An Improved Small Object Detection Method in Remote Sensing Images Based on YOLOv8

Wang Hao
Dept. of Mechanics and Mathematics
Belarusian State University
Minsk, Belarus
ahcenewang@gmail.com

Sergey Ablameyko
Dept. of Mechanics and Mathematics
Belarusian State University
Minsk, Belarus
ablameyko@bsu.by

Abstract. **Small object detection has long been a difficulty and research hotspot in computer vision. Driven by deep learning, small object detection has made major breakthroughs and has been successfully used in fields such as national defense security, intelligent transportation, and industrial automation. In our research, we conduct a comprehensive analysis and improvement of the YOLOv8-n algorithm for object detection, focusing on the SE Attention and detection heads of small object. Through detailed ablation studies to assess its contribution to model performance, each strategy is systematically evaluated individually and collectively. The results show that each strategy uniquely enhances the performance of the model, significantly improving mAP when the two strategies are integrated.**

Keywords: YOLOv8, Object detection, SE Attention

I. INTRODUCTION

Object detection is an important research direction in the field of computer vision and is the basis for other complex vision tasks. As the cornerstone of image understanding and computer vision, object detection is the basis for solving higher-level vision tasks such as segmentation, scene understanding, object tracking, image description, and event detection. Small object detection has long been a difficult point in object detection.

Small objects refer to object imaging sizes that are small. There are usually two ways to define small objects. One is the absolute size. In the COCO [1] dataset, objects smaller than 32 × 32 pixels are considered small objects; the other is relative size, according to the definition of the International Society of Optical Engineering, a small object is a object with an imaging area of less than 80 pixels in a 256 × 256 pixels image, that is, if the size of the object is less than 0.12% of the original image, it can be considered a small object. Examples of small objects are shown in Figure 1. bulls and dogs marked in the area in Figure 1(a) represent regular-sized objects, and cars marked in the area in Figure 1(b) represent small objects. Compared with regular-sized objects, small objects occupy fewer pixels in the image, have lower resolution, and have weaker feature expression capabilities.

(a) regular size objects (b) small objects

Fig. 1. Comparison of regular objects and small objects

In recent years, the rapid development of deep learning technology has injected fresh blood into small object detection, making it a research hotspot. However, compared to regular-sized objects, small objects usually lack sufficient appearance information, making it difficult to distinguish them from the background or similar objects. Driven by deep learning, although object detection algorithms have made major breakthroughs, the detection of small objects is still unsatisfactory. Small object detection is still full of challenges, for satellite remote sensing images, the objects in the image, such as cars and boats, may only have dozens or even a few pixels. Precisely detecting tiny objects in satellite remote sensing images will help government agencies curb drug and human trafficking, find illegal fishing vessels and enforce regulations prohibiting illegal transshipment of cargo.

In order to solve the problem of small object detection in remote sensing datasets, many researchers have improved the network models of the Faster RCNN and YOLO series. Zhang et al. [2] proposed a detection approach for aircraft based on Faster-RCNN, called MFRC. Their research used the K-means algorithm to cluster aircraft data in remote sensing images, improved anchor points based on the clustering results, reduced the pooling layer in the network from four layers to two layers, and used the Soft-NMS algorithm to optimize the aircraft bounding box. Zhou et al. [3] proposed an improved detection algorithm YOLO-SASE. Taking the super-resolution reconstructed image as input, combined with the SASE module, SPP module and multi-level receptive field structure, the number of detection output layers is adjusted by exploring feature weights to improve feature utilization efficiency. Gong et al. [4] proposed a novel concept, fusion factor, to control information that deep layers deliver to shallow layers, for adapting FPN to tiny object detection. Their results show that when configuring FPN with a proper fusion factor, the network is able to achieve significant performance gains over the baseline on tiny object detection datasets. Bosquet et al. [5] introduced STDNet and ConvNet as approaches for identifying tiny objects smaller than 16 ×16 pixels based on regional concepts. Ji et al. [6] introduced a small object detection algorithm based on YOLO v4 and Multi-scale Contextual information and Soft-CIOU loss function. Ren et al. [7] introduced an improved algorithm based on Faster R-CNN that sets appropriate anchor points to exploit high-resolution single high-level feature maps by designing a similar architecture employing top-down and skip connections, combined with context information and introduces a data enhancement method called "random rotation" to further improve the detection performance of small remote sensing objects. Hao et al. [8] they proposed a vehicle small object detection algorithm based on residual networks. Based on the original SSD (Single Shot MultiBox

Detector) algorithm, a residual network with stronger feature extraction capabilities and fewer parameters can be used to effectively improve the detection accuracy of small objects. Xin et al. [9] introduced DIOU-NMS and Alpha-IoU, based on the YOLOv5 network structure, the non-maximum inhibition in the original YOLOv5 is replaced by the non-maximum inhibition based on DIOU_Loss, and the original IoU system is replaced by the α-IoU system, significantly improved mAP and accuracy. Wang et al. [10] used Wise-IoU (WIoU) v3 as a bounding box regression loss, introduced BiFormer to optimize the backbone network, which improves the model's attention to critical information and designed a feature processing module named Focal FasterNet block (FFNB), effectively improved the performance of the model.

In terms of small object detection, most of the existing research is based on early mature models such as YOLO and Fast RCNN. Although these studies have slightly improved small object detection, deep learning is now developing rapidly, with new frameworks and models As they continue to be created and the hardware becomes more powerful, new and improved models will achieve better performance. As the latest model, YOLOv8 [11] has significant improvements in all aspects, including the ability to handle small objects. By conducting improvement research based on YOLOv8, we may further improve the model's detection accuracy for small objects. Therefore, in this paper, we made two improvements to the YOLOv8-n model for small object detection based on remote sensing data.

Improvements include the following points:

- Added a detection head for small objects to enhance detection capabilities for small objects.
- Added the SE Attention [12] module to the network improves the detection capability of the model.

II. The Proposed Methodology

Our model is based on YOLOv8, combined with the SE Attention module and the new small object detection head, which can improve the accuracy of model in detecting small objects. The following two parts are about the principle of the SE Attention module, how it is inserted into YOLOv8, and the structure and concept of the detection head for smaller objects.

The main principle of SE Attention as shown in Figure 2, SE module first performs a Squeeze operation on the feature map obtained by convolution to obtain channel-level global features, and then performs an Excitation operation on the global features to learn the relationship between each channel and obtain the weights of different channels, and finally multiply them by the original feature map to get the final features.

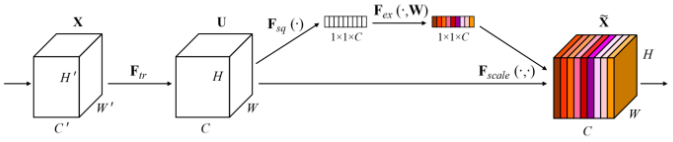

Fig. 2. A Squeeze-and-Excitation block.

There are three operations in SE module:

1) Squeeze operation:

- Assume that the input feature map is X and its size is $C \times H \times W$, where C is the number of channels, H and W are the height and width respectively.
- In the Squeeze operation, we perform global average pooling on the feature map to compress it into a feature vector. This can be achieved by averaging the feature maps of each channel.
- Mark the feature vector after the pooling operation as $Z \in R^C$, where Z_C represents the compression feature of the channel c.

2) Excitation operation:

- In the Excitation operation, we use a fully connected layer and nonlinear activation function to learn the weight of each channel to capture the relationship between channels. Assume that the parameters of the fully connected layer are $W_1 \in R^{C' \times C}$ and $W_2 \in R^{C \times C'}$, where C' is a smaller dimension.
- Input the feature vector C into a fully connected layer: $Y_1 = \sigma(W_1 Z)$, where σ represents the nonlinear activation function (ReLU).
- Input the output Y_1 of the fully connected layer to another fully connected layer: $Y_2 = \sigma(W_2 Y_1)$, where σ represents the nonlinear activation function (Sigmoid).
- The resulting output $Y_2 \in R^C$ represents the weight vector of each channel.

3) Scale operation:

- Apply the learned weight vector Y_2 to each channel on the input feature map X.
- For each channel X, multiply its corresponding feature map X_C and the weight Y_{2C} to obtain the weighted feature map: $X'_C = X_C \cdot Y_{2C}$.
- Recombine all weighted feature maps to obtain the final output feature map X'.

The process of the entire SE module can be expressed as: $X' = Scale(X) = X \cdot sigmoid(W_2 \operatorname{Re} LU(W_1 \cdot Pool(X)))$, where $Pool$ represents the global average pooling operation, $\operatorname{Re} LU$ represents the ReLU activation function, and $sigmoid$ represents the sigmoid activation function. This formula can be automatically back propagated for training, adjusting the values of W_1 and W_2 via gradient descent to optimize the performance of the model. By learning the weight of each channel, the SE module is able to adaptively adjust the feature map.

In essence, the SE module performs attention or gating operations in the channel dimension. This attention mechanism allows the model to pay more attention to the channel features with the largest amount of information and

suppress those unimportant channel features. In addition, the SE module is universal, which means it can be embedded into existing network architecture. In Figure 3, it shows how SE Attention is integrated into the C2f or Conv module.

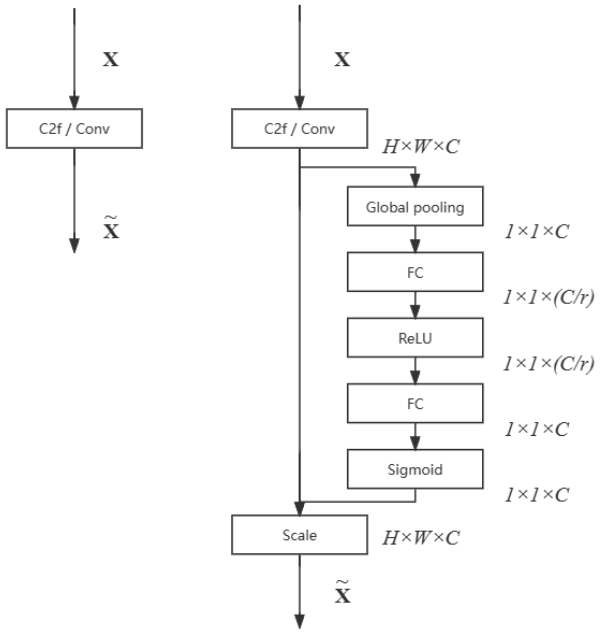

Fig. 3. The schema of the original module (left) and the SE module (right)

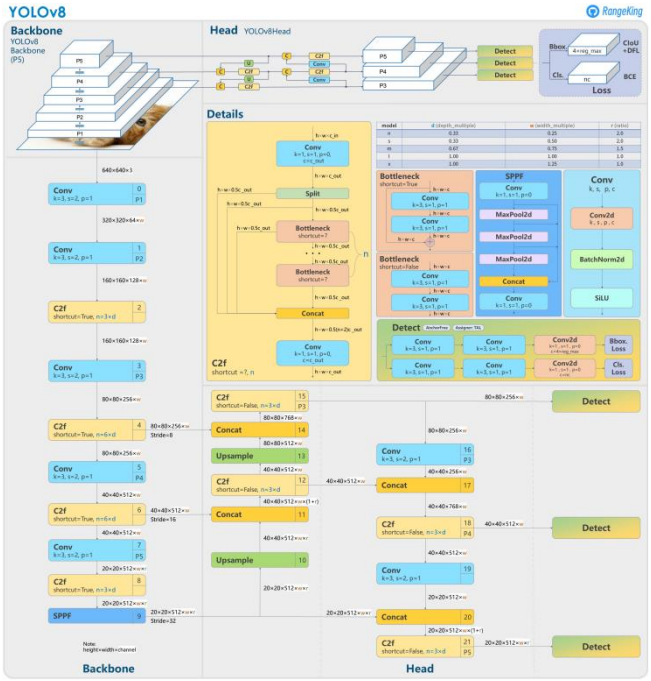

Fig. 4. Structure of YOLOv8

Figure 4 shows the complete structure [13] of YOLOv8. For the addition of small object detection heads, we only need to pay attention to its head structure. For the original YOLOv8, the downsampling multiple is large, and it is difficult to learn the learning information of small objects with deeper feature maps. Therefore, we add a small detection head, which can make YOLOv8 pay more attention to the characteristics of small objects and enhance the detection accuracy of small objects. The original three detection feature map sizes are 80 × 80, 40 × 40 and 20 × 20, which correspond to the three detection heads that can detect objects with sizes above 8 × 8, 16 × 16 and 32 × 32 pixels. Now we have added a 160 × 160 detection feature map, which is used to detect objects above 4 × 4 pixels. Smaller detection pixels can extract more features about small objects.

In the head part of YOLOv8, we have added an additional object detection head to achieve better detection results for small objects. Figure 5 shows how to add a small object detection head structure to the YOLOv8 head structure. The red box is the newly added structure.

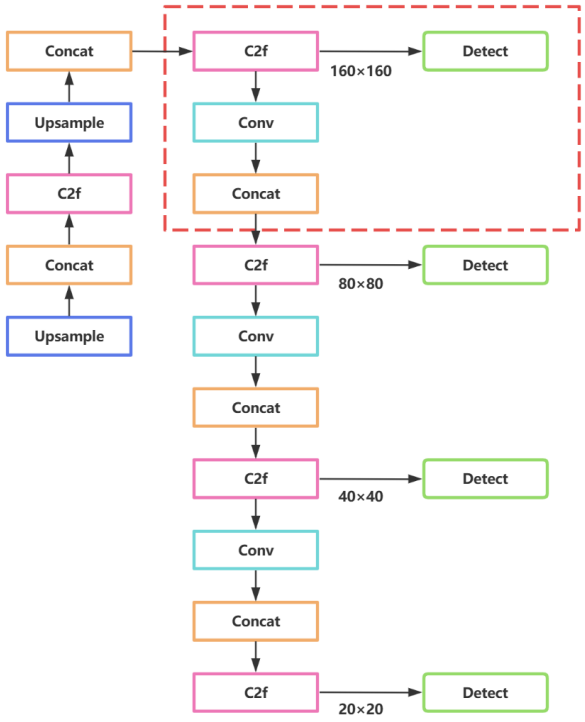

Fig. 5. The structure after adding new detection head.

III. TRAINING SET

In experiments we use DOTA-v2.0 [14] as our dataset, which is a large-scale dataset for object detection in aerial images. It can be used to develop and evaluate object detectors in aerial images. The images are collected from different sensors and platforms. 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. The instances in DOTA images are annotated by experts in aerial image interpretation by arbitrary (8 d.o.f.) quadrilateral. DOTA-v2.0 collects more Google Earth, GF-2 Satellite, and aerial images. There are 18 common categories (planes, helicopters, ships, cars, playgrounds, etc.), 11,268 images and 1,793,658 instances in DOTA-v2.0. The 11,268 images of DOTA are split into training, validation sets. To avoid the problem of overfitting, the proportion of training and validation set is smaller than the test set. Training contains 1,830 images and 268,627 instances. Validation contains 593 images and 81,048 instances.

In Figures 6 and 7 we can see the types of objects that can be detected in this dataset and the instances of each type of object.

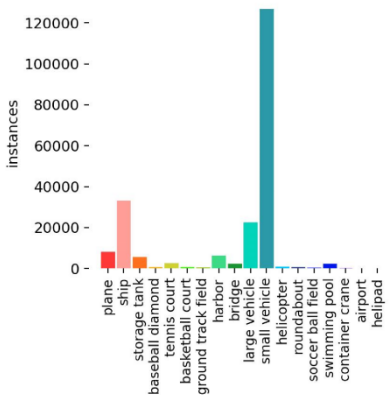

Fig. 6. Number of each category in DOTA-v2.0

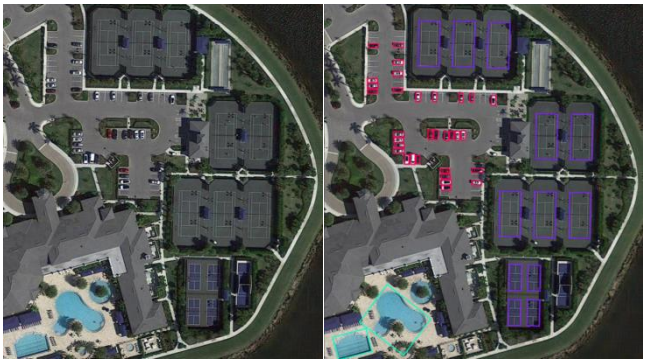

Fig. 7. Picture from training set

IV. Results

The experiments were done on the AutoDL platform with a NVIDIA GeForce RTX 3090 (24268MiB), each experiment was performed for 300 epochs. Experimental results are shown in the Table below, and the example of detection is shown in Figure 8 and 9.

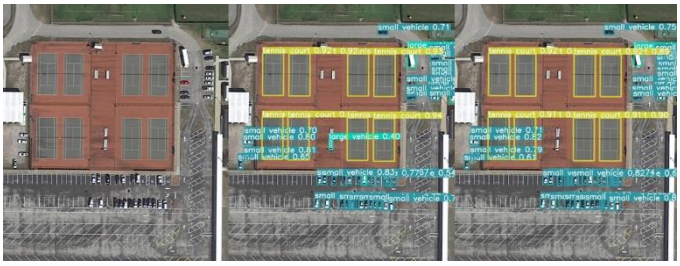

Fig. 8. Original image, result of YOLOv8n and our method

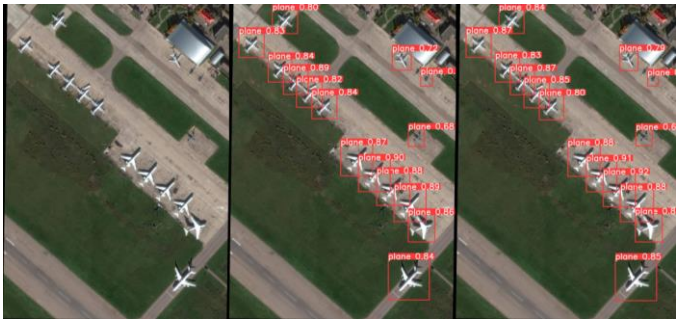

Fig. 9. Original image, result of YOLOv8n and our method

The experiments show that when the SE Attention module and a new detection head are added to YOLOv8n, better indicators are achieved in terms of accuracy.

TABLE I. Comparison results

Detector	*mAp50(%)*	*mAP50-95(%)*	*Recall(%)*
YOLOv8n	29.76	17.54	27.57
YOLOv8n+SE Attention	31.03	18.54	28.28
YOLOv8n+one more head	31.28	18.51	29.69
Our method	31.99	19.12	29.87

In experiments, our method is called YOLOv8_4DH+SE, the algorithm with the SE module added is YOLOv8_SE, and the algorithm with a small object detection head added is YOLOv8_4DH. Simply put, in the object detection model, the higher the mAP and Recall indicators mean the better the model performance. In Figure 10 are the results about mAP and Recall. We can see that under the following three indicators, YOLOv8 using our method has achieved good improvements.

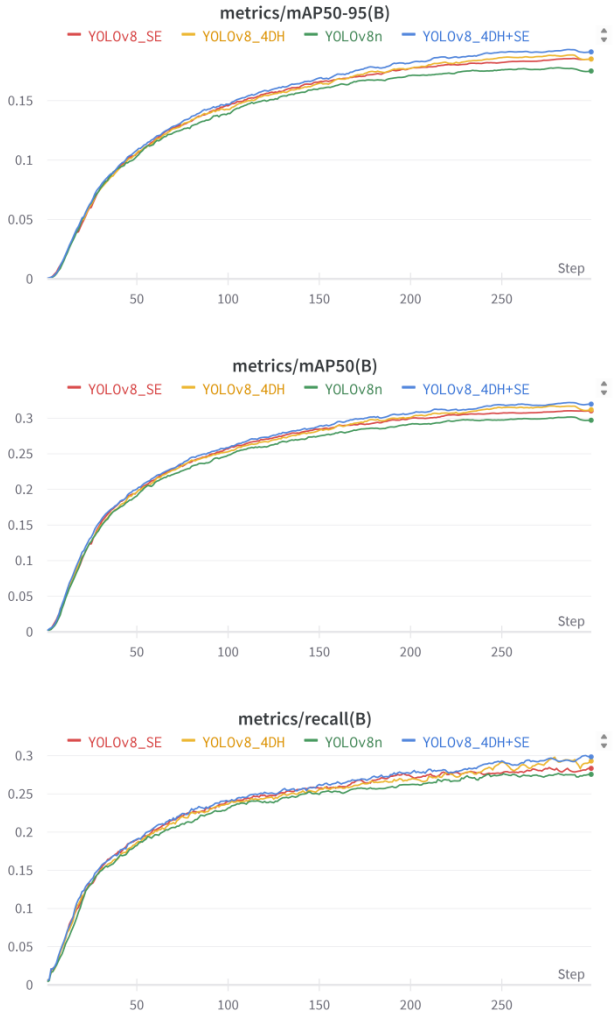

Fig. 10. Result of mAP50, mAP50-95 and Recall

V. Conclusion

This article proposes a method, which adds a SE Attention module and a detection head for small objects in YOLOv8. Our method improves the performance of YOLOv8 in small object detection on the remote sensing

image DOTA-v2.0 dataset, and improves the interference of small object detection in noisy data to a certain extent.

There are many classic algorithms in the field of deep learning. Due to limited research time, it is impossible to try and practice them all. The main shortcomings of current research include that the model is too complex, the training time is too long, and the categories in the dataset are unbalanced, which affects the final accuracy of the model.

More complex networks may not be conducive to high-quality feature representation of small objects, and we also need to avoid high computational costs and information loss. So, in the future, it is planned to work on a method for more lightweight or general in small object detection.

REFERENCES

[1] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," in European Conference on Computer Vision, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:14113767

[2] Y. Zhang, C. Song, and D. Zhang, "Small-scale aircraft detection in remote sensing images based on Faster-RCNN," Multimedia Tools and Applications, vol. 81, pp. 18091–18103, 2022.

[3] X. Zhou, L. Jiang, C. Hu, S. Lei, T. Zhang, and X. Mou, " YOLO-SASE: An Improved YOLO Algorithm for the Small Targets Detection in Complex Backgrounds," Sensors (Basel, Switzerland), vol. 22, 2022, [Online]. Available: https://api.semanticscholar.org/CorpusID:249907623

[4] Y. Gong, X. Yu, Y. Ding, X. Peng, J. Zhao, and Z. Han, "Effective Fusion Factor in FPN for Tiny Object Detection," 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1159–1167, 2020.

[5] B. Bosquet, M. Mucientes, and V. M. Brea, "STDnet: Exploiting high resolution feature maps for small object detection," Eng. Appl. Artif. Intell., vol. 91, p. 103615, 2020.

[6] S.-J. Ji, Q. Ling, and F. Han, "An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information," Comput. Electr. Eng., vol. 105, p. 108490, 2023.

[7] Y. Ren, C. Zhu, and S. Xiao, "Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN," Applied Sciences, vol. 8, p. 813, 2018.

[8] Z. Hao, C. L. Wang, P. Chen, D.-H. Fan, and Y. Shang, "A detection method for small target in remote sensing image based on improved SSD," in International Conference on Electronic Information Engineering and Computer Science, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258280251

[9] L. Xin, X. Xie, and J. Lv, "Research on Remote Sensing Image Target Detection Algorithm Based on YOLOv5," 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), vol. 5, pp. 1497–1501, 2022.

[10] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, " UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios," Sensors (Basel, Switzerland), vol. 23, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:260939243

[11] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8." 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[12] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, " Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141, 2017.

[13] Brief summary of YOLOv8 model structure. [Online]. Available: https://github.com/RangeKing

[14] J. Ding et al., "Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1 – 1, 2021, doi: 10.1109/TPAMI.2021.3117983.