

# DETECTION OF AIRCRAFTS USING RCNN IN REMOTE SENSING IMAGES

Bhavani Sankar Panda <sup>1</sup>, Kakita Murali Gopal <sup>1</sup> and Rabinarayan Satpathy <sup>2</sup>

<sup>1</sup> Giet University, Gunupur, Odisha, India

<sup>2</sup> Sri Sri University, Cuttack, Odisha, India

## Abstract

Modern remote sensing images have developed, and the resultant of the sophisticated technology makes the detection of objects a tedious task. Detection of aircraft from remote sensing imagery has some challenges that are often not addressed. The image captured by remote sensing technology would have more information about the objects in the scene. Smaller objects around the aircraft and similar unknown objects could easily affect the detection model. And there will be numerous objects in the wide-open space of the airfield and airports. Although several research studies showed aircraft detection from remote sensing images, the models didn't solve the overlap problem. And therefore, the present study proposed a method to resolve the overlap problem using the Region-based convolutional neural network architecture from the YOLOv3 model to detect the aircraft from the remote sensing images. The model would use the Interaction of unit and the ground truth data from the images to select the appropriate bounding boxes for the aircraft. The model was trained, and the results produced the model could perform at an accuracy of 98.5 %. The RCNN model could detect the aircraft appropriately, and the metrics analysis shows the model had less error while detecting the aircraft. The performance comparison with other aircraft detection models showed that the proposed model performed better.

## Keywords

Aircraft detection, Region of interest, remote sensing images, RCNN

## 1. Introduction

Recently, object detection based on remote sensing images has been developed, and the necessity of collecting information from remote sensing images has also grown. Especially for the military and defence applications, the detection of objects in the remote sensing images is important, and aircraft detection from the remote sensing images is among them. Detecting aircraft from the airports and airspace could provide crucial information and is often considered an important aspect of military Intel. Aircraft detection has important aspects, and several research studies have tried to implement deep learning models from remote sensing images. The remote sensing images with sophisticated technology can now produce a higher resolution of images. These high-resolution images have more information about the object on the surface, and the deep learning models couldn't differentiate the objects. As often, the objects become harder to distinguish. The images could provide information about the people in the field, so detecting the only aircraft from the airspace and airfield has become a complex task. To discuss further on the detection of objects from the remote sensing images, the features are harder to be extracted. The selection of features becomes difficult during the bounding box creations. Although the studies have discussed this, an appropriate solution was not

---

ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States

EMAIL: [bs.panda@giet.edu](mailto:bs.panda@giet.edu) (A. 1); [kmgopal@giet.edu](mailto:kmgopal@giet.edu) (A. 2); [rabinarayan.satpathy@gmail.com](mailto:rabinarayan.satpathy@gmail.com) (A. 3)

ORCID: [0000-0002-0746-5156](https://orcid.org/0000-0002-0746-5156) (A. 1);



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

found yet. Extracting information from the remote sensing images need to be improved, especially for the detection of a single object.

Some earlier studies have shown that the RCNN model could detect aircraft. The Region-based convolutional neural networks were known for the quick and appropriate selection of features from the images. Many image processing applications and detection models use the RCNN model to perform feature extraction and detection. To employ the RCNN models, the dataset has to be properly annotated.

## 2. Literature Review

The research study on detecting aircraft through remote sensing [1] aimed to resolve the aeroplane detection, including inappropriate detection of small objects while developing the ROI and proposed an algorithm to improve aeroplane detection using remote sensing images. The author used multilayer features and a new, improved non-maximal suppression to detect the aeroplane from the remote sensing images. The multilayer feature fusion method made the model include features from the different layers and fuse them to select the appropriate features from the images. The non-maximal suppression improved the detection box overlap problem and appropriately selected the specified feature. The proposed method improved the selection and detection of aeroplanes from remote sensing images. Another study on detecting aeroplanes, [2] using remote sensing images made a deep learning-based. The region-based convolutional neural network is used for the detection, built from scratch. The RCNN performs the detection and identification of the aircraft. A huge dataset comprising remote sensing images of airports and aircraft are used for training, and the validation of the model was based on the images collected from the Turkey airport. The proposed model displayed accuracy of 98.34 %. And the model was able to detect the aircraft using bounding boxes, and the prediction results demonstrated aircraft detection.

Here the study demonstrated [3] a tomato detection using YOLO-tomato to resolve the overlap challenges in the detection of tomato. This improved YOLOv3 model with an altered bounding box in a circular shape, which improved the calculation of non-maximum suppression. And the model showed effective results and performance compared to other studies. Also, the study adopted [4] a new method to create bounding boxes based on the height and width of the cluster centres. The bounding boxes are created based on the Markov chains values, and Intersection over union method is followed to measure the distance between clusters and individual points. The results showed the method could converge faster and the performance metrics was better compared to the conventional YOLOv3 model.

The study on Platform Screen Doors [5] used a deep learning technique to identify the object and classify it. And as stated by the Authors it was the first attempt to detect objects on platform screen doors in metros. The study used 984 images at  $600 \times 480$  pixels and manually labelled for six different objects. The object detection performance was evaluated by using three famous detection models, YOLOv3, Single-Shot MultiBox-Detector (SSD), and Center-Net through transfer learning. The YOLOv3 model was able to detect foreign objects at an exceptional speed which is about 200 FPS (frames per second), and CentreNet produced accurate results consistently. The study on detecting aircraft [6] used reinforcement learning and a CNN model. The reinforcement learning adopted here to resolve the challenge in detecting the number of objects from the remote sensing images. The detection agent by the reinforcement learning accurately detects the aircraft using a bounding box and performs better. The CNN model is used to detect the features and analyze the probability of finding the aircraft in the image.

Moreover, another study on Aircraft detection [7] using a novel method from remote sensing images depends on the deep-residual network (ResNet) and Super Vector (SV). The ResNet is used to merge the convolutional features and improves the resolution of features which in turn produces a better proposal for detection. Meanwhile, SV coding is used to extract the histogram of oriented

gradient (HOG) from the ROI to complete the detection and classification process. The model could produce better results even in the presence of irregular backgrounds.

## 2.1 Research Gap

Although there were enough studies to detect aircraft from the remote sensing images, aircraft detection is still complex, and the overlapping of other objects from the images is not yet solved completely. The remote sensing images also has objects that might cause improper detection of aircraft [2]. The deep learning models have overlap issues and smaller objects around the aircraft.

## 2.2 The objective of the study

Based on the literature review and the research gap, the following objectives are followed for the proposed study,

1. To detect aircraft from the remote sensing images using the Region-based Convolutional Neural Network model.
2. To use Intersection over a unit and ground truth data parameters to select the appropriate features to overcome the irrelevant smaller object detection.

## 3. Proposed Methodology

The study proposed an aircraft detection model based on the remote sensing images, using the YOLOv3 model. This section discusses the overall proposed methodology for the study. First, the dataset details were discussed compared to earlier studies on detecting aircraft with the images and then followed by the neural network model adopted for the study, which is YOLOv3 and transfer learning was adopted here for customized objects.

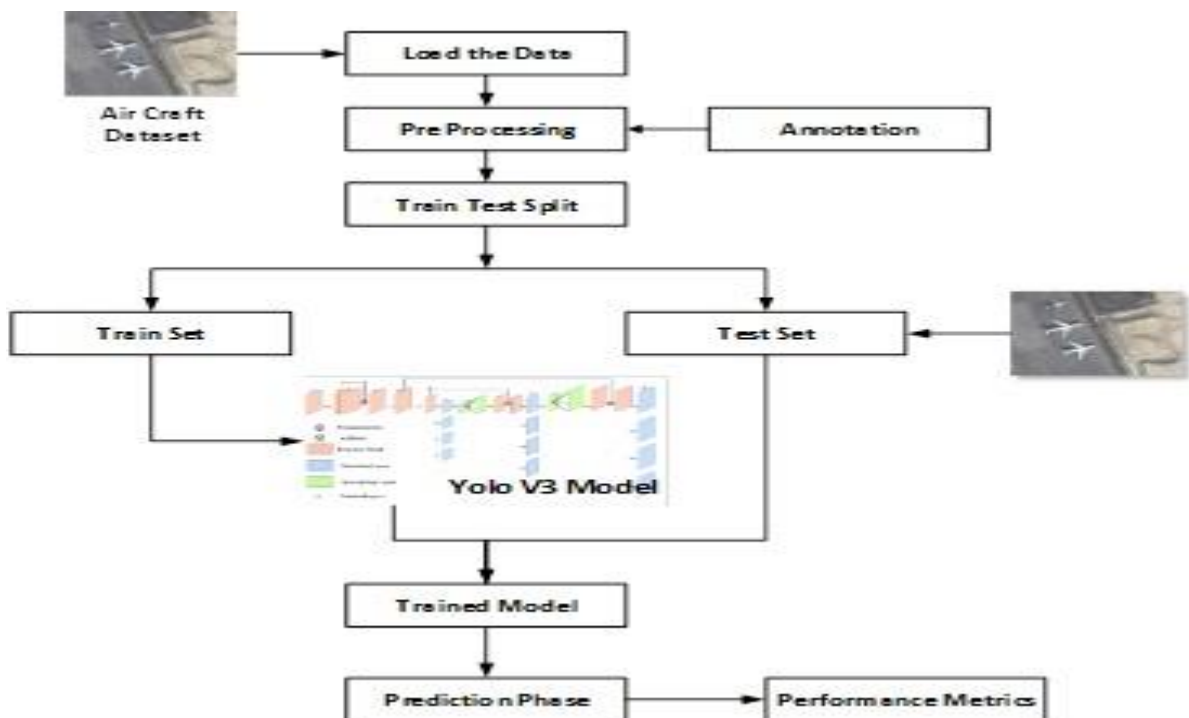


Fig. 1. Overall flow of proposed methodology

### 3.1. Dataset description

From the flow diagram, it can be observed that the process starts with a dataset, which was collected from the kaggle open-source[8]—the dataset comprised of images with and without aircraft. The images were first initially pre-processed and annotated for the presence of aircraft. The dataset is then split into two sets with aircraft and without crafts. While splitting the images, it is observed that 8000 images have aircraft, and 24,000 images don't have aircraft. The images were colour (RGB) based, and the details are listed in table 1,

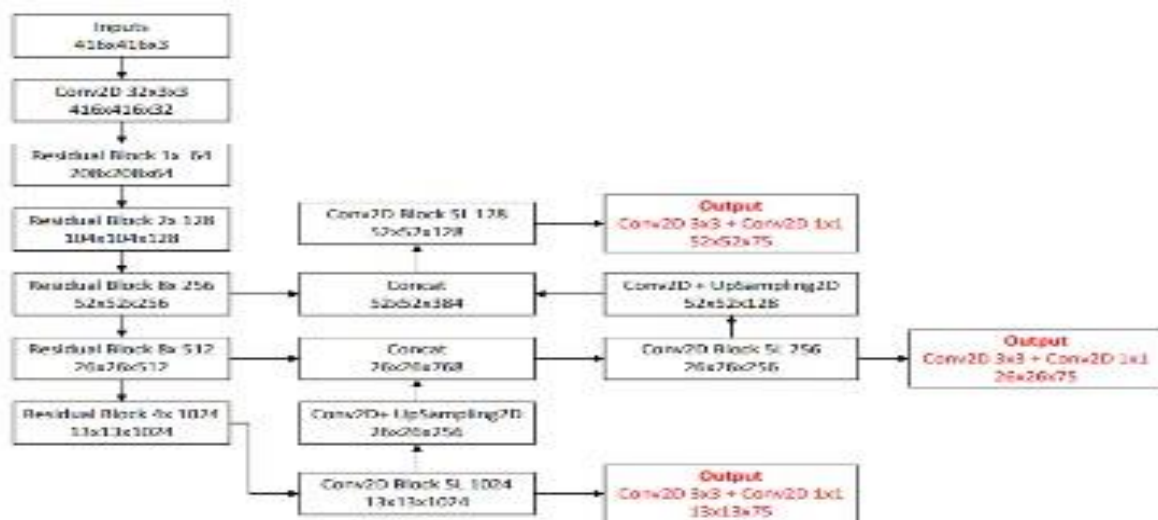
**Table 1.** Details about the dataset

Class (Annotations)	Total images	Image properties
Plane	8,000	20×20 / RGB
Not Plane	24,000	

The selected images for the training were 8000 images with a resolution of 20 x 20 pixels. The annotations were predefined with the dataset yet verified once manually. Then the images were scaled to 416 x416 to fit the YOLOv3 architecture.

### 3.2. Neural network Architecture

The Neural network architecture adopted here is the YOLOv3 model, one of the famous object detection tools among the deep learning model. The reason to choose the YOLOv3 model is that the earlier studies showed the complexity in detecting the bounding boxes for the aircraft and the inconsistent accurate results showed that from the remote sensing images, detection of aircraft could be a challenge without overlapping. And as for the YOLOv3 architecture, the resultant model could resolve the overlapping issue as the Region based convolution selection were followed. The transfer learning method is adopted here, and for the final object detection, the final layer functions are based on detecting only aircraft from the input images. Every other object detection from the conventional YOLOv3 model is removed.



**Fig. 2.** The architecture of the neural network Model

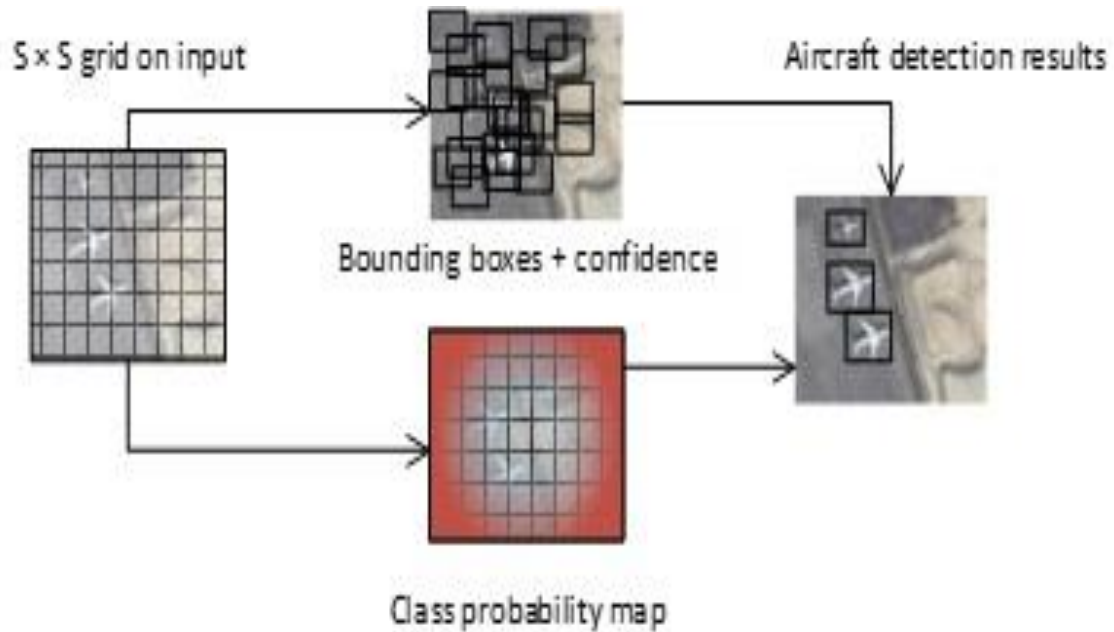
As for the layers, The RCNN model uses 53 convolutional layers, and there will be no pooling layers. The model is a complete, fully convolution network model as only convolutional layers are used. Initially, images were loaded into the convolutional layer with 416 x 416, filtered and got a reduced

shape of 208 x208. Then, a res-block consist of two convolutional layers and a residual layer is used to search for the features that are to be detected from the images. Again another res-block was used two times and then upsampled, producing downscaled images of 52 x 52. Following the up sampling, the images are further passed onto res-block eight times which in turn produces a reduced image of about 26 x26. And the repeated convolutions by the res-block number of times and the up sampling finally detect the set of the regions for the input image.

### 3.3. ROI based feature selection

The YOLOv3 model adopts a single-stage object detector for the detection problem. The image is split into grids equally, as shown in the image. Then each cell would be analyzed to create the bounding box. The boxes were created based on the confidence parameter, which utilizes the ground truth and the Interaction of the units from the image and the equation (1).

$$Confidence = Probability(object) \times Interaction\ of\ Unit\ (Ground\ truth, prediction) \quad (1)$$



**Fig. 3.** Region of Interest Detection

The binary cross-entropy for predicting the object score and truth object score is used as a loss function as shown in the equation (2).

$$E_1 = \sum_{i=0}^{S^2} \sum_{j=0}^{BX} W_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) - (1 - C_i^j) \log(1 - C_i^j)] \quad (2)$$

The coordinates of the bounding box is given by the parameter  $b_x$  and  $b_y$ , as by the equation (3) & (4) where, the  $C_x$  and  $C_y$  are the offset of the grid cell. The position of the bounding box is given by the  $t_x$  and  $t_y$ .

$$b_x = \text{Sig}(t_x) + c_x \quad (3)$$

$$b_y = \text{Sig}(t_y) + c_y \quad (4)$$

Where, the width and height of the bounding box is given by the parameter  $b_w$  and  $b_h$ , as by the equation (5) & (6) where, the  $p_w$  and  $p_h$  are the offset of the grid cell. The position of the bounding box is given by the  $t_w$  and  $t_h$ .

$$b_w = p_w e^{t_w} \quad (5)$$

$$b_h = p_h e^{t_h} \quad (6)$$

$$\text{Sig}(\hat{t}_x) = g_x - c_x \quad (7)$$

$$\text{Sig}(\hat{t}_y) = g_y - c_y \quad (8)$$

$$\hat{t}_w = \log \left( \frac{g_w}{p_w} \right) \quad (9)$$

$$\hat{t}_h = \log \left( \frac{g_h}{p_h} \right) \quad (10)$$

$(t_x^\wedge, t_y^\wedge, t_w^\wedge, t_h^\wedge)$  are the truth values and  $(g_x, g_y, g_w, g_h)$  are the ground truth box parameters in the equation (7), (8), (9) & (10).

$$E_2 = \sum_{i=0}^{S^2} \sum_{j=0}^{BX} W_{ij}^{obj} [(\text{Sig}(t_x)_i^j - \text{Sig}(t_x)_i^j)^2 + (\text{Sig}(t_y)_i^j - \text{Sig}(t_y)_i^j)] + \sum_{i=0}^{S^2} \sum_{j=0}^{BX} W_{ij}^{obj} [(\text{Sig}(t_w)_i^j - \text{Sig}(t_w)_i^j)^2 + (\text{Sig}(t_h)_i^j - \text{Sig}(t_h)_i^j)] \quad (11)$$

The model uses two equations for the coordinate prediction to generate the bounding box, equation (1) to generate the probability in terms of confidence and equation (11) uses the square error for the prediction of coordinates. The above mentioned equations are the representation of the calculations to create the appropriate bounding box in the given image.

The object detection in the YOLOv3 model was based on three different scales. The three different scale detection could reduce the detection of smaller objects problem. The 8000 images are used for training the model, and for the model training 120 batch size and 300 epochs. The model was trained by using CPU - Intel Core i7 – 7700 @ 2.80 GHz, GTX 1050, 16GB RAM and in Windows 8/Python 3.7 environment.

## 4. Results and Discussion

This section discusses the results and the performance analysis for the proposed model to detect the aircraft. As discussed earlier, the study has used the YOLOv3 model to detect aircraft from remote sensing images. Initially, the dataset contained both aircraft and no aircraft set of images. The images with aircraft are filtered as presented in fig. 4 & 5.



**Fig. 4.** Images with aircraft from the dataset





**Fig. 5.** Images without aircraft from the dataset

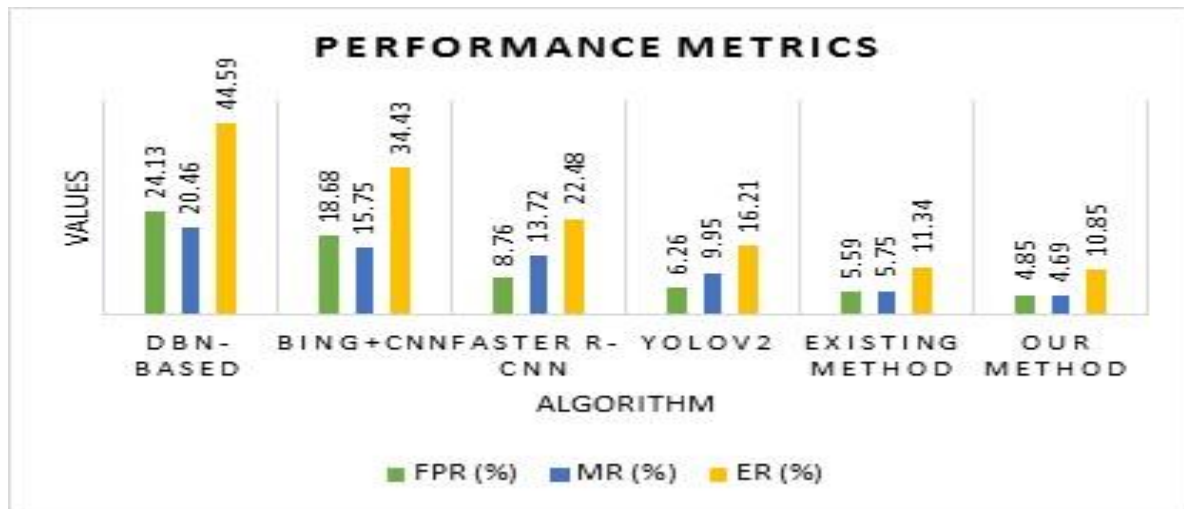
#### 4.1. Performance analysis

The model was trained with 8000 images using the YOLOv3 model, and the proposed method was compared with other studies to evaluate the model and its detection ability.

**Table 2.** Comparison of metrics with other Aircraft detection models.

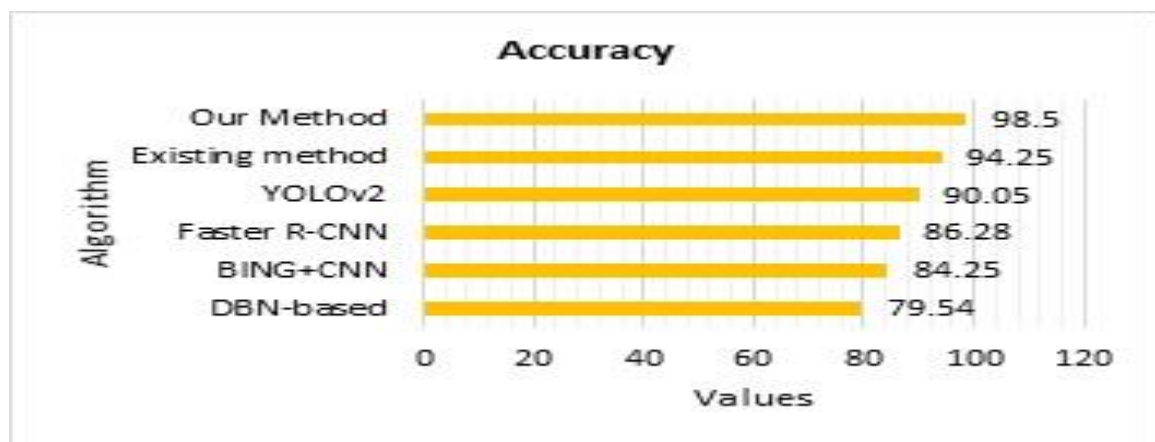
Model	Accuracy of the model	False Positive ratio (FPR) (%)	Missing Ratio (MR) (%)	Error ratio (ER) (%)	Time (s)
DBN-based	79.5	24.1	20.4	44.5	171.2
BING+ CNN [9]	84.2	18.6	15.7	34.4	6.4
Faster R-CNN	86.2	8.7	13.7	22.4	0.15
YOLOv <sub>2</sub>	90.0	6.2	9.9	16.2	0.03
Recent study[1]	94.2	5.5	5.7	11.3	0.16
Propose method	98.5	4.85	4.69	10.8	0.29

Above table 2 shows the performance comparison for detecting aircraft from remote sensing images. From the summarized data, it could be observed that the proposed method has produced 98.5 % accuracy. But compared to the YOLOv2 based model and the study [1], the time taken to detect the aircraft is less for the proposed model with minor differences. However, the accuracy of detection is better in the proposed method.



**Fig. 6.** Comparison of FPR, MR and ER for the proposed model

The comparison of metrics FPR, MR and ER are shown in fig 6, and the graphical representation of accuracy comparison for the proposed model is shown in fig.7.



**Fig. 7.** Accuracy comparison proposed model

**Table 3.** Comparison of metrics with conventional CNN [2]

Model	Accuracy	Specificity	Sensitivity	False Positive rate	False Negative rate
CNN [2]	98.4	98.13	98.62	0.0187	0.0138
Proposed	98.5	97.72	97.45	0.0168	0.0129

The comparison of the conventional CNN model for the proposed RCNN model shown in Table 3 compares confusion matrix metrics. Although the CNN model showed better accuracy, specificity, and sensitivity, the FP rate and FN rate was better in the proposed RCNN model, indicating the model has less error in detection. And the graphical representation of the comparison of the false positive and false negative rates are shown in fig.8





**Fig. 8.** Performance metrics comparison between CNN and the proposed method.

**Table 4.** Comparison of metrics with previous studies

Method	Recall (%)	Precision (%)
Method 1 [10]	59.2	67.3
Method 2 [11]	63.6	80.6
F-RCNN [12]	82.4	87.8
F-RCNN+ResNet-50 [13]	83.7	88.6
Only conv4_x+HOG [7]	84.1	89.6
Existing Method [7]	91.9	93.5
Proposed method	97.13	97.65

Precision and the Recall metrics show the model's ability to reproduce the results with consistent accuracy. Table 4, for the comparison of metrics with previous studies, shows how the model's performance increased over the years. The proposed model showed a 97.65 % precision and 97.13 % recall which is good enough for the proposed model. The comparison of the proposed model with the earlier studies for the different metrics showed us the proposed model for detecting the aircraft from the remote sensing images.

## 5. Conclusion

The detection of aircraft from the remote sensing images had complex challenges such as detection of smaller objects and overlapping of objects. Detecting objects from the remote sensing images has several limitations, which result in the improper detection of the aircraft. To overcome the underlying issues in aircraft detection, the proposed method adopts the YOLOv3 model to identify the aircraft from the remote sensing images. Based on the RCNN model and confidence based selection of ROI they made, the model to choose the appropriate features for the detection of aircraft. This made the model identify the aircraft effectively from the remote sensing images. The performance comparison with earlier studies for the proposed model showed that the model performed better in detecting the aircraft.

## References

1. M. Zhu, Y. Xu, S. Ma, S. Li, H. Ma, and Y. Han, "Effective airplane detection in remote sensing images based on multilayer feature fusion and improved nonmaximal suppression algorithm," *Remote Sensing*, vol. 11, p. 1062, 2019.
2. F. Ucar, B. Dandil, and F. Ata, "Aircraft detection system based on regions with convolutional neural networks," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 8, pp. 147-153, 2020.
3. G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3," *Sensors*, vol. 20, p. 2145, 2020.
4. L. Zhao and S. Li, "Object detection algorithm based on improved YOLOv3," *Electronics*, vol. 9, p. 537, 2020.
5. Y. Dai, W. Liu, H. Li, and L. Liu, "Efficient foreign object detection between PSDs and metro doors via deep neural networks," *IEEE Access*, vol. 8, pp. 46723-46734, 2020.
6. Y. Li, K. Fu, H. Sun, and X. Sun, "An aircraft detection framework based on reinforcement learning and convolutional neural networks in remote sensing images," *Remote Sensing*, vol. 10, p. 243, 2018.
7. J. Yang, Y. Zhu, B. Jiang, L. Gao, L. Xiao, and Z. Zheng, "Aircraft detection in remote sensing images based on a deep residual network and super-vector coding," *Remote Sensing Letters*, vol. 9, pp. 228-236, 2018.
8. K. a. dataset. <https://www.kaggle.com/rhammell/planesnet>.
9. H. Wu, H. Zhang, J. Zhang, and F. Xu, "Fast aircraft detection in satellite images based on convolutional neural networks," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4210-4214.
10. W. Zhang, X. Sun, K. Fu, C. Wang, and H. Wang, "Object detection in high-resolution remote sensing images using rotation invariant parts based model," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 74-78, 2013.
11. W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 137-141, 2016.
12. S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
13. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.