Article

SAR Ship Detection Based on Improved Libra RetinaNet

Haomiao Liu[1,&], Haizhou Xu[1,&], Lei Zhang[3], Weigang Lu[4], Fei Yang[1, 2*] and Jingchang Pan[1]

[1] School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264200, China; 201800800522@mail.sdu.edu.cn (H.L.); 201800800552@mail.sdu.edu.cn (H.X.); pjc@sdu.edu.cn (J.P.)
[2] School of Computer Science and Technology, Shandong University, Qingdao 266000, China;
[3] Diagnostic Radiology and Nuclear Medicine, University of Maryland, Baltimore, MD 21201, USA; cszhanglei@gmail.com
[4] Department of Educational Technology, Ocean University of China, Qingdao 266000, China; luweigang@ouc.edu.cn
* Correspondence: feiyang@sdu.edu.cn
& These authors contributed equally to this work and should be considered co-first authors

Abstract: Maritime ship monitoring plays an important role in maritime transportation. Fast and accurate detection of maritime ship is the key to maritime ship monitoring. The main sources of marine ship images are optical images and synthetic aperture radar (SAR) images. Different from natural images, SAR images are independent to daylight and weather conditions. Traditional ship detection methods of SAR images mainly depend on the statistical distribution of sea clutter, which leads to poor robustness. As a deep learning detector, RetinaNet can break this obstacle, and the problem of imbalance on feature level and objective level can be further solved by combining with Libra R-CNN algorithm. In this paper, we modify the feature fusion part of Libra RetinaNet by adding a bottom-up path augmentation structure to better preserve the low-level feature information, and we expand the dataset through style transfer. We evaluate our method on the publicly available SAR dataset of ship detection with complex backgrounds. The experimental results show that the improved Libra RetinaNet can effectively detect multi-scale ships through expansion of the dataset, with an average accuracy of 97.38%.

Keywords: synthetic aperture radar; deep learning; data augmentation; object detection; ship detection

1. Introduction

With the development of maritime transportation, the incidence of maritime violations such as environmentally damaging ship accidents, piracy, illegal fishing and illegal cargo transportation has also increased. Therefore, more intensive higher demands are put forward for maritime ship detection.

Synthetic aperture radar (SAR)

images are unaffected by daylight and weather conditions, and the resolution of SAR images remains constant when it is far away from the observation objects, as shown in Fig. 1. Therefore, SAR images are widely used in ship detection to ensure maritime transportation safety. Despite the springing up of SAR ship detection methods, extracting and identifying the ship objects is still a challenge work due to the influence of background noise and the variety of objects in the sea.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

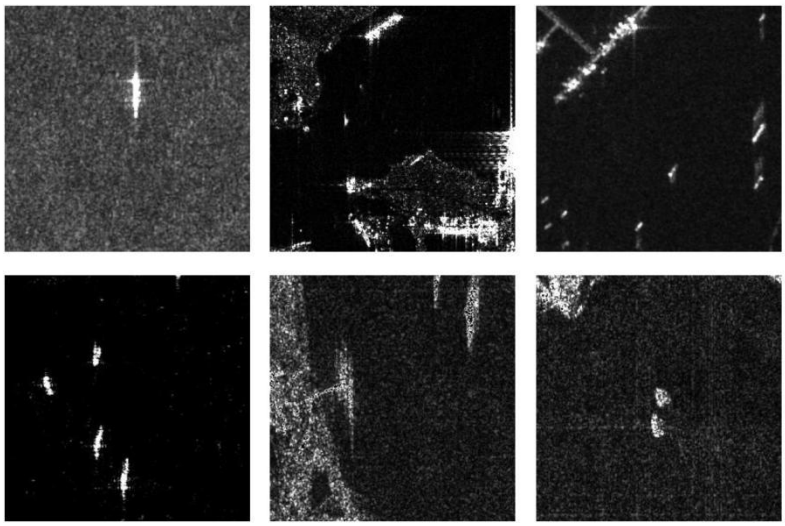


Figure 1. Representative images in the SAR ship dataset. They are different in resolution, incidence angle, polarization, and backgrounds.

Two traditional methods of ship detection are statistical models and features extraction, which are highly dependent on statistics and handcrafted features [2]. However, due to the different surroundings of ships and the various orientations of ships, the application of statistical models and features extraction methods is limited.

Deep learning has a strong ability to automatically learn layer-wise representations, and it has been widely used in SAR ship detection. Object detectors based on deep learning are mainly divided into one-stage detectors and two-stage detectors. One-stage detectors mainly include SSD [3], RetinaNet [4], YOLO series [5–8], and the two-stage detectors are represented by R-CNN series [9–12]. Inspired by the aforementioned works, we propose a practical state-of-the-art ship detection method in this paper. We present the following technical contributions in this work:

1. We expand the SAR ship dataset through transfer learning. We use the CycleGAN [13] to transfer the satellite image collection into SAR image collection. By enriching the dataset, we can achieve higher detection accuracy on the SAR ship dataset and improve the performance of detection models.
2. We improve the feature fusion scheme of Libra RetinaNet [14]. The improved Libra RetinaNet reduces low-level features loss in the process of feature transfer, and thus can be more efficiently used for tiny object detection in SAR ship images under complex backgrounds.

2. Related Work

2.1. SAR datasets of ship detection

In recent years, there have been many publicly available natural image datasets in the field of object detection. Due to the large difference between SAR images and natural images, the application of deep learning models in SAR images is hindered. Therefore, the requirement of labeled SAR ship dataset is proposed.

In 2018, Yuanyuan Wang et al. implemented ship detection method based on SSD and transfer learning and evaluated the method on the ship dataset of Sentinel-1 SAR images [2]. In 2019, Yuanyuan Wang et al. built ship slices in multi-resolution SAR images based on Gaofen-3 SAR images [15]. Also, Yuanyuan Wang et al. built a SAR image ship detection dataset under complex backgrounds based on 102 Gaofen-3 SAR images and 108 Sentinel-1 SAR images [1].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

2.2. Object Detection Algorithms Based on Deep Learning

Object detectors based on deep learning have achieved good performance in terms of detection accuracy and speed, such as R-CNN series [9–12], YOLO series [5–8], SSD [3], and RetinaNet [4]. These deep learning detectors can be categorized into two-stage detectors and one-stage detectors.

Two-stage detectors contain two stages. The first stage uses external modules to generate candidate proposals, and the second stage is to classify the proposals. Due to the consideration of hard samples, the two-stage detectors generally have higher detection accuracy comparing with one-stage detectors. The typical two-stage algorithm is R-CNN, which adopts the region proposal scheme instead of the sliding-window paradigm. Then Region Proposal Network (RPN) was proposed in the Faster R-CNN framework integrates proposal generation with the second stage classifier into a single convolution network to achieve the end-to-end object detection. And this framework has been improved by other researchers, e.g., [12].

One-stage detectors have been tuned for speed but their accuracy trails that of two-stage detectors. Recently, a lot of research works have focused on improving the detection accuracy of the one-stage detectors. Single shot multibox detector (SSD) as a typical one-stage detector proposes the pyramid feature hierarchy to make predictions. Then a number of studies on YOLO series also obtain higher speed and accuracy. In [4], the authors propose the focal loss to improve the detection accuracy of one-stage detectors.

2.3. Ship detection algorithms on SAR images

SAR images are all-weather, full-time, and wide coverage. These images are increasingly used for ship detection to ensure the safety of marine transportation. Currently there are three typical ship detection methods which are statistical models, features extraction, and deep learning [2] respectively. However, the statistical model and features extraction are sensitive to ship's location and environment. The detection algorithms based on deep learning are more robust compared with the other two and thus has been widely used in SAR datasets of ship detection in recent years.

In 2017, Kang, M et al. improved Faster R-CNN based on CFAR and applied it to the ship detection of SAR images [16]. Based on CNN, Liu, Y et al., based on CNN, implemented SAR image ship detection using land and sea segmentation in 2017 [17]. In 2019, Wang Y et al. achieved ship detection on Gaofen-3 SAR images based on RetinaNet [15].

3. Materials and Methods

3.1. Data Augmentation

Supervised object detection algorithms are data-driven tasks, which means that richer datasets tend to bring better results. Data augmentation is the way to enrich datasets.Generally,data augmentation can be achieved by performing various transformations on the original dataset, such as contrast transformation, brightness transformation, seasonal transformation, etc. In this paper, since SAR images are not affected by light and weather conditions, it would not be significant if SAR ship dataset is enriched only applying contrast transformation, brightness transformation or seasonal transformation. We observed that the satellite ship dataset publicly available on Kaggle and the SAR ship dataset have strong comparability in terms of the ship size and view for capturing the images. And the main contents of the images in these two datasets are both ships in the sea. However, compared with SAR ship dataset, satellite ship dataset contains a rich variety of complex scenes. And it is inappropriate to implement the ship detection task of the SAR images on the satellite ship dataset directly because of their differences in the imaging mode. In this paper, we consider transforming satellite ship images to SAR ship images through transfer learning. CycleGAN [13] is then adopted to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

transfer the style of the satellite ship images, so as to expand the training set of SAR ship detection.

CycleGAN is an approach to learn to translate an image from a source domain X to a target domain Y by capturing correspondences between high-level appearance structures. The goal is to learn the mapping G: X→Y between two image collections. In this paper, we regard the satellite ship image collection as X, and regard the SAR ship image collection as Y. The model also contains other mapping function F: Y→X and two associated adversarial discriminators D_Y and D_X.

The objective of CycleGAN contains adversarial losses and cycle consistency losses [13]. D_Y encourages G to generate images G(X) which are indistinguishable from the domain Y. It adopts adversarial loss:

$$L_{GAN}(G, D_Y, X, Y) = \mathrm{E}_{y\sim p_{data}(y)}[\log D_Y(y)] + \mathrm{E}_{y\sim p_{data}(x)}[\log(1 - D_Y(G(x)))], \quad (1)$$

3

The adversarial loss for F and D_X is defined as L_{GAN} (F, D_X ,Y, X), which is demonstrated as following:

$$L_{GAN}(F, D_X, Y, X) = \mathrm{E}_{x\sim p_{data}(x)}[\log D_X(x)] + \mathrm{E}_{y\sim p_{data}(y)}[\log(1 - D_X(G(y)))], \quad (2)$$

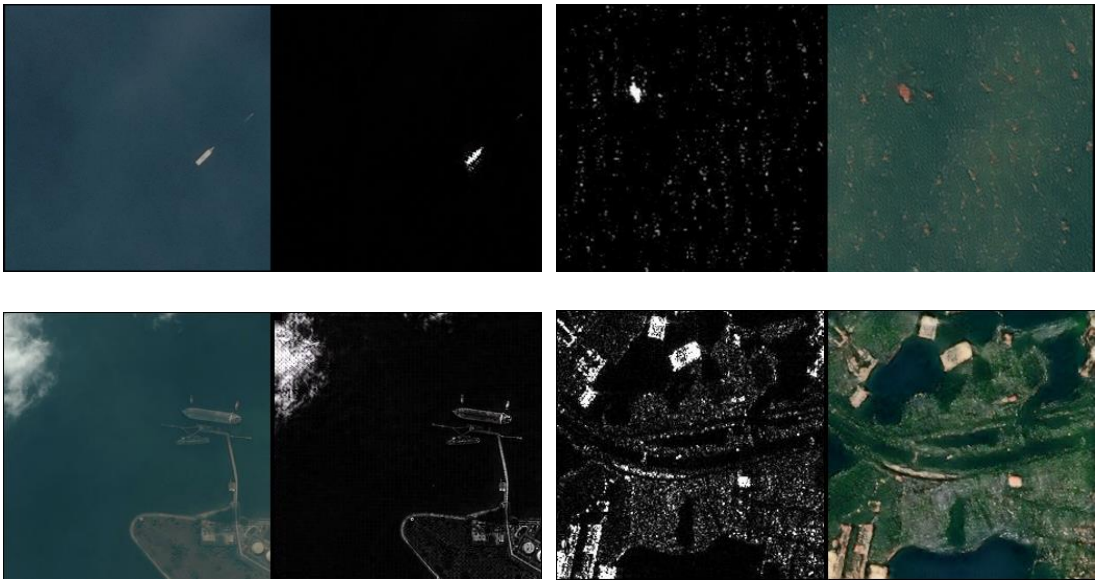
C

Cycle consistency loss is used to prevent the contradiction between mappings G and F [13], it is expressed as Eq. (3):

$$L_{cyc}(G, F) = \mathrm{E}_{x\sim p_{data}(x)}[\| F(G(x)) - x \|_1] + \mathrm{E}_{y\sim p_{data}(y)}[\| F(G(y)) - y \|_1]. \quad (3)$$

.

Eq. (3) contains forward cycle consistency and backward cycle consistency, which means that the input image should be able to be brought to the original image in every image translation cycle. Results on the translation between satellite ship images and SAR ship images are shown in Fig. 2. And it can be seen from Fig. 2 that we will obtain SAR ship images with complex scenes through transfer learning.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

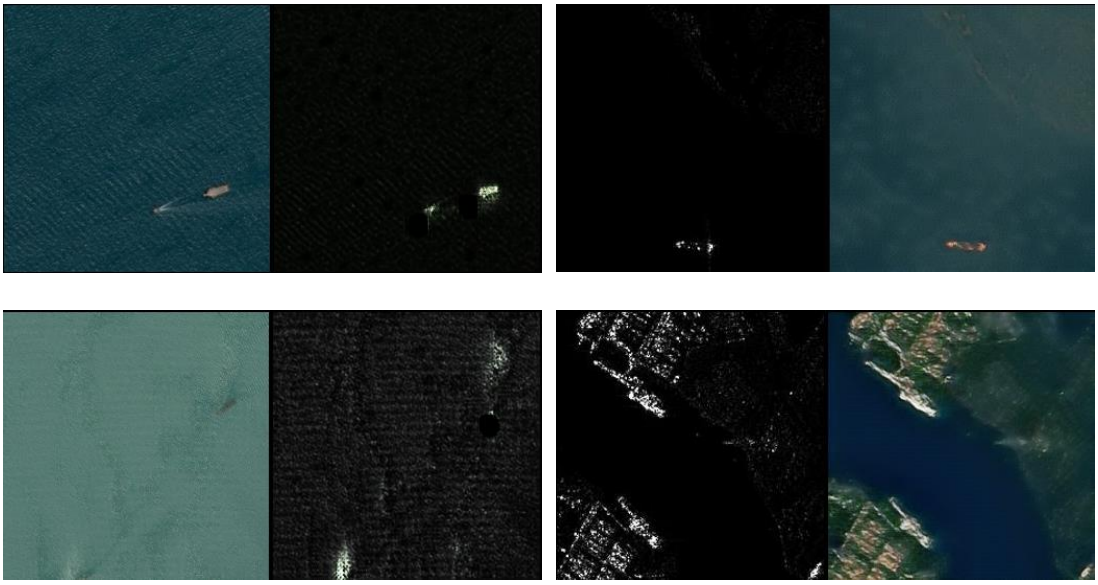


Figure 2. Representative results of CycleGAN on the style translation of satellite ship images and SAR ship images: (left) the satellite image is translated to SAR image; (right) the SAR image is translated to satellite image.

3.2. Improved Libra RetinaNet

Tiny object detection with the complex scene is a great challenge in the SAR ship detection task. In this paper, we propose the Improved Libra RetinaNet to detect tiny ship in the intricate SAR ship images.

3.2.1. The Architecture of Improved Libra RetinaNet

The improved Libra RetinaNet in our work is presented based on the Libra RetinaNet [14], which is a combination of RetinaNet and Libra R-CNN. RetinaNet as a one-stage detector address the extreme foreground-background class imbalance with Focal Loss [4]. The architecture of original Libra RetinaNet is shown as Fig. 3. In the framework of Libra R-CNN, multi-level features are fused and enhanced by FPN and BFP respectively. Classification loss is optimized by focal loss and regression loss is optimized by balanced L1 loss. Libra R-CNN can work as complementary with RetinaNet by mitigating the feature level imbalance and objective level imbalance, and the integrated Libra RetinaNet is thus able to be employed for detecting ships in SAR images.

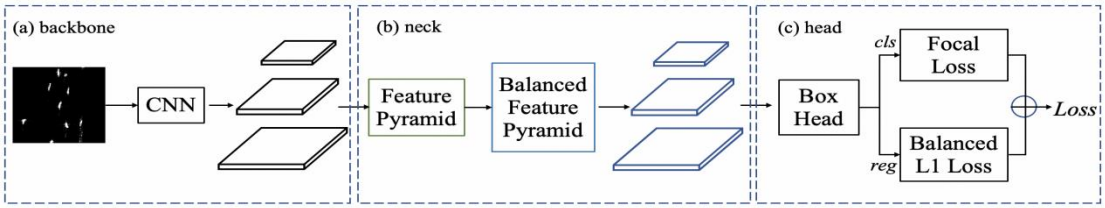


Figure 3. The architecture of Libra RetinaNet. (a) is used to extract features from an image. (b) is responsible for fusing the multi-scale features. (c) contains bounding box classification network and bounding box regression network

To approach the problem of low-level features loss in the Libra RetinaNet which is critical for detecting tiny object in the SAR ship images with the complex scenes, we introduce a bottom-up path augmentation to improve Libra RetinaNet. The architecture of our improved Libra RetinaNet is shown as Fig. 4. It contains three modules: backbone, neck, and head.

Backbone is the module that extracts feature maps from an image. It generally selects deep convolutional neural networks to extract features, such as VGG [18], ResNet [19].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

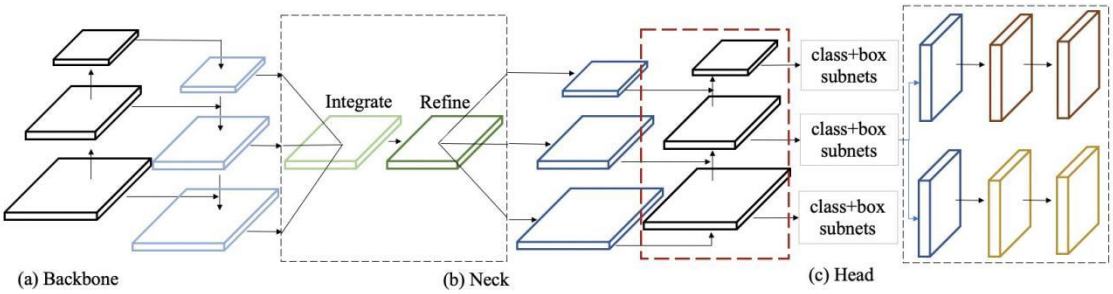


Figure 4. The architecture of our improved Libra RetinaNet. (a) is the part that extracts features from an image. (b) is the part that fuses the low-level features and high-level features. (c) is responsible for bounding box classification and bounding box regression. The red dotted box indicates the bottom-up path augmentation.

Neck is the module that connects the backbone with head to fuse multilevel features. The neck module of original Libra RetinaNet contains two parts: Feature Pyramid Networks (FPN) [20] and Balanced Feature Pyramid (BFP) [14], as shown in Fig. 5. FPN obtains high-resolution and strong semantic feature by fusing low-level features and high-level features, which are extracted by the backbone via a top-down path and lateral connections. Then BFP works as complementary with FPN. It resizes the multi-level features to the same size of the intermediate level by interpolation or max-pooling. Once the features are rescaled, the balanced semantic features are obtained by adding these features of the same size and averaging them. Then the embedded Gaussian non-local attention [21] or direct convolutions is used to refine the balanced semantic features.

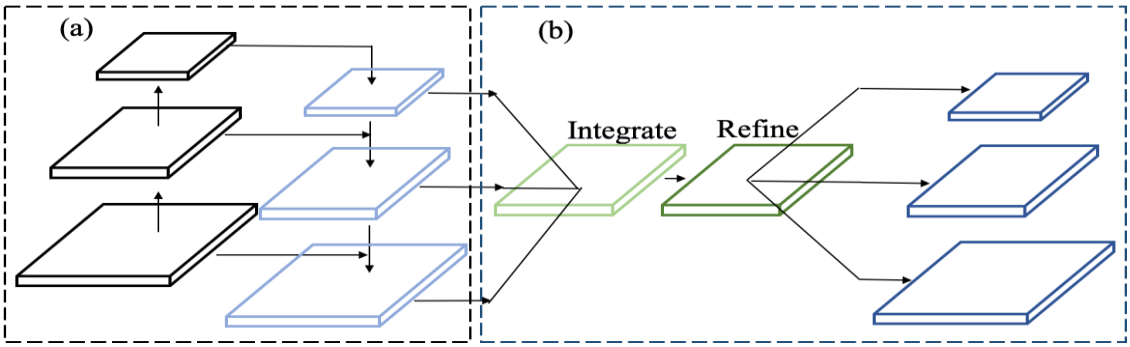


Figure 5. The architecture of the neck module in Libra RetinaNet. (a) is feature pyramid network (FPN). (b) is balanced feature pyramid network (BFP).

Head is responsible for bounding box classification and regression. It contains two parts: bounding box classification network and the bounding box regression network. The two parts are both composed of three convolutional layers.

In the process of features extraction, low-level features containing more location and detail information is conducive to tiny object detection, while high-level features mining semantic information help to detect the large object well. However, the features extraction in Libra RetinaNet is generally achieved in backbone using a deep convolutional neural network, such as VGG [18], ResNet [19], etc. This means that the transfer of the features from the low-level to high-level requires passing through dozens or even hundreds of layers. The loss of low-level features is therefore serious during
feature flowing among different layers, which is obviously a big hindrance for tiny object detection. The study in this paper considers solving the problem of low-level features loss in the process of feature transfer. Inspired by [22], we bring a bottom-up path augmentation into the neck module of Libra RetinaNet. The final architecture of our improved Libra RetinaNet is shown in Fig.4. By adding the bottom-up path augmentation structure, and the lateral connecting with the enhanced feature maps of BFP in the neck module of Libra RetinaNet as well, the transfer of low-level features to high-level features only requires a few layers instead of dozens or hundreds of layers. In addition, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

bottom-up path augmentation does not increase the number of parameters and calculations too much. For the improved Libra RetinaNet, in the case of almost no increase in training costs, the path of low-level features transmission is shortened, and the low-level features are better preserved, which is more conducive to tiny object detection in SAR ship detection task.

The head module is responsible for bounding box classification and regression. It contains two parts: bounding box classification network and the bounding box regression network. The two parts are both composed of three convolutional layers.

3.2.1. Loss Function

The improved Libra RetinaNet adopts the multi-task loss, which is defined as the sum of classification loss and localization loss.

The focal loss [4] is proposed o address the class imbalance between easy and hard samples. It is expressed as Eq. (4):

$$FL(\mathrm{p}_t) = -\alpha(1-p_t)^{\gamma}\log(p_t), \tag{4}$$

in which parameters α_t and γ are hypermeters to moderate the weights between easy and hard samples. And p_t is defined as Eq. (5):

$$\mathrm{p}_t = \begin{cases} p & \text{if } \mathrm{y}=1 \\ 1-p & \text{otherwise} \end{cases} \tag{5}$$

R
e

here p is the estimated probability by model and y = 1 species the ground-truth.

The balanced L1 loss [14] is designed to enhances the gradients of samples with loss less than 1.0, so as to achieve more balanced training in accurate location. It is expressed as Eq. (6):

$$\mathrm{L}_b(x) = \begin{cases} \dfrac{\alpha}{b}(b|x|+1)\ln(b|x|+1) - \alpha|x| & \text{if } |\mathrm{x}|<1 \\ \gamma|x|+C & \text{otherwise} \end{cases} \tag{6}$$

where α is used to control the gradients of easy samples, γ is a moderate factor to control the upper bound of regression errors, parameter b is to ensure the continuity of function. And γ, α and b are constrained by Eq. (7):

$$\alpha\ln(b+1) = \gamma \tag{7}$$

4. Results

4.1. Dataset and Evaluation Metric

Our experiments are implemented on the SAR ship dataset under complex backgrounds. The dataset was created using 102 Chinese Gaofen-3 images and 108 Sentinel-1 images [1]. It consists of 43,819 ship chips of 256 pixels, with differences in polarization, resolution, incidence angle and backgrounds. This dataset is a voc-style [23] dataset and each ship chip correspond to an annotation file, which indicates the ship location, the

 doi:10.20944/preprints202108.0509.v1

ship chip name, and the image shape. The entire dataset is randomly split into a training dataset (70%), a validation dataset (20%), and a test dataset (10%). All experiments results follow the standard evaluation protocol used in object detection, i.e., Intersection-over-Union (IoU). And the evaluation metric used for all reported results is mean average precision (mAP).

4.2. Implementation Details

We train the CycleGAN model for 200 epochs, then we use the model to translate 29070 satellite images to SAR images, which means that we expand 29070 images based on the original SAR ship dataset. During detection experiments, for fair comparisons, we implement all experiments on PyTorch and mmdetection [24]. The backbone used in our experiments is the publicly available ResNet-50. We train models on 2 GPUs (2 or 4 images per GPU) for 12 epochs. And we set different input size for different models due to the GPU's limited memory. According to the Linear Scaling Rule [25], the initial learning rate is set to 0.01 or 0.005 which is depended on the batch size. All other hyper-parameters are set as the default settings in mmdetection [24].

4.3. Main Results

The experimental results on SAR ship test dataset are shown in Table 1. Among these models, Faster R-CNN [11] is a two-stage detector, Cascade R-CNN [26] is a multi-stage detector. RetinaNet [4] and Libra RetinaNet [14] are both one-stage detectors. In Table 1, the symbol '*' means results in [1]. Compared with the result in [1], our re-implenmented result on Faster R-CNN with the improved neck module based on the augmented data have mAP of 96.75%, which is 8.49 points higher than the mAP of Faster R-CNN in [1]. And the result on RetinaNet in the same way achieves 96.53% mAP, which is 5.17 points higher than the mAP of RetinaNet in [1].We also employed the Cascade R-CNN to detect the ship of SAR image data and the mAP is 96.36%. After implementing transfer learning and improving the neck module of Cascade R-CNN, the mAP increases by 0.82 points. From Table 1, we can observe that the detection performance of these detectors is evidently improved after adopting our proposed data augmentation and introducing the bottom-up augmentation in the neck module in the structure of these detectors. The proposed data augmentation strategy increases mAP by 0.4-0.8 points, and our improved neck module increases mAP by s0.2-0.3 points. Our re-implemented result on Libra RetinaNet achieves 96.53% mAP, which outperforms three original detector models in the table. Subsequent way of data augmentation increases mAP by 0.53 points. And the highest mean average precision (mAP) is 97.38% that is achieved by our improved Libra RetinaNet, which is an integration of Libra RetinaNet and our improved neck module. Overall the proposed data augmentation method increases mAP by 0.4-0.8 points, and our improved neck module increases mAP by 0.2-0.3 points on the basis of data augmentation. The substantial improvements validate that enriching data and better preservation of low-level features can boost the performance of the detectors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

Table 1. Ship detection mean average precision(mAP) of models. The symbol "*" means results in [1].

Method	Input Size	Data Augmentation	Improved Neck	mAP (%)
Faster R-CNN*	600 * 800	-	-	88.26
Faster R-CNN	1024 * 1024	-	-	95.81
Faster R-CNN	1024 * 1024	√	-	96.53
Faster R-CNN	1024 * 1024	√	√	96.75
RetinaNet*	800 * 800	-	-	91.36
RetinaNet	1024 * 1024	-	-	95.68
RetinaNet	1024 * 1024	√	-	96.29
RetinaNet	1024 * 1024	√	√	96.53
Cascade R-CNN	800 * 800	-	-	96.36
Cascade R-CNN	800 * 800	√	-	96.93
Cascade R-CNN	800 * 800	√	√	97.18
Libra RetinaNet	1024 * 1024	-	-	96.53
Libra RetinaNet	1024 * 1024	√	-	97.05
Improved Libra RetinaNet (our)	1024 * 1024	√	√	97.38

The SAR ship detection result images are shown in Fig.6. In Fig. 6, the ground truths data example are shown in the first column, and the second column shows the corresponding detection results of original Libra RetinaNet. The detection results after data augmentation is presented in the third column. The last column shows the detection results of our improved Libra RetinaNet. The detection results in the first three rows show that since we translate satellite images containing complex scenes into SAR images and expand them to the original data set, the model learns the features more accurately on richer data and correctly detects the exact ships. From the first row and the third row we can observe that our proposed data augmentation strategy can help reduce the false detection, and the second row shows that our method addresses the miss detection of ship objects. The last three rows in Fig. 6 show that compared to the original Libra RetinaNet, our improved Libra RetinaNet achieves higher scores on the bounding boxes in the SAR image, which means that the improved Libra RetinaNet can achieve significant performance for multi-target detection based on complex backgrounds.

5. Discussion and Conclusion

In this paper, we propose a practical approach for boosting the ship detection in SAR images. We first propose the data augmentation strategy to expand the SAR ship dataset. We use the CycleGAN to perform the transfer learning and transfer the satellite ship images into the SAR ship images, which can not only enlarge the original dataset, but also increase the diversity of samples, so as to effectively improve the robustness of the model. Then we improve the neck module of Libra RetinaNet which is applied to the tiny ship detection of SAR images under complex backgrounds. With FPN and BFP for multi-level features fusion and enhancement, Libra R-CNN can be combined with RetinaNet to detect ships in SAR images. Due to the serious loss of low-level features during the feature transfer process of Libra RetinaNet, we improve the neck of Libra RetinaNet by adding the bottom-up path augmentation structure to better preserve low-level features, which tremendously raise the performance of tiny ship detection with an average accuracy of 97.38%. This implies that our method can be accepted for approaching the tasks involving the ship objects of variant scale in the SAR ship dataset with the complex backgrounds and can facilitate further research.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 August 2021 doi:10.20944/preprints202108.0509.v1

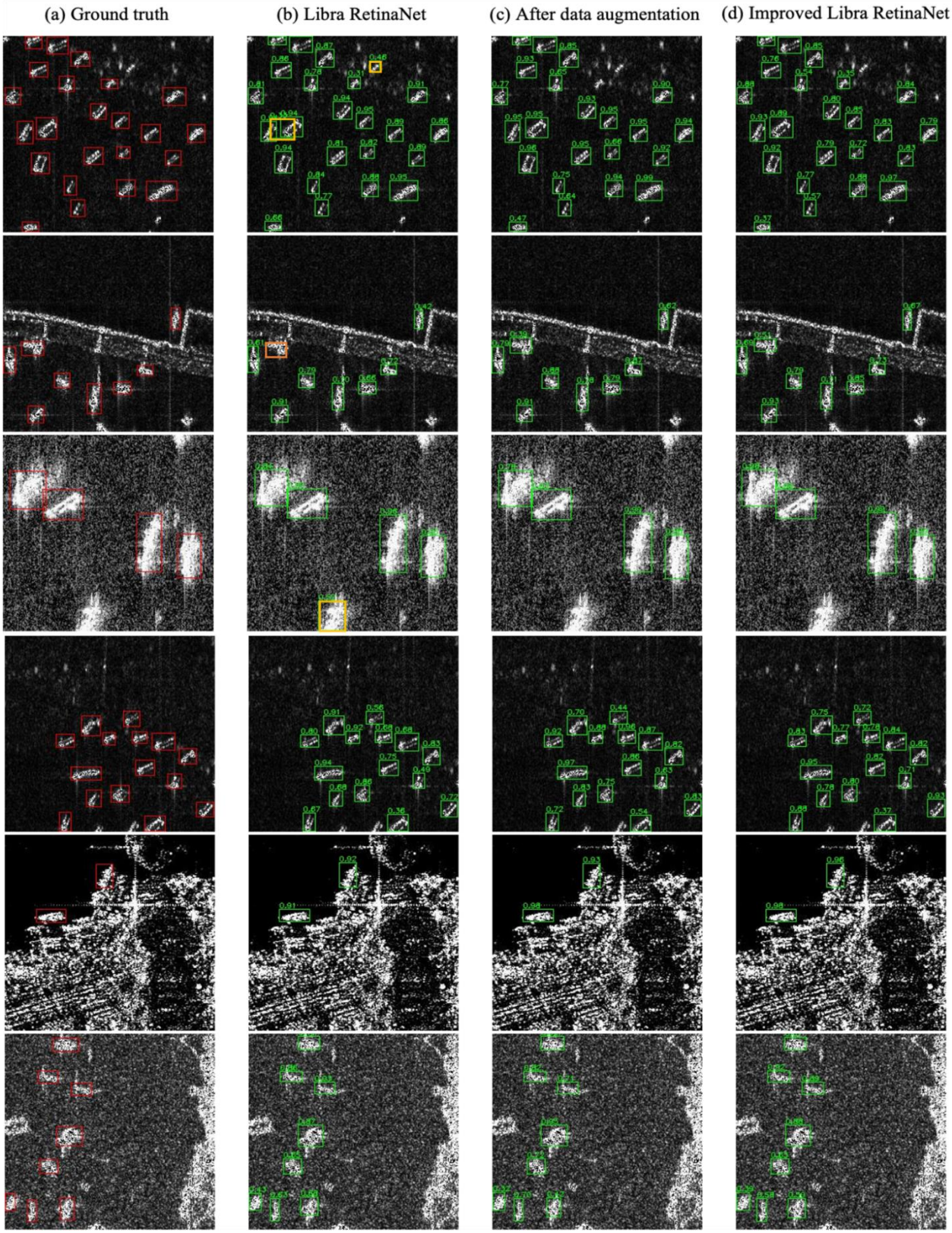


Figure 6. Ship detection results under complex backgrounds. (a) shows the ground truths, (b) shows the detection results of original Libra RetinaNet, (c) shows the detection results after data augmentation, and (d) shows the detection results of our improved Libra RetinaNet. Orange is miss-detection and yellow is false alarm.

Author Contributions: Conceptualization, Fei Yang; Investigation, Lei Zhang and Weigang Lu; Methodology, Haomiao Liu, Haizhou Xu and Fei Yang; Software, Haomiao Liu and Haizhou Xu; Writing – review & editing, Haomiao Liu, Haizhou Xu, Lei Zhang, Weigang Lu, Fei Yang and Jingchang Pan.

All authors will be informed about each step of manuscript processing including submission, revision, revision reminder, etc. via emails from our system or assigned Assistant Editor.

Funding: This research was funded by Natural Science Foundation of Shandong Province, grant number ZR2019MF011 and Postdoctoral Science Foundation of China, grant number 2017M622210.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work has been supported by the Natural Science Foundation of Shandong Province (No. ZR2019MF011), and the Postdoctoral Science Foundation of China (No. 2017M622210). This work was also supported in part by National Natural Science Foundation of China (No.61502275) and the Natural Science Foundation of Shandong Province (No. ZR2017MF051). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V used for this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A sar dataset of ship detection for deep learning under complex backgrounds," IEEE Trans. Ind. Electron., vol. 11, no. 7, 2019.
2. Y. Wang, C. Wang, and H. Zhang, "Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 sar images," IEEE Trans. Ind. Electron., vol. 9, no. 7-9, pp. 780–788, 2018.
3. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision. Springer, 2016, pp. 21–37.
4. T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," IEEE Trans. Ind. Electron., vol. PP, no. 99, pp. 2999– 3007, 2017.
5. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
6. J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," IEEE Trans. Ind. Electron., 2016.
7. ——, "Yolov3: An incremental improvement," IEEE Trans. Ind. Electron., 2018.
8. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," IEEE Trans. Ind. Electron., 2020.
9. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.
10. R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international con- ference on computer vision, 2015, pp. 1440–1448.
11. S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99.
12. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceed- ings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.
13. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "MaskJ.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223– 2232 r-cnn," in Proceed- ings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.
14. J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra r-cnn: Towards balanced learning for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2019, pp. 821–830.
15. Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on retinanet using multi-resolution gaofen-3 imagery," IEEE Trans. Ind. Electron., vol. 11, no. 5, p. 531, 2019.
16. M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster r-cnn based on cfar algorithm for sar ship detection," in 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP). IEEE, 2017, pp. 1–4.
17. Y. Liu, M.-h. Zhang, P. Xu, and Z.-w. Guo, "Sar ship detection using sea- land segmentation-based convolutional neural network," in 2017 Interna- tional Workshop on Remote Sensing with Intelligent Processing (RSIP). IEEE, 2017, pp. 1–4.
18. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," IEEE Trans. Ind. Electron., 2014.

19. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
20. T.-Y. Lin, P. Dollár, 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, 2017, pp. 2117–2125.
21. X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7794–7803.
22. S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8759–8768.
23. M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," IEEE Trans. Ind. Electron., vol. 88, no. 2, pp. p.303–338, 2010.
24. K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng,Z. Liu, J. Xu et al., "Mmdetection: Open mmlab detection toolbox and benchmark," IEEE Trans. Ind. Electron., 2019.
25. P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola,A. Tulloch, Y. Jia, and K. He, "Accurate, large mini-batch sgd: Training imagenet in 1 hour," IEEE Trans. Ind. Electron., 2017.
26. Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6154–6162.