© October 2023| IJIRT | Volume 10 Issue 5 | ISSN: 2349-6002

Real-Time Object Detection using DL with YOLO V7

S.Samyukta[1], D.M.V.Priya[2]
[1]Assistant Professor, Baba Institute of Technology and Science, Visakhapatnam.
[2]Assistant Professor, Gayatri Vidya Parishad, Visakhapatnam.

***Abstract*— In this article, it will shows that how human life is getting much easier day by day. With the use of intelligent technologies such as mobile applications, humans have a more flexible lifestyle that allows them to stay anywhere in the world. Real-time object detection using YOLO just look once. YOLO is one of the most widely used techniques in deep learning for pervasive real-time object detection. Since YOLOv5 and YOLOv7 both belong to his YOLO family, we included the empathy between them in his COCO dataset. Because it is one of the better, more accurate and faster algorithms with advanced DL techniques for object detection. Object recognition is one of the most difficult tasks in computer vision. The purpose is to recognize objects more accurately. From this, he found YOLOV7 to be the fastest and most accurate compared to YOLOV5 on pre-trained images.**

***Index Terms*—API, SSD, FRCNN, CNN, RCNN, YOLO (yolo only look once),DL(deep learning).**

I. INTRODUCTION

Object recognition using computer vision is becoming a challenging task these days. It used to be difficult for humans to identify and locate the object. It used to take them a lot of time to recognize and classify the object. YOLO is used here, which belongs to one-stage detection with high performance. The image is used for single objects detected by a grid bounding box. For object detection we used the YOLOV7 technique (you only look once) with FCN and Open CV[1]. To identify speed bumps, pits, barriers and obstacle objects with pre-trained image. We do a classification based on the single pass of the image SSD, then the category is displayed at the top of the bounding box. Classify images with YOLOV7. Because it was easier and faster to be precise. This allows users to find safe driving roads. Classify Images: Images are predicted based on pre-trained images[3]. Localization of images: Where the object is in the image is located using the bounding box. Recognizing Images: After recognizing the object, there will be a category with the corresponding class object.

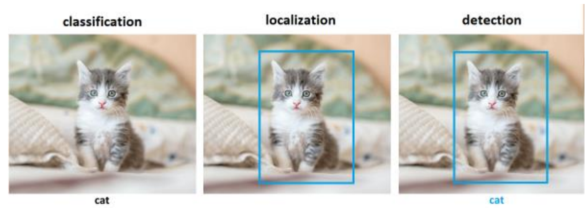

FIG:1.1 IMAGE CLASSIFICATION AND LOCALIZATION

II. WORKING OF YOLO

The YOLO algorithm divides the image into NXN grids. Each image has an equal dimensional area with height and width. These NXN grids are responsive and help detect and locate the object within. These grids extract the feature object and predict the B-bounding box using their cell coordinates along with the object label and the probability that the object is present in the cell by computing the parameters[7]. This process plays a big role as both detection and detection of cells are taken from the image, but it leads to many duplicate predictions as multiple cells predict the same object with different bounding box predictions. With non-maximal suppression, the lower probability is suppressed[5].

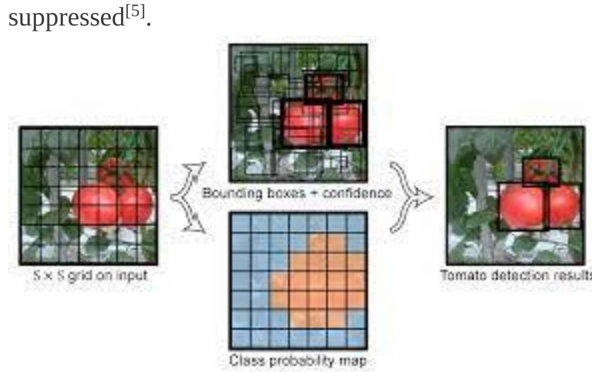

FIG: 2.1 FEATURE EXTRATION

A. Residual blocks

This method helps in implementing grids on capture images. The grids are formed on the input image. In form of SxS dimension grids. This helps in detecting the object by subtracting unwanted data from the input picture.

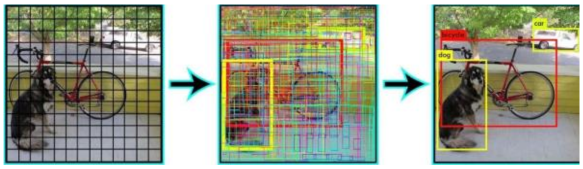

FIG: 2.2 RESIDUAL BLOCK

B. Bounding box

The detected object is instance and highlights with rectangular box. With this parameter like Class (object), x, y (centre of the rectangular box), width, height with help of these bounding box was form.

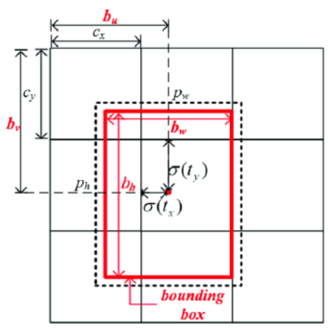

FIG: 2.3 BOUNDING BOX

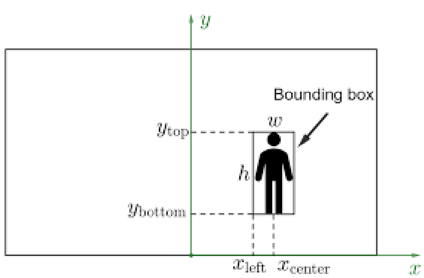

FIG: 2.4 PARAMETER CALCULATIONS

C. Anchor box

Anchor box is used to detect multi object in one single image then it will create multi bounding box to that images even small objects can be predicted with aspect ratio of class object. The anchor boxes bases on k-means configure.

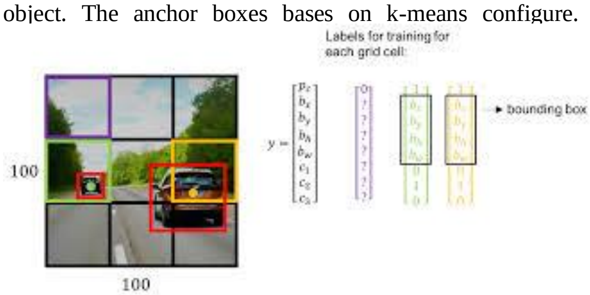

FIG: 2.5 CENTER PIONT OF IMAGE

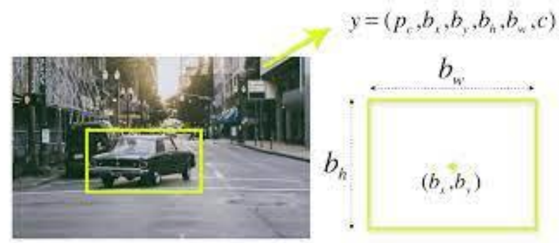

FIG:2.6 OBJECT DETECTION

D. IOU

The bounding box which overlaps it will be predicting the object in correct manner. It will eliminate unequal part. Confidence score is predicted by ground truth value. It will be represented by confidence scores by 1 when it predicts the object in correct box. Otherwise it was represented by 0.

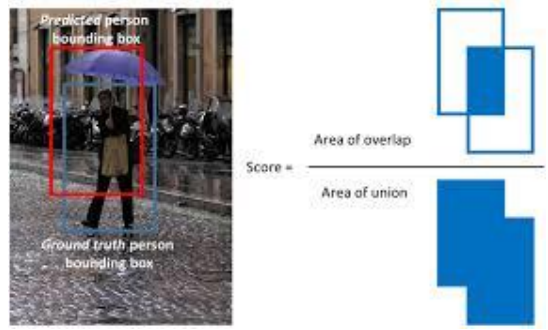

FIG:2.7 IOU

E. Transfer learning

Transfer learning which help to use the pre trained model to extract the feature of the object. With this we can detect any kind of object for the image.

F. Cost function:

When we find complicate to predict the object then cost function is used. Which measure the performance between input and output data.

III. LITERATURE SURVEY

Gowsikraja P et.al [1] proposed that object detection was done by using Haar Cascade classifier is one of the machine learning algorithm it will extract the feature based on size and position on both positive as well as negative training dataset. It will classify the object based on characteristic of object. It was also says about conditional random field and markov field.

Yuan Zhang et.al [2] proposed that YOLO and RAR(traffic congestion problem occur in our urban and cities area due to heavy traffic the driver can't able to recognize the object in front of him so to avoid the fatigue of driver HOG algorithm, Adaboost algorithm, Haar Cascade are used to make the drive safe.

Dawn Wilson et.al [3] proposed that object detection uses the YOLO algorithm with CNN as the supporting deep learning neural network. The CNN layers are arranged in a sorted 3D (dimension) form based on width, height and depth. So CNN is less efficient than YOLO. It consisted of a Spatial Pyramid Pooling (SPP) module and a Path Aggregation Network (PAN).

Geethapriya. S et.al [4] proposed compassion between CNN and FRCNN using YOLO algorithm for detecting real time objects denotation the object class. The FRCNN which comes under two stage so it is more efficient than one stage of object detection.

Peiyuan Jiang et.al [5] proposed the compassion between various Yolo versions.as Yolo was found as efficiency, fast and accurate than other techniques of deep learning. Target image is recognized based on method by feature selection and feature extraction. It was done by was done by regression problems with pre trained images target images are classified.

Ziad El Khatib et.al [6] proposed the yolo neural network with GPU based lora enable cube sat neural using CUDA (Compute Unified Device Architecture) framework was used for classifying the image. Parameter used here epoch and mean average precision point –to –point object detection using FCNN which was the second stage of the yolo.

Mannem Ponika et.al. [7] proposed compassion between traditional method and deep learning method. CNN is less efficient than yolo version human can able to recognise but not accurately and efficiently where machine learning is interconnected with deep learning SPP,RCNN,FCNN in second stage of detecting object. First stage object detection SSD single pass image for the purpose of detecting the instance and localised, classification the images.

Md. Bahar Ullah et.al. [8] proposed that CPU based real time object detection using non GPU which comes under non neural approach defines feature techniques SIFT,HOG,method which describe about the template colour match, shape non neural network is less efficient then neural network.

Abhinandan Tripathi et.al.[9] this was proposed as comparison between traditional model of detecting images and deep learning object detection. The traditional methods are FRCNN,CNN,RCNN,etc where as deep learning object detection techniques are YOLO, SSD.CNN is more accurate and speed too when compare with traditional methods. Yolo is more accurate than CNN.

Jiayi Fan et.al.[10] proposed as comparison between faster RCNN and YOLOV2 real time object detection RCNN comes under two stage detection of the object by fusing between RCNN andYOLOV2. Feature extraction can be done by two method of networks classification network and other one is RPN (Region Proposal Network).

Milad Ahmadiet.al.[11] proposed video detection image on BFAN (blur-aid feature aggregation networkit is one of the good feature collector network. with high-precision moving object detection has become a major issue. It can detect multiple objects in form of light process. Moved towards IOT &sensors.

Jing Tao et.al. [12] proposed that OYOLO(Optimized YOLO) is more accurate than yolo and RFCN incase of pre-processing used histogram of gradient image of grey scale number of pedestrians crossing a line. A type of FPGA with high resolution camera video pictures the yolo v3 was inspiration of FPN(feature pyramid network and Darknet-53 inaccurate position and recall and precision area which is Google net is superior than vgg16.

Chengji Liu[13]proposed that removing noise blur jitter yolo network was used and average precision mean calculate the point in the weighted means of precision at each threshold of traffic scenes. dataset included are ImageNet ,COCO,VOC, coordinated points of Gaussian noise image degradation images contain traffic sign dataset. Candidate box and feature extraction and feature selection to input image. Which predict the traffic sign label as class object are detected the output based on localization.

IV. Proposed system

The goal of object detection is to find an object and precisely locate it in a given image in terms of width and height, and mark the object with a rectangular shape called a bounding box. The field has been labeled with the class category corresponding to the object. More specifically, the problem of object detection is to identify and predict the right object. Solving is about determining where the object is and what it is. Therefore, it is not easy to identify and solve object problems. Unlike the human eye, a computer processes images in two dimensions. In addition, the size of the object, its orientation in space, its pose and its position in the picture vary. Technically, the above task involves image recognition and positioning. It is about 88% smaller than YOLOv4 (27MB vs. 244MB). It is about 180% faster than YOLOv4 (140 FPS vs. 50 FPS). It's about as accurate as YOLOv4 on the same task (0.895mAP vs. 0.892mAP).

A. System Architecture

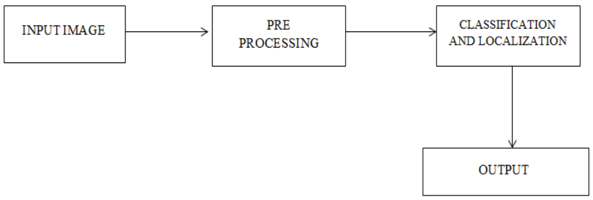

FIG:4.1 FLOW CHART

B. Pre-Processing:

- The process of pre-processing an image that predicts the desired data and suppresses the undesired data from the given input images.
- It normalizes and scales the image size to 448*448 and also reduces the brightness effect. The image is also cropped and adjusted to the aspect ratio so that feature extraction can be easily performed.

V. YOLO Object Detection Algorithm

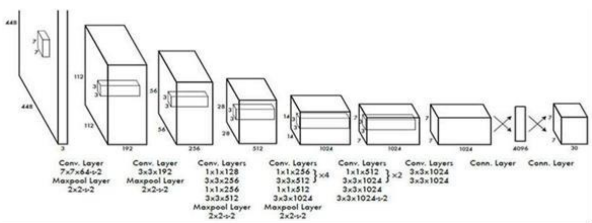

FIG: 5.1 YOLO ARCHITECTURE

YOLO (You Only Look Once) stands for "You Only Look Once". YOLOV5 is a new Convolutional Neural Network (CNN) launched in June 2020 from Ultra lytic. It's one of the fastest object detection algorithms available in the open source computer vision community. It is a real-time, high-accuracy, and "just-look-once" approach to object detection. A single neural network (YOLOV5) processes the entire image, then divides it into parts, and predicts the "bounding boxes" and the "probability" of each component. Each component is weighted by its expected probability. The YOLO method only looks once at the image and makes predictions after one forward propagation through the network. This model splits the image into an S*S (S-grid) and predicts the B bounding box, confidence for each of the S-grid cells, and the C class probabilities for each of the C-grid cells. The predictions are encoded as the S*S (B*5 + C) tensor.

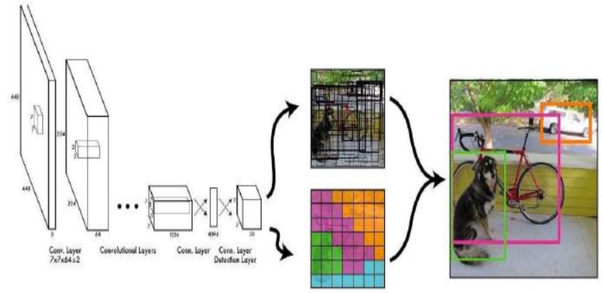

FIG: 5.2 FEATURE EXTRATION

In this article we saw object detection using YOLOV7 technology. YOLO is a state-of-the-art real-time object detection algorithm because it is much faster compared to other algorithms while maintaining good accuracy. The YOLO network understands the general representation of objects; however, space limitations limit accuracy for nearby and smaller objects. We have a newer version of this algorithm, YOLO v7, which solves this problem and is more accurate and faster.

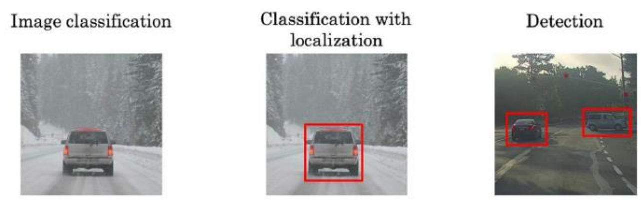

FIG: 5..3 IMAGE DETECTION

A data flow diagram (DFD) is a traditional visual representation of the data flows of a system. A clean and clear DFD can graphically describe the exact amount of system requirements. The purpose of a DFD is to show the scope and limitations of the system as a whole. It can be used as a communication tool between the system analyst and everyone involved in the order as a starting point for system redesign. 1. All names must be unique. This makes it easy to refer to DFD elements. 2. Remember that a DFD is not a flowchart. Arrows are a flowchart that represents a sequence of events; The arrows in the DFD represent the flow data. A DFD does not contain a sequence of events. 3. Suppress logical decisions. If we ever feel the urge to draw a diamond-shaped box on the DFD, suppress that urge! Flowcharts use a diamond-shaped box to represent decision points where there are multiple available paths from which only one is chosen. This means an order of events that does not make sense in the DFD. 4. Don't get bogged down in details. Move error situations and error handling to the end of the analysis.

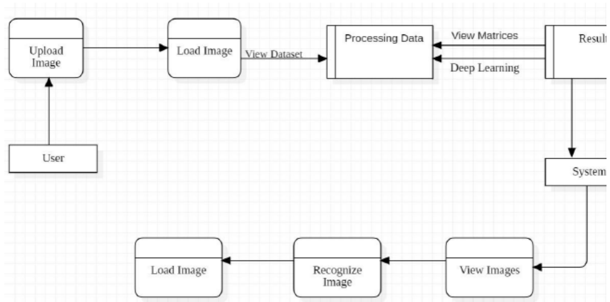

FIG: 5.4 IMAGE DETECTION FLOW CHART

Data Set Prepartion And Training

COCO dataset most commonly used to detect objects. It is an extensive and fast computer vision. It is used for detection, segmentation and labeling of the object. The single-shot multi-box detector is used to detect the multi-object bounding box. By averaging the objects in a single pass. It also works in sematic segmentation of objects like a mask and again was represented by a bounding box. It includes 330,000 images, each with annotations for 80 object categories and 5 captions.

Filename	Size	SHA-1
train2017.zip	18 GB	10ad623668ab00c62c096f0ed636d6aff41faca5
val2017.zip	778 MB	4950dc9d00dbe1c933ee0170f5797584351d2a41
annotations_trainval2017.zip	241 MB	8551ee4bb5860311e79dace7e79cb91e432e78b3
stuff_annotations_trainval2017.zip	401 MB	e7aa0f7515c07e23873a9f71d9095b06bcea3e

TABLE: 3.1 COCO DATASET

VI. RESULTS AND ANALYSIS

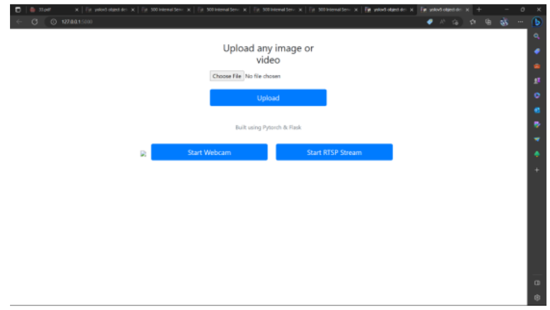

FIG: 6.1 WEBSITE RESULT

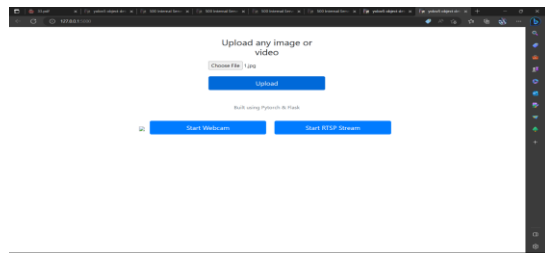

FIG:6.2 UPLOAD IMAGE

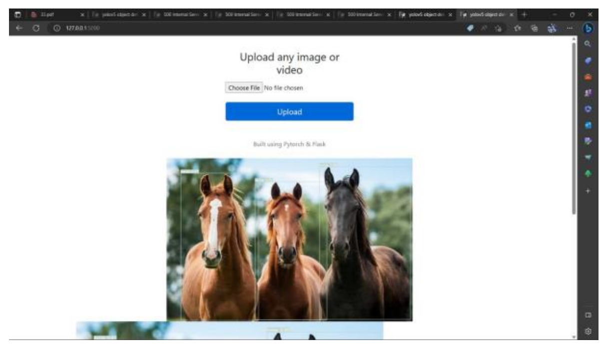

FIG: 6.3 OBJECT DETECTION

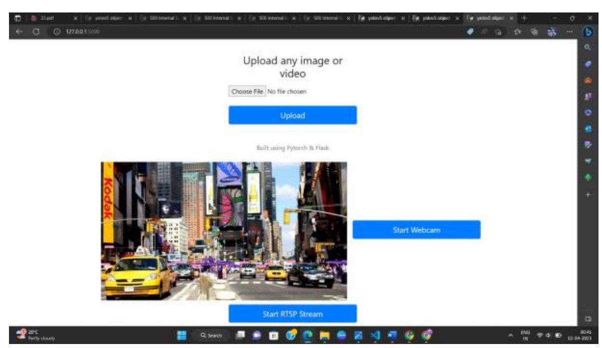

FIG 6.4 VIDEO IMAGE DETECTION

YOLOV7- 320 FRS AP of 28.2 with inference of 22ms. (COCO dataset). Speed of yolov7 is 46.6/1.5/48.1ms inference for 5000 images. When compare with YOLOv5 the speed of pre-process images is 0.7ms. So YOLOV7 is 3 X times faster than SSD object detection technique. YOLOv7 is much better than SSD and has similar performance as DSSD.

VII. CONCLUSION

Using the YOLOV7 approach, we have seen how to detect objects. The most advanced real-time object detection method is YOLO since it is substantially faster than previous algorithms while still being able to recognize objects accurately. The YOLO network is capable of understanding generalized object representation, however accuracy is limited when dealing with close and smaller objects due to spatial restrictions. This issue is addressed in a more recent version of the algorithm called YOLO v7, which is also faster and more accurate.

VIII. FURTHER SCOPE

The current YOLO V7 model outperforms all the existing models by over 40% in speed and with good accuracy. As a future enhancement we can train the yolov7 model for a particular task, we can achieve extremely accurate and faster results than the current existing models. Adding authentication module and database can be future enhancements. . It is difficult to travelling in night time and that too in heavy rains on road by our vehicles then we wouldn't identify the speed breaker, pits and bar-gate. The person who is going to drive this vehicle he will be staring toward the road continuously. From this we can develop face recognition attendance, detecting speed breaks, pits, bar-gate. grocer detection.

REFERENCES

[1] Gowsikraja P , Thevakumaresh T , Raveena M , Santhiya.J , Vaishali.A.R.R.et.al. Object Detection Using Haar Cascade Machine Learning Algorithm IJCRT. Volume 10, Issue 6 June 2022.

[2] Yuan Zhang, Youpeng Sun, Zheng Wang and Ying Jiang YOLOv7-RAR for Urban Vehicle Detection Sensors 2023, 23, 1801.

[3] Dawn Wilson, Dr. Manusankar C , Dr. Prathibha P H et.al. Analytical Study on Object Detection using YOLO Algorithm. IJISRT; Aug 2022. Volume 7, Issue 8, pg 587-589.

[4] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam et.al. Real-Time Object Detection with Yolo IJEAT; Feb 2019. Volume-8 ,Issue-3S , pg 578-580.

[5] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, et.al. A Review of Yolo Algorithm Developments Procedia Computer Science 199 (2022) 1066–1073.

[6] Ziad El Khatib, Adel Ben Mnaouer, Sherif Moussa, et.al. LoRa-enabled GPU-based CubeSat Yolo Object Detection with Hyperparameter Optimization. ISNCC ; Sep 2022 978-1-6654-8544.

[7] Mannem Ponika, Kopalli Jahnavi, P S V S Sridhar et.al. Developing a YOLO based Object Detection Application using OpenCV. ICCMC ; June 2023978-1-6654-6408.

[8] Md. Bahar Ullah, et.al CPU Based YOLO: A Real Time Object Detection Algorithm. TENSYMP; June 2020.

[9] Abhinandan Tripathi , Chaynika Srivastava , Shrawan Kumar Pandey et.al.Object Detection using YOLO: A Survey. IC3I; July 2022 979-8-3503-9826.

[10] Jiayi Fan, JangHyeon Lee, InSu Jung , YongKeun Lee .et.al.Improvement of Object Detection Based on Faster R-CNN and YOLO ITC-CSCC ; June 2021 978-1-6654-3553.

[11] Milad Ahmadi, Zichun Xu, Xinli Wang et.al.fast multi object detection and counting by yolo v3CAC; Mar 2021 978-1-6654-2647.

[12] Jing Tao, Hongbo Wang, Xinyu Zhang, Xiaoyu Li, Huawei Yang ,et.al.An Object Detection System Based on YOLO in Traffic Scene. ICCSNT; oct 2017 978-1-5386-0493-9.

[13] Chengji Liu ,Yufan Tao,Jiawei Liang ,Kai Li ,Yihang Chen et.al. Object Detection Based on YOLO Network. ITOEC; 2018 978-1-5386-5373-9.

[14] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database.

Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:248-255.

[15] Girshick R. Fast R-CNN. IEEE International Conference on ComputerVision. IEEE,2015:1440-1448.

[16] Sorel M, Flusser J. Space-Variant Restoration of Images Degraded by Camera Motion Blur. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2008, 17(2):105-116.

[17] Gong, B., Ergu, D., Cai, Y., & Ma, B. (2020). A Method for Wheat Head Detection Based on YOLO V4.

[18] Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 30(11), 3212-3232.

[19] Rachel Huang, Jonathan Pedoeem, Cuixian Chen, et.al. YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers .https://arxiv.org/pdf/1811.05588.pdf YOLO-LITE

[20] P. Viola and M. Jones. et.al. 2001. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, IEEE Comput. Soc, Kauai, HI, USA, I-511-I-518.

[21] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu.2019. A Survey of Deep Learning-Based Object Detection. IEEE Access 7, (2019), 128837—128868

[22] Hao Zhang and Xianggong Hong. et.al. 2019. Recent rogresses on object detection: a brief review. Multimed Tools Appl 78, 19 (October 2019), 278