International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

OPENCV BASED REAL TIME OBJECT DETECTION USING NEURAL NETWORK

SHATAKSHI PANDEY[1]

Mtech Scholar[1]

MRS SHUBHANGI DUBEY[2]

Assistant Professor[2]

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, DR. C.V.RAMAN UNIVERSITY, BILASPUR ,(C.G)

Abstract - Computer Vision is a field of study that helps to develop techniques to recognize images and displays. It has different features like image recognition, object detection and image creation, etc. Object detection is used in face detection, vehicle detection, web images, and safety systems. The Objective is to distinguish of objects utilizing You Only Look Once (YOLO) approach. This technique has a few focal points when contrasted with other object detection algorithms. In different algorithms like Convolutional Neural Network, Fast-Convolutional Neural Network the algorithm won't take a gander at the image totally yet in YOLO the algorithm looks the image totally by anticipating the bounding boxes utilizing convolutional network and the class probabilities for these boxes and identifies the image quicker when contrasted with different algorithms. Using these techniques and algorithms, based on deep learning which is also based on machine learning require lots of mathematical and deep learning frameworks understanding by using dependencies such as OpenCV we can detect every single object in image by the area object in a highlighted rectangular box and recognize every single object and assign its tag to the object. This additionally incorporates the exactness of every strategy for distinguishing objects.

Key Words: YOLO, Convolution neural network (CNN), Fast-CNN, OpenCV.

I. INTRODUCTION

With the advancement in the video surveillance and image processing, object detection has known a rising interest in the computer visualization industry. However, achieving high performance and a near-real-time object detection is a key concern in both large-scale systems and embedded platforms. Therefore, a reliable and accurate near real-time object detection application, running on an embedded system, is crucial, due to the rising security concerns in different fields. The application can be deployed in different platforms; it can be deployed on a high-performance platform as well as in mobile platform. This application can be used in surveillance systems with distributed cameras and a backend server in which the detection takes place. It can also be used in mobile devices equipped with camera and processor. A high response time in terms of detection is essential for such systems. We are particularly interested in designing a system able to simultaneously detect multiple objects on a scene. This detection information will help surveillance cameras send Realtime information about the detected objects to the back end central system. The information sent to the back-end system can be used to detect the presence of an object in the covered area or to recognize a specific person (blacklisted person) from the detected face in the area. In particular, we are able to train our application in order to detect any object the surveillance system is interested in such as guns or dogs… This paper presents an overview of the cascade object detection algorithm as well as Haar-Like feature selection used by cascade classifier. Then, it proposes an OpenCV-based solution for multiple object detection, and finally, presents the results of the comparison of performances in a regular platform and an embedded device. The detection principle used in the application is based on object detection proposed by Viola et al. [4] for facial detection where the Haar-like feature-based cascade classifier for the detection, is adopted. The OpenCV library is described in section II then the adopted object detection is detailed in section III. Section IV presents the system, the developed application and the results obtained after porting the application on regular platform and embedded platform.

II. LITERATURE SURVEY

In the year 2017 Tsung-Yi Lin, Piotr Dollar, Ross Girshick, KaimingHe,BharathHariharan, and SergeBelongie proposed Feature Pyramid Networks for Object Detection. With the launch of Faster-RCNN, YOLO, and SSD in 2015, it seems like the overall structure an object identifier is resolved. Analysts begin to take a gander at improving every individual pieces of these networks. Highlight Pyramid Networks is an endeavor to improve the identification head by utilizing highlights from various layers to frame a feature pyramid. This feature pyramid thought isn't novel in computer vision research. In those days when highlights are still physically planned, feature pyramid is now a powerful method to recognize patterns at various levels. Utilizing the Feature Pyramid in deep learning is likewise not a groundbreaking thought: SSPNet, FCN, and SSD all showed the advantage of aggregating multiple layer highlights before classification. Nonetheless, how to share the feature pyramid among RPN and the region-based detector is still yet to be resolved.

In the year 2017 Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick proposed Mask R-CNN.In this paper Mask RCNN is certainly not a commonplace object detection network. It was intended to settle a difficult example division task, i.e, making a mask for each object in the scene. Nonetheless, Mask R-CNN indicated an incredible augmentation to the Faster R-CNN framework, and furthermore thusly motivated object location research. The fundamental thought is to add a binary mask prediction branch after ROI pooling alongside the current bounding box and characterization branches. Obviously, both perform multiple tasks preparing (division + detection) and the new ROI Align layer add to some improvement over the bounding box benchmark.

In the year 2017 NavaneethBodla, Bharat Singh, Rama Chellappa, Larry S. Davis proposed Soft-NMS – Improving Object Detection with One Line of Code. In this paper Non maximum suppression (NMS) is broadly utilized in anchor baseded object detection networks to diminish copy positive proposition that are close-by. All the more explicitly, NMS iteratively wipes out applicant boxes on the off chance that they have a high IOU with a surer applicant box. This could prompt some sudden conduct when two objects with a similar class are to be sure near one another. SoftNMS rolled out a little improvement to just downsizing the certainty score of the overlapped applicant boxes with a boundary. This scaling boundary gives us more control when tuning the localization execution, and furthermore prompts a superior exactness when a high review is likewise required.

In the year 2017 ZhaoweiCai UC San Diego, Nuno Vasconcelos UC San proposed Cascade R-CNN: Delving into High Quality Object Detection. While FPN investigating how to plan a superior R-CNN neck to utilize backbone highlights Cascade R-CNNexamined an upgrade ofR-CNNgrouping and regression head. The basic assumption that is straightforward yet sagacious: the higher IOU rules we utilize while planning positive focuses on, the less false positive predictions the networkwill figure out howto make. In any case, we can't just increment such IOU threshold from regularly utilized 0.5 to more forceful 0.7, in light of the fact that it could likewise prompt all the more overpowering negative models during training. Cascade R-CNN's answer is to chain various recognition head together, eachwill depend on the bounding box recommendations from the past detection head.

In the year 2017 Tsung-Yi Lin PriyaGoyal Ross GirshickKaiming He Piotr Dollar proposed Focal Loss for Dense Object Detection. To comprehend why one-stage locators are typically not comparable to two-stage detectors, RetinaNet explored the frontal area foundation class unevenness issue from a one-stage detectors dense predictions. Take YOLO for instance, it attempted to predict classes and bounding boxes for all potential areas meanwhile, so the majority of the yields are coordinated to negative class during training. SSD tended to this issue by online hard model mining. YOLO utilized an objectiveness score to certainly prepare a closer view classifier in the beginning phase of training. RetinaNet thinks the two of them didn't get the way in to the issue, so it developed another loss function work called Focal Loss to assist the network with realizing what's significant. Focal Loss added a power γ to Cross-Entropy loss. The α boundary is utilized to adjust such a focusing effect.

In the year 2018 Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, JiayaJia proposed Path Aggregation Network for Instance Segmentation. In this paper Occurrence division has a close relationship with object detection, so regularly another case segmentation network could likewise profit object recognition research in a roundabout way. PANet targets boosting data stream in the FPN neck of Mask R-CNN by adding an extra base up path after the first top-down path. To picture this change, we have a ↑↓ structure in the first FPN neck, and PANet makes it more like a ↑↓↑ structure prior to pooling highlights from various layers. Likewise, rather than having separate pooling for each element layer, PANet added an "adaptive feature pooling" layer after Mask RCNN's ROIAlign to merge multi-scale features.

In the year 2018 ChengjiLiu, Yufan Tao, JiaweiLiang, Kai Li, Yihang Chen proposed Object Detection Based on YOLO Network. In this paper YOLO v3 is the latest form of the YOLO versions. Following YOLOv2's convention, YOLOv3 acquired more thoughts from past exploration and got a powerful incredible one-stage finder like a beast. YOLO v3 adjusted the speed, exactness, and execution unpredictability really well. Also, it got truly mainstream in the business as a result of its quick speed and basic parts. Basically, YOLO v3's success comes from its all the more impressive backbone include extractor and a RetinaNet-like identification head with a FPN neck. The new spine networkDarknet-53 utilized ResNet's skip connections with accomplish a precision that is comparable to ResNet-50 yet a lot quicker.

In the year 2020 Mingxing Tan, Ruoming Pang, Quoc V Le proposed EfficientDet: Scalable and Efficient Object Detection. In this paper Efficient DE indicated us some all the more energizing advancement in the object detection area. FPN structure has been end up being an amazing technique to improve the identification network performance for objects at various scales. Popular detecting network, for example, RetinaNet and YOLO v3 all received a FPN neck before box regression and arrangement. Afterward, NAS-FPN and PANet both showed that a plain multi-layer FPN structure may profit by more plan enhancement. Efficient Det kept investigating toward this path, in the end made another neck called BiFPN. Essentially, BiFPN highlights extra cross layer associations with energize include aggregation to and for. To legitimize the proficiency part of the network, this BiFPN additionally eliminated some fewer valuable associations from the first PANet plan. Another creative improvement over the FPN structure is the weight feature fusion. BiFPN added extra learnable loads to highlight aggregation so the network can get familiar with the significance of various branches. Besides, much the same as what we found in the image characterization network EfficientNet, EfficientDet likewise acquainted a principled path with scale an object identification network. The φ parameter in the above formula controls both width (channels) and depth (layers) of both BiFPN neck and detection head.

III. OBJECT DETECTION AND TRACKING PROPOSED METHOD:

1. OpenCV

OpenCV is a fairly wide resource for image recognition, deep learning, and image analysis that is becoming increasingly important in contemporary networks. OpenCV can recognize objects, faces, and even human handwriting from photos and videos. Object detection is a subset of machine learning, signal processing, and big data that deals with identifying features in photos and

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

images. The first step entails using a large number of negative and positive labelled images to train a cascade function. After the classifier has been trained, the training images are used to extract identifying features, known as "HAAR Features." HAAR features are basically rectangular features with bright and dark pixels in different areas. The value of each function is determined by subtracting the amount of pixel intensity in the bright region from the pixel intensity in the dark region. These attributes are calculated using all of the image's potential sizes and locations. Many irrelevant features can be present in an image, while only a few ways define the object, a few related features may be used. The classifier is learned to extract useful features from the pre-labeled dataset and add sufficient weights to each feature to obtain the lowest possible errors. Poor function refers to a single feature. The weighted sum of the weak features is the final classifier.

Fig. 1 - Image Recognition Using OpenCV

The context takes up a large portion of the image; The item to be viewed is only a small portion of the picture. Cascaded classifiers are used to speed up the detection process. If even a single negative feature is detected in a region of an image during this step, the algorithm goes on to the next region after ignoring the region for further processing. The requisite object in the image is the only area that contains all of the identifying features. The requisite object in the image is the only area that contains all of the identifying features.

2. CNN ARCHHITECTURE

A convolution neural network comprises of information and an output layer, just as various hidden layers. The hidden layers of a CNN commonly comprise of a progression of convolution layers that ev with an increase or other dot product. The activation function is generally a RELU layer, and is in this way followed by extra convolutions, for example, pooling layers, completely associated layers and normalization layers, referred to as hidden layers on the grounds that their sources of input and output are masked by the activation function and last convolution.

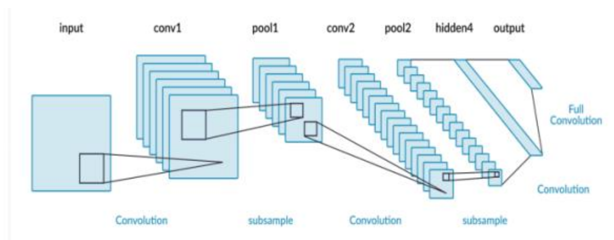

Fig. 2: Convolution Neural Networks (CNN)

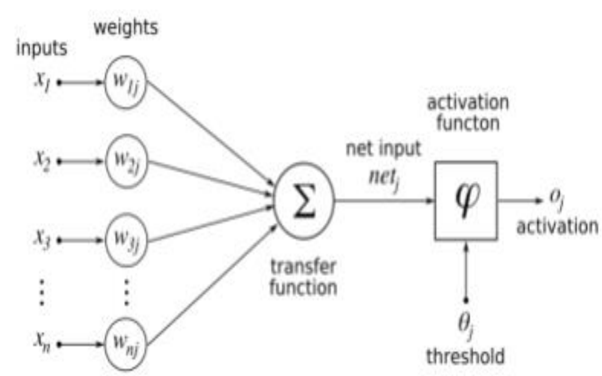

Fig. 3 - The Basic Component of Artificial Neural Networks, The Artificial Neuron, has the Following Structure

3. SVM CLASSIFIER

Support Vector Machines are a kind of supervised machine learning algorithm that gives analysis of knowledge for classification and multivariate analysis. While they will be used for regression, SVM is usually used for classification we feature out plotting within the n-dimensional space. Value of every feature is additionally the worth of the precise coordinate. Then, we discover the perfect hyper-plane that differentiates between the 2 classes. The basic principle behind the working of Support vector machines is straightforward –

Create a hyper-plane that separates the data-set into classes allow us to start with a sample problem. Suppose that for a given data-set, you've got to classify red triangles from blue circles. Your goal is to make a line that classifies the info into two classes, creating a distinction between red triangles and blue circles.

4. Single Shot Detector (SSD) algorithm

SSD is a popular object detection algorithm that was developed in Google Inc. [1]. It is based on the VGG-16 architecture. Hence SSD is simple and easier to implement.

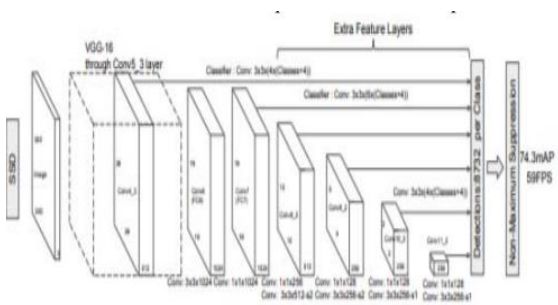

Fig. 4. VGG-16 SSD Model.

Fig. 4 shows VGG 16 SSD model. A set of default boxes is made to pass over several feature maps in a convolutional manner. If an object detected is one among the object classifiers during prediction, then a score is generated. The object shape is adjusted to match the localization box. For each box, shape offsets and confidence level are predicted. During training, default boxes are matched to the ground truth boxes. The fully connected layers are discarded by SSD architecture. The model loss is computed as a weighted sum of confidence loss and localization loss. Measure of the deviation of the predicted box from the ground truth box is localization loss. Confidence is a measure of in which manner confidence the system is that a predicted object is the actual object. Elimination of feature resampling and encapsulation of all computation in a single network by SSD makes it simple to train with MobileNets. Compared to YOLO, SSD is faster and a method it performs explicit region proposals and pooling (including Faster R-CNN).

5. MobileNets algorithm

MobileNets uses depth wise separable convolutions that helps in building deep neural networks. The MobileNets model is more appropriate for portable and embedded vision-based applications where there is absence of process control. The main objective of MobileNets is to optimize the latency while building small neural nets at the same time. It concentrates just on size without much focus on speed. MobileNets are constructed from depth wise separable convolutions. In the normal convolution, the input feature map is fragmented into multiple feature maps after the convolution [2].

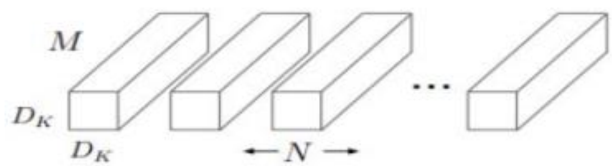

Fig. 5. Normal Convolution [2]

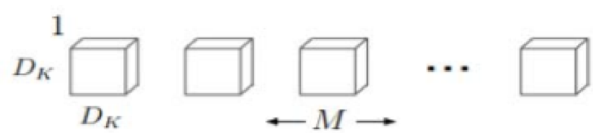

Fig. 6. Depth wise Convolution Filters [2]

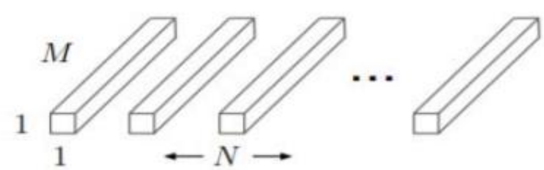

Fig. 7. 1×1 Convolutional Filters called Pointwise Convolution in the context of Depth wise Separable Convolution [2].

The number of parameters is reduced significantly by this model through the use of depth wise separable convolutions, when compared to that done by the network with normal convolutions having the same depth in the networks. The reduction of parameters results in the formation of light weight neural network as shown in fig 5 to 7.

IV. RESULT:

Based on CNN architecture algorithm, a python program was developed for the algorithm and applied in Open-CV. OpenCV is run in PyCharm IDE. Total 20 objects were trained in this model. The following results got after effective scanning, detection delivered by camera with accuracy level. The proposed algorithm was also capable of detection multiple object in single frame.

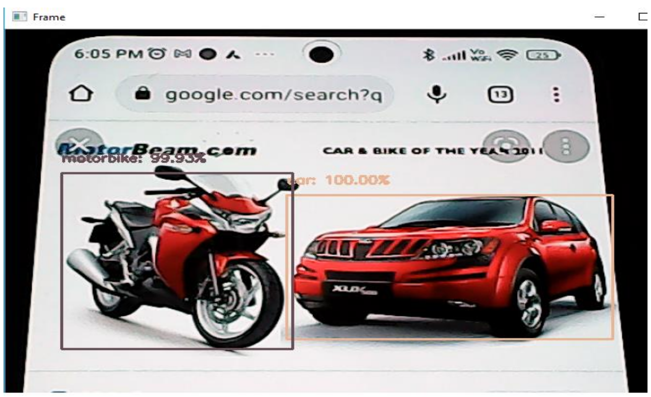

Fig 8: Proposed system detect motorbike and car with accuracy of 99%

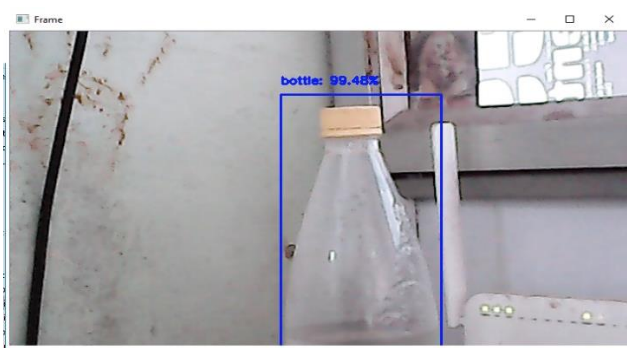

Fig 9: Proposed system detect bottle accuracy of 99.48%

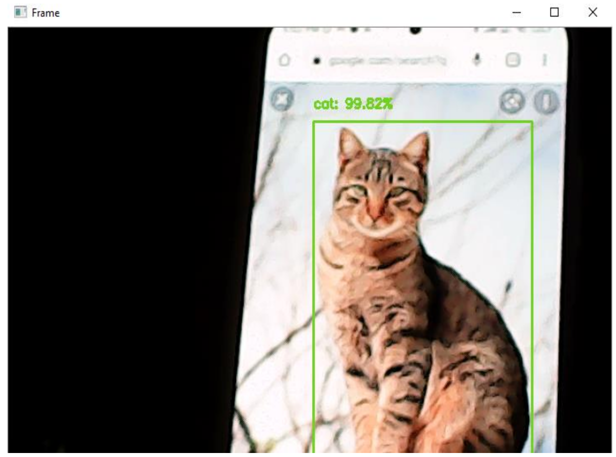

Fig 10: Proposed system detect cat accuracy of 99.62%

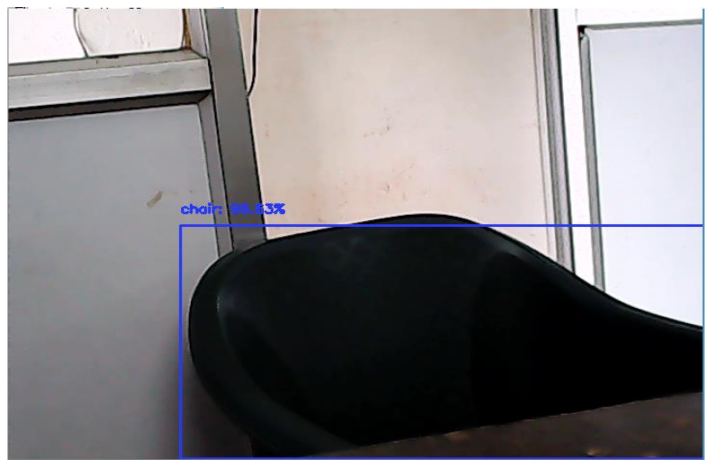

Fig 11: Proposed system detect chair with accuracy of 96.63%

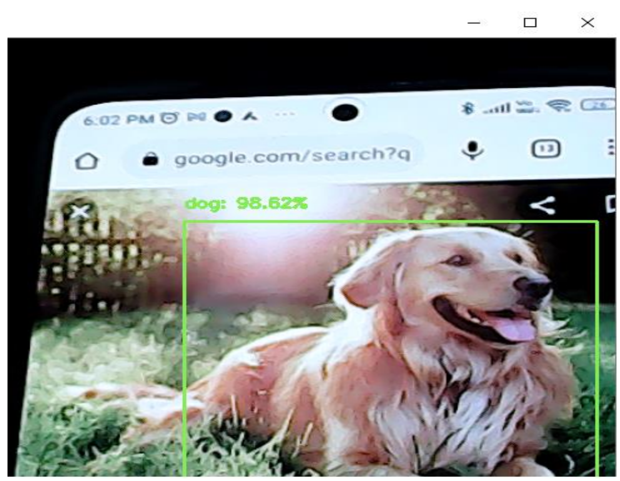

Fig 12: Proposed system detect dog with accuracy of 96.61%

Fig 13: Proposed system detect Person with accuracy of more than 90%

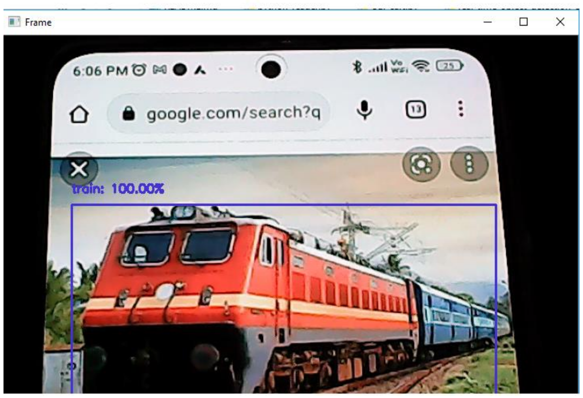

Fig 13: Proposed system detect train with accuracy of 100.

V. CONCLUSION:

Objects are detected using SSD algorithm in real time scenarios. Additionally, SSD have shown results with considerable confidence level. Main Objective of SSD algorithm to detect various objects in real time video sequence and track them in real time. This model showed excellent detection and tracking results on the object trained and can further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video surveillance. This real time analysis of the ecosystem can yield great results by enabling security, order and utility for any enterprise. Further extending the work to detect ammunition and guns in order to trigger alarm in case of terrorist attacks. The model can be deployed in CCTVs, drones and other surveillance devices to detect attacks on many places like schools, government offices and hospitals where arms are completely restricted.

REFERENCES

[1] Chengji Liu1, Yufan Tao1, Jiawei Liang1, Kai Li1, Yihang Chen1 "Object Detection Based on YOLO Network" in 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference(ITOEC), https://ieeexplore.ieee.org/document/8851911.

[2] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia," Path Aggregation Network for Instance Segmentation" in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, https://ieeexplore.ieee.org/document/8579011

[3] Kaiming He, Georgia, Gkioxari Piotr, Dollar Ross, Girshick, "Mask R-CNN" in 2017 IEEE International Conference on Computer Vision (ICCV), https://ieeexplore.ieee.org/document/8237584

[4] Zhaowei Cai, Nuno Vasconcelos, "Cascade R-CNN: Delving into High Quality Object Detection" in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, https://ieeexplore.ieee.org/document/8578742/authors#authors

[5] Navaneeth Bodla,Bharat Singh,Rama Chellappa, Larry S. Davis,"Soft-NMS – Improving ObjectDetectionwithOne Line of Code" in 017 IEEE International Conference on Computer Vision (ICCV), https://ieeexplore.ieee.org/document/8237855

[6] Liguang Yan, Baojiang Zhong Weigang Song, "REGIONBASED FULLY CONVOLUTIONAL NETWORKS FOR VERTICAL CORNER LINE DETECTION" in 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), https://ieeexplore.ieee.org/document/8266465

[7] Koen E. A. van de Sande, Jasper R. R. Uijlingst, Arnold W. M. Smeulders, "Segmentation as Selective Search for Object Recognition"in 2011 International Conference on Computer Vision, https://ieeexplore.ieee.org/document/6126456

[8] Mingxing Tan, Ruoming Pang, Quoc V. Le,"EfficientDet: Scalable and Efficient Object Detection" in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), https://ieeexplore.ieee.org/document/9156454

[9] Andrew Edie, Johnson and Martial Hebert,"Recognizing Objects by Matching Oriented Points" in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, https://ieeexplore.ieee.org/abstract/document/609400

[10] Tsung-Yi, Lin Priya Goyal, Ross Girshick Kaiming, He Piotr Dollar, "Focal Loss for Dense Object Detection" in 2017 IEEE International Conference on ComputerVision (ICCV), https://ieeexplore.ieee.org/document/8237586

[11] Joseph Redmon, Ali Farhadi,"YOLO9000: Better, Faster, Stronger"in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), https://ieeexplore.ieee.org/document/8100173

[12] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature Pyramid Networks for Object Detection" in 2017 IEEE Conference on ComputerVision and PatternRecognition (CVPR), https://ieeexplore.ieee.org/document/8099589

[13] Chengcheng Ning, Huajun Zhou, Yan Song, linhui Tang, "INCEPTION SINGLE SHOT MULTIBOX DETECTOR FOR OBJECT DETECTION" in 2017 IEEE International Conference on Multimedia&ExpoWorkshops (ICMEW), https://ieeexplore.ieee.org/document/8026312

[14] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi,"You Only Look Once:Unified,Real-TimeObject Detection"in 2016 IEEE Conference on ComputerVision and Pattern Recognition (CVPR), https://ieeexplore.ieee.org/document/7780460

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" in IEEE Transactions on Pattern Analysis and Machine Intelligence, https://ieeexplore.ieee.org/document/7485869