iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: VIII Month of publication: August 2024

DOI: https://doi.org/10.22214/ijraset.2024.63581

Review on Object Identification Using Mobile Camera System

Sheetal R. Dehare[1], Pravin L. Yannawar [2]
Department of Computer Science and IT, Dr. B.A.M.U, Aurangabad, Maharashtra, India

Abstract: Over the past 20 years, object detection has developed into a productive research area and an extensively used task in computer vision. It aims to rapidly and simply find many different things in a given image using specified categories. Based on the training model approach, the algorithms can be separated into two groups: One is the detection single-stage algorithms. The other is the detection of two-stage algorithms. In this study, sample processes for every stage are briefly announced. Following a review and discussion of several fundamental techniques, this article introduces public and proprietary datasets that are often used in object detection. Thus, the challenges in object recognition are noted.

Keyword: object detection, Convolutional neural networks- (CNN), deep learning, single-stage object detection Model, and Two-Stage Object Detection Model.

I. INTRODUCTION

For humans, detecting objects is a simple task. A baby as young as a few months old can start recognizing common objects, but until the turn of the previous few years, teaching computers to do just that was difficult. It involves locating and identifying every occurrence of an object (such as vehicles, people, street signs, etc.) in the field of vision. Similar problems have existed with various tasks including classification, segmentation, motion estimation, scene interpretation, etc. in computer vision. Early Feature extractors like the Jones-Viola detector [1]. the histogram of oriented gradients (hog) [2], and others were used to create early object identification models. These models performed badly on unknown datasets and were slow and inaccurate. The field of visual perception was changed by the reintroduction of convolution neural networks (CNNs) and image classification for deep learning. Further research on using it in computer vision was inspired by using Image Net's 2012 Large Scale Visual Recognition Challenge (ILSVRC) by Alex Net [3]. Currently, object detection is used for security, medical, self-driving cars, identity identification, and other purposes. It has experienced exponential growth in recent years along with the quick development of new tools and approaches. Deep learning has been used for datasets due to technological advancements and the availability of powerful graphics processing units (GPUs); researchers in fields including object classification, detection, and recognition have shown state-of-the-art results. Deep learning requires powerful computer resources and larger data sets to complete both training and testing. Image classification is the most widely studied subject of computer vision, and it has achieved outstanding success in international competitions with the use of deep learning algorithms like PASCAL, ILSVRC, VOC, and MS-COCO [3].

Even while object detection might be valuable, it can be difficult to solve all of its issues. However, despite the advancement of techniques, issues still arise as a result of various lighting, pose, foreground geometry, and other aspects. There are significant efforts made to obtain the best outcomes from every input source. The methods of implementation and the outcomes of each technique vary based on the various factors as new techniques and improvements in object identification are introduced. These factors also determine which techniques can be used for certain applications to achieve the best outcomes. While some methods are more accurate, they can also take longer to process. While some achieve results with less accuracy and in far less time. The selection of techniques may change depending on these criteria.

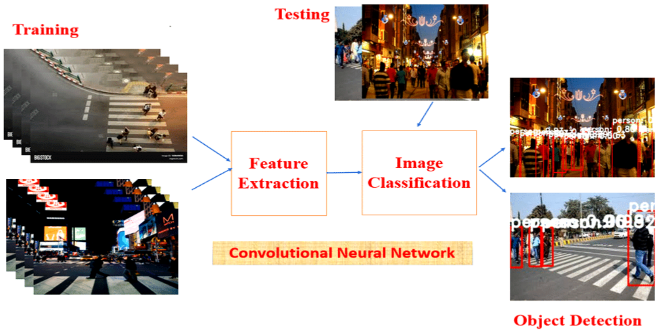

Figure 1. Basic Block Diagram of Object Detection [4]

II. LITERATURE SURVEY

In this section, the works carried out by various researchers are as follows:

1) Wei Liu [2016] SSD: Single Shot MultiBox Detector. The purpose of this study. The SSD model is simple in comparison to approaches that require object proposals since it completely eliminates proposal production and the subsequent pixel or feature resampling stage, concentrating all processing in a single network. Experimental results on the PASCAL VOC, MS COCO, and ILSVRC datasets illustrate that SSD has comparable accuracy to methods that use an additional object proposal step while being significantly faster. This is despite the fact that SSD provides a single framework for both training and inference. SSD offers noticeably higher accuracy even with smaller input image sizes than other single-stage methods. SSD for 300x300 input produces 72.1% mAP on an Nvidia Titan X at 58 frames per second. [5]
2) Bardia Doosti [2016] Localized object detection with Convolutional Neural Networks. In this, the PASCAL VOC 2007 dataset used, 20 images of dogs, airplanes, and cars. The dog-identified region received 61%, which is a great achievement. When it came to the car category, the algorithm did poorly, coming in the bottom with a 38% ranking, followed by a 19% ranking for the accuracy of the region. [6]
3) Joseph Redmon, [2016] You Only Look Once: Unified, Real-Time Object Detection, In this paper, they released YOLO, a unified model for object detection that is easy to build and can be trained on entire photos. The entire model is jointly trained, and a loss of function that approximately matches detection performance is used to train YOLO. [7]
4) Sandeep. Kumar [2017] In object detection and recognition of images, the EasyNet model is very simple to implement and build The EasyNet model can recognize many items and perform various forms of identification. In the future, the work can be extended by detecting moving objects with non-static backgrounds. [10]
5) Joseph. Redmon [2017] In this paper they introduced Yolov2 receives a 76.8 map at 67 frames per second on VOC 2007. Yolov2 gets 78.6 maps at 40 FPS. Even though just 44 of the 200 classes have detection data, 19.7 mAPs are awarded to the YOLO9000 on the ImageNet detection validation set. On the 156 classes, YOLO9000 received 16.0 mAP. That is not in COCO. However, YOLO predicts detections for over 9000 different item categories, indicating that it is able to detect more than simply 200 classes. It continues to work in real time. [8]
6) Andrew G. Howard [2017] In the study MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, SSD, Faster-RCNN, Squeezenet, AlexNet, Inception V2 Compared to well-known models, MobileNets has greater size, speed, and accuracy features, being used for a variety of purposes, MobileNet. [9]
7) Kaiming. He [2017] This paper introduced the Mask R-CNN Algorithm. It's easy to learn this. On every test, All currently available single-model entries, including the winners of the COCO 2016 contest, are outperformed by Mask R-CNN. [11]
8) Rohini. Goel [2019] Object Recognition Using Deep Learning. The uses of object recognition techniques for specific sorts of items (such as faces, buildings, etc., and plants) are also highlighted. In this, they evaluate the advantages and disadvantages of present techniques and future use. [12]
9) Shrey Srivastava [2019] Comparative analysis of deep learning image detection algorithms. This Paper compares there are three main methods for processing images: 1. SSD 2. Faster R-CNN, and 3. YOLO to determine which of the three is the quickest and most efficient. The study's findings show that the use of any algorithm relative to the other two is strongly impacted by the use cases for which it is most appropriate. The best of the three algorithms, YOLO-v3 outperforms SSD and Faster R-CNN in a similar testing scenario. [13]
10) Fan. Zhang [2020] DetReco: Object-Text Detection and Recognition Based on Deep Neural Network. After research, it is revealed in this study that the suggested technique YOLOV3, CRNN (Convolutional Recurrent Neural Network), and NMS (no maximum suppression) COCO, attain a mean average precision (mAP) for similar objects of 78.3mAP. In terms of detecting performance, it also receives 72.8 AP for the (Average Precision) of texts. [14]
11) Jun Deng [2020] A review of research on object detection based on deep learning. Although deep learning has been widely used in many sectors, they have developed detection techniques based on this technology in this work. However, deep learning still has some issues that need to be resolved.1) Reduce the requirement for data. 2) To obtain effective small object detection. 3) Understanding multi-category object detection [15]
12) Ms. K.Suitha [2021] Human and Object Detection Using Surveillance System, In this paper after review on different object detection. They find, that MobileNet, and SSD, these techniques need to concentrate on handling sudden illumination. They have identified & discussed the limitations/future scope of various methods. [16]

13) Syed Sahil Abbas Zaidi [2022] A Survey of Modern Deep Learning-based Object Detection Models, In this study, they describe the evolution of single-stage and two-stage detectors' successors. As compared to both it is identified that in general, two-stage detectors are more specific. Because of its slowness, it cannot be utilized for real-time applications like security or autonomous vehicles. One-stage detectors, on the other hand, are now faster and equally accurate as the former in the previous few years. [17]

Table1. Performance comparison among object detection networks[5]

System	VOC2007 TEST mPA	FPS	Number of Boxes	Input resolution
Faster R-CNN (VGG16)	73.2 7 1	7	~ 6000	~1000x600
Fast YOLO YOLO(VGG16)	52.7 66.4	7 21	98 98	448x448 448x448
YOLO (customized)	63.4	45	98	448x448
SSD300	74.3	46	8732	300X300
SSD512	76.8	19	24564	512x512
SSD300*(VGG16)	77.2	46	8732	300x300
SSD12	76.8	22	2456	12x512
SSD512*(VGG16)	79.8	19	24564	512x512

III. METHODOLOGY

Taking the visual as an input, either by an image or a video data set by mobile, webcam, etc.

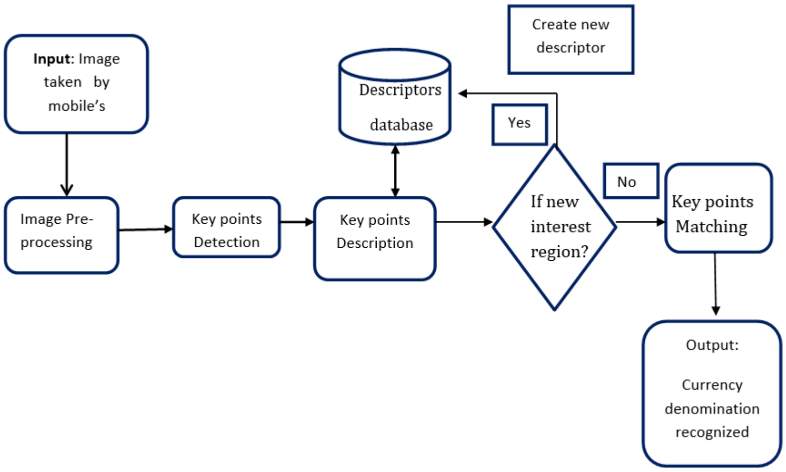

Figure 2. Object Identification Flowchart

Any of the methods mentioned in the previous section can be implemented using the same three key steps, namely the identification of feature points, the description of each area surrounding those points as a feature vector using a descriptor, and finally the use of a function that enables the comparison of descriptors in order to carry out the matching. Figure.2 illustrates the process used to find and identify the objects.

Deep Learning Base Object Recognition Methods: There are two types of region-based convolutional neural network models:

1) Two-Stage object detection model
2) Single-Stage object detection model

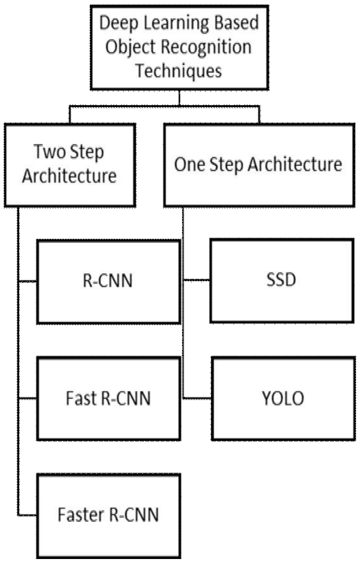

Figure 3. Steps Architecture

A. Two-Stage Target Detection Algorithm

1) R-CNN

The R-CNN [18] was the first research work on R-CNNs, and it showed how CNNs may be used to increase and extend detection performance. To convert detection into classification, R-CNN combines CNNs and a module for class-neutral region proposals. Girshick introduced the R-CNN algorithm in 2014 as the first interesting model to use Convolution Neural Networks (CNN) for object detection. For classification, the model creates 2000 region proposals per image and resizes them. Furthermore,

CNN is used for feature extraction and model training, and an SVM classifier is used to identify objects. Finally, a bounding box regression process is performed using a trained linear regression model. R-CNN does significantly improve accuracy when compared to a traditional detection method, but it requires a lot of calculations and does it inefficiently. Second, converting a fixed-length feature vector and the region proposal directly might damage the objects.

Drawbacks of the R-CNN Model

The algorithm for selective search is not flexible. Due to its slow speed, it cannot be used for real-time processing. Training takes a long period because there are several stages. R-CNN is rarely employed in real-world applications.

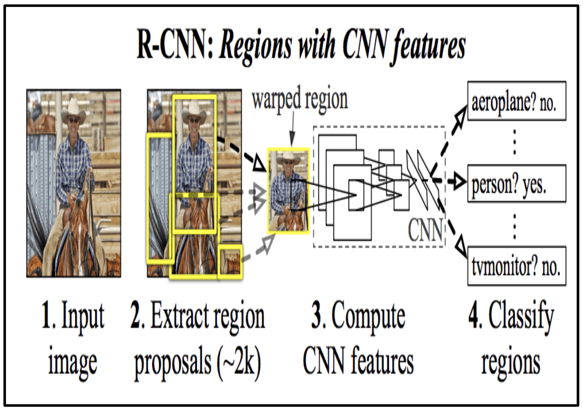

Figure 4. R-CNN architecture [18]

2) *FAST R-CNN*

The replacement for the original R-CNN is called Fast R-CNN. For accurate object detection, a number of proposed regions are developed using the selective search algorithm.

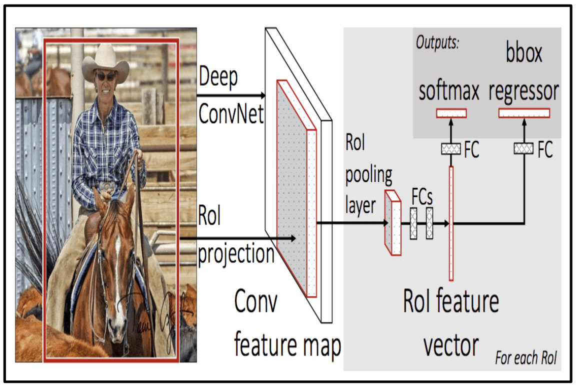

Figure.5. Fast RCNN architecture [19]

In 2015, Girshick presented the Fast R-CNN [19] model. A Fast R-CNN design takes the complete image as input along with item suggestions, as illustrated in Fig. 5. Max-pooling and convolutional layers are used to process the image, creating a conv feature map. The layer of ROI pooling pulls the feature map's fixed-length feature vector for the respective object proposal. The various layers receive as input all feature vectors. Two separate outputs are produced: Softmax Each object class has four real-valued numbers as well as probability for different object classes. Fast R-CNN uses less processing power while improving detection accuracy. CNN receives the entire image as input in order to extract features.
Uses: Instead of SVM, Fast R-CNN uses SoftMax as in R-CNN, for classification. The model is faster and end-to-end trainable because of a pooling layer called Region of Interest (ROI).
Limitations: The process is lengthy and complex. Calculations take a very long time. The model's computation time is approximately 2 seconds.

3) *Faster R-CNN*

Regional convolutional neural network (R-CNN) Faster R-CNN is the next part in the series. [20] The model proposed by Ren et.al, replaces the previous Selective Search method to create region proposals with region proposal networks, as seen in Fig. 6. The model is made of 2 modules: the Fast R-CNN detection method and a full convolution neural network used to generate all region proposals. These two modules share a set of convolution layers. The shared convolution layer receives the input image and passes it through the CNN network at the very end. In order to create a higher-dimensional feature map, the picture is passed forward to the specified convolution layer, and the feature map is used as the RPN network's input, respectively. Even though Faster R-CNN has good detection accuracy, real-time detection is still not possible with it.

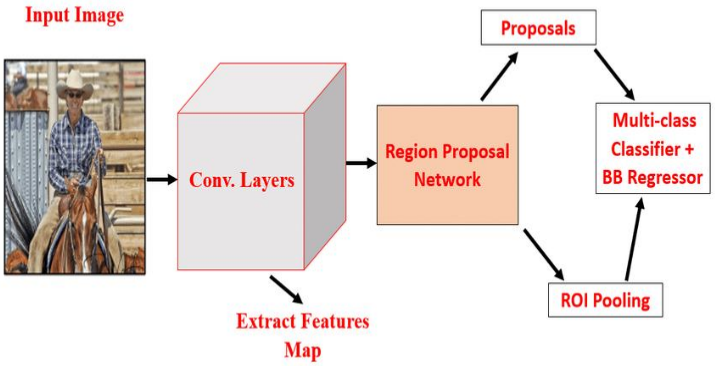

Figure 6. Faster RCNN architecture [20]

B. One-Stage Target Detection Algorithm:

1) SSD

In 2016, W. Liu et al [21] proposed the Single-Shot Detector (SSD) concept, which is seen in Figure.7. SSD uses a single pipeline for completing object detection, making it a one-stage approach. Since SSD only requires one step to complete, it has an increase in performance over the R-CNN series method and is hence more likely to be used in engineering. Using a convolutional neural network features from an image, which is then sent into the SSD process to create a feature map. Six layers' worth of feature maps are extracted, and the default box is then generated for each feature map point. The default box is then filtered and output once all the created default boxes have been gathered and placed into the Non-Maximum Suppression. The feature maps of some layers only need to be trained for object detection at a matching scale since the visual fields of perception of the different feature maps are scale-dependent. As a result, SSD combines the upper and lower-level feature maps and uses multi-scale regional features for regression. SSD also uses the previous framework of different scales and aspect ratios. It somewhat solves the issue of incorrect positioning and problems in identifying small things.

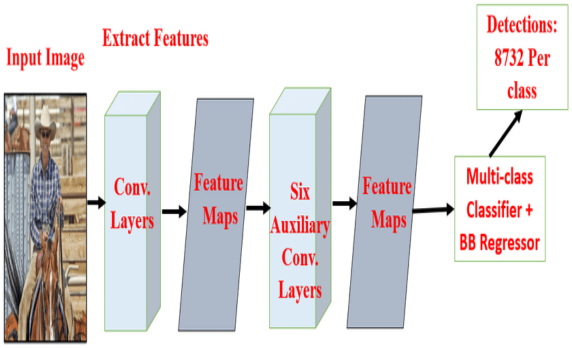
Input Image
Extract Features
Detections: 8732 Per class
Conv. Layers
Feature Maps
Six Auxiliary Conv. Layers
Feature Maps
Multi-class Classifier + BB Regressor

Figure 7. Single-shot multi-box detector architecture [21]

2) YOLO

Previous object detectors using deep learning are multi-stage, sequential, two-stage finders that have difficulties with real-time applications. A single-stage network with quick frame processing speed is needed to overcome this. In a pipeline as shown in Fig. 8 such a model was YOLO, which combined the tasks of classification and regression to determine the class of an object and its location. Later, YOLO-v2 [8], YOLO-v3 [22], and YOLO-v4 [23] were released in successive waves as improvements to the real-time application for object detection. The YOLO versions have higher generalization capabilities on other unknown datasets and are accurate, quick, and easy to use.

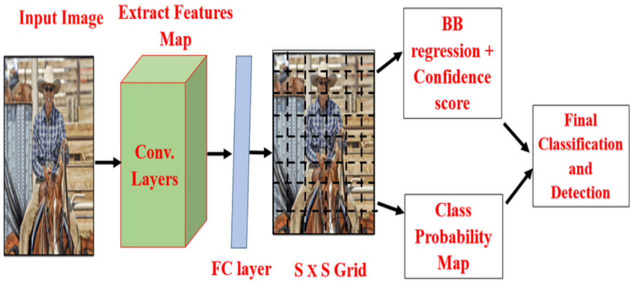

Figure 8. You only look once architecture

IV. DISCUSSION

In recent times one of the toughest and most basic subjects in computer vision is object detection though it is tough yet it has gotten a lot of attention in the current period. After studying these review papers according to my point of view I recognized that there are variations of algorithms for object detection, i.e Region proposal-based, and the other regression classification-based frameworks, whereas comparison both the Region-based is composed of 3 correlated stages including region-based, proposal generation, Feature extraction, and classification.

The Regression-based it is one step framework mapping straightly from image pixels to bounding box coordinates and class probabilities that can reduce time expenses. SSD, RCNN, FAST-RCNN, FASTER-RCNN, YOLO, etc. this is the different algorithms used by the researchers in their research in this most of the algorithms which are used are SSD and YOLO. SSD one of the main rewards is its speed and efficiency. Because a single network is used, objects can be detected in real-time. But still, there are some issues it may struggle to see items that are much smaller or larger than the objects in the training sample. Although DL-based detection algorithms have come to be common in various fields of diligence, there still exist a number of problems that must be addressed in order to achieve a high level of learning efficiency for small objects and proposed an improved-category object discovery. DL requires a large amount of data in order to produce good results.

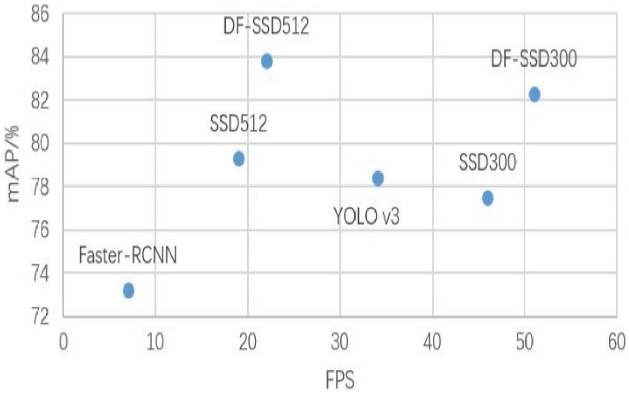

Figure 9. Improved single-shot multi-box detector target detection method based on deep feature fusion [26]

V. CONCLUSION

Though object detection has advanced significantly over the last year, the finest detectors are still a long way from reaching performance capacity. The demand for lightweight models that may be used on mobile and embedded devices will rise rapidly as their real-world applications increase. Although there has been a rise in interest in this area, there are still many unresolved questions. In this study, we show the evolution of single-stage and two-stage detectors over their predecessors. This study provides a detailed overview based on deep learning object identification frameworks also classification mechanisms. Finally, from the aforementioned methodologies, it can be inferred that Single Shot Detector (SSD) has the greatest mAP of all the techniques, at 76.9%. This review is especially useful for improvements in learning systems connected to neural networks since it offers useful understandings and also suggestions for upcoming advancement. So our approach is using Android, and the MobileNet SSD model we will be performing future work using this we will try to improve the accuracy of Object detection and work on the Real Time Application, which will detect the object and speak the object name.

REFERENCES

[1] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. I–511–I–518, (2001).

[2] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. 886– 893, ISSN:1063-6919, (2005-06).

[3] Krizhevsky A., Sutskever I., Hinton, G.E. “Imagenet classification with deep convolutional neural networks”, Neural Information Processing Systems; CurranAssociates, Inc, SanDiego, CA, pp. 1097–1105 (2012).

[4] Murthy C. B., Hashmi M. F., Bokde N. D., & Geem Z. W. “Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review”, Applied sciences, 10(9), 3280. (2020).

[5] Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C. Y., & BergnA. C. “SSD: Single Shot Multibox detector”, European conference on computer vision (pp. 21-37). Springer, Cham. (2016- October).

[6] Doosti, B., & Avula, V. H. “Computer Vision Final Report: “Localized object detection with Convolutional Neural Networks”, (2016).

[7] Redmon J., Divvala S., Girshick R., & Farhadi A. “You only look once: Unified, real-time object detection”, IEEE conference on computer vision and pattern recognition, pp. 779-788, (2016).

[8] Redmon Joseph and Ali Farhadi, “YOLO9000: better, faster, stronger”, IEEE conference on computer vision and pattern recognition (2017).

[9] Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., & Adam, H., “Mobilenets: Efficient convolutional neural networks for mobile vision applications”, preprint arXiv:1704.04861, (2017).

[10] Kumar S., Balyan A., & Chawla M., “Object Detection and Recognition in Images”, International Journal of Engineering Development and Research, 5(4), 1029-1034, (2017)

[11] He K., Gkioxari G., Dollár P., & Girshick R., “Mask R-CNN”, IEEE international conference on computer vision (pp. 2961-2969) (2017).

[12] Goel R., Sharma A., & Kapoor R., “Object Recognition Using Deep Learning”, Journal of Computational and Theoretical Nanoscience, 16(9), 4044-4052,

(2019)

[13] Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V., “Comparative analysis of deep learning image detection algorithms”, Journal of Big Data, 8(1), 1-27, (2021)

[14] Zhang F., Luan J., Xu Z., & Chen W., “DetReco: object-text detection and recognition based on deep neural network”, Mathematical Problems in Engineering, 2020.

[15] Deng J., Xuan X., Wang W., Li Z., Yao H., & Wang Z., “A review of research on object detection based on deep learning” Journal of Physics: Conference Series (Vol. 1684, No. 1, p. 012028). IOP Publishing. (2020, November).

[16] Suitha M. K., Naidu V. B., Shankar P. G., & Kumar P. H. “Human & Object Detection Using Surveillance System”. (2021)

[17] Zaidi S. S. A., Ansari M. S., Aslam A., Kanwal, N., Asghar M., & Lee B., “A survey of modern deep learning-based object detection models”, Digital Signal Processing, 103514. (2022).

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (June 2014).

[19] Girshick, R. “Fast R-CNN”, IEEE international conference on computer vision. Santiago, pp. 1440-1448, (2015).

[20] Ren S.Q., He, K.M., Girshick R., Sun, J., “Faster R-CNN: towards real-time object detection with region proposal networks”, Advances in neural information processing systems. Montreal, pp. 91-99, (2016).

[21] Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.Y., Berg A.C., “SSD: Single shot multibox detector”, European Conference on Computer Vision, Amsterdam, The Netherlands, Springer: Berlin, Germany, pp. 21–37, (8–16 October 2016)

[22] Redmon Joseph and Ali Farhadi, “Yolov3: An incremental improvement”. arXiv (2018).

[23] Jeong, Jisoo, Hyojin Park, and Nojun Kwak. "Enhancement of SSD by concatenating feature maps for object detection." *arXiv preprint arXiv:1705.09587*,2017

[24] C Bhagya and A Shyna, “An Overview of Deep Learning Based Object Detection Techniques”, 1st International Conference on Innovations in Information and Communication Technology (ICIICT) (2019).

[25] Ashwani Kumar, Zuopeng Justin Zhan, Hongbo Lyu, “ Object detection in real-time based on improved single shot multi-box detector algorithm”, Kumar et al. EURASIP Journal on Wireless Communications and Networking,2020https://doi.org/10.1186/s13638-020-01826.

[26] Bai D, Sun Y, Tao B, Tong X, Xu M, Jiang G, Chen B, Cao Y, Sun N, Li Z. Improved single shot multibox detector target detection method based on deep feature fusion. Concurrency and Computation: Practice and Experience. 2022 Feb 15;34(4):e6614.