iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VII **Month of publication:** July 2020

DOI: https://doi.org/10.22214/ijraset.2020.30430

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

Object Detection using Region based Convolutional Neural Network: A Survey

Rekha B. S[1], Dr. G N Srinivasan[2], Chandana EP[3], Achala N Gowda[4]
[1]Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India
[1]Assistant Professor, RV College of Engineering, Bangalore, India
[2]Professor, Dept. of ISE, RV College of Engineering, Bangalore, India
[3, 4]II Sem, M. Tech IT, Dept. of ISE, RV College of Engineering, Bangalore, India

Abstract: Region based CNN system tries to locate objects in an image. Region proposals are just smaller parts of the original image. Object detection can be a subset of computer vision which is also a programmed method for locating engrossing objects in an image with respect to the background which involves detecting instances of objects from a particular class in an image. In a distributed Edge-Cloud RCNN pipeline, the object detection pipeline will be split into various components. These components are dynamically distributed in the cloud, and real time object detection will be enabled by optimal performance. Using Bayesian Optimization, a sequential search algorithm proposes better bounding boxes of RCNN helps to reduce localization error by tackling the localization problem by converting region proposal into an optimization problem and iteratively solve it using Bayesian optimization as a black box optimizer. The part filters of DPM is appropriate for detecting occluded objects.
Key Terms: Edge Computing, computer vision, Deformable part-based model (DPM), Bayesian optimization, Region proposal, Distributed computing.

I. INTRODUCTION

Region based convolutional neural networks (RCNNs) or Regions with CNN consists of two steps: Using selective search, initially it identifies a manageable number of bounding box region candidates of an object and then it extract's CNN features from each region of object independently for classification.

A. *Selective search Algorithm*

1) Many candidate regions are generated by generating initial sub segmentation.
2) Greedy algorithm is used to recursively join similar into larger ones.
3) Final candidate region proposals are produced using the generated regions.

A convolutional neural network (CNN) is majorly for image classification. While a R-CNN, with the R standing for region, is basically for object detection. A typical CNN can only tell the class of the objects but not where they are located. If multiple objects are in the visual field then the CNN bounding box regression cannot work well, due to interference. In R-CNN, interference is minimized when the CNN is enforced to make a single region as centre of activity, because it is anticipated that only a single object of interest will dominate in a given region. Selective search algorithm is used to detect the regions in the Region based CNN, followed by resizing so that the regions are of equal size before they are fed to a CNN for bounding box regression and classification. Convolutional Neural Networks is a class of neural networks that processes an input image and classifies the object based on values using layers; Filter depth as an input layer, RELU (Rectified linear Unit) which helps to increase the non-linearity. The function of the pooling layer is to continuously reduce the spatial size of the representation to decrease the number of parameters and computations in the convolutional neural network. The last layer is a fully connected layer, which uses the output of the last pooling layer that acts as an input. There can also be one or more fully connected layers.

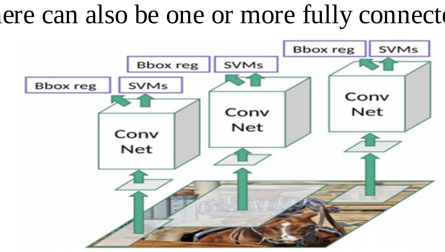


Fig 1: Diagram of RCNN for object detection. [9]

There are three models of RCNN for the entire process of object detection: Feature extraction can be achieved by Convolutional neural network, identifying objects is done by linear SVM classifier and Tightening the bounding boxes is achieved using a regression model as shown in Figure 1.

Initially, a pre-trained CNN is taken, which is restrained on the fully connected layer that is built on the number of classes which are needed to be identified. Later, the ROI for every image is taken & later the regions are remodelled to match as per the size of CNN. Now regions are obtained, hence a Linear Support Vector Machine (SVM) classifier is trained to classify the background and objects, i.e., one binary SVM is trained for each class. In the last step, once the object has been classified in the image, a linear regression model is trained to get an output tighter coordinate for the box.

II. RELATED WORK

In [1], a distributed Edge-Cloud RCNN pipeline is proposed. The object detection pipeline can be broadly split into components. These components are dynamically distributed in the cloud, using the same, able to obtain optimal performance, which helps real time object detection. Evaluating the performance is based on a distributed computing platform..

In [2], with imposing outcomes on various object detection's standards, a revolutionized technology; Deep Neural Networks is very essential for the task of detecting objects. An important merit of DNN is the secure differentiation potentiality of the models which are trained and high capacity, but has an indefinite localization which is a major error source. Using Bayesian Optimization, a selective search algorithm is proposed.

In [3], a substructure that combines the Region based CNN and multiple objects can be detected using DPM is presented. Furthermore, a new filter is proposed, which is decided upon the dense subgraph discovery algorithm that helps to refine the candidate proposals generated by the deformable part-based model. Each single object must be detected with higher accuracy among every object in an image, exceptionally in the cases where objects are close together which can be obtained by combining these two models.

III. SYSTEM ARCHITECTURE

Figure 1 is an outline of Region based CNN architecture. The system takes an image as input, later makes use of the selective search to suggest around 2000 regions, then passes every region to a huge convolution neural network to calculate the feature vectors, and then using specific linear SVMs, every feature vector will be classified. Resizing the regions to the expected size of a network can be done by randomly selecting the needed regions of different dimensions and featured ratios. The last component SVM will be trained for each class, which can be done using the extracted features from convolutional neural networks. Selective search proposes around two thousand regions per image, then they propagate through Convolutional neural network for extracting features and later train using SVM classifier, during test time. A non-maximum suppression technique is used to reject the candidates whose regions overlap having higher scores. The intersection-over-union (IoU) metric helps to calculate the score of overlapping regions. Also, each proposed region of interest is used to extract the convolutional neural network features, at test time and expensive computation had to be done.

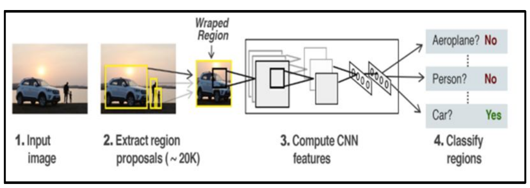


Figure 2: RCNN using Bayesian Optimization [1]

An enhanced structure is built, including the RCNN along with DPM, which helps to detect multiple objects as shown in Figure 3. The Deformable Part-based model with part filters are used to refine the potentiality for handling obstruction and the Region based-CNN has greater capability in feature learning which cannot only do feature extraction in each object, but can also separate the wrong bounding boxes created by the DPM, which helps in better accuracy. Furthermore, the candidate proposals obtained from a deformable part-based model can have some cases where an object must be added by any two bounding boxes whose predefined threshold is greater overlapped intersection-over-union. With the help of DPM and the R-CNN, object detection will first produce the outlines from an input image through the Deformable part-based model and the selective search algorithm individually. Later, derived outlines are individually fed into a convolutional neural network model for extracting features. Later on, convolutional neural network is processed and SVM allocates score to each outline.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

Then, with help of the new Dense Subgraph discovery, filter refining of the proposals which are generated by a deformable part-based model can be done. Later, both the proposals i.e. one from selective search algorithm and other one is proposals which must be sent to non-maximum suppression algorithms for decreasing the total number of outlines that makes Intersection-over-Union to overlap. Finally, the position of proposals comprises objects that can be obtained as output and is the final result as described in Figure 3.

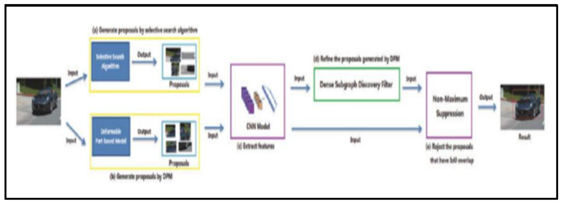


Figure 3: RCNN with DPM [2]

In order to achieve a distributed pipeline, three basic components are essentially required; An algorithm to obtain region proposals, a convolutional neural network to extract features, and any additional layers which will be used for classification and bounding box regression at final stage. The initial implementation of distributed pipeline is emulated after the simplest Region based-CNN model. Many possible distribution cases for region-based-CNN models are illustrated in the fig4. In order to accelerate an algorithm for proposal of regions, a high computing power is provided by a CPU bound cloud. The Raspberry Pi's limited computing abilities may present a support to that system.

Selective Search makes use of the theory of super pixels to separate an image at various levels of granularity. The original pixels of an image must be merged to indicate the existence of an object, which is called Super pixels. The likeliness of a bounding box containing an object is defined by objectness. To generate bounding boxes, the algorithm to predict objectness is used which has various kind of indicators such as edge density, colour contrast. EdgeBoxes is a tentatively new robotic process automation, which makes bounding box predictions rely on an edge map which is built by structured Forests. Structured Forests makes use of an aimed model to produce edges from an input image, which is later given to EdgeBoxes to predict bounding box.

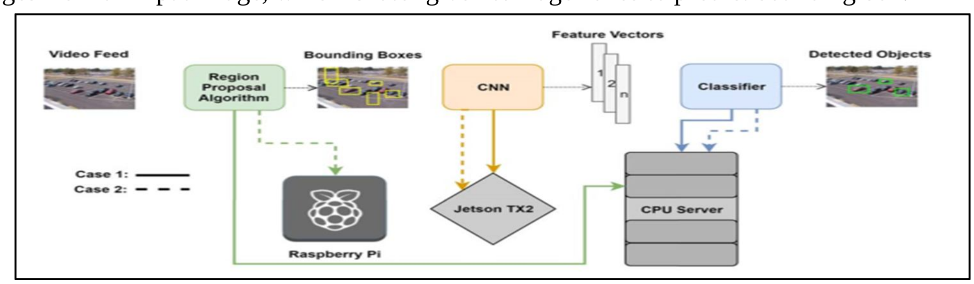


Figure 4: RCNN with Distributed edge cloud[3]

IV. IMPLEMENTATION

Let x and y are the two inputs received by detection score function f(x, y), where x is an image and y is a region coordinate. Finding the local minimum of f with respect to y on a new image x is the job of an object detection framework. With the solution of images, it is vital to find an efficient looking calculation for the competitor locales as assessment of the scoring function at numerous areas must be finished.

Algorithm 1 Bayesian Optimization based Region Proposal

Require: Image x, classifier f_{cnn}, a set of regions with classification scores D_n.

Ensure: A set of newly proposed regions with detection scores.

function PROPOSE REGIONS
- $D \leftarrow D_n$
- **for** $i \leftarrow 1, t$ **do**
  - $D_{proposal} = \varphi$
  - **for all** $(y_{best}, f_{best}) \in D$ **do**
    - $D_{local} = \{(y, f) \in D : IoU(y, y_{best}) > threshold\}$
    - $\hat{y} = \operatorname{argmax}_y \alpha(y|D_{local})$
    - $\hat{f} = f_{cnn}(x, \hat{y})$
    - $D_{proposal} \leftarrow D_{proposal} \cup \{(\hat{y}, \hat{f})\}$
  - $D \leftarrow D \cup D_{proposal}$

In region-based convolutional neural network object detectors, detection scoring function f is required to be evaluated by the detection framework on a fewer number of regions. One potential issue of object location pipelines dependent on district propositions is having no right discovery if no area is proposed in sufficient vicinity to the ground truth bouncing box. To alleviate this hazard region-based CNN recognition pipelines thickly propose bouncing boxes covering the whole picture, yet again this will significantly build the location time. Inside the Bayesian Optimization setting which is utilized as a global maximiser of an obscure function f, without significantly increasing the number of proposed regions a higher detection score can be developed by a sequential searching algorithm which uses previously proposed regions. The fundamental thought is to fit a probabilistic surrogate model.

Assume, {y1,y2,…. Yn} is a set of solutions of the detection score function f(x, y).

$$P(f\ D_n) \propto p(D_n \mid f)p(f) \qquad (1)$$

Where $D_n = \{(Y_j, f_j)\}n\ j=1$, and $f_j = f(x, y_j)$

Is a probabilistic model from which the Bayesian optimization framework f(x, y) is thought to be drawn.

Once fit a probabilistic model on f, It will augment the estimation of fn + 1 with high chance by sampling a new solution Yn +1. Here acquisition function $\alpha(yn{+}1|D_n)$ represents chance, that is utilized to exchange off among variance and mean of the fitted probabilistic model on f. Amgad Muhammad *et al.* [1] create the set

$$D_n+1 = D_n \cup \{yn+1\} \qquad (2)$$

By using Bayesian Optimization algorithm recursively once the model and the acquisition function are fitted.

Junliang Li *et al.* [2] proposed DPM and RCNN for multiple object detection, to extract from each part of the whole object the DPM includes a part filter, in any event, when an object is in part blocked by different objects. Unique in relation to the DPM, without lessening their number and anticipated bounding boxes can be spared by recognition windows as the produced. To be explicit, a lot of n recommendations as D = {Di}ni =1 from the DPM, and every one of them empowers to cover parts or the entire item in the picture as indicated by the element map from part channels can be represented.

As indicated by Joshua Herrera et al. [3] in appropriated Edge-Cloud RCNN, Edge Boxes was utilized for area propositions on a video of a parking area. The Edge Boxes RPA comprises two stages. Initial, an edge guide and direction map are created for an information picture utilizing Structured Forests. This can be found in Fig. 5, where the upper left picture is the first video feed and the upper right picture is the edge map created by Structured Forests. The subsequent stage nourishes the edge guide and direction map into the Edge Boxes calculation for bouncing box proposition. This can be found in the base picture of Figure 5. It ought to be noticed that, for representation purposes, just 100 boxes were proposed, while the genuine number of boxes proposed will be a request for size more prominent.

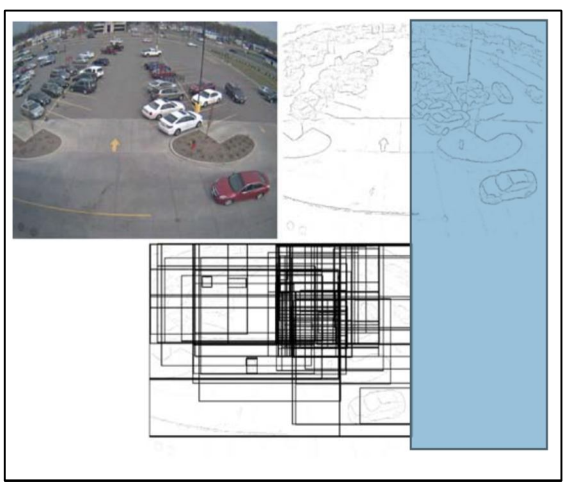


Figure 5: Edge map generation and bounding box proposals [4]

V. DATASET

For the above study, Pascal Visual Object Classes offers standardized image datasets for object recognition. Pascal VOC additionally offers a typical arrangement of devices for getting to the informational collections and comments and empowers correlation and assessment of various strategies. In Pascal VOC, a file for each of the images in the database is created.

Pascal Visual Object Classes provides a training set of labelled images, and it is basically a supervised learning problem.

The twenty object classes that have been chosen are:

1) *Animal:* Dog, cat, horse, cow, sheep, bird.
2) *Indoor:* Chair, potted plant, dining table, bottle, tv/monitor, sofa.
3) *Person:* Person
4) *Vehicle:* Bicycle, bus, motorbike, car, train, aeroplane

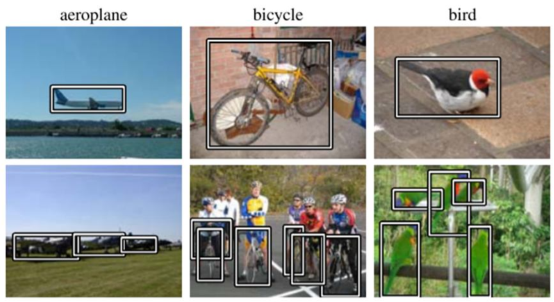


Figure 6: Example for Pascal VOC dataset [5]

VI. SOFTWARE ARCHITECTURE

RCNN architecture can be viewed as a layered style software architecture. Layered style of software architecture bolsters configuration dependent on expanding levels of deliberation and enables implementers to separate a complex problem into a sequence of incremental steps.

Here, RCNN is partitioned based on region of interest and then as a convolutional layer, pooling layer, fully connected input layer, fully connected layer, fully connected output layer. Region based convolutional neural network architecture supports enhancement and reuse because different techniques and approaches such as Bayesian Optimization, DPM and edge cloud are used to enhance selective search in RCNN

VII. CONCLUSION

For multiple object detection, DPM with RCNN integration framework has been studied. To produce the recommendations containing portions of the article for object impediment dealing with, incredible part channels of DPM are used, those propositions considers each single item yielded as the conclusive outcome. Thus, DPM and RCNN obtains higher accuracy.

Also, RCNN utilizing Bayesian advancement improves the objectivity of the proposed districts and obtains higher classification likelihood, hence produces the better region proposals for classification of the objects. An RCNN architecture with distributed edge cloud uses a pipeline as the component of the region proposal algorithm.

Table 1: Summary of various papers referred

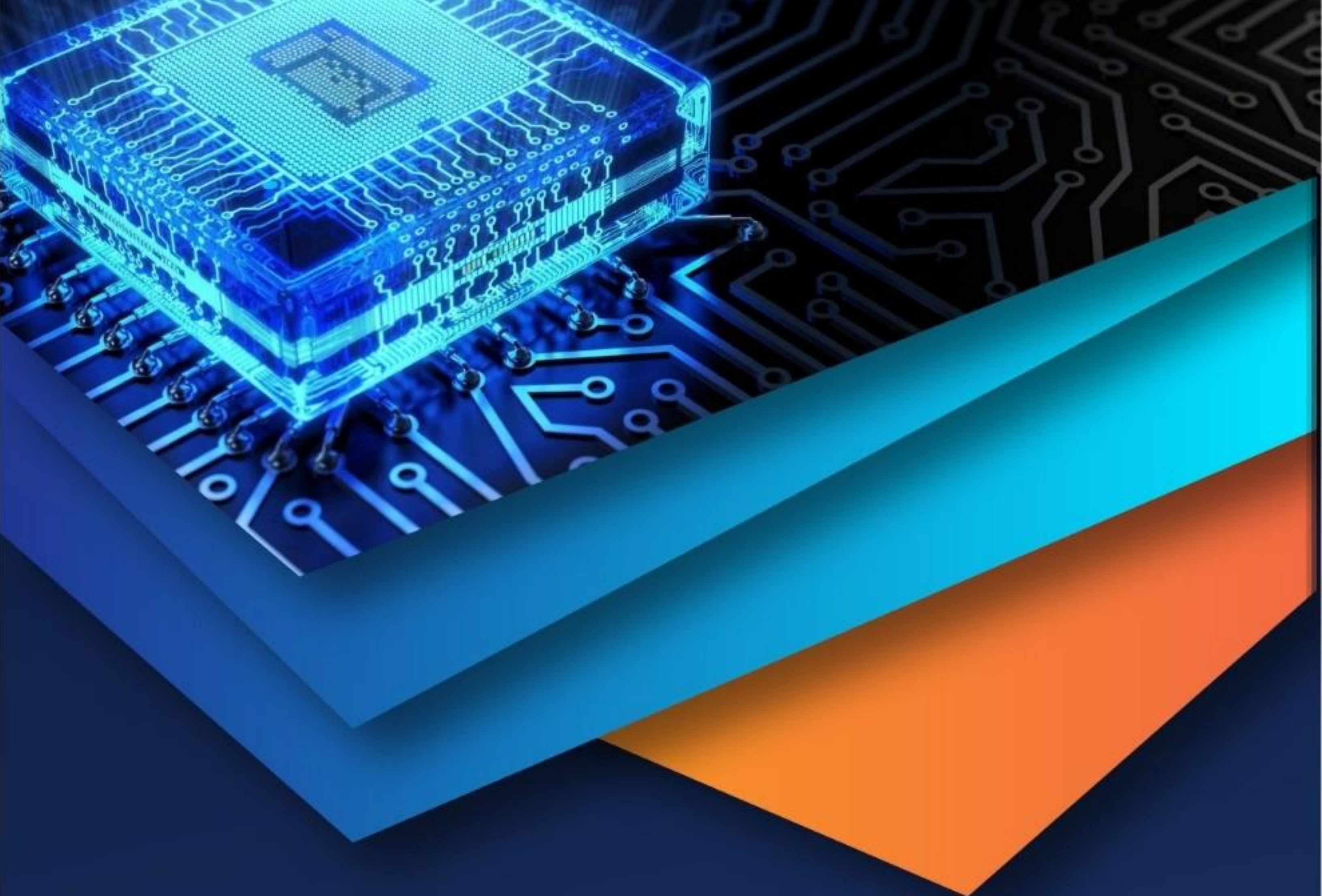
Year of publication	Author(s)	Title	Domain	Conclusion
2014	Ross Girschik, Jeff Donahue, Trevor Darrell, Jitendra Malik	Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation	Object Detection	The model proposed is highly effective to pre-train the network, with supervision, for an auxiliary task with abundant data and then to fine tune the network for the target task where data is scarce.
2018	Amgad Muhammad, Mohamed Moustafa	Improving Region Based CNN Object Detector Using Bayesian Optimization	Object Detection	A sequential Algorithm using Bayesian Optimization is used to provide better bounding boxes. Thus, helps to reduce the Localization error by iteratively solving the region proposals of the objects.
2018	Junliang Li, Hon-Cheng Wong	Multiple Object Detection by a Deformable Part-Based Model and an R-CNN	Object Detection	The DPM and the RCNN are integrated together to build a framework for multiple object detection which provides the powerful part filters which enables the object occlusion handling. The generated proposals are filtered and those can detect each single object are outputted as the final result.
2018	Joshua Herrera, Mevlut A Demir, Parsa Yousefi, John J Prevost, and Paul Rad	Distributed Edge Cloud R-CNN for Real Time Object Detection	Object Detection	A distributed edge cloud RCNN pipeline achieves optimal performance which enables the real time object detection. The model evaluates the performance on a distributed computing platform as a proof of the concept.

VIII. FUTURE WORKS

Further improvement and evaluation of accuracy of the combined pipeline of object detection while implementing individual components of distributed pipeline is desirable. Also, further execution advancements can be accomplished by receiving a fast-RCNN architecture. Different other advancement techniques may yield better outcomes by finding various models as the item works earlier, for instance Bayesian neural systems and random forests. Notwithstanding picking various models, still, there may be various bits with the Gaussian procedure as the capacity earlier and additionally hyper-parameter enhancement.

REFERENCES

[1] Amgad Muhammad, Mohamed Moustafa" Improving Region Based CNN Object Detector Using Bayesian Optimization" IEEE International Conference on Image Processing, Applications and Systems (IPAS), 2018.
[2] Junliang Li, Hon-Cheng Wong "Multiple Object Detection by a Deformable Part-Based Model and an R-CNN" IEEE Signal Processing Letters, Vol. 25, No. 2, February, 2018
[3] Joshua Herrera, Mevlut A Demir, Parsa Yousefi, John J Prevost, and Paul Rad" Distributed Edge Cloud R-CNN for Real Time Object Detection" Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.
[4] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik "Rich feature hierarchies for accurate object detection and semantic segmentation" IEEE conference on Computer vision and pattern recognition, 2014.
[5] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," Int. J. Compute. Vision, vol. 88, no. 2, pp. 303–338, 2010.
[6] Xue Yang, Hao Sun, Xian Sun, Menglong Yan, Wenhui Diao, Zhi Guo "Object detection with head direction in remote sensing images based on rotational region CNN" IEEE International Geoscience and Remote Sensing Symposium, 2018.
[7] Anitya Kumar Gupta, "Time Portability Evaluation of RCNN Technique of OD Object Detection – Machine Learning (Artificial Intelligence)" International Conference on Energy, Communication, Data Analytics and Soft Computing, 2017.
[8] Wei Zhang, ShihaoWang, Sophanyouly Thachan, Jingzhou Chen, Yuntao Qian "Deconv R-CNN for small object detection on remote sensing images" IEEE International Geoscience and Remote Sensing Symposium, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY