PERFORMANCE ASSESSMENT OF OBJECT DETECTION FROM MULTI SATELLITES AND AERIAL IMAGES

Mahmoud Ahmed[1*], Naser El-Sheimy[2], Henry Leung[1], Ahmed M. Kamel[3], Adel Moussa [2,4]

[1] Dept. of Electrical and Software Engineering, University of Calgary, 2500 University Dr NW, Calgary, Alberta T2N 1N4 Canada – (leungh@ucalgary.ca, mahmoud.ahmed2@ucalgary.ca)

[2] Dept. of Geomatics Engineering, University of Calgary, 2500 University Dr NW, Calgary, Alberta T2N 1N4 Canada – (elsheimy@ucalgary.ca, amelsaye@ucalgary.ca)

[3] Dept. of Guidance Navigation and Control, Military Technical College, Kobry Elkobbah, Cairo, Egypt - a.kamel@mtc.edu.eg

[4] Department of Electrical and Computer Engineering, Port-Said University, Port-Said 42523, Egypt - amelsaye@ucalgary.ca

KEY WORDS: Object detection, heterogeneous sensors, anchor box, dedicated model.

ABSTRACT:

Object detection in remote sensing imagery plays an important role in many applications, such as tracking and change detection. With the development of deep learning algorithms and advancement in hardware systems, improved accuracies have been achieved in the detection of various objects from remote sensing images. However, object detection across heterogeneous remote sensing imagery remains an important issue, particularly for satellite and aerial imagery. The colour variation for the same ground objects, variable resolutions, different platform heights, the parallax effect, and image distortion brought on by diverse shooting angles are the biggest hurdles in satellite-aerial detection applications. The research aims to obtain successful model for detecting aircrafts from satellite and aerial images and reduce cost and the gap of revisit time between sensors. The networks were tested using aerial, GF-2, Jilin-1 (JL-1) and Pleiades satellites test sets after being trained individually using the RGB high-resolution aerial set and panchromatic low-resolution GF-2 satellite set to validate the efficiency of the trained models. Also, the aerial-trained model and GF-2 satellite-trained model as dedicated models were compared with each other, and model trained by all dataset for Object Detection in Aerial Images (DOTA). It is observed that the anchor sizes and augmentation methods can enhance the performance of detection models. k-means algorithm and data augmentation were applied to produce better anchor box selection and avoid overfitting, atmospheric conditions problems, respectively. The accuracy assessment results demonstrate that the aerial-trained model outperforms the GF-2 satellite-trained model. In addition, the results of two dedicated detection models show improved accuracy compared to the DOTA-trained model.

1. INTRODUCTION

1.1 General object detection review

Since remote sensing images from satellite sensors are taken from high altitudes and include atmospheric interference, viewpoint fluctuation, background clutter, and lighting differences, they are more complex than computer vision images (Cheng and Han, 2016). The visual interpretation approach, which benefits from the expertise of specialists for the identification of various objects/targets, is still commonly utilized in object detection investigations of satellite imagery. This approach is time consuming because it involves a manual process, and the accuracy of the method depends on the level of competence of the specialist. In order to decrease human mistakes, save time, and improve efficiency, several studies have been conducted on the automatic recognition of various targets, such as buildings, aircraft, ships, etc. (Zhao et al., 2019; Zhou et al., 2016).

Images from multiple remote sensing platforms or sensors are referred to as heterogeneous images(Zhan et al., 2018) (Ansari et al., 2020; Tian, 2020). Remote sensing image change detection include the analysis of multitemporal and multi resolution information at the same time, and the outcome is highly significant for a wide range of applications, including tracking urban growth(Classification et al., n.d.), monitoring land use, evaluating disasters, and assessing damage.

Resolution (both spatial and temporal) is the main determinant of a constellation's value and cost, although other aspects also play a role such as object detection performance, segmentation accuracy, change detection fidelity, crop cover recall, etc. However, automatic detection is difficult for satellite images because of the complexity of the background, differences in data collecting geometry, terrain, and illumination conditions, and the diversity of objects. The classification of the objects and their location in the images are two essential tasks that make up the object detection task. The improvement of these two tasks, either alone or jointly, has been the focus of many studies to date (Gidaris and Komodakis, 2016).

There are many methods for detecting aircraft that have been presented in the literature. (Gao et al., 2013) applied a circular frequency filter to determine the location of the aircraft before using a multilayer feature to define local and spatial information aircraft layouts. But this method is restricted to patch-level identification. In (Wu et al., 2015), a method for identifying planes from satellite data that combines CNN and binarized norm gradients demonstrates the importance of being rotation invariant. The binarized norm gradients (BING) technique aids in the production of weaker candidates for prediction, however, the CNN extracts feature from the raw images. Deep belief networks (DBNs) were among the first deep neural networks to be utilized for airplane detection. Multiple global criteria along with DBNs are applied by Chen et al. (Chen et al., 2013), to pinpoint the aircraft precisely. The tests revealed that DBNs produced significantly higher results than Histogram of Oriented Gradients (HOG), Wavelet, and Gabor. Multilayer feature fusion and a faster R-CNN framework were used together to detect aircraft (Zhu et al., 2019). The network's capacity to recognize smaller aircraft was improved by the integration of multilayer features using L2 normalization, feature connection, scaling, and feature dimension reduction. This method improved the candidate regions suggested by the Region Proposal Network while reducing the execution time.

A deep patch orientation network (DON)(Maher et al., 2018) took advantage of target patches in the identification process and raised the likelihood that undiscovered targets will be found. By considering the targets' orientations, DON defined more details

about the targets. The DON network enhances the performance of the frameworks, such as Faster RCNN, in terms of detection.

1.2 Paper Contributions

The main objective of this paper is the detailed evaluations of the object detection techniques for aircraft detection from multiple datasets. This main objective us achieved through the following contributions:

• Assessing the accuracy of the results using the dedicated object detection models which outperform object detection trained model by all datasets of sensors.

• The proposed convenient anchor sizes could be solution to enhance results when detecting objects from heterogenous sensors. Furthermore, the distribution of objects in datasets affect the performance of training.

• Investigating the best scenario that can detect objects from different sensors to overcome revisit time and cost problems.

•Improving the learning with augmentation aerial and satellite datasets, especially radiometric augmentation for colour images and hyperparameters tuning to improve the performance of object detection model on the satellite and aerial image domain.

• Demonstrating a comparative analysis of aerial and satellite trained object detection models across various object sizes and IOUs, as well as performing an independent investigation with aerial and satellite test sets with a different resolution specification than the training dataset to assess the transferability.

2. DEEP LEARNING ARCHITECTURE

In this section, the general architecture of the Faster R-CNN framework is presented. The loss function and Resnet 50 as backbone network are explained in detail.

2.1 Deep learning-based object detectors

Faster R-CNN is one of the most well-known object detection networks that makes use of CNN architecture to get accurate and timely results. These characteristics led to its first use in applications requiring processing almost instantly, including video indexing tasks. R-CNN has increasingly gotten faster over time The R-CNN, its initial implementation, uses a hierarchical grouping strategy and a selective search method to produce item recommendations. The 2000 window are given to a pre-trained CNN model in the form of rectangular boxes. As the rectangular boxes, and they are sent to a CNN model that has already been trained. Following that, the CNN model's feature maps for them are extracted in order to send them to an SVM for classification (Girshick et al., 2014).

The Fast R-CNN was developed in 2015 by Girshick R. et al. and advances the R-CNN solution. Fast R-CNN differs from R-CNN in that it produces object suggestions from the CNN feature map rather than obtaining them from the entire input image.

In this manner, feature maps can be extracted without having to use the CNN technique 2000 times. The region of interest (ROI) pooling is then used to ensure that an output size that is uniform and predetermined is obtained. Finally, a SoftMax classifier is used to categorize the future maps, and linear regression is used to construct bounding box localizations. Finally, a SoftMax classifier is used to categorize the future maps, and linear regression is used to construct bounding box localizations.

A region proposal network takes the place of the selective search strategy in the Faster R-CNN (RPN). The objective of this network is to learn an object's proposition using feature maps as in Fig.2. This object detecting technique starts with the RPN. The RPN is given the feature maps that were taken from a CNN in order to suggest the regions. The region proposals are produced using k anchor boxes for each place on the feature maps. Given the three distinct scales and three different aspect ratios used in the original research, the anchor box number k is defined as 9 (Ren et al., 2015) as shown in Figure 1. There are k anchor boxes overall with a size of Wx H feature map, which contain the negative (not object) and positive (an object) anchor boxes. The RPN learns to produce region suggestions during the training phase using these anchor boxes. The RPN's bounding box classification layer generates two scores for each k, indicating whether an object or not there is an object.

Predicting the four k coordinates (box center, width, and height), a regression layer is applied as shown in Figure 2. At the second stage of the network, the ROI pooling operation is carried out as in the Fast R-CNN after production of the region suggestions. An ROI feature vector is created from fully connected layers, just like in Fast R-CNN, and this vector is categorized by SoftMax.

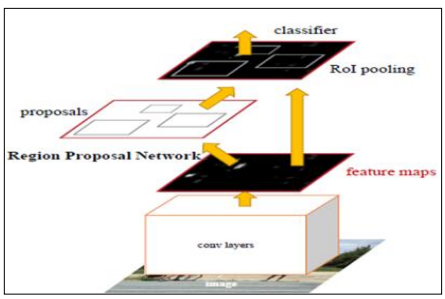


Figure 1. General Architecture of Faster R-CNN(Ren et al., 2015).

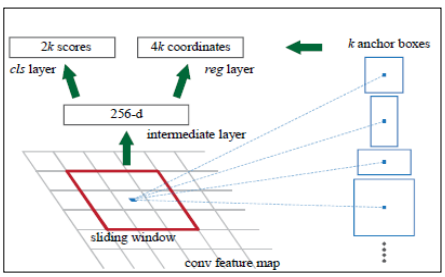


Figure 2. Region Proposal Network (RPN)(Ren et al., 2015)

2.2 Loss function

The following objective function is minimized in Fast R-CNN for an image (Girshick, 2015):

$$L(p_{i,}\ t_i) = \frac{1}{N_{cls}}\sum_i L_{cls}\,(p_i, p_i^*) + \lambda \frac{1}{N_{reg}}\sum_i p_i^*\, L_{cls}\,(t_i, t_{i,}^*) \quad (1)$$

where i = index of an anchor
pi = the prediction probability of anchor i
pi* = the ground truth label
Lcls = the classification loss
Lreg = the regression loss
t_i = vector representation of predicted bounding box
ti* = ground truth bounding box

λ = balancing the loss function terms
Ncls , Nreg = the normalization parameters of the classification and regression losses

2.3 Backbone networks

Degradation problems may arise when CNN networks are created with a deeper structure. The higher-level layers might simply serve as an identity function as the architecture gets deeper. The feature maps produced, which constitute their output, resemble the raw data. This results in accuracy saturation, which is quickly followed by degradation. This issue can be resolved by utilizing the residual blocks as shown in Figure 3. The residual blocks can be utilized to alter the function as equation 2 (He et al., 2016):

$$H(x) = F(x) + x \quad (2)$$

where F(x) = stacked non-linear layers
x = identity function
H(x) = mapping to a function

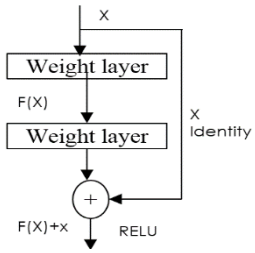


Figure 3. Residual block diagram

The deep residual networks (Resnet), a novel type of convolutional neural network design that is far deeper (up to 152 layers) than those previously implemented, were proposed by The et al. Resnet uses residual or skip connections to simplify network training. In 2015, Resnet took first place in several computer vision challenges, including COCO detection and ImageNet detection. Resnet50 replaces each 2-layer block in the 34-layer net with this 3-layer bottleneck block, resulting in a 50-layer ResNet (He et al., 2016) as shown in Figure 4.

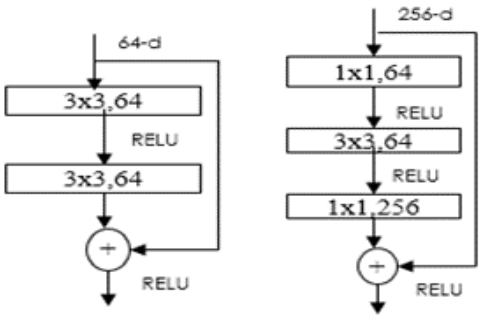


Figure 4. Difference between Residual block of Resent 34 and 50.

3. PROPOSED METHODOLOGY

In this section, two main experiments are implemented using Faster RCNN with Resnet-50 as the foundation CNN. Faster RCNN has shown to be better framework for detecting aircraft and is ideal for real-world with little training data situations (Alganci et al., 2020; Azam et al., 2022). In the first experiment, Faster RCNN model is trained by aerial set as shown in Figure 5. Second experiment, Faster RCCN is trained by GF-2 Satellite set as shown in Figure 6. Each experiment has two scenarios: the default anchor sizes are used in the first scenario and k-mean algorithm is used to define the anchor sizes in the second scenario. Each model is then tested by aerial, GF-2, JL-1, Pleiades satellites test sets. Information about the used satellite and aerial images of DOTA, Airbus Aircraft Detection dataset and image split process are given initially. A detailed description of the steps and parameterization of the training process are also given.

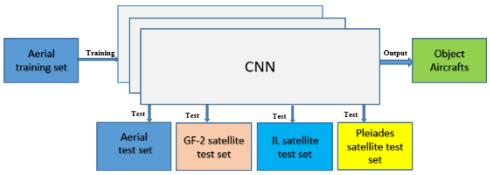


Figure 5. Faster RCNN trained by aerial set, testing by aerial, GF-2, JL-1 and Pleiades satellite test set.

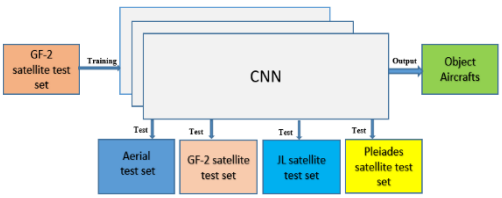


Figure 6. Faster RCNN trained by GF-2 satellite set, testing by aerial and GF2- satellite test sets.

3.1 EXPERIMENTDATASETS

The aircraft detection algorithms must be applied to images of various sizes, resolutions, and sources. As a result, the experiments implemented two aircraft datasets: a DOTA and an Airbus Aircraft Detection dataset. Two datasets are used to validate the efficacy of the trained models.

3.1.1 DOTA dataset: Although object detection in natural scenes has made significant progresses in the past decade, aerial and satellite imagery have exposed to the enormous variation in scale, orientation, and shape of object instances on the earth's surface as well as the lack of well-annotated datasets of objects in aerial scenes. A large-scale dataset for object detection in aerial images promotes object detection research in Earth Vision, also known as Earth Observation and Remote Sensing (DOTA)(Xia et al., 2018).

For testing and training, the DOTA dataset is used. It is an open-source dataset for using remote sensing photos to identify objects. The DOTA images are collected from the Google Earth, GF-2 and JL-1 satellite provided by the China Centre for Resources Satellite Data and Application, and aerial images provided by CycloMedia B.V. DOTA consists of RGB images and grayscale images. The RGB images are from Google Earth and CycloMedia, while the grayscale satellite images. There are 15 different categories, including a roundabout, storage tank,

baseball diamond, tennis court, basketball court, ground track field, harbor, and aero plane, a field and a pool. The images are collected from different sensors and platforms as in Table 1. Each image is of the size in the range from **800 × 800 to 20,000 × 20,000** pixels and contains objects exhibiting a wide variety of scales, orientations, and shapes.

Item	GF-2	JL-1	Aerial
Resolution	1 meter	70 cm	10:15 cm
Bands	grayscale	grayscale	RGB

Table 1. Specification of DOTA Dataset

3.1.2 Airbus Aircraft Detection dataset

The dataset contains 103 Pleiades images with a resolution of about 50 cm. Each image is saved as a JPEG file with a resolution of 2560 x 2560 pixels (i.e. , 1280 metres on ground). It has various types of aircrafts. Furthermore, for variety, some airports captured multiple times at different acquisition dates with different atmospheric shooting conditions such as fog or cloud (https://www.kaggle.com).

3.1.3 Image split: In this research, custom tool for DOTA V1.5 is applied to divide dataset into three categories according source aerial or GF-2, JL-1 satellites images, which contain aircrafts. DOTA and Airbus Aircraft Detection images resized to 600 x 600. Labelled DOTA and Airbus Aircraft Detection dataset are converted into COCO format for training and testing. The number of images in aerial, GF-2 training sets and aerial, GF-2, JL-1, Pleiades satellites test sets are in listed Table 2.

Item	Aerial	GF-2	JL-1	Pleiades
Training images	719	800	-	-
Testing images	121	293	181	200

Table 2. number of training and testing sets

3.2 Training

In this work, two experiments are performed with the PyTorch open-source deep learning framework. Transfer learning technique is applied by using the pre-trained Faster RCNN network with the COCO dataset(Alganci et al., 2020). Fine-tuning of the parameters and extending the training set with the DOTA datasets are also applied.

3.2.1 Hyperparameters good selection of parameters, such as optimizer function, number of epochs, and learning rate is very important to train architecture. The Stochastic Gradient Descent Method (SGDM) typically delivers good results for transfer learning and Adam performs better when starting from scratch (Azam et al., 2022).

The Nesterov momentum is applied during the training process. The traditional momentum strategy makes a huge jump in the direction of the updated accumulated gradient before first calculating the gradient at the current position. In contrast, Nesterov momentum jumps significantly in the direction of the previously collected gradient, measures the gradient upon arrival, and then makes adjustments. The Nesterov momentum approach reduces number of iterations to converge for the global minimum (Huang et al., 2019).

The Anchor boxes are crucial settings for Faster RCNN object optimization. value of the box size varies depending on the dataset. The shape, scale, and the number of anchor boxes impact the efficiency and precision of the detectors(Ren et al., 2015). For In the first scenario anchor sizes 128^2, 256^2, and 512^2 pixels, and 3 aspect ratios of 1:1, 1:2, and 2:1 are used. While, in the second scenario, K means algorithm is applied to cluster the dimensions (length and width) of objects of training and testing datasets into three clusters. Figures 5, 6, 7 show samples of clustering objects the training and test sets. Areas of objects in each cluster of the training and testing datasets are calculated in Table 3, where the smallest and the largest areas were 34^2, 376^2 pixels, respectively. Hence, the convenient sizes of anchor boxes 32^2, 64^2 ,128^2, 256^2, and 512^2 pixels are selected.

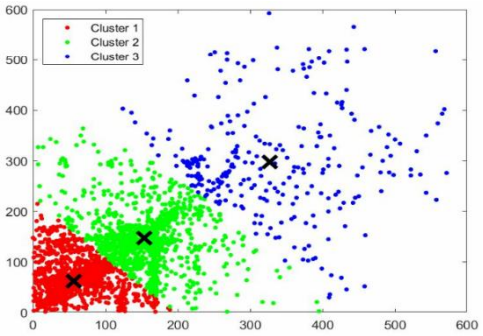


Figure 5. K means clusters aerial training set

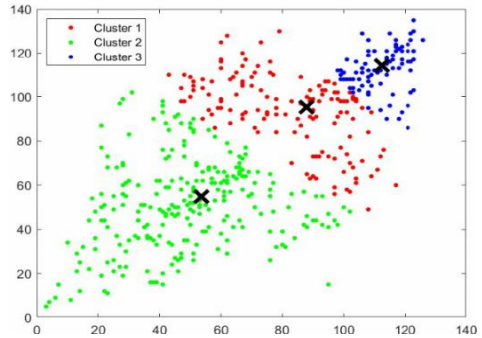


Figure 6. K means clusters GF-2 satellite training set

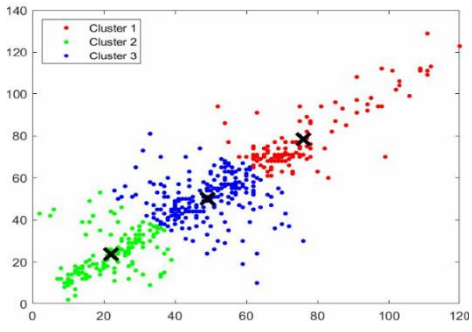


Figure 6. K means clusters JL-1 satellite test set

Area of clusters	Aerial	GF-2	JL-1	Pleiades
Small	90	56	**34**	60
Medium	208	96	64	108
Large	**376**	116	101	161

Table 3. Area of clusters in training and test set

Weight decay is a regularization technique in deep learning. It works by adding a penalty term to the cost function of a neural network which has the effect of shrinking the weights during

backpropagation. This prevents the network from overfitting the training data as well as the exploding gradient problem. The stochastic gradient with a batch size of 4 is applied, beginning with momentum term of 0.9, learning rate 10^{-2}, also the number of epochs 13, are used for training models.

3.2.2 Augmentation: There are numerous approaches that use deep convolutional networks to tackle computer vision problems to outperform current benchmarks. One of the hardest problems is enhancing these models' capacity for generalization. A model's performance on data it has previously seen (training data) vs data it has never seen before is measured by its generalizability (testing data). Poorly generalizable models have overfitted the training set (Shorten and Khoshgoftaar, 2019). The effectiveness of data augmentations is applied during training from simple transformations such as horizontal and vertical flipping, color space augmentations, and random brightness to avoid overfitting problems. Moreover, in the first experimental, second scenario hue saturation value (HSV) and RGB shift methods are applied to imitate lightning and atmospheric conditions.

4. ACCURACY ASSESSMENT

In this section, we discuss evaluation metrics and analysis results of our work for aircraft detection in aerial and satellite images. COCO metrics is used to measure mean average precision (mAP) and mean average recall (mAR) with different intersection over union (IOU) and visual inspection of the data.

4.1 Evaluation metrics: The average precision (AP) and the average recall (AR) score are two often used performance indicators in object detection. At each iteration of the training process, a detector compares the predicted bounding boxes with the ground truth bounding boxes using the intersection over union (IOU). Accordingly, the network considers a prediction to be accurate if the predicted object's bounding box overlaps the ground truth box by at least 50%. The recall is the ratio of the number of successfully identified items to the total number of ground truth samples, and the accuracy is the percentage of correct matches among all objects that are detected as matches. Because the recall rate and precision rate alone are insufficient to assess the effectiveness of the framework. By defining the true positive (TP) as truly detected objects, the false negative (FN) as non-detected objects, and the false positive (FP) as falsely detected objects, the precision, recall was calculated as The COCO metric API is also utilized to evaluate the characteristics and effectiveness of the object detection algorithms using 12 different metrics shown in TABLE 4.

The average recall (AR) and average precision (AP) are determined by averaging over 10 different IOU ranging from 0.5 to 0.95 with 0.05 intervals, unless otherwise specified. Additionally, the values for AP where IOU is 0.5 and 0.75 are computed. According to all categories and IOU values, AP represents the calculation's average precision. There is only one object category which is airplane in this study. Averaging across categories and IoUs, AR represents the percentage of correct objects per image. The interpretation of the enclosing box areas is used to further verify these calculations. AR is the maximum number of detections per image, averaged over categories and IoUs. These calculations are also checked by interpreting the bounding box areas. According to COCO, objects with a size smaller than 32^2 pixels are defined as small, between 32^2 and 96^2 as medium, and more than 96^2 pixels as large. The metric calculations are performed according to all scale levels and for separate scales and individual scales are taken into consideration while doing the metric computations (Alganci et al., 2020).

Metric	Calculated for
M1	AP for [IoU = 0.50:0.95 | area = all | maxDets = 100]
M2	AP for [IoU = 0.50 | area = all | maxDets = 100]
M3	AP for [IoU = 0.75 | area = all | maxDets = 100]
M4	AP for [IoU = 0.50:0.95 | area = small | maxDets = 100]
M5	AP for [IoU = 0.50:0.95 | area = medium | maxDets = 100]
M6	AP for [IoU = 0.50:0.95 | area = large | maxDets = 100]
M7	AR for [IoU = 0.50:0.95 | area = all | maxDets = 1]
M8	AR for [IoU = 0.50:0.95 | area = all | maxDets = 10]
M9	AR for [IoU = 0.50:0.95 | area = all | maxDets = 100]
M10	AR for [IoU = 0.50:0.95 | area = small | maxDets = 100]
M11	AR for [IoU = 0.50:0.95 | area = medium | maxDets = 100]
M12	AR for [IoU = 0.50:0.95 | area = large | maxDets = 100]

Table 4. COCO metrics Evaluation

4.2 Evaluation with Accuracy Metrics: The performances of aerial and satellite trained model are examined with the COCO metric. According to the COCO metrics, the aerial trained model has the best results in two scenarios when considering the mean of the precision for different IoU values. GF-2 satellite trained model provided promising results in the two scenarios for mAP of 0.5 IoU with GF-2 satellite test set, while aerial trained model is better if the high mAP of 0.5 IOU for aerial and GF-2, JL-1 and Pleiades satellite test set is desired. For metrics 4, 5, and 6 aerial model in the second scenario provides the best mAP result for different IOUs in small, medium, and large objects for the aerial and satellite test sets compared with that in the first scenario. However, in aerial, Pleiades and JL-1 satellite test sets, the GF-2 satellite trained model performs poorly for small objects in the two scenarios. The seventh, eighth, and ninth metrics provide information about the recall rates for all different IOUs according to the detection number per image. Similarly, the aerial trained model provides better results according to these metrics. When the AR results are investigated according to metric 10, it shows that the recall rates of Gf-2 satellite model are worse than aerial models for small object aerial and satellite test set in two scenarios. Coco metrics of first experiment of two scenarios aerial trained model and second experiment of two scenarios are shown in Tables 5,6 and Tables 7,8 respectively.

Metric	Aerial	GF-2	JL-1	Pleiades
M1	0.611	0.541	0.121	0.174
M2	0.902	0.831	0.227	0.579
M3	0.731	0.657	0.107	0.059
M4	0.297	0.216	0.022	0
M5	0.679	0.557	0.141	0.168
M6	0.522	0.767	0.56	0.235
M7	0.145	0.203	0.074	0.101
M8	0.492	0.613	0.166	0.25
M9	0.67	0.622	0.167	0.25
M10	0.415	0.292	0.031	0.006
M11	0.725	0.627	0.191	0.228
M12	0.603	0.80	0.642	0.331

Table 5. COCO metrics of first experiment first scenario, trained model by aerial datasets according to aerial and Satellite test sets.

Metric	Aerial	GF-2	JL-1	Pleiades	
					+ HSV
M1	0.493	0.656	0.253	0.094	0.243
M2	0.898	0.977	0.506	0.378	0.656
M3	0.468	0.793	0.213	0.028	0.118
M4	0.199	0.231	0.039	0.005	0.001
M5	0.532	0.661	0.324	0.092	0.204
M6	0.134	0.815	0.646	0.131	0.346
M7	0.423	0.226	0.109	0.072	0.136
M8	0.57	0.691	0.297	0.182	0.329
M9	0.321	0.698	0.306	0.187	0.339
M10	0.312	0.3	0.064	0.019	0.028
M11	0.599	0.708	0.388	0.164	0.295
M12	0.57	0.828	0.711	0.251	0.462

Table 6. COCO metrics of first experiment second scenario, trained model by aerial datasets according to aerial and Satellite test sets.

Metric	Aerial	GF-2	JL-1	Pleiades
M1	0.353	0.708	0.224	0.153
M2	0.681	0.96	0.394	0.484
M3	0.324	0.855	0.22	0.033
M4	0.106	0.212	0.01	0
M5	0.45	0.721	0.282	0.132
M6	0.18	0.81	0.576	0.222
M7	0.07	0.221	0.102	0.109
M8	0.304	0.745	0.256	0.216
M9	0.428	0.753	0.261	0.216
M10	0.144	0.227	0.013	0
M11	0.543	0.769	0.329	0.18
M12	0.217	0.843	0.65	0.308

Table 7. COCO metrics of second experiment, first scenario, trained model by GF-2 according to aerial and Satellite test sets.

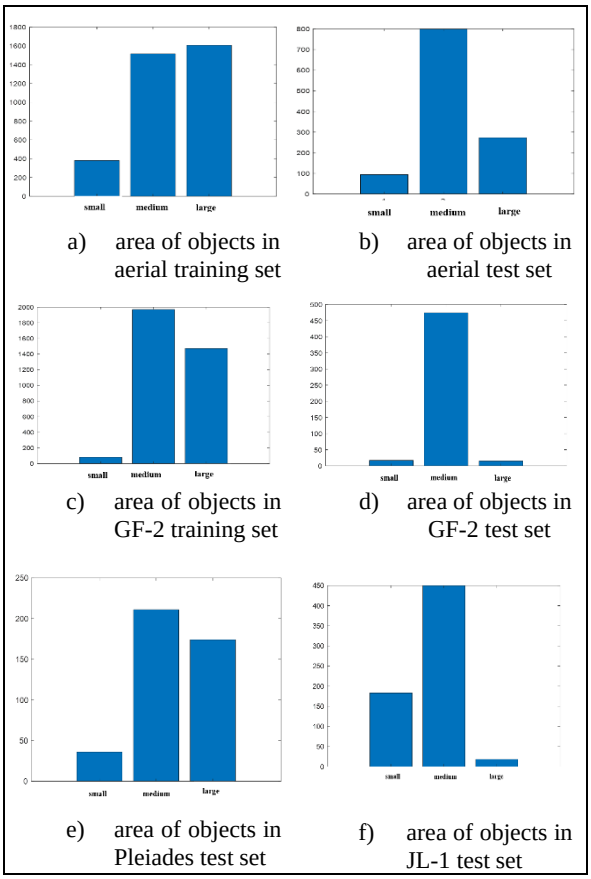
Metric	Aerial	GF-2	JL-1	Pleiades
M1	0.244	0.685	0.268	0.094
M2	0.547	0.969	0.516	0.371
M3	0.203	0.854	0.254	0.017
M4	0.046	0.171	0.015	0.001
M5	0.351	0.698	0.355	0.098
M6	0.047	0.885	0.682	0.138
M7	0.05	0.225	0.124	0.069
M8	0.245	0.724	0.32	0.194
M9	0.348	0.733	0.327	0.198
M10	0.085	0.232	0.025	0.022
M11	0.473	0.746	0.433	0.189
M12	0.106	0.887	0.733	0.249

Table 8. COCO metrics of second experiment, second scenario, trained model by GF-2 datasets according to aerial and Satellite test sets.

4.3 Analysis of the results

When the results of two experiments, first scenario for aerial and GF-2 satellite trained models are compared for metrics 4, 10 of small objects, the results in test set for the GF-2 satellite trained model provided low results. There is a performance gap for detecting small aircraft compared to medium and large aircraft. The mAP results of aerial trained model according to aerial, GF-2, JL-1 and Pleiades satellite test set are 0.902, 0.831, 0.227, 0.579 respectively, however, the mAP results of GF-2 trained model according to aerial, GF-2, JL-1 and Pleiades satellite test sets are 0.68, 0.96, 0.394 and 0.484 respectively. Also, the fourth and tenth metrics of small object for aerial trained model is higher than the fourth and tenth metrics of small objects for GF-2 satellite trained model.

The bounding box area distributions of aircraft samples for the aerial and GF-2 training, aerial, GF-2, JL-1 and Pleiades satellite test datasets are examined. It is found that the aerial train set contains nearly the same distribution as the aerial, GF-2, JL-1 and Pleiades satellite test datasets, with small, medium and large areas, while GF-2 satellite train set does not include the same number of small aircraft areas less than 32^2 pixels in the aerial, GF-2, JL-1 and Pleiades satellite set. The main reason behind the performance gap is that the dimensions of small aircrafts inside the GF-2 satellite training set. They are distributed differently than the aerial, GF-2, JL-1 and Pleiades satellite sets contain different types of aircraft, as shown in Fig.10.



a) area of objects in aerial training set
b) area of objects in aerial test set
c) area of objects in GF-2 training set
d) area of objects in GF-2 test set
e) area of objects in Pleiades test set
f) area of objects in JL-1 test set

Figure 7. The distribution areas of objects in datasets

When the results of the two scenarios in the first experiment aerial satellite trained model are compared with the mAP, results in aerial, GF-2, JL-1, Pleiades satellite test sets for the first scenario has poorer performance. A key to object detection quality is the selection of anchor boxes that cover every possible combination of object sizes found in a dataset. The proposed anchors should naturally include the data's varying aspect ratios and scales. K-means is used to estimate the ideal bounding boxes. The choices of anchors size in the second scenarios improves the accuracy of detection. Moreover, the proposed anchor sizes with (HSV) and RGB shift augmentation methods, especially for colour images enhances the results as Pleiades test set. This case study also confirms the importance of HSV and RGB shift

augmentation for testing and training colour images, where the mAP of Pleiades test set increases from 0.378 to 0.656. The mAP results of the first experiment aerial trained model second scenario according to aerial, GF-2, JL-1 and Pleiades satellite test sets are 0.898, 0.977, 0.506, 0.656 respectively, however, the mAP results of first scenario according to aerial, GF-2, JL-1 and Pleiades satellite test sets are 0.902, 0.831, 0.227, 0.579 respectively.

The mAP of two scenarios for the second experiment GF-2 satellite trained model are compared according to the results in test sets. The first scenario has a high mAP than the second scenario in aerial and Pleiades satellite test sets. However, the second scenario gives a higher mAP than the first scenario in JL-1satellite test set. The mAPs of both scenarios are the same for GF-2 test set. The results indicate that the distribution of objects in datasets could improve the accuracy of GF-2 satellite trained model besides the proposed convenient anchor sizes.

The results shows that the accuracy of the two experiments aerial trained model according to aerial, GF-2, JL-1 and Pleiades satellite test set are 0.898, 0.977, 0.506, 0.656 respectively, and GF-2 trained model according to aerial, GF-2, JL-1 and Pleiades satellite test set are 0.68, 0.96, 0.394 and 0.484 respectively. Both experiments gave higher accuracy compared to the DOTA trained model according to DOTA and Pleiades satellite test sets 0.717, 0.364 respectively(Alganci et al., 2020). Hence, the aerial and GF-2 trained models are more likely to perform well in aerial and satellite tests than the DOTA-trained model.

Finally, this section presents some samples images of the two experiments. The detection results from the aerial and satellites test set are inspected visually to assess the performance of two different trained models. The detection of aerial trained model in the first experiment for aerial and satellites test set provide selected samples in Figures 8,9,10,11 which include different sized aircrafts, and the image patches have illuminance difference, background complexities, and different band information. The detection results of the second experiment using two the Gf-2 satellite trained model for aerial and satellites test set are depicted in, Figures 12, 13,14, 15.

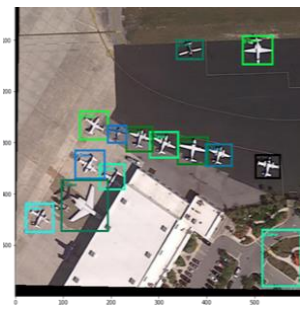


Figure 8. Aerial image.

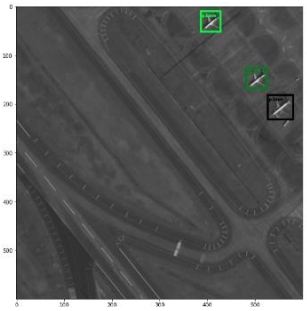


Figure 9. GF-2 image.

Figure 10. JL-1 image.

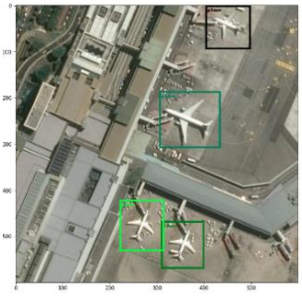


Figure 11. Pleiades image.

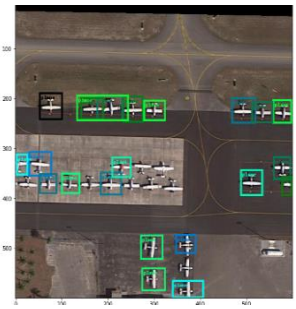


Figure 12. Aerial image.

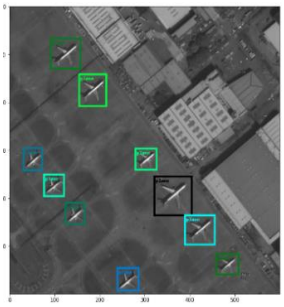


Figure 13. GF-2 image.

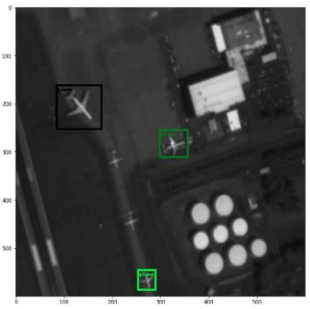


Figure 14. JL-1image.

Figure 15. Pleiades image.

5. CONCULSION

Modern deep learning-based object detection frameworks and convolutional neural networks have opened up many new ways to improve the accuracy, robustness, and detection speed of aircraft detection algorithms. The networks trained by the aerial and Gf-2 satellite sets separately from the DOTA datasets and the performance of them was evaluated with the aerial, GF-2, JL-1 and Pleiades satellite test datasets. The best results are obtained from the aerial trained network according to the COCO metrics. The satellite trained model also provides promising results. Results were also impacted by the object sizes, diversities and anchor sizes. Hence, using K means algorithm is the effective approach to estimate appropriate anchor sizes for object detection models. In summary, transfer learning and parameter tuning approaches on pre-trained object detection networks generate promising results for airplane detection from satellite and aerial images. Additionally, because R, G, and B bands are primarily created for natural images, the object detection network frequently uses them. However, a satellite detection model that trained by one band (grayscale) images and low resolution 1 m, shows promising results for both training and testing sets. The approaches of data augmentation consist of altering the intensities of the RGB channels in training images improve the performance of detection model, especially when using RGB images from heterogenous sensors. Faster RCNN with Resnet-50 as backbone CNN turned out to be a promising framework for aircraft detection, which is suitable for detection aircrafts. with high precision. Also, object detection from heterogeneous sensors overcomes the gap of revisit time between sensors and cost.

A comparative study is conducted between the dedicated object detection model and generalized object detection model on optical various aircrafts detection datasets. The intent is to investigate the effectiveness of the dedicated model under

various conditions (such as various resolutions, different bands, sizes, heterogenous sources). The results show that the dedicated object detection model achieves better performance, even under different sensors datasets. In addition, the performance of the dedicated training models with appropriate anchor sizes selection for aircraft detection are tested in large various images, which have good robustness.

6. ACKNOWLEDGMENTS

This work was partially supported by Dr. Naser El-Sheimy research funds from NSERC and Canada Research Chairs programs. Thanks also goes to the funding of the first author by the Egyptian government.

7. REFERENCES

Alganci, U., Soydas, M., Sertel, E., 2020. Comparative Research on Deep Learning Approaches for Airplane Detection from Very High-Resolution Satellite Images. Remote Sensing 2020, Vol. 12, Page 458 12, 458. https://doi.org/10.3390/RS12030458

Ansari, R.A., Buddhiraju, K.M., Malhotra, R., 2020. Urban change detection analysis utilizing multiresolution texture features from polarimetric SAR images. Remote Sens Appl 20, 100418. https://doi.org/10.1016/J.RSASE.2020.100418

Azam, B., Khan, M.J., Bhatti, F.A., Maud, A.R.M., Hussain, S.F., Hashmi, A.J., Khurshid, K., 2022. Aircraft detection in satellite imagery using deep learning-based object detectors. Microprocess Microsyst 94. https://doi.org/10.1016/J.MICPRO.2022.104630

Chen, X., Xiang, S., Liu, C.L., Pan, C.H., 2013. Aircraft detection by deep belief nets. Proceedings - 2nd IAPR Asian Conference on Pattern Recognition, ACPR 2013 54–58. https://doi.org/10.1109/ACPR.2013.5

Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing 117, 11–28. https://doi.org/10.1016/J.ISPRSJPRS.2016.03.014

Classification, R.F., Change, R., Analysis, V., n.d. remote sensing Long-Term Land Use / Land Cover Change Assessment of the Kilombero Catchment in Tanzania Using Random Forest Classification and Robust Change Vector Analysis.

Gao, F., Xu, Q., Li, B., 2013. Aircraft detection from VHR images based on circle-frequency filter and multilevel features. The Scientific World Journal 2013. https://doi.org/10.1155/2013/917928

Gidaris, S., Komodakis, N., 2016. LocNet: Improving Localization Accuracy for Object Detection.

Girshick, R., 2015. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision 2015 Inter, 1440–1448. https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-Decem, 770–778. https://doi.org/10.1109/CVPR.2016.90

Huang, W.R., Emam, Z., Goldblum, M., Fowl, L., Terry, J.K., Huang, F., Goldstein, T., 2019. Understanding Generalization through Visualizations 1–11.

Maher, A., Gu, J., Zhang, B., 2018. Deep-Patch Orientation Network for Aircraft Detection in Aerial Images. Communications in Computer and Information Science 757, 178–188. https://doi.org/10.1007/978-981-10-7389-2_18/FIGURES/9

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 39, 1137–1149. https://doi.org/10.48550/arxiv.1506.01497

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 1–48. https://doi.org/10.1186/S40537-019-0197-0/FIGURES/33

Tian, S., 2020. Hi-UCD : A Large-scale Dataset for Urban Semantic Change Detection in Remote Sensing Imagery 1–6.

Wu, H., Zhang, H., Zhang, J., Xu, F., 2015. Fast aircraft detection in satellite images based on convolutional neural networks. Proceedings - International Conference on Image Processing, ICIP 2015-December, 4210–4214. https://doi.org/10.1109/ICIP.2015.7351599

Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 3974–3983. https://doi.org/10.1109/CVPR.2018.00418

Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X., 2019. Object Detection with Deep Learning: A Review. IEEE Trans Neural Netw Learn Syst 30, 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865

Zhou, P., Cheng, G., Liu, Z., Bu, S., Hu, X., 2016. Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. Multidimens Syst Signal Process 27, 925–944. https://doi.org/10.1007/s11045-015-0370-3

Zhu, M., Xu, Y., Ma, S., Li, S., Ma, H., Han, Y., 2019. Effective Airplane Detection in Remote Sensing Images Based on Multilayer Feature Fusion and Improved Nonmaximal Suppression Algorithm. Remote Sensing 2019, Vol. 11, Page 1062 11, 1062. https://doi.org/10.3390/RS11091062