Defense-related Object Detection in Aerial Images

Luciano Severo Bittencourt [1]and Paulo André Lima de Castro [1]
[1] Instituto Tecnológico de Aeronáutica, São José dos Campos - SP, Brasil

***Abstract*—Object detection in aerial images (ODAI) is an important task in computer vision and has applications in several areas, such as defense, environmental monitoring, land use surveillance and even track-ing of maritime routes. Recently, researchers directed their efforts to the ODAI, which requires detectors capable of dealing with arbitrary orientations, large variations in aspect ratios, densely clustered objects, multiple classes and instances per image. Our work aimed to detect planes, ships, harbors, storage tanks and helicopters that are relevant to defense systems. Such defense-related objects may present special challenges for detection and a reliable detector may be very useful as information source for defense systems. We have used publicly available aerial images and implemented some detectors based on Rotation-equivariant Detector - ReDet, which presented a very good performance for a broad class of objects. We tested such detectors using only defense-related objects. Our tests included dataset with and without data augmentation. The results achieved are consistent with the results published in some previous competitions.**

***Index Terms*—Object detection in aerial images, Earth vision, Rotation-equivariant detector.**

I. Introduction

Object detection in aerial images (ODAI) [1] is applicable in areas of defense, environmental monitoring, land use control and tracking of maritime routes. ODAI gained greater attention from researchers after convolutional neural networks achieved high levels of accuracy with natural images [2] and after sufficiently large datasets [1], [3] were made available for the training of detectors.

Object detectors with the highest accuracies, such as the Rotation-equivariant Detector - ReDet [4] and detectors that implement or Regions of Interest Transformer [5], work in two stages.

In the first stage, an image scan is performed to select a large number of regions that may contain objects of the classes to be detected, considering pre-established anchors in terms of scales and aspect ratios. The detector stores regions with the intersection over union (IoU) metric greater than an established threshold. The IoU metric is calculated according to the equation below:

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (1)$$

B_p means the bounding box of a prediction and B_{gt} means the bounding box of a ground truth [5]. The ground truth is available in the The large-scale Dataset for Object deTection in Aerial images (DOTA) training and validation data annotations.

In the second stage, the detectors analyse each proposed region and perform the classification of existing objects and their locations in the image [6].

In the detection of objects in natural images the location is performed with horizontal detection contours, called horizontal bounded box - HBB, while the task of detecting objects in aerial images is performed with HBB and also with oriented bounded box - OBB. The objective of OBB is to avoid ambiguities in the tasks of classifying and locating objects [5]. The detection performed with OBB is the most indicated when there is agglomeration of instances in the same image, which is common in the case of some classes, such as vehicles, trucks, ships, planes and helicopters.

The automatization of the defense-related object detection process is essential due to the territorial dimensions of Brazil, the fifth largest country in the world in territorial area [7], with more than 15,700 km from land borders [8]. Even a large number of experts locating and classifying objects in aerial images obtained by planes, drones and satellites constitute a time-consuming and inefficient solution.

The application of a high accuracy detector in defense-related classes is the main contribution of this work, as it can accelerate decision-making in security and defense operations.

Automating the detection of defense-related objects can significantly decrease the processing time of aerial images. For example, to locate and classify suspicious objects, such as river harbors used to support the trafficking of narcotics or minerals in the Amazon. In other words, automatization improves the command and control cycle performed during military operations [9].

To illustrate the results that can be achieved with a high-precision object detector, with defense-related classes, we performed two trainings with ReDet, with and without data augmentation, using the DOTA-v1.5 dataset.

II. Background

A. Dataset DOTA

The large-scale Dataset for Object deTection in Aerial images, DOTA, was produced by researchers at Huazhong University of Science and Technology. In its first version, it contained 2,806 images, 15 classes and 188,282 instances [3]. It was built to boost research with object detection, in the area of remote sensing, also called Earth Vision.

The detection of objects in natural images had advanced a lot with the use of large datasets, such as MSCOCO [10], ImageNet [11] e Places [12]. However, these datasets are not suitable for training object detectors in aerial images, since they do not contain OBB annotations and they don't have instances with wide variations in scale, orientation and aspect ratios.

The construction of the DOTA dataset considered the following requirements:

- A large number of images;
- Many instances per categories;
- Properly oriented object annotation; and
- Many different classes of objects.

At the time of publication, the DOTA-v1.0 was the largest annotated object dataset with a wide variety of categories.

The DOTA-v1.5 dataset is an enhancement of DOTA-v1.0. The DOTA-v1.5 uses the same images as DOTA-v1.0. However, it gained a new category, container cranes. The development team also added the extremely small instance annotations (with less than 10 pixels). It contains 403,318 instances in total [13].

B. Rotation-equivariant Detector

The Rotation-equivariant Detector - ReDet [4] is an oriented object detector that reached the state of the art, in February 2021, on the test data of the DOTA-v1.5 dataset.

ReDet incorporates rotation-equivariant networks into the backbone instead of traditional convolutional neural networks to extract the features because the regular CNNs are not equivariant to the rotation. That is, compared with convolutional neural networks, which share translation weights, rotation-equivariant networks share translation and rotation weights. ReDet also uses ResNet with Feature Pyramid Networks - FPN [14] as the backbone to implement a rotating equivariant backbone network, named Rotation-equivariant ResNet (ReResNet) so to extract the features of the rotation equivariant, which can accurately predict the orientation and significantly reduce the model size.

In addition, the ReDet has a novel Rotation-invariant RoI Align (RiRoI Align), which produces RoI-wise rotation-invariant features from rotation equivariant feature maps.

Compared with ordinary backbones, the rotation-equivariant backbone has the following advantages:

- Higher degree of weight sharing;
- Enriched orientation information; and
- Smaller model size.

The ReDet achieved mean Average Precision (mAP) of 66.86 on the test data of the dataset DOTA-v1.5, when trained without data augmentation (single-scale), and reached mAP of 76.80, when trained with data augmentation (multi-scale).

C. Related Works

In the research to prepare this article, we did not find similar works focusing on the detection of defense-related objects using DOTA data. However, there are publications that concentrate detections on the class vehicles and planes [15], [16], or only on the class planes [17], using the UCAS-AOD dataset [18].

There are publications on the detection of objects of the class ships [19], [20], [21], [22], annotated with OBB in the HRSC2016 dataset [23].

III. Defense-related Objects

When evaluating the DOTA-v1.5 dataset, the defense-related classes considered most important for the specific context of Brazil were: harbors, ships, storage tanks, planes and helicopters. These classes are directly related to the sovereignty of airspace, the logistics of illicit activities in border regions and the country's critical infrastructure.

A. Harbors

This object category is a defense-related class because small harbors, used for small ships, are also included in the dataset. Small harbors can support vessels used for drug trafficking and support illegal mining activities.

The detection of this type of object in aerial images can contribute to the planning of joint operations carried out with the participation of the Armed Forces, Federal Police and government agencies [24].

B. Ships

Similar to harbors, ships support drug trafficking and illegal mining. In addition, its detection is very important for the defense of the integrity of a country, as it can help the tracking of sea and river routes.

An efficient algorithm could have located the origin of the oil spill that occurred in 2019 [25], as the selection of vessels sailing with the transponder off can direct the inspections to be carried out by the Navy and mitigate the risk of similar occurrences on the Brazilian coast.

C. Storage Tanks

Storage tanks are fundamental elements in the logistics of a country. Typically, its location is associated with a refinery, a maritime oil terminal or the ends of pipelines.

These types of structures are part of a country's critical infrastructure and receive special attention from a nation's security forces [26].

D. Planes and Helicopters

As well as ships, planes and helicopters transport drugs and precious metals. The rapid detection of these objects in aerial images is essential for successful ground inspections and interceptions [27].

IV. Training Process

The ReDet detector training process requires a computer equipped with at least one graphics processing unit (GPU), with a minimum of 12 GB of memory, compatible with the Compute Unified Device Architecture (CUDA).

The memory requirement stems from using the GPU with the CUDA and the Pytorch library to process a large volume of data. Attempts to train computers equipped with lower-capacity GPUs resulted in repeated out-of-memory (OOM) failures, even with adjustments to data batch sizes.

In addition, the computer must have at least 260 GB available for storage, which is the space occupied by the dataset prepared for training without data augmentation and with data augmentation.

After setting up the environment [28], successfully compiling the CUDA samples and downloading the DOTA 1.5 dataset, it is possible to test the operation with the execution of the constant code in demo_large_image.py, which uses the model trained by the project developers.

If the tests were successful, the next step is to prepare the images for training. In this step, the images are divided into standardized 1024 x 1024 images, with an overlap of 200 pixels, in the case of training without data augmentation, and with an overlap of 512 pixels and multiple scales in the factors of 0.5 and 1.5, in the case of training with data augmentation. At the end of the data preparation process, the directory structure must be identical to the one mentioned on the project website [28].

The training execution is performed according to the combination of pre-trained residual network - ResNet selected as the detector backbone, according to the network that addresses the scale variations of the instances, called neck, according to the network that selects image regions that are likely to contain instances, called the head, and according to the combination of the number of GPUs and number of the images per GPU.

The combination of the number of GPUs and the number of images per GPU establishes the batch size, used as a basis for setting the learning rate [29].

After training is complete, it is necessary to verify that the model file was created correctly, and then proceed to inference on the test set.

At the end of the inference, the model generates a file with contour detections for each class. As the DOTA test data does not contain ground truth annotations, the operator will be able to evaluate the new trained model by submitting the files with the detections for each class to the evaluation server maintained by the developers [13].

The evaluation results contain the average precision achieved by the model in each class and contain the mean average precision, called mAP. In calculating the average precision, the division of the intersection between the prediction and the ground truth by the union of the prediction and the ground truth (IoU) and the correct classification in the prediction is considered.

V. Results Obtained for the Defense-related Classes

The results obtained in the complete training with single-scale (without data augmentation), using a node with 2 x CPU Intel Xeon Ivy Bridge 2.4GHZ, 64GB DDR3 RAM, and 2 x Nvidia K40, for the defense-related classes are in Table I.

The difference column confirms that the results obtained in our training were consistent with the best published results [4] up to March 2021.

However, data augmentation is fundamental to obtaining better results.

Therefore, new training attempts with data augmentation were performed, also using a node with 2 x CPU Intel Xeon Ivy Bridge 2.4GHZ, 64GB DDR3 RAM, and 2 x Nvidia K40.

The results with data augmentation are in Table II. Data augmentation significantly improved Average Precision (AP).

The table III details the percentage gain for each class, with emphasis on the storage-tank class, where the data augmentation provided a 20,28 % gain.

In addition to the numerical results, to emphasize the importance of data augmentation, we present the results in test images of the DOTA-v1.5 dataset.

A. Results for Harbors Class

The Fig. 1 contains the first example of the results for the harbors class.

The prediction, with the single-scale model, did not detect all harbors. In addition, it also classified a harbor as a plane.

On the other hand, with the multi-scale model, all harbors were detected, despite a positive-negative detection for harbor class and one for plane class.

The Fig. 2 contains the second example of the results for the harbors class.

In the second example, the single-scale model detected four harbors as being just one, two other harbors as being one, and it did not detect the harbor next to the largest boat in the image. With the multi-scale model, detection failures did not occur.

B. Results for Ships Class

In the first example of the ships class, in Fig. 3, the single-scale model failed to detect fifteen instances, including small ships. With the multi-scale model, the detector failed in nine instances.

In the second example of the ship's class, in Fig. 4, the single-scale model failed to identify a parking lot and the streets around it as a ship. The multi-scale detector did not make the same mistake.

C. Results for Storage Tanks Class

In Fig. 5, the single-scale model incorrectly located three buildings as small-vehicles.

TABLE I
Single-scale results

Class	AP Train (ours)	AP ReDet	Difference (%)
Harbors	$74,0788$	$73,3601$	$0,97965$
Ships	$88,6530$	$80,9204$	$9,5558$
Storage-tanks	$64,5253$	$68,6393$	$-5,9937$
Planes	$80,6705$	$79,2033$	$1,8524$
Helicopters	$67,9064$	$63,3306$	$7,2253$

TABLE II
Multi-scale results

Class	AP Train (ours)	AP ReDet	Difference (%)
Harbors	$77,9017$	$78,3211$	$-0,5354$
Ships	$89,9945$	$90,0025$	$-0,0089$
Storage-tanks	$77,6117$	$75,3295$	$3,0297$
Planes	$88,0037$	$88,5074$	$-0,5691$
Helicopters	$78,6568$	$76,0987$	$3,3616$

TABLE III
GAINS WITH MULTI-SCALE

Class	Gain (%)
Harbors	5, 1606
Ships	1, 5132
Storage-tanks	20, 2810
Planes	9, 0903
Helicopters	15, 8312

With the multi-scale model, the detector did not repeat the error. Both models correctly identified the storage tanks.

In Fig. 6, the single-scale model failed to detect one of the tanks and classified smaller storage tanks as small-vehicle.

With the multi-scale model, the detector did not repeat the errors but also did not detect smaller storage tanks.

D. Results for Planes Class

The Fig. 7 contains the first example of the results for the planes class.

The single-scale model did not detect any of the seven instances in the image.

On the other hand, in Fig. 7, the multi-scale model found four instances of the plane class and one false positive.

In the second example of the planes class, the single-scale trained model detected only one of the five instances in the image.

With the multi-scale model, the detector correctly located the five instances.

E. Results for Helicopters Class

In Fig. 9, the single-scale model detected only two of the seven instances in the image. The model trained with data augmentation found seven instances.

Both models doubly detected one of the helicopters as also being an airplane.

In Fig. 10, the detector failed to locate five of the eighteen helicopters in the image. Furthermore, with the single-scale model, eleven instances were incorrectly classified as ships.

With the multi-scale model, the detector located the eighteen instances, with only one of them doubly classified as an airplane.

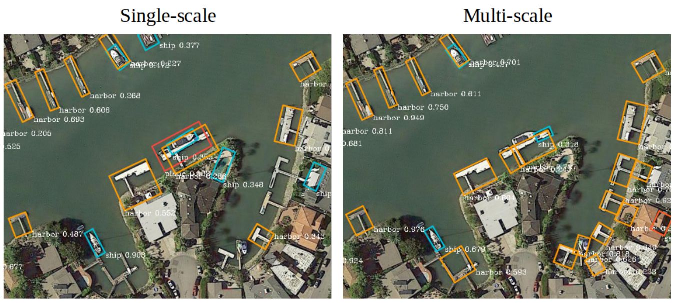
Single-scale
Multi-scale

Fig. 1. First example of the results for the harbors class.

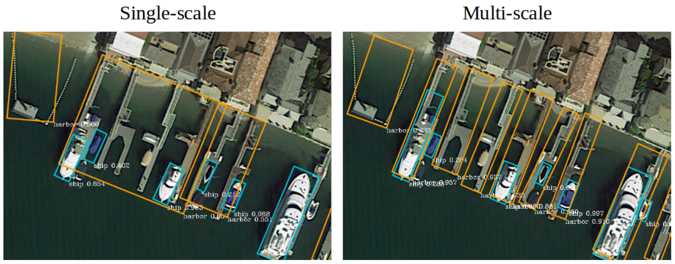
Single-scale
Multi-scale

Fig. 2. Second example of the results for the harbors class.

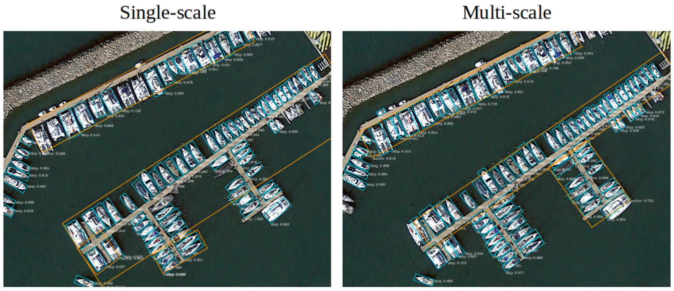
Single-scale
Multi-scale

Fig. 3. First example of the results for the ships class.

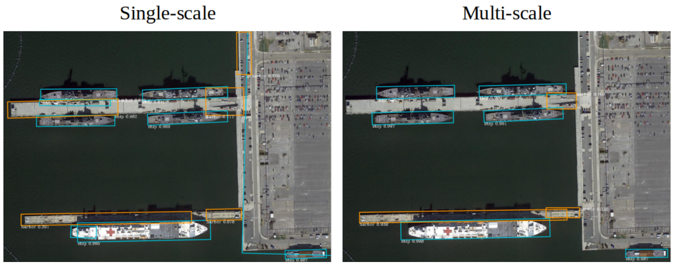
Single-scale
Multi-scale

Fig. 4. Second example of the results for the ships class.

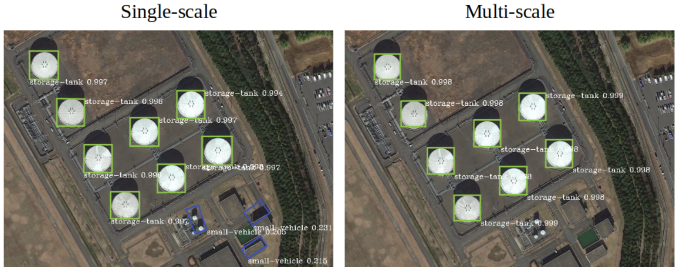
Single-scale
Multi-scale

Fig. 5. First example of the results for the storage tank class.

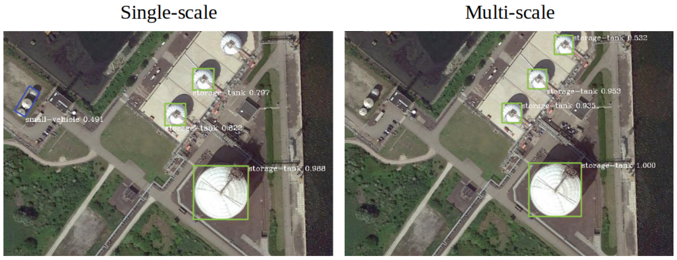
Single-scale
Multi-scale

Fig. 6. Second example of the results for the storage tank class.

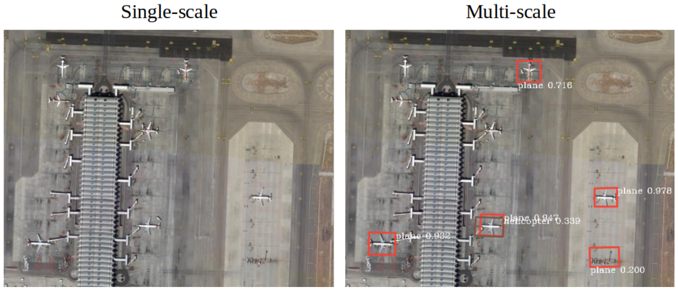

Fig. 7. First example of the results for the planes class.

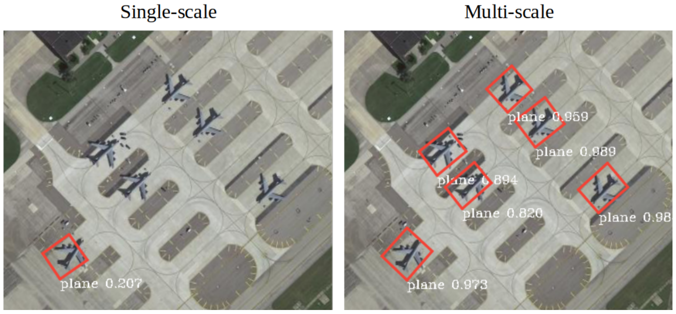

Fig. 8. Second example of the results for the planes class.

VI. Conclusion

The work presented average precision similar to the best published results [4] up to March 2021. In addition, the results presented in this work emphasize the importance of data augmentation in the training process of object detectors in aerial images. In the specific case of defense-related classes, the difference was very significant, especially when verifying the results in the images.

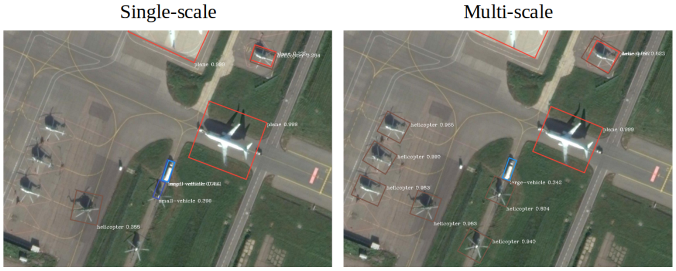

Fig. 9. First example of the results for the helicopters class.

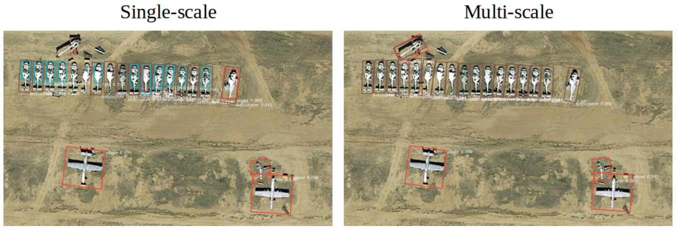

Fig. 10. Second example of the results for the helicopters class.

The average precision (AP) values achieved are high. However, as in the case of harbors and helicopters, with 77.90 and 78.65 AP, respectively, algorithms and methodologies can still be researched to achieve higher values. In future works, it is possible to evaluate the impact of fine-tuning the hyperparameters used in ReDet in the eventual gains obtained in the defense interest classes. To guide fine-tuning, we are developing a methodology based on the design of experiments, similarly to a case study carried out with random forest [30].

References

[1] J. Ding, N. Xue, G.-S. Xia, X. Bai, W. Yang, M. Yang, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[3] G.-S. Xia, X. Bai, L. Zhang, S. Belongie, J. Luo, M. Datcu, and M. Pelillo, "DOTA: A Large-scale Dataset for Object Detection in Aerial Images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 06 2018.

[4] J. Han, J. Ding, N. Xue, and G.-S. Xia, "Redet: A rotation-equivariant detector for aerial object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2786–2795.

[5] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning Roi Transformer for oriented object detection in aerial images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2849–2858.

[6] 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.

[7] Federal Ministry for Economic Cooperation and Development, "German development cooperation with Brazil. March 2021," Acessed: Sep. 12, 2022. [Online]. Available: https://www.bmz.de/en/countries/brazil

[8] Encyclopedia Britannica, "Brazil. september 2022," Acessed: Sep. 12, 2022. [Online]. Available: https://www.britannica.com/place/Brazil

[9] J. R. Boyd, "Patterns of conflict," December 1986, accessed: Aug. 10, 2022. [Online]. Available: https://www.colonelboyd.com/boydswork.

[10] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[12] B. Zhou, À. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning Deep Features for Scene Recognition using Places Database," in *NIPS*, 2014.

[13] DOAI 2019, "Challenge-2019 on Object Detection in Aerial Images. June 2019," Acessed: Apr. 12, 2022. [Online]. Available: https://captain-whu.github.io/DOAI2019/dataset.html

[14] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 07 2017, pp. 936–944.

[15] H. Zhang and J. Liu, "Direction Estimation of Aerial Image Object based on Neural Network," *Remote Sensing*, vol. 14, p. 3523, 07 2022.

[16] F. Liu, W. Zhao, G. Zhou, L. Zhao, and H. Wei, "SR-Net: Saliency region representation network for vehicle detection in remote sensing images," *Remote Sensing*, vol. 14, pp. 1313–1333, 03 2022.

[17] Y. Wang, H. Wu, L. Shuai, C. Peng, and Z. Yang, "Detection of plane in remote sensing images using super-resolution," *PLoS ONE*, pp. 1–19, 2022.

[18] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, "Orientation robust object detection in aerial images using deep convolutional neural network," *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 3735–3739, 2015.

[19] S. He, H. Zou, R. Li, X. Cao, F. Cheng, and J. Wei, “Teacher-Student Network for low-quality remote sensing ship detection,” in *2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE)*, 2021, pp. 283–287.

[20] X. Tan, T. Tian, and H. Li, “Inshore ship detection based on improved faster R-CNN,” in *MIPPR 2019: Automatic Target Recognition and Navigation*, J. Liu, H. Hong, and X. Hua, Eds., vol. 11429, International Society for Optics and Photonics. SPIE, 2020, pp. 1–9. [Online]. Available: https://doi.org/10.1117/12.2536638

[21] D. Zhang, C. Wang, and Q. Fu, “DD-Net: A dual detector network for multilevel object detection in remote-sensing images,” *Journal of Sensors*, vol. 2022, p. 1–12, jul 2022.

[22] Y. Wu, W. Zhao, R. Zhang, and F. Jiang, “AMR-Net: Arbitrary-oriented ship detection using attention module, multi-scale feature fusion and rotation pseudo-label,” *IEEE Access*, vol. 9, pp. 68 208–68 222, 2021.

[23] Z. Liu, H. Wang, L. Weng, and Y. Yang, “Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 1074–1078, 2016.

[24] Small Wars Journal, “Illicit confluences: The intersection of cocaine and illicit timber in the amazon. October 2016,” Acessed: Apr. 30, 2022. [Online]. Available: https://smallwarsjournal.com/jrnl/art/illicit-confluences-intersection-cocaine-and-illicit-timber-amazon

[25] BBC News, “Brazil oil spill: Where has it come from? November 2019,” Acessed: Apr. 30, 2022. [Online]. Available: https://www.bbc.com/news/world-latin-america-50223106

[26] Cybersecurity and Infrastructure Security Agency, “Critical infrastructure sectors,” Acessed: Apr. 30, 2022. [Online]. Available: https://www.cisa.gov/chemical-sector

[27] Insight crime, “Small aircraft feed illegal mining operations in Brazil’s Amazon. October 2021,” Acessed: Apr. 30, 2022. [Online]. Available: https://insightcrime.org/news/small-aircraft-feed-illegal-mining-operations-brazil-amazon/

[28] ReDet, “A Rotation-equivariant Detector for Aerial Object Detection,” Acessed: Apr. 18, 2022. [Online]. Available: https://github.com/csuhan/ReDet

[29] F. Zhang, X. Wang, S. Zhou, and Y. W. 0002, “DARDet: A Dense Anchor-Free Rotated Object Detector in Aerial Images,” *IEEE Geosci. Remote Sensing Lett.*, vol. 19, pp. 1–5, 2022.

[30] G. A. Lujan-Moreno, P. R. Howard, O. G. Rojas, and D. C. Montgomery, “Design of Experiments and Response Surface Methodology to Tune Machine Learning Hyperparameters, with a Random Forest Case-Study,” *Expert Systems with Applications*, vol. 109, no. C, p. 195–205, nov 2018.