Detection of Multiclass Objects in Satellite Images Using an Improved Algorithmic Approach

Abhimanyu Singh[1],
Computer Science and Engineering Department
Defense Institute of Advanced Technology Pune, India
[1] abhiman_pce@yahoo.co.in,

Manisha J Nene[2]
Computer Science and Engineering Department
Defense Institute of Advanced Technology Pune, India
[2] mjnene@diat.ac.in

***Abstract*— Object Detection (OD) in natural images has made tremendous strides during the last ten years. However, the outcomes are infrequently adequate when the natural image OD approach is used straight to Satellite Images (SI). This results from the intrinsic differences in object scale and orientation introduced by the omniscient viewpoint of the SI. Detecting objects is a challenging task especially when small object areas and complicated backgrounds appear in satellite images under analysis. Occlusion and intense item overlap have a further negative effect on the detection performance. The self-attention mechanisms are proposed to search for minute details in an image. However such searches mechanism come with complexity or high computational cost due to uncertainty induced in visual resolutions. The study in this research paper addresses the problems experienced in the accuracy and precision and the efficacy of the proposed model is demonstrated with the result in this paper.**

Keywords—Object Detection, iSAID, SGD, SGDM, YOLO, OD, FC.

I. Introduction

In order to observe the surface of the globe, high-resolution SIs is often taken utilizing earth satellite technology. For existing interpretation algorithms, processing a high number of SIs poses a significant challenge, though. OD, or the precise and efficient recognition of preset objects from images, is one of the most fundamental issues in computer vision. It can be used in a variety of applications, including traffic control in cities and precision farming [1][3].

Data-driven deep learning algorithms have considerably aided recent breakthroughs in segmentation and OD tasks [4][6]. The training dataset's size and quality both affect the detection precision [7]. However, it is still difficult to identify things in optical SIs. Below is a list of the causes, firstly comparing to natural images captured by terrestrial cameras with horizontal perspectives. Secondly various imaging conditions such perspectives, lighting, and occlusion, objects in SIs commonly have different visual appearances and optical properties [11][12].

Feature extraction is crucial for OD, dealing with large datasets often involves a lot of time and staffing. In contrast, by autonomously extracting trustworthy feature representations from the raw data, data-driven deep learning algorithms outperform conventional extraction techniques. Additionally, they can decrease the stress brought on by conventional feature engineering and modeling.

The aforementioned methods are partial frameworks; however there would be limitations on the detection's speed. OD and categorization require pre-processing, such as the application of a matching filter [28] and to improve the classification accuracy, additional classifiers, such as 3-dimesional segmentation, can be applied [29]. SIs generally has high resolutions so the time complexity of the image must be kept very low.

This paper discusses numerous experiments conducted to propose a model that is both improved and also meets all the requisites for object detection. The outline of the paper is structured as preliminaries and notations followed by the proposed methodology, application scenarios, results achieved and the proposed future work that can be undertaken.

II. PRELIMINARIES AND NOTATIONS

A. Nature of Data

The number and variety of training datasets can be increased using a variety of data augmentation approaches, which also enhances the detection model's robustness and generalizability. Using distortion, rotation, and scaling, early data augmentation approaches improved the accuracy of picture classification. Additionally, photometric adjustments are frequently used. Increasing the dataset, for instance, by changing the colour, saturation, and value of the training data, the training dataset's images from multiple categories are randomly combined and weighted.

B. Cutout

Cutout randomly removes portions of the sample image and replaces them with blank pixels. The advantages of the aforementioned methods are combined in mosaic. The fundamental idea is to use four photographs that were randomly cropped and then stitched together to create one image as training data. It was initially suggested in YOLOv4. It dramatically enhances the backgrounds of the pictures. Four photographs will inevitably result in a larger batch size. Therefore, batch normalisation must be used to calculate the four images.

(978-1-6654-9291-1/22/$31.00 ©2022 IEEE)

C. *Mixup -Mosaic*

By effectively enlarging the satellite dataset, Mixup and Mosaic greatly improve OD performance, especially for small objects. Therefore, added Mixup was used.

In this work, referring to the observations by expanding the data it gives good label and yields good results. This paper, works towards establishing a performance baseline and enable the identification of improvement areas for multi object detection in satellite images.

III. APPLICATION SCENARIOS & PROPOSED TECHNIQUE

A. *Complexity of Data and Data set*

A portion of the iSAID Dataset was used to conduct the experiment. The photos used in iSAID are identical to those in the DOTA-v1.0 dataset and are primarily gathered from Google Earth. Some of the images were captured by the satellite JL-1, while others were by the satellite GF-2 of the China Centre for Resources Satellite Data and Application.

The primary data set as a benchmark for object detection in aerial photos, DOTA is a sizable dataset. Plane, ship, tank, baseball diamond, tennis court, basketball court, ground track field, harbor, bridge, large vehicle, small vehicle, helicopter, roundabout, soccer field, and swimming pool are among the object classes in iSAID. This massive iSAID dataset consists of 2,806 high-resolution photos with 655,451 object instances for 15 categories.

B. *Data Pre-conditioning*

In the beginning, the dataset's photos were divided into tiles with n*n dimensions. As a result, n^2 tiles in total were obtained from the iSAID pictures. For the training set random sub-sampling was conducted. The exact same dataset was used in the experiments.

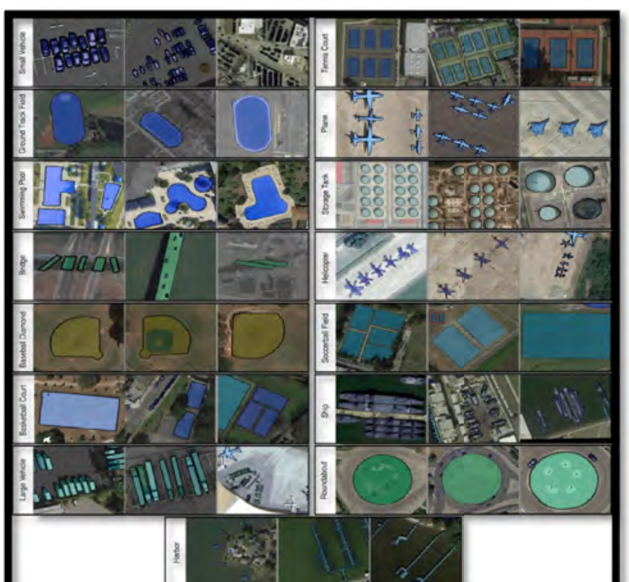

Fig 1: Collection of Classes in iSAID Data Set

Figure 2 depicts the data set's collection of classes. These classes include a wide range of classes such as ships, aircrafts, helicopters, tanks, and huge vehicles. Aside from the items, the data collection includes photographs of places, water bodies, small and big lawns, and badminton courts.

IV. EXPERIMENT & OBSERVATIONS

A. *Step 1- : Training the model with SGD optimizer*

The initial learning rate was set to 0.01 and the optimizer was set to SGD. a baseline assessment was required to make comparisons with next experiments. It was carried out using 10 epochs on the iSAID Dataset's sub-sampled tiles. Figure 3 represents the result achieved which was mAP@0.5 | F1 max 0.44 at 0.403 confidence | PR AUC 0.487. a model was generated which can be used to refine future experiments.

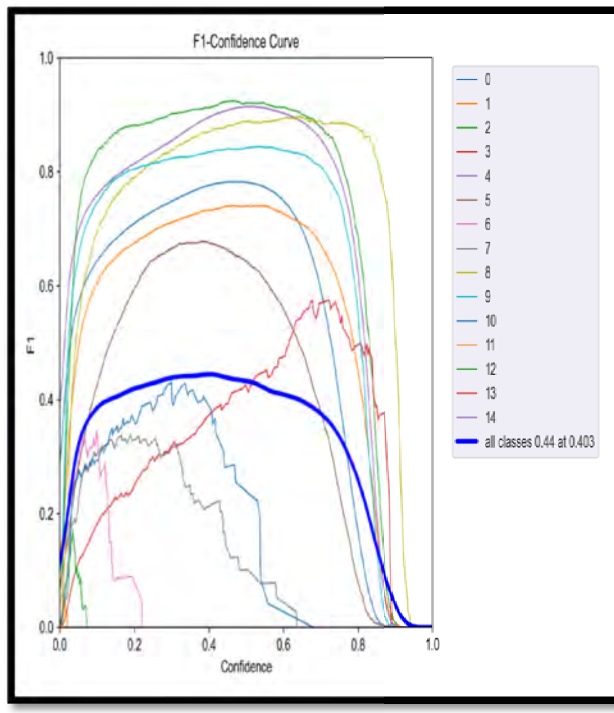

Fig 2. F1 curve with SGD Optimizer

B. *Step 2 : Training with decreased SGD Optimizer at 0.917*

The model's SGD momentum was decreased to 0.917 to verify that whether slowing down the momentum leads to better outcomes. It was carried out using 10 epochs on the iSAID Dataset's sub-sampled tiles. Figure 4 represents the result which is mAP@0.5 | F1 max 0.46 at 0.459 confidence | PR AUC 0.496. It was observed the model reflected enhanced performance.

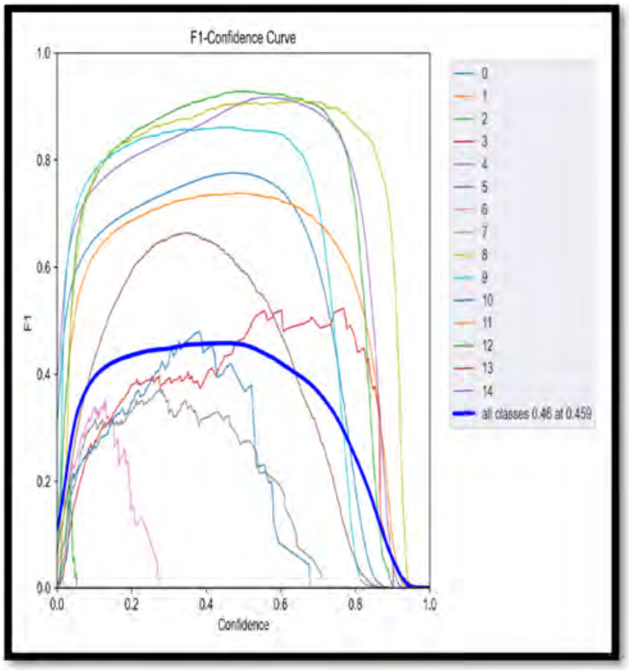

Fig 3. F1 curve for SGD Optimizer at 0.97

C. Step 3- Reducing the weight decay of SGDM Model

The model's SGD optimizer weight decay was lowered to 0.00049 to verify that by slowing down weight deterioration outcomes improves or not. It was carried out using 10 epochs on the iSAID Dataset's sub-sampled tiles. Figure 5 reflects the result which is mAP@0.5 | F1 max 0.46 at 0.422 confidence | PR AUC 0.493. It was observed that the performance was not enhanced so the weight decay of 0.0005 was maintained, which produced the superior performance.

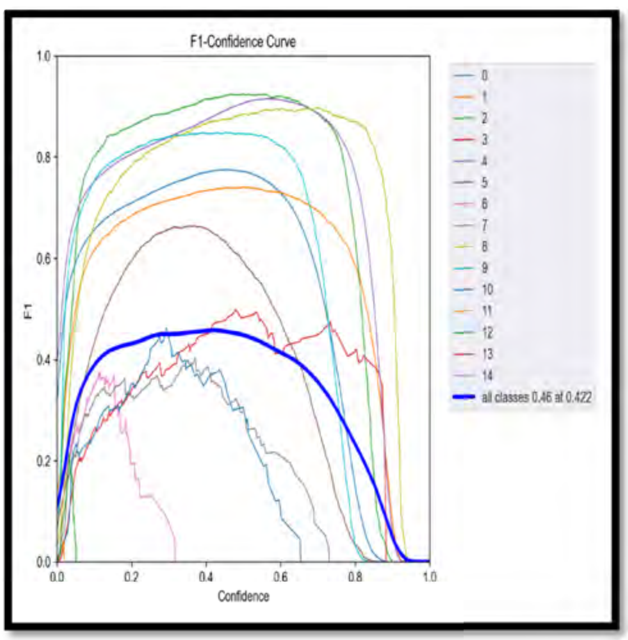

Fig 4. F1 curve for Lowered SGDM Model

D. Step 4: Reduce the Warmup Bias of SGDM Model

The model's warm-up bias learning rate was reduced to 0.099. It was carried out using 10 epochs on the iSAID Dataset's sub sampled tiles. Figure 6 reflects the result which is mAP@0.5 | F1 max 0.46 at 0.335 confidence | PR AUC 0.496. The performance of the model was not further improved so the next experiment was conducted with decreasing the number of epochs.

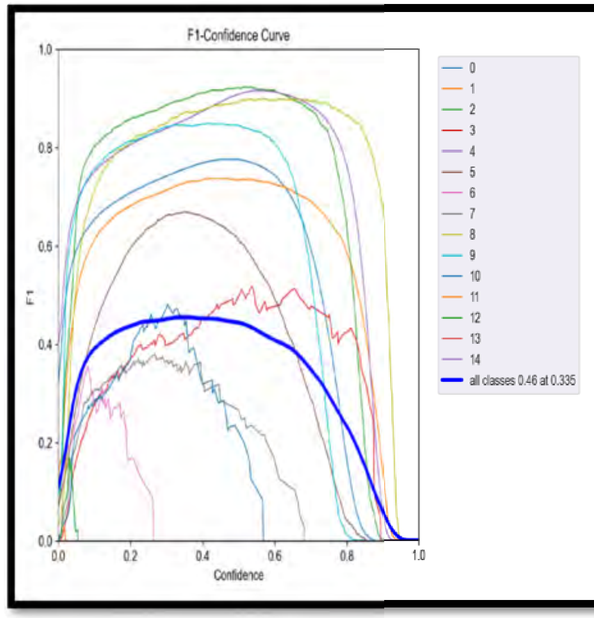

Fig 5. F1 curve for Reduced Warmup Bias

E. Step 5- Decreased Epochs for SGDM Model

The model's warm-up epochs were reduced to 2.0.It was carried out using 10 epochs on the iSAID Dataset's sub-sampled tiles. Figure 7 shows the result which is mAP@0.5 | F1 max 0.47 at 0.384 confidence | PR AUC 0.502. It was learnt that the performance was successfully improved.

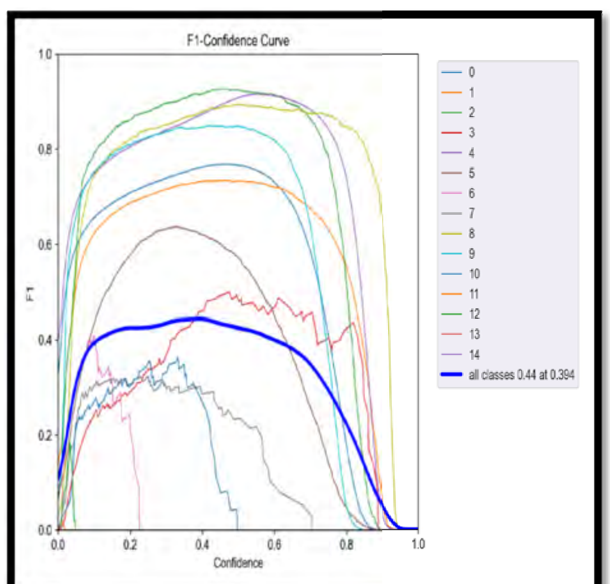

Fig 6. F1 curve for Decreased Epochs

F. Step 6 :Increased Warmup Epochs for SGDM Model

The model's warm-up epochs were raised to 4.0 in an effort to determine whether doing so will improve the models performance. It was carried out using 10 epochs on the iSAID Dataset's sub- sampled tiles. figure 7 shows the result which is mAP@0.5 | F1 max 0.44 at 0.394 confidence | PR AUC 0.482. It was noticed and discovered that the performance was successfully improved.

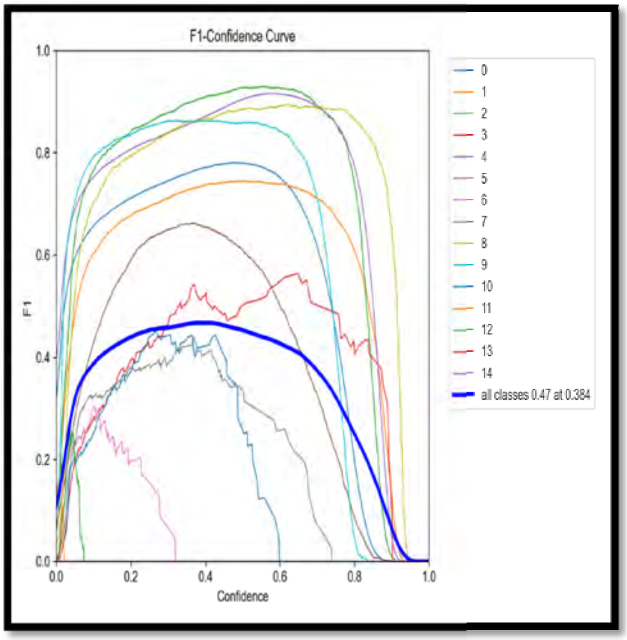

Fig 7. F1 curve for Increased Epochs

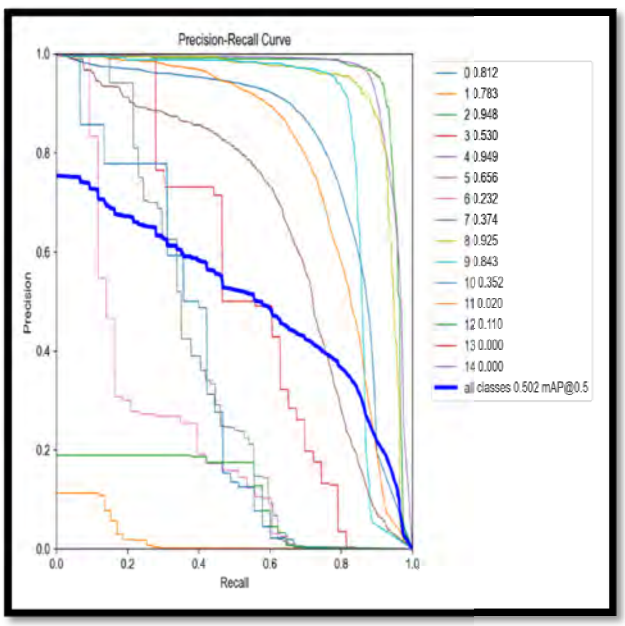

Fig 8. Precision – Recall for Increased Epochs

V. OBSERVATIONS & INFERENCES

The learning rate was set to 0.01 and the optimizer set to SGD, which could then be improved upon by adjusting the parameters. The SGD momentum of the model was reduced to 0.917, and it was found that it showed improved performance.

It was also discovered that by cutting the model's warm-up epochs to 2.0, successfully enhanced the performance. Further, the experiment was carried out with fewer epochs, because the model's performance was not further improved after the warm-up bias learning rate was decreased to 0.099

A. Merit of the Proposed Work

Figure 10 shows the predicted object detection by the model. The images in the column C1 correctly represents predicted ships, aircrafts. The objects like areas, grounds, helipads, small and large vehicles are predicted in columns C2, 3 & 4. Figure 11 represents an input satellite images and corresponding result depicting the aircrafts of different spatial parameters which is possible only after the cumulative results obtained in the experiments conducted.

Traffic cones are recognized as an OD scenario by the technique being used, which leverages a deep learning framework. The model was able to manage the identification task and earn good scores. Results show that our method is successful at finding objects with maybe varying colors and slightly different shapes.

VI.CONCLUSION & FUTURE WORK

Detecting objects in satellite image is a difficult problem because of the variety of target sizes, perspectives, and complicated backdrops. With deep learning's significant success in large-scale ground scene object recognition, object detection research based on deep learning has substantially increased in recent years.

However, because of the diversity of the scale, the direction of the object, and the complexity of the backdrop, the detection impact of aerial photos is comparably less efficient when existing methods are used directly. This paper presents an end-to-end object recognition approach based on an innovative approach towards the neural networks to address numerous problems, including dense objects, objects with variable orientation, and variety in aspect ratios.

Following a thorough assessment with numerous cutting-edge object identification algorithms, the suggested technique is shown to be successful in detecting multiclass artificial objects in aerial photos. The findings on the iSAID datasets reveal that our technique outperforms various state-of-the-art methods in terms of achieving mean average accuracy by tuning various parameters of the algorithm.

The study in this research paper addresses the problems experienced in the accuracy and precision and the efficacy of the proposed model is demonstrated with the result in this paper.

The same tests should be carried out using more latest versions of YOLO, which proved to be the most recent and accurate object identification approach, to further analyze and examine the algorithm's performance on the data set for multiple class objects.

Fig 9. Predicted Object in the Image

(a) *(b)*

Fig 10. SI as an (a) Input and (b) result of Object Detectio

VII. REFERENCES

[1] Sadgrove,, .W. Real-time object detection in agricultural/remote environments using the multiple- expert colour feature extreme learning machine (mec-elm). Comput. Ind. 2018, 98, 183–191.

[2] Reilly, V.; Detection and tracking of large number of targets in wide area surveillance. In Computer Vision— ECCV 2010, Proceedings of the 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 186–199.

[3] Kussul, N , A. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 2017, 14, 778–782.

[4] Zhang, . Stagewise Unsupervised Domain Adaptation with Adversarial Self-Training for Road Segmentation of Remote-Sensing Images. IEEE Trans. Geosci. Remote Sens. 2022, 60, 1–13.

[5] Lin,. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 42, 318–327.

[6] Liu, : Single shot multibox detector. In Computer Vision—ECCV 2016, Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2022; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37.

[7] Wang, Y.; Bashir, S.M.A.; Khan, M.; Ullah, Q.; Wang, R.; Song, Y.; Guo, Z.; Niu, Y. Remote sensing image super-resolution and object detection: Benchmark and state of the art. Expert Syst. Appl. 2022, 197, 116793.

[8] Everingham, M, The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. 2010, 88, 303–338.

[9] Lin, T.Y.. Microsoft coco: Common objects in context. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.

[10] Cheng, G.; Lang, C.; Wu, M.; Xie, X.; Yao, X.; Han, J. Feature enhancement network for object detection in optical remote sensing images. J. Remote Sens. 2021, 2021, 9805389.

[11] Long, Yet Accurate object localization in remote sensing images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 2017, 55, 2486–2498.

[12] Ke, L Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS J. Photogramm. Remote Sens. 2020, 159, 296–307.

[13] Redmon, J A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

[14] Redmon, J Better, faster, stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–27 July 2017; pp. 6517–6525.

[15] Joseph, R Yolov4: An incremental improvement. arXiv 2018, arXiv:1804.02767.

[16] Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. arXiv 2020, arXiv:2004.10934.

[17] Roy, A.M.; Bose, R.; Bhaduri, J. A fast accurate fine-grain object detection model based on yolov4 deep neural network. Neural Comput. Appl. 2022, 34, 3895–3921.

[18] Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. Cspnet: A new backbone that can enhance learning capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1571–1580.

[19] Liu, S aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.

[20] Liu, Z . Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.

[21] Dalal, N Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.

[22] A. Van Etten, et al "You only look twice: Rapid multi-scale object detection in satellite imagery," 2018. [Online]. Available: arXiv: 1805.09512.

[23] Y. Gu, B. Wang, and B. Xu, "A FPN-based framework for vehicle detec-tion in aerial images," in Proc. 2nd Int. Conf. Video Image Process.,Hong Kong, Dec. 2018, pp. 60–64.

[24] W. Zhou "Layer-weakening feature fusion network for remote sensing detection," in Proc. 10th Int. Conf. Internet. Multimedia Comput. Service, Nanjing, China, Aug. 2018, pp. 1–6

[25] C.Li,, "Feature-attentioned object detection in remote sensing imagery," in Proc. 26th IEEE Int. Conf. Image Process., Taipei, Taiwan, Sep. 2019,pp. 3886–3890.

[26] Girshick, R Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

[27] Z. Deng, "Multi-scale object detection in remote sensing imagery with convolutional neuralnetworks," ISPRS J. Photogrammetry Remote Sens., vol. 145, pp. 3–22,Nov. 2018.

[28] Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Berlin, Germany, 11–14 March 2015; pp. 1440–1448.

[29] Ren, S : Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 1137–1149.

[30] A. S. Ladkat, , "Modified matched filter kernel for classification of hard exudate," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-6, doi: 10.1109/INVENTIVE.2016.7830123.

[31] Ajay S. Ladkat, Sunil "Deep Neural Network-Based Novel Mathematical Model for 3D Brain Tumor Segmentation", Computational Intelligence and Neuroscience, vol. 2022, Article ID 4271711, 8 pages, 2022. https://doi.org/10.1155/2022/4271711