Recognition of Objects Using Fast RCCN-Hybrid Particle Swarm Firefly Algorithm

Satish Kumar Patnala [1], K Suresh Kumar [2], Killi Bhushan Rao [3], K Praveen Kumar [4], Venkata Anusha Kolluru [5], Parimila Garnepudi [6], Shaik Mahaboob Basha [7], Ravi Kumar Munaganuri [7], Mamatha Maddu [7], and Narasimha Rao Yamarthi [7,*]

[1] Department of CSE (AIML), GMRIT, Rajam, AP, India
[2] Department of CSE, Narasaraopet Engineering College, Narasaraopeta, Guntur, AP, India
[3] Department of CSE, KLU Deemed to be University, Guntur, AP, India
[4] Department of Mechanical Engineering, RVR&JC College of Engineering, Guntur, AP, India
[5] Department of CSE (AIML), RVR&JC College of Engineering, Guntur, AP, India
[6] Department of CSE, VFSTR deemed to be University, Vadlamudi, Guntur, AP, India
[7] School of Computer Science and Engineering, VIT-AP University, Amaravati, AP, India
Email: satishkumar.p@gmrit.edu.in (S.K.P.); sureshkunda546@gmail.com (K.S.K.); kchbhushanarao@kluniversity.i (K.B.B.); kpraveen717@gmail.com (K.P.K.); kvanusha@rvrjc.ac.in (V.A.K.); parimala@gmail.com (P.G.); mahaboob.23phd7064@vitap.ac.in (S.M.B.); ravi.22phd7123@vitap.ac.in (R.K.M.); mamatha.22phd7103@vitap.ac.in (M.M.); y.narasimharao@vitap.ac.in (N.R.Y.)
*Corresponding author

Abstract—**In this study, we propose an enhanced object detection framework by integrating Faster Recurrent Convolutional Neural Network (F-RCNN) with a hybrid Particle Swarm-Firefly Algorithm (PSFA) for optimization. The model aims to improve both detection accuracy and computational efficiency for large-scale image datasets. The framework begins with preprocessing and segmentation using K-means clustering, followed by feature extraction via F-RCNN. To optimize the network's training process, we incorporate PSFA, leveraging the exploration capabilities of Particle Swarm Optimization (PSO) and the adaptive brightness-based search mechanism of the Firefly Algorithm (FA). The proposed model is evaluated using PASCAL VOC 2007 dataset, COCO dataset and ImageNet dataset. Objects like car, ship, cat, dog, horse, person etc are focussed and recognition of the objects in images are done from different class of objects. The integration of optimization with deep learning method gives a promising improvement in obtaining the objects of various classes. The mean average precision in detection of objects is evaluated using matlab software tool. The overall average precision rate achieved is 84.6% for PASCAL VOC 2007 DS, 86.8% for COCO DS and 87.4% for ImageNet DS which is higher compared to other techniques like faster recurrent convolutional neural network (RCNN), Mask recurrent convolutional neural network (M-RCNN) and Region-Based Fully Convolutional Network (R-FCN).**

Keyword—**K-mean clustering, deep learning, convolutional neural network, fast recurrent CNN, particle swarm optimization, firefly algorithm**

I. Introduction

In the fields of image processing and computer vision, object detection and recognition are a new and rapidly developing issue. A person can quickly and simply analyse an image. With only a single look, humans can comprehend images and extract all relevant information from them. The images consist of a greater number of objects with same class and the positions of the objects are located nearly. So, nowadays deep learning methods have gained more importance in detection of objects [1]. Large variations in the objects that are present in remote sensing images need to be recognized in a deeper manner due to the visual quality of images and background disturbances caused in the images. The machine learning method like deep neural networks have been implemented by author in [2]. Further region based convolutional neural networks have been proposed where it performs convent forward pass for every object which appears in the image were each layer has its own specific role, hence the process is slow and rate of recognition is not found to be effective and for the purpose of visual based recognition a spatial pyramid type pooling in deep neural networks is proposed in [3].

Machine learning based objection detection methods are emerging in present scenario, where feature extraction and classification plays a key role. Some of the existing methods used for extraction of features in histograms of gradients [4], deep neural networks [5]. Most of the methods follow traditional methods in recognition of objects. In this paper, neural networks and meta-heuristic techniques is been introduced to improve the detection rate and make it fast for the users in applications where the model is required.

The paper provides a deep discussion of existing model in section II. The main contribution is integrating hybrid optimization model i.e., particle swarm-firefly algorithm

Manuscript received December 3, 2024; revised February 5, 2025; accepted March 17, 2025; published June 12, 2025.

doi: 10.18178/joig.13.3.293-303

with faster recurrent convolutional neural network. This paper designs a specific model showing the process of object detection using Faster Recurrent Convolutional Neural Network Particle Swarm-Firefly Algorithm (FRCNN-PSFA) model. The suggested model can be integrated with 6G network for improving the detection accuracy by leveraging latency and higher bandwidth. The problem while processing 6G Internet of Things (IOT) object detection is its hardware structure which is very costly infrastructure and power utilization is high. The model is tested using PASCAL VOC 2007 dataset, MS COCO dataset and ImageNet dataset.

II. Related Work

With ongoing technological advancements, detection of objects has been thoroughly and aggressively researched. Most early detectors submit their features to a well-trained classifier for classification after immediately extracting them from the deformable component model. Review on object detection conducted by author in [6]. The Histogram of Gradient (HOG) model detection [7], Region based Convolutional Neural Network (CNN) object detection [8]. However, many manually created features struggle to satisfy the requirements of the application and have poor generalization capabilities.

Object detection models are performed depending on various conditions. The author in [9] suggested detection of objects in foggy weather condition. Restoration of image is performed. In [10] underwater based sonar images are considered. A sliding window pre-processing is performed for object detection. The author in [11] considered snowfall condition for detection of objects.

For the classification module to reuse the convolutional network features independent of the input image's resolution and increase detection speed, the author in [12] later proposed SPPNet by integrating traditional Spatial Pyramid Pooling (SPP) into the R-CNN architecture. Although Fast R-CNN [13] is an enhancement of SPPNet that enables simultaneous training of classification and regression tasks, it still depends on external area recommendations. Following that, the first detector that permits end-to-end training was Faster R-CNN [14]. Several aspects of object identification are combined into a single deep learning framework by implementing an efficient and accurate Region Proposal Network (RPN) [15] to generate region suggestions. Furthermore, efficient extension techniques have been put out to improve the prediction performance even more.

The study produced significant improvements in mean Average Precision (mAP), showing an increase of almost 30% above the previous state-of-the-art findings on the PASCALVOC dataset. The author in [16] introduced a technique that combines contextual understanding and real-time item detection. Their approach uses Deep Neural Networks (DNN) with various parameters to effectively identify and recognise things. Faster R-CNN [17] and R-Fully Convolutional Network (FCN) [18] are the best feature maps among object detectors for forecasting candidate area proposals of different sizes and aspect ratios. However, because the receptive field in the top feature map is fixed, objects of different scales clash with the fixed receptive field in the natural image, limiting the ability to predict objects that are too big or too small and potentially impairing object detection performance.

The author in [19] suggested a review on Object Detection Using Deep Learning, CNNs and Vision Transformers. Deep learning techniques are utilized in some of the object detection applications. The author in [20] utilized sonar images for detection of objects in underwater navigation by using GoogleNet CNN model. The 6G IOT for future intelligence and automatic object detection is proposed by author in [21]. As a core partition of future 6G networks, Space-Air-Ground Integrated Networks (SAGIN) [22] have been envisioned to provide countless real-time intelligent applications. The author in [20] suggested the vision of using Fuzzy Logic/ Quantum Fuzzy Logic (FL/QFL) in SAGINs and presented a few representative applications enabled by the integration of FL and QFL in SAGINs. The author in [23] discussed about ccombining edge computing architecture, which provides low computing costs and latency, with strong Artificial Intelligence (AI) [24] detection and prediction skills might result in smart healthcare applications. The object detection in healthcare is very important model. The roles of 6G in a wide range of prospective IoT applications via five key domains, namely Healthcare Internet of Things, Vehicular Internet of Things and Autonomous Driving, Unmanned Aerial Vehicles. To better understand the deep learning concepts and stepping into 6G IoT [25]. The use of optimization techniques in deep learning is discussed in this paper. The enhancement of Deep Learning (DL) and optimization model can also be incorporated in IoT model. The involvement of optimization techniques in challenging tasks help is identifying a solution for the proposed task. In this paper a hybrid optimization model is combined with deep neural network to achieve the best results in object recognition [26].

III. Methodology

The proposed model designed is a combination of deep learning and optimization model which effectively recognize the objects. Initially the input dataset is pre-processed, perform segmentation in which the image labels can be identified. Further perform feature extraction and detect the object. The proceedings of proposed work are shown in Fig. 1.

A. Preprocessing

It is one of the important steps which help in improving the quality level of image by removing the environmental noise present in the image. The images are enhanced using sharpening filters for better recognition of objects. In this process a Gaussian Filter (GF) [27] is used to remove the noise present in the image. Smoothing is performed before applying to the filter. The image consists of three colours Red Green Blue (RGB) which passes through the GF to improve the quality of image. The GF function is given in Eq. (1).

$$GF\ (x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\frac{-[(x-\mu_x)^2+(y-\mu_y)^2]}{2\sigma_x\sigma_y}} \tag{1}$$

where x and y are the horizontal and vertical axis distances from the center, respectively while μ is the mean value of image and σ is standard deviation of the images which depend on the pixels of the images.

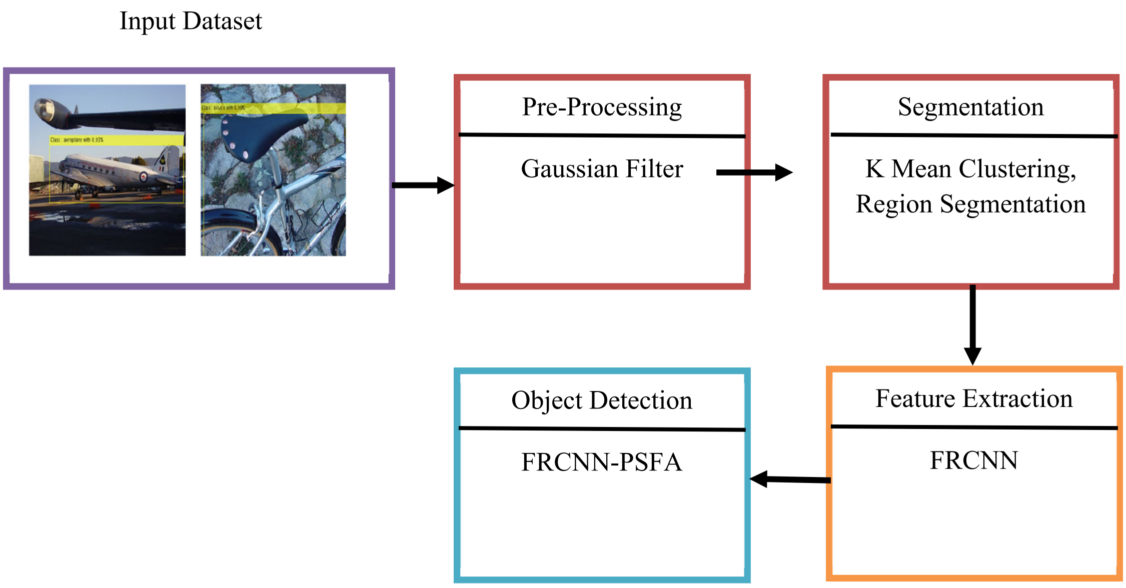

Fig. 1. Process flow of proposed model.

B. Segmentation

After preprocessing the images in the dataset are properly segmented before fetching to CNN model [28]. The segmentation step helps the model to meet the cost effectives of the design. In this process two methods are utilized for segmentation and is been implemented. One is Clustering using K mean and other is regions-based segmentation. The output both segmentation techniques are combined and fetched the image for FRCNN.

1) Clustering using K mean

After the completion of pre-processing stage, the objects need to identified based on the level of intensity, colour, and region. A dataset's components are grouped according to how similar they are using the K-means approach [29]. Homogeneous colour patches are clustered using the k-mean approach, which requires no previous knowledge other than the initial number of clusters, k. White is the colour when all three values are 255; black is the colour when all three values are muted or zero. Consequently, we shall obtain a certain pixel colour shade by combining these three. The values vary from 0–255 as each integer is an 8-bit number. Eq. (2) describes K-means clustering, which uses Euclidean distance to determine how similar two images are.

$$\mathrm{ED_{xy}} = \sqrt{(\mathrm{x_2} - \mathrm{x_1})^2 + (\mathrm{y_2} - \mathrm{y_1})^2} \tag{2}$$

where ED_{xy} is represents the distance between two data points x and y respectively. Every cluster in K-mean has a centroid. First, each cluster's centroids are selected at random, and the Euclidean distance between each item and the cluster centroid is calculated. The object will thus become a member of the nearest cluster [30]. When an image enters a cluster, the mean is used to determine the cluster's new centroid. This procedure is continued until every object in the same cluster is still there.

2) Region oriented segmentation

Applications for image segmentation are many and have been used to almost every associated field of image processing as well as a wide variety of picture types [31]. Two careful task which need to be execute are object detection and multi-object detection can be greatly improved by working on them simultaneously and passing information from one to the other. Using region-based segmentation, the commonalities between nearby pixels are seen. A separate zone will be formed by pixels that have comparable properties [32–33]. In the article, reference intensity levels for the area at each pixel are compared to neighbouring pixels in a picture. We employ similarity metrics, such grey level discrepancies, for areas with uniform grey levels. To prevent joining different parts of the image, we use connectivity. The neighbouring pixel is chosen if the difference, as shown by Eq. (3), is less than or equal to the difference threshold.

$$|I[x(i)] - [x(j)]| < Th \tag{3}$$

These segmentation models are performed on the input dataset which helps to improve the performance of R-CNN model. The task will be easy if the image is segmented as the K-Mean groups all the pixel with similar traits which helps in identifying significant features. To guarantee that the CNN concentrates on the most pertinent characteristics, region-based segmentation separates important items or regions of interest within an image.

C. Extraction of Features

The FRCNN model is utilized to extract the features. Initially input image and the objects which are need to be detected are taken as a set and given to faster R-CNN network. The feature maps need to be assessed by processing the entire image which is given as input with all layers of convolutional and pooling layer. As RCNN is a region-based method the regions are extracted based on the length of fixed vectors. Next step is featuring extraction, where the features of RGB image with a fixed resolution is forwarded through five layers of convolutional and two connected layers. The model of FRCNN is shown in Fig. 2. To pass through convolutional layers the image data need to be converted to CNN compatible data [34]. Our object detection system consists of three modules. The first generates category-independent region proposals. These proposals define the set of candidate detections available to our detector. The second module is a large convolutional neural network that extracts a fixed-length feature vector from each region. Finally, third module is we perform classification. The entire process is performed on PASCAL VOC 2007 dataset, MS COCO dataset and ImageNet dataset.

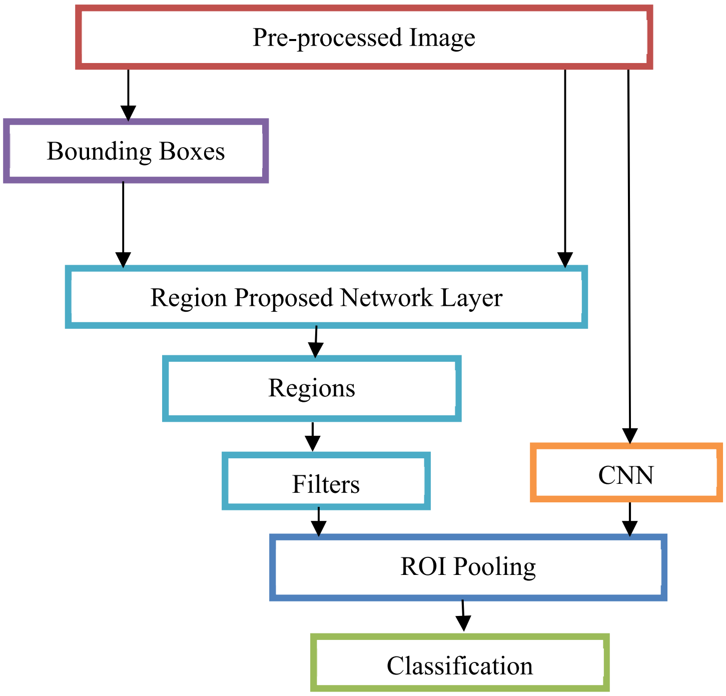

Fig. 2. Process flow of Faster-RCNN.

The weights updated by using FRCNN [35] is fed to hybrid firefly so that more accurate and better results can be obtained. The hybrid model is combination of particle swarm and firefly algorithm. The hybrid model is discussed in below section. In the faster RCNN anchors which are termed as bounding boxes are present. Here each object is surrounded with a bounding box. The trained data set consists of bounding boxes and which works well for the image datasets. The use of bounding boxes increases the speed of detecting the object and improves in providing high rate of accuracy. The required object will be surrounding with a bounding box and the feature extraction performed based on the objects, were all the similar features of the objects present in the bounding boxes are extracted, and regression process is performed in different layers so that major feature vectors are extracted.

D. Particle Swarm Firefly Algorithm (PSFA)

Both the techniques Swarms and flies are approximately near to each other. FA, it mimics the response of fireflies with respect to the nature behaviour in the atmosphere. Fireflies communicate with each other, try to find the pray and identifies mates using the light pattern which gets flashed. Firefly algorithm is based on some of the ideal traits of fireflies.

- All fireflies are defined as unisex so that the fireflies get attracted irrespective of their sex.
- Here we compared the relation between attractiveness and brightness which are proportional to each other, thus for any two blinking fireflies, as the proportionality is considered between attractiveness and brightness the distance between the flies increases gradually and the less bright one will move towards the brighter one. When there is no firefly which is brighter than the firefly takes a step towards the other fly.
- The brightness of a firefly is affected or determined by the landscape of the objective function.

The firefly algorithm is improved by combining swarm optimization which can be termed as improved version of flirefly algorithm. In this model we go with three main executions. One, is improving the distance between the files. Second, is to improve the fitness function and third is involving Swarm search criteria [36]. The parameters and function involved in this process is discussed. Initially the fitness function needs to evaluated which implicates the quality of the boundary boxes detection. The fitness function for object detection utilized in this work is given in Eq. (4).

$$Fit_f = \alpha.IoU + \beta.C_s - \gamma.L_{BB} - \delta.E_{error} \quad (4)$$

where, $\alpha, \beta, \gamma, \delta$ are the weight coefficients. Here, IoU is the intersection over union which is the accuracy of the bounding box; C_s is the confidence score; L_{BB} is the loss associated with the bounding box; E_{error} is error in the alignment of the edges which ensure the alignment of object boundaries.

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union} \quad (5)$$

$$C_s = Prob_{BB}(object) \times Prob_{CC}(object) \quad (6)$$

$Prob_{BB}(object)$ is the probability of the object that is the appeared in the bounding box; $Prob_{CC}(object)$ is the probability of the object detected which belongs to the correct class.

$$L_{BB} = |x_{pre} - x_{GT}| + |y_{pre} - y_{GT}| + |w_{pre} - w_{GT}| + |h_{pre} - h_{GT}| \quad (7)$$

where, x_{pre}, y_{pre}, w_{pre}, h_{pre} are the value of the pixels in the predicted bounding box and x_{GT}, y_{GT}, w_{GT}, h_{GT} are the ground-truth values of the pixels.

$$E_{error} = \sum |Img_{edges}(x, y) - BB(x, y)| \quad (8)$$

where, $Img_{edges}(x, y)$ is the edge coordinates of the image and $BB(x, y)$ is the mask of the pixels in the bounding box.

The weight coefficients are given as $\alpha = \frac{1}{1+e^{-(IoU-0.5)}}$; $\beta = 1 - \frac{e^{-C_s}}{1+e^{-C_s}}$; $\gamma = \frac{1}{1+e^{-L_{BB}}}$; $\delta = \frac{1}{1+e^{-E_{err}}}$.

The search efficiency will be improved by combining any two optimization models. In this paper, a hybrid model which is combination of PSO and FA termed as PSFA. Every particle is represented as bounding box and is based on Eq. (9).

$$V_{pi}(t+1) = w_i V_{pi}(t) + c_1 r_1 (P_{ibest} - X_{pi}) + c_2 r_2 (G_{best} - X_{pi}) \quad (9)$$

where, V_{pi} is the velocity of the swarm particles, X_{pi} is the position of bounding box coordinates, w_i is the weight of inertia, $c_1 c_2$ are the social coefficients of the particle, $r_1 r_2$ are the random values, P_{ibest} is the best position of i^{th} particle and final global best of the particle is G_{best}.

The firefly-based functioning of the objects depends on the concept of brightness (B) and attractiveness (A). The brightness of the i^{th} firefly that represent position of bounding bx is given in Eq. (10) and the best position is given in Eq. (11).

$$B_i = e^{-\gamma d_{ij}^2} \quad (10)$$

$$X_{ibest} = X_i + \beta e^{-\gamma d_{ij}^2}(X_j - X_i) + \alpha(ramd - 0.5) \quad (11)$$

where the term γ is the coefficient of light absorption; d_{ij} is the distance between the i^{th} and j^{th} firefly; The attraction factor is termed as β and the random movement factor is α.

Hybrid PSFA Algorithm

Step1. Initialize the swarm particles and the fireflies (bounding boxes)
Step2. Evaluate the fitness function using Eq. (4)
Step3. Update the Position and velocity of PSO using Eq. (9)
Step4. Update the firefly best position using Eq. (11)
Step5. Checking the termination condition
    Best position reached
    Maximum number of iterations
    end
Step6. **Return** the best bounding box

E. Detection Process: F-RCNN- Hybrid PSFA Process

In this considered work a hybrid optimization technique is proposed by combining particle swarm with firefly algorithm. This hybrid optimized technique is combined with faster RCNN to obtain high level of rate of precision. The below flowchart shows the proposed model process.

In the proposed methodology first, Faster RCNN is trained with the help of proposed weights which are predefined in neural networks. The dataset consists of images and all the images are tuned into even dimension; n number of nodes are taken as input. The training is performed in the layers and one encoded vector will be consider as output. After the process of training these weights will be optimized and converted into particles. The PSO further optimizes the weights as the weights are converted as particles. After getting optimized Fireflies are generated. In which the swarm particles are replaced with Fireflies and the search position is updated with respect to intensity of light to obtain the best solution. Finally update Faster RCNN with the weights obtained after applying hybrid optimization technique. Hence the output prediction is calculated and the results are shown. The main aim is to increase the prediction accuracy rate. The process of hybrid algorithm is shown in Fig. 3.

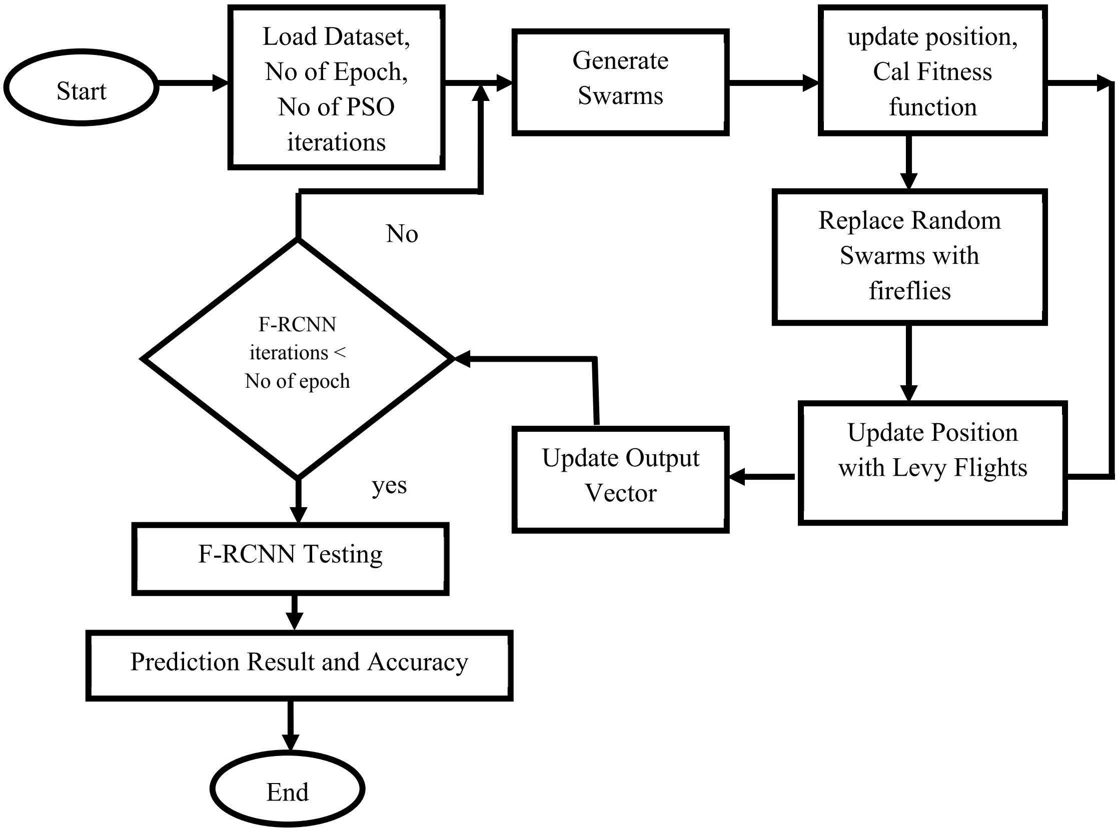

Fig. 3. Proposed flowchart.

IV. Experimental Results

A. System Environment

The results are evaluated using windows 11 OS with Intel i7 processor, 16Gb RAM and 8Gb Graphic card. The configuration of the system helps to improve the processing speed with the large input dataset. The data is evaluated using the proposed model using Matlab software environment. Here the dataset utilized is divided were 70% is given to training and 30% is given for testing.

B. Dataset

PASCAL Visual Object Classification (PASCAL VOC) 2007 is a familiar and widely used dataset for object detection with about 10,000 training and validation images with objects and bounding boxes. There are 20 different categories in the PASCAL VOC dataset. The dataset contains different types of objects such as train, cycle, ship, van, chair, cat etc. Some of the images in dataset contains multiple objects. The MS COCO dataset consists of 330K images with 80 different object categories. The ImageNet dataset consists of 14197122 images. Since 2010 the dataset is used in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [37], a benchmark in image classification and object detection. The publicly released dataset contains a set of manually annotated training images.

C. Analysis

Testing of dataset is performed, where the features of the objects are extracted based on the bounding boxes. Once features are extracted and labels are applied for the training data and compute the RCNN features for classification. The features will ignore the bounding boxes irregular shapes of the region. Similar regions are extracted, so that object in the image will be recognised and detected. The training and testing of faster RCNN has no loss in detection of performance. The experimental results achieved using the matlab software tool are shown below (Figs. 4–7).

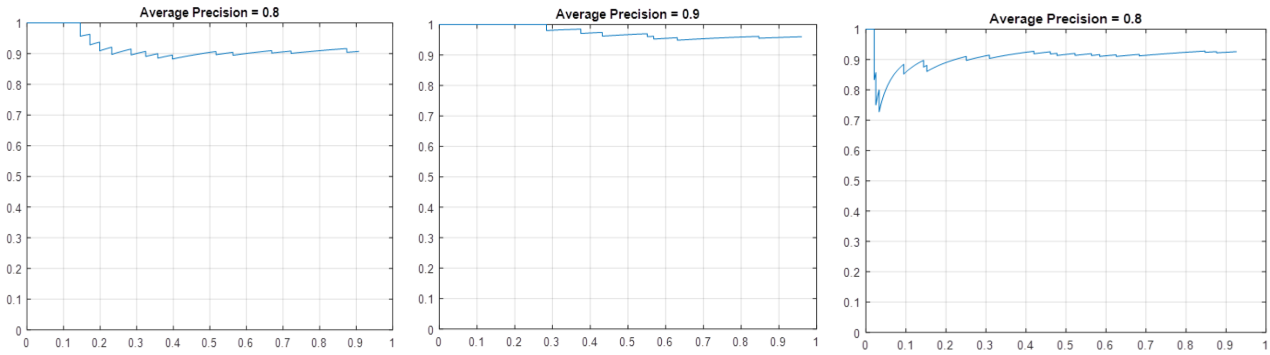

Fig. 4. Images and their respective average precisions.

Fig. 5. Ship, glass and van detection and its precisions.

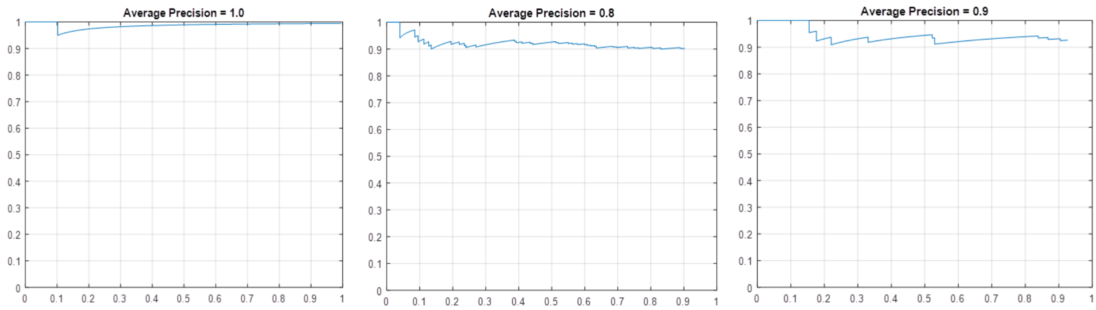

Fig. 6. Cat, chair and cow detection and its precision.

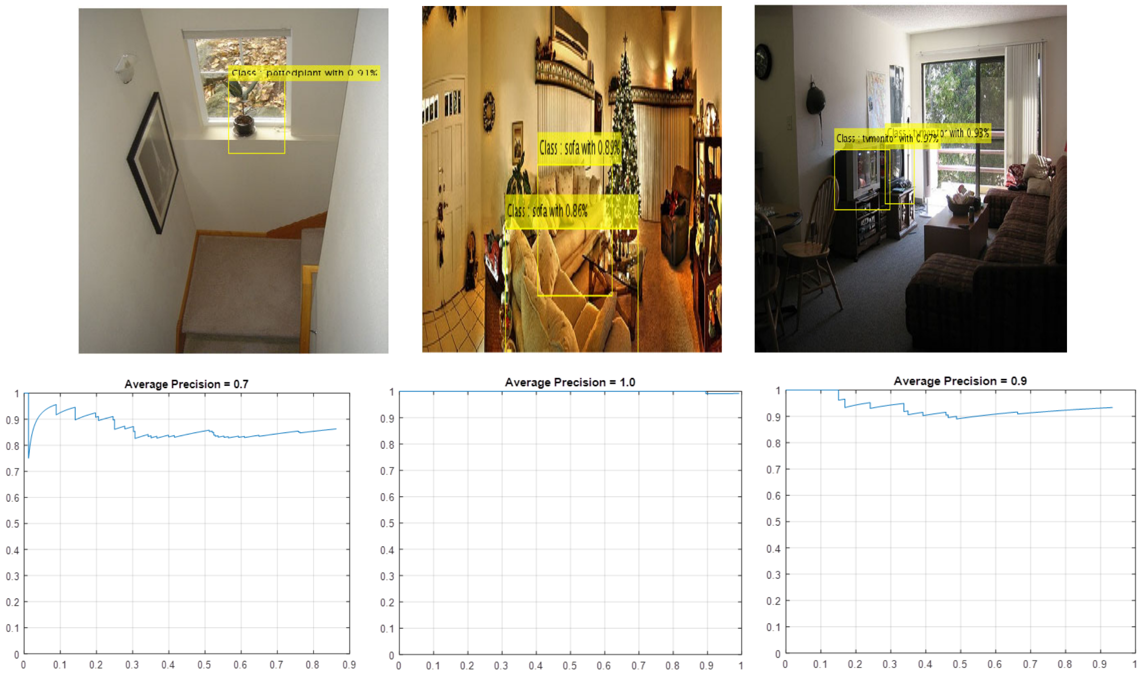

Fig. 7. Plant pot, sofa, TV detection and its precision.

The recognition precision parameter in terms of percentage for pascal VOC 2007 data using various technique is shown in Table I. Among which the proposed hybrid algorithm the rate of accuracy is proved to be the best.

TABLE I. RESULTS OF PRECISION OBTAINED USING DIFFERENT TECHNIQUES

Pascal VOC 2007 DS/Method	DPM-ST [24]	DPM-HSC [22]	DPM-V5 [23]	RCNN- TNET [25]	MR-CNN [26]	R-FCN [27]	Proposed model
Aero	23.8	32.2	33.2	64.2	80.3	79.9	**83.62**
Bike	58.2	58.3	60.3	69.7	84.1	87.2	**93.5**
Bird	10.5	11.5	10.2	50	78.5	81.5	**84.0**
Boat	8.5	16.3	16.2	41.9	70.8	72	**79.3**
Bottle	27.1	30.6	27.3	32	68.5	69.8	**47.6**
Bus	50.4	49.9	54.3	62.6	88	86.8	**93.2**
Car	52.2	54.8	58.2	71	85.9	88.5	**67.7**
Cat	7.3	23.2	23.0	60.7	87.8	89.8	**98.3**
Chair	19.2	21.5	20.0	32.7	60.3	67	**83.2**
Cow	22.8	27.7	24.1	58.5	85.2	88.1	**87.4**
Table	18.0	34.0	26.7	46.5	73.7	74.5	**99**
Dog	8.0	13.7	12.7	56.1	87.2	89.8	**95.2**
Horse	55.9	58.1	58.1	60.7	86.5	90.6	**98.5**
M-Bike	44.8	51.6	48.2	66.8	85	79.9	**89.8**
Person	33.2	39.9	43.2	54.2	76.4	81.2	**76.2**
Plant	13.2	12.4	12.0	31.5	48.5	53.7	**74.8**
Sheep	15.9	23.5	21.1	52.8	76.3	81.8	**58.1**
Sofa	22.8	34.4	36.1	48.9	75.5	81.5	**99.1**
Train	46.2	47.4	46.0	57.9	85	85.9	**99.3**
TV	44.9	45.2	43.5	64.7	81	79.9	**87.2**
MAP	**29.1**	**34.3**	**33.7**	**54.2**	**78.2**	**80.5**	**84.6**

The proposed model of object detection is performed on three different datasets. All the three results are compared and analysed. The analysis of PASCAL VOC 2007 dataset model using different techniques is shown in Fig. 8.

The analysis of MS COCO dataset model using different techniques is shown in Fig. 9 and the analysis of ImageNet dataset model using different techniques is shown in Fig. 10.

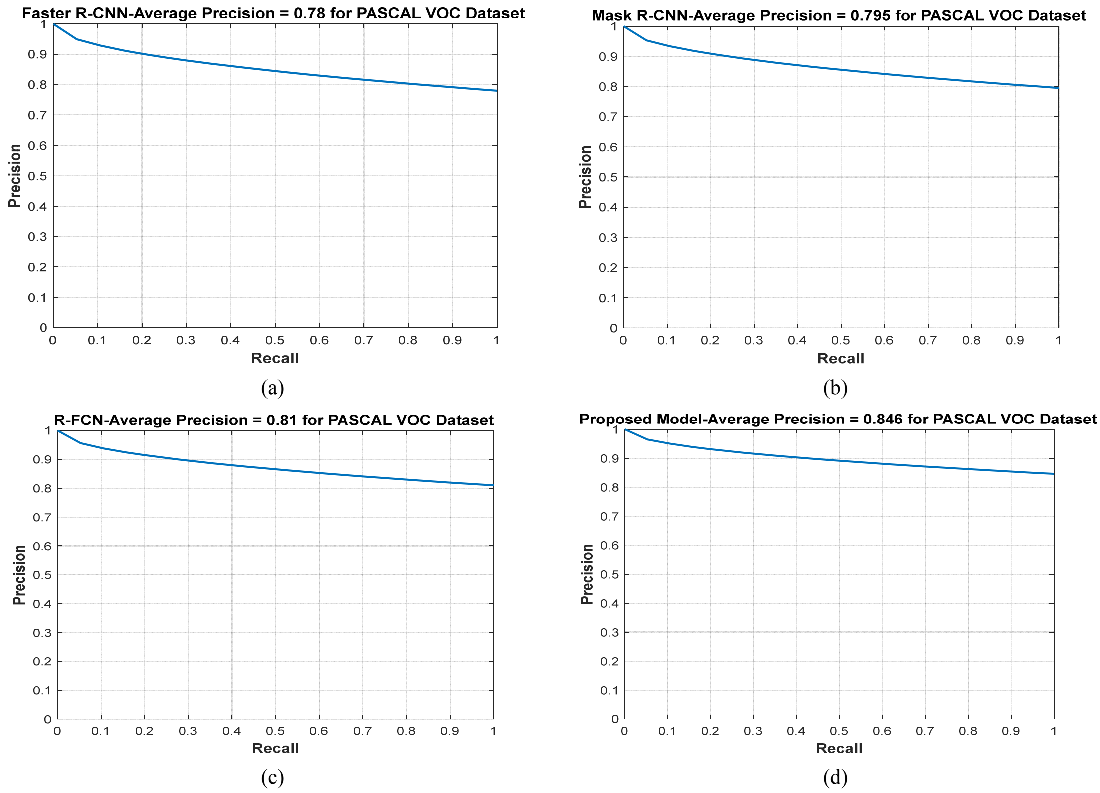

Fig. 8. Mean Average Precision (MAP) analysis for PASCAL VOC 2007 DS. (a) Using Faster R-CNN; (b) Using Mask R-CNN; (c) Using R-FCN; (d) Using proposed FRCNN-PSFA.

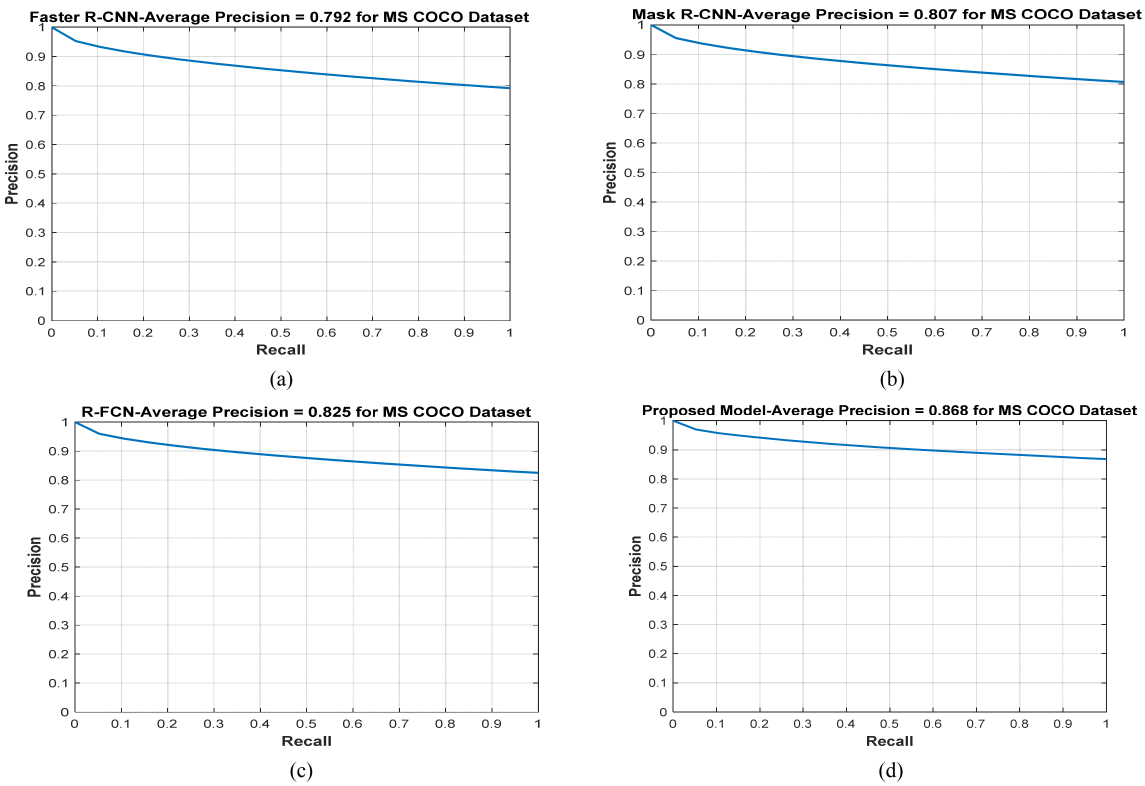

Fig. 9. MAP analysis for MS COCO DS. (a) Using Faster R-CNN; (b) Using Mask R-CNN; (c) Using R-FCN; (d) Using proposed FRCNN-PSFA.

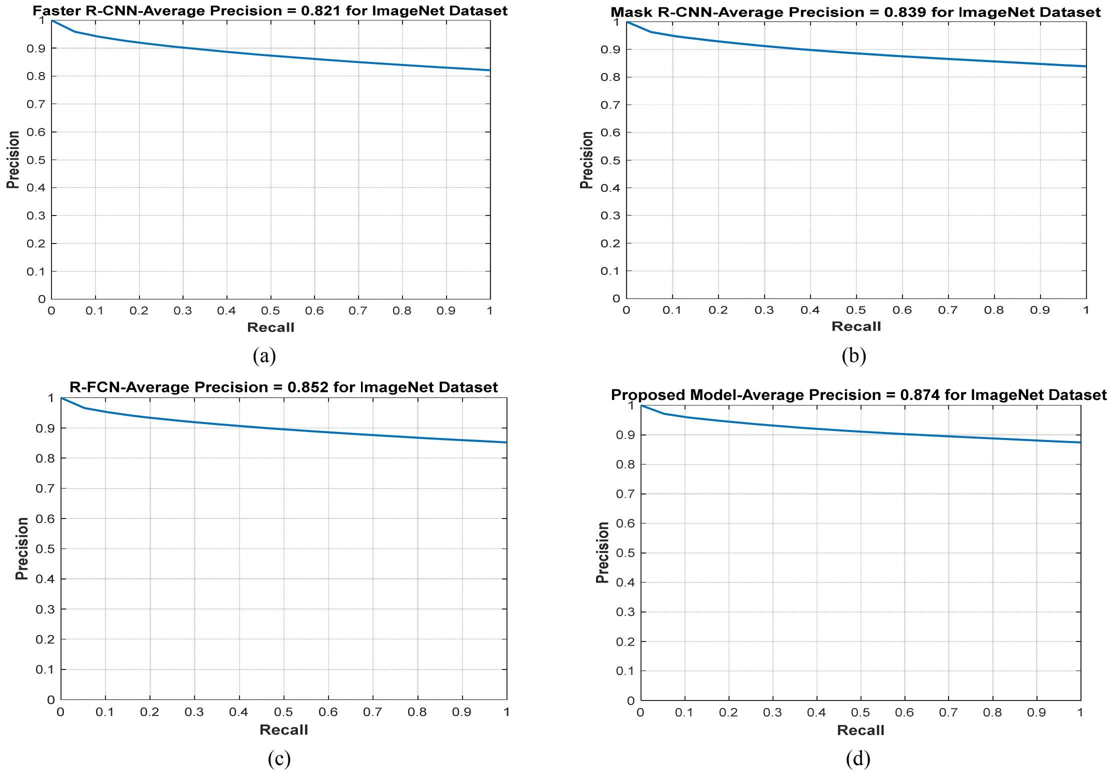

Fig. 10. MAP analysis for ImageNet DS. (a) Using Faster R-CNN; (b) Using Mask R-CNN; (c) Using R-FCN; (d) Using proposed FRCNN-PSFA.

The model designed is suitable for all dataset and achieved good results compared with other techniques and are shown in Table II.

TABLE II. MAP USING DIFFERENT TECHNIQUES FOR DIFFERENT DATASETS

Method/Dataset	PASCAL VOC 2007 MAP (%)	MS COCO MAP (%)	ImageNet MAP (%)
Faster RCNN	78.0	79.2	82.1
Mask RCNN	79.5	80.7	83.9
R-FCN	81.0	82.5	85.2
Proposed FRCCN-PSFA	**84.6**	**86.8**	**87.4**

V. CONCLUSION

In this paper, object detection framework has been proposed with two different optimization techniques and are combined with deep neural network to obtain high level of detection rate. The proposed technique is termed as FRCN-PSFA, here PSFA which is a hybrid optimization algorithm. The segmentation technique involves helps in guiding the deep learning model. Training and testing of dataset are performed and finally object has been detected and recognized. The results achieved conclude that rate of precision is improved compared to other existing techniques, the proposed method obtained a highest precision of 99.3% in case of train detection and 99.1% in case of sofa detection. The experimental results are been evaluated using the PASCAL VOC 2007 dataset with a Mean Average Precision (MAP) of 84.6%, MS COCO dataset with an MAP of 86.8% and Image Net dataset with an MAP of 87.4%. The accuracy of detecting object is low and still need to be improved. The model when considered for real time applications the accuracy plays an important role.

Existing techniques were created to address the fundamental challenge of object detection. However, there is still a great deal of potential for creating new processes and object detection as fundamental services in real-time applications, such deep-sea bases, autonomous vehicles, robots moving on planets, industrial facilities, and drone cameras, where activities requiring a high degree of precision are required. There is still a significant speed difference between machine vision and human eyes, especially when it comes to identifying certain microscopic things. AutoML, or creating a detection model that requires less human interaction, is the way of the future for object detection.

Intelligent IoT object systems increase accuracy rates, and 6G will offer high bandwidth and ultra-low latency (<1 ms), enabling real-time object recognition at the edge without the need for cloud computing. Object detection will be used by IoT devices (such as robots, drones, and smart cameras) for automation and decision-making in practical applications.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

S.K.P., K.S.K., K.B.B and K.P.K, V.A.K: Conceptualization, Data Curation, Formal Analysis, Investigation, Resources, Software, Writing original draft: N.R.Y, P.G., S.M.B., R.K.M, M.M: Methodology, Project administration, Supervision, Validation, Visualization, Writing-Review & editing, Funding acquisition. All authors had approved the final version.

REFERENCE

[1] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[2] D. Ciresan, A. Giusti, J. Schmidhuber *et al.*, "Deep neural networks segment neuronal membranes in electron microscopy images," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *ECCV*, 2014.

[4] L. Zhu, Y. Chen, A. Yuille, and W. Freeman, "Latent hierarchical structural learning for object detection," *Computer Vision and Pattern Recognition*, 2010.

[5] D. C. Ciregan, J. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. Comput. Vis. Pattern Recognition*, 2012, pp. 3642–3649.

[6] K. Li and L. Cao, "A review of object detection techniques," in *Proc. 2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*, Nanchang, China, 2020, pp. 385–390.

[7] S. Guo, F. Liu, X. Yuan, C. Zou, L. Chen, and T. Shen, "HSPOG: An optimized target recognition method based on histogram of spatial pyramid oriented gradients," *Tsinghua Science and Technology*, vol. 26, no. 4, pp. 475–483, Aug. 2021.

[8] T. M. K. Kumar, S. Aluvala, and P. Mandal *et al.*, "Improving accuracy in object detection using region-based convolutional neural network," in *Proc. 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, Bengaluru, India, 2023, pp. 1–6.

[9] J. Ma, M. Lin, G. Zhou, and Z. Jia, "Joint image restoration for domain adaptive object detection in foggy weather condition," in *Proc. 2024 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 2024, pp. 542–548.

[10] G. Baatar, L. Ritzau, D. Karimanzira, and T. Rauschenbach, "Enhanced underwater object detection in sonar images with range-setting embedding and sliding window preprocessing," in *Proc. OCEANS 2024—Halifax*, Halifax, NS, Canada, 2024, pp. 1–7.

[11] Z. Wang, G. Zhou, J. Ma, T. Xue, and Z. Jia, "Beyond the snowfall: Enhancing snowy day object detection through progressive restoration and multi-feature fusion," in *Proc. ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, Republic of, 2024, pp. 3315–3319.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal.Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[13] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448.

[14] H. Jeong, S. Choi, S. Jang, and Y. Ha, "Driving scene understanding using hybrid deep neural network," *Journal of Photogrammetry and Remote Sensing*, vol. 8, no. 2, pp. 1–4, Feb. 2019.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, Dec. 2015, pp. 91–99.

[16] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, Dec. 2016, pp. 379–387.

[17] F. S. Hassan and A. Gutub, "Improving data hiding within colour images using Hue component of HSV colour space," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 56–68, Mar. 2022.

[18] Z. Cai, X. Zhu, P. Gergondet, X. Chen, and Z. Yu, "A friction-driven strategy for agile steering wheel manipulation by humanoid robots," *CyborgBionic Syst.*, vol. 4, p. 64, Jan. 2023. doi: 10.34133/cbsystems.0064

[19] H. Jiang, S. Chen, Z. Xiao, J. Hu, J. Liu, and S. Dustdar, "Pa-Count: Passenger counting in vehicles usingWi-Fi signals," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, Mar. 2023. doi: 10.1109/TMC.2023.3263229

[20] Z. Xiao, J. Shu, H. Jiang, G. Min, H. Chen, and Z. Han, "Perception task offloading with collaborative computation for autonomous driving," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 457–473, Feb. 2023. doi: 10.1109/JSAC.2022.3227027

[21] B. R. Chughtai and A. Jalal, "Object detection and segmentation for scene understanding via random forest," in *Proc. 4th Int. Conf. Advancement Comput. Sci. (ICACS)*, Feb. 2023, pp. 6–13.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," arXiv preprint, arXiv:1506.02640, 2015.

[23] J. J. Lim, C. L. Zitnick, and P. Doll´ar, "Sketch tokens: A learned mid-level representation for contour and object detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3158–3165.

[24] X. Ren and D. Ramanan, "Histograms of sparse codes for object detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3246–3253.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint, arXiv:1409.1556, 2014.

[26] S. Gidaris and N. Komodakis, "Object detection via a multi-region and semantic segmentation-aware CNN model," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1134–1142.

[27] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, Dec. 2016, pp. 379–387.

[28] B. Amjoud and M. Amrouch, "Object detection using deep learning, CNNs and vision transformers: A review," *IEEE Access*, vol. 11, pp. 35479–35516, 2023.

[29] H. T. Nguyen, E. H. Lee, and S. Lee, "Study on the classification performance of underwater sonar image classification based on convolutional neural networks for detecting a submerged human body," *Sensors,* vol. 20, no. 1, p. 94, 2020.

[30] D. C. Nguyen *et al.*, "6G Internet of things: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[31] V. K. Quy *et al.*, "From federated learning to quantum federated learning for space-air-ground integrated networks," in *Proc. 2024 IEEE Conference on Standards for Communications and Networking (CSCN)*, Belgrade, Serbia, 2024, pp. 402–407.

[32] Q. V. Khanh, A. Chehri, V. A. Dang, and Q. N. Minh, "Federated learning approach for collaborative and secure smart healthcare applications," *IEEE Transactions on Emerging Topics in Computing*, vol. 13, no. 1, pp. 68–79, 2025.

[33] F. S. Hassan and A. Gutub, "Improving data hiding within colour images using Hue component of HSV colour space," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 56–68, Mar. 2022.

[34] Z. Cai, X. Zhu, P. Gergondet, X. Chen, and Z. Yu, "A friction-driven strategy for agile steering wheel manipulation by humanoid robots," *Cyborg Bionic Syst.*, vol. 4, p. 64, Jan. 2023. doi: 10.34133/cbsystems.0064

[35] H. Jiang, S. Chen, Z. Xiao, J. Hu, J. Liu, and S. Dustdar, "Pa-Count: Passenger counting in vehicles usingWi-Fi signals," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, Mar. 2023. doi: 10.1109/TMC.2023.3263229.

[36] Z. Xiao, J. Shu, H. Jiang, G. Min, H. Chen, and Z. Han, "Perception task offloading with collaborative computation for autonomous driving," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 457–473, Feb. 2023. doi: 10.1109/JSAC.2022.3227027

[37] B. R. Chughtai and A. Jalal, "Object detection and segmentation for scene understanding via random forest," in *Proc. 4th Int. Conf. Advancement Comput. Sci. (ICACS)*, Feb. 2023, pp. 6–13.