEMS

Challenges in Object Detection

In object detection, the bounding boxes are always rectangular. As a result, if the object contains the curvature part, it does not help determine its shape. In order to find precisely the shape of the object, we should use some of the image segmentation techniques.

Some non-neural methods may not detect objects with high accuracy or may produce a large number of falsepositive detections. Although neural network methods are more accurate, there are some drawbacks. For example, they require a large amount of annotated data for training. Training is often expensive in time and space and, as a result, prolonged on standard computers.

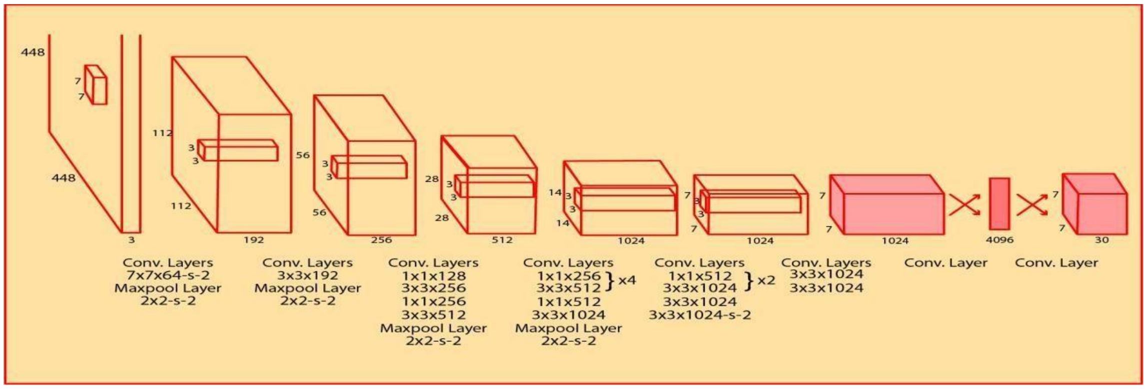
In order to solve these challenges, we can use the YOLO algorithm. Thanks to the transfer learning capabilities, we would be able to use already pre-trained models or spend some time fine-tuning models with our data. Furthermore, the YOLO algorithm is one of the most popular methods for performing object detection in real-time because it achieves high accuracy on most real-time processing tasks while maintaining a reasonable speed and frames per second, even on devices accessible to almost everyone.

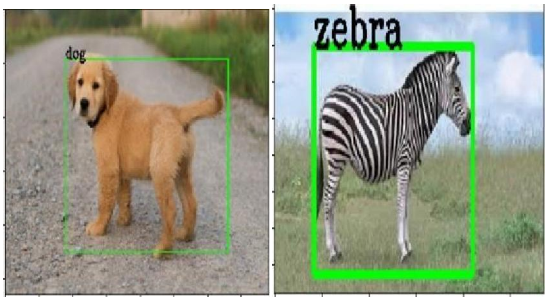
V. ARCHITECTURE OF YOLOv2

YOLO v2 is trained on different architectures such as VGG-16 and GoogleNet. The paper also proposed an architecture called Darknet-19. The reason for choosing the Darknet architecture is its lower processing requirement than other architectures *5.58 FLOPS*(as compared to 30.69 FLOPS on VGG-16 for *224 * 224* image size and *8.52 FLOPS* in customized GoogleNet). The structure of Darknet-19 is given below:

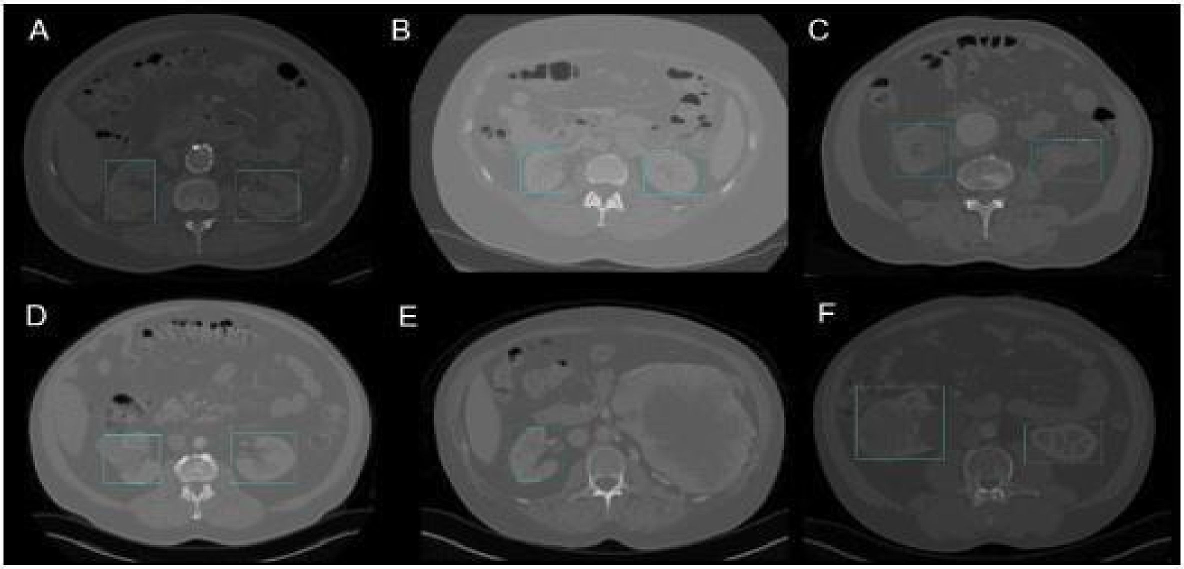
For detection purposes, we replace the last convolution layer of this architecture and instead add three *3 * 3* convolution layers every *1024 filters* followed by *1 * 1* convolution with the number of outputs we need for detection.

For VOC we predict 5 boxes with 5 coordinates (t_x, t_y, t_w, t_h, t_o (objectness score)) each with *20 classes* per box. So total number of filters is 125.





Detection and labelling of single image



2D Kidney detection by YOLOv3 (Image from Kidney Recognition in CT using YOLOv3)

VI. CONCLUSION

This paper proposes YOLOv2 algorithm for the detection of objects in images with localization and video records. The main aim of this paper is to detect the objects in real time i.e. live detection using webcam and also through video records. GPU version is extremely fast which helps the functionalities perform accurate using anchor boxes. The dataset used in this paper is COCO which consists 80 classes. Using the model YOLOv2 it is easy to detect objects with grids and boundaries prediction and also it helps in predicting with very small objects or objects which very far in the image. In video records detection of moving objects are easier using darknet and it produces .avi file with detections. In live detection system uses webcam to detect live objects. Pretrained datasets helped to detect in efficient way and classifying the objects in less time.

REFERENCES

[1]. Redmon Joseph, et al. "You only look once: Unified, real-time object detection." proceedings arXiv in May 2016.

[2]. Geethapriya. S, et al. "Real-Time Object Detection with Yolo" proceedings of the International Journal of Engineering and Advanced Technology (IJEAT) inFeb 2019

[3]. Swetha M S, et al. "Object Detection and Classification in Globally Inclusive Images Using Yolo" proceedings of the International Journal of Advance Research in Computer Science and Management Studies (IJARCM) in Dec 2018

[4]. Keerthana T, et al. "A REAL TIME YOLO HUMAN DETECTION IN FLOOD AFFECTED AREAS BASED ON VIDEO CONTENT ANALYSIS" proceedings of the International Research Journal of Engineering and Technology (IRJET) in Jun 2019

[5]. Sandeep Kumar, et al. "Object Detection and Recognition in Images" proceedings of the International Journal of Engineering Development and Research (IJEDR)in 2017

[6]. M R Sunitha, etal. "A Survey on Moving Object Detection and Tracking Techniques" proceedings of the International Journal of Engineering and Computer Science (IJEAC) in 2016.

[7]. Jifeng Dai, et al. "R-FCN: Object Detection via Region-based Fully Convolutional Networks", proceeding of the Advances in Neural Information Processing Systems 29 (NIPS) in 2016.

DOI: 10.48175/IJARSCT-17515