Communication

Mid-low Resolution Remote Sensing Ship Detection Using Super-Resolved Feature Representation

Shitian He, Huanxin Zou *, Yingqian Wang, Runlin Li , Fei Cheng and Xu Cao

[1] College of Electronic Science and Technology, National University of Defense Technology; heshitian19@nudt.edu.cn (S.H.); wangyingqian16@nudt.edu.cn (Y.W.); lirunlin14@nudt.edu.cn (R.L.); chengfei08@nudt.edu.cn (F.C.); cx2020@nudt.edu.cn (X.C.)

* Correspondence: hxzou2008@163.com; Tel.: +86-1397-313-3366

Abstract: Existing methods enhance mid-low resolution remote sensing ship detection by feeding super-resolved images to the detectors. Although these methods marginally improve the detection accuracy, the correlation between image super-resolution (SR) and ship detection is under-exploited. In this paper, we propose a simple but effective ship detection method called *ShipSR-Det*, in which both the output image and the intermediate features of the SR module are fed to the detection module. Using the super-resolved feature representation, the potential benefit introduced by image SR can be fully used for ship detection. We apply our method to the *SSD* and *Faster-RCNN* detectors and develop *ShipSR-SSD* and *ShipSR-Faster-RCNN*, respectively. Extensive ablation studies validate the effectiveness and generality of our method. Moreover, we compare *ShipSR-Faster-RCNN* with several state-of-the-art ship detection methods. Comparative results on the HRSC2016, DOTA and NWPU VHR-10 datasets demonstrate the superior performance of our proposed method.

Keywords: Ship detection; image super-resolution; mid-low resolution remote sensing images

1. Introduction

Optical remote sensing ship detection plays an important role in port management, marine rescuing and military reconnaissance. With the advances of deep learning, recent methods [1–3] generally use deep convolution neural networks (DCNNs) for remote sensing ship detection, and achieve significant improvements over traditional paradigms. As a key factor for ship detection, high-resolution (HR) images (with ground sample distance (GSD) smaller than 10 m/pixel) can provide abundant appearance information and thus introduce benefits to the detection task [4]. However, obtaining an HR image posts a high requirement on the satellite sensors and generally results in an expensive cost. In contrast, mid-low resolution images (with GSD larger than 10 m/pixel) can be acquired more cheaply, but their insufficient details post great challenges to ship detection. To achieve a better trade-off between detection accuracy and resource consumption, performing image super-resolution (SR) on mid-low resolution remote sensing images to recover their missing details has become a popular research topic and has been widely investigated in recent years [5–8].

In the field of remote sensing object detection, several methods [9–12] perform image SR as a pre-processing approach, and feed the super-resolved image to detection network to improve the detection performance. In these methods, image SR and object detection are performed as two separate processes, and the connection between these two processes is the super-resolved image only. Although the super-resolved images contain more details, the informative features extracted by the SR module cannot be fully used by the detection module, which hinders the further improvement of detection accuracy.

To fully use the informative feature representation provided by the SR network, in this paper, we propose an SR-based ship detection method for mid-low resolution

Figure 1. Visual results achieved by *Faster-RCNN* and our proposed *ShipSR-Faster-RCNN* on the HRSC2016 dataset. "HR" denotes high-resolution images and "GT" represents groundtruth labels. Our method recovers missing details in the input image and achieves promising detection performance.

remote sensing images. Our method, named *ShipSR-Det*, consists of an SR module and a detection module. Different from previous methods that only feed super-resolved images to the detector, in our method, both output images and intermediate features produced by the SR module are fed to the detection module. By the assistance of the super-resolved feature representation, the detection module can extract more informative features to achieve accurate ship detection. We adopt *SSD* [13] and *Faster-RCNN* [14] as our detection module to develop *ShipSR-SSD* and *ShipSR-Faster-RCNN*, respectively, and demonstrate their effectiveness through extensive ablation studies and visualizations. As shown in Fig. 1, our *ShipSR-Faster-RCNN* can well recover the missing details in the input images to enhance the detection performance. Moreoever, we compare our *ShipSR-Faster-RCNN* with several popular detectors on the HRSC2016 [15], DOTA [16] and NWPU VHR-10 [17] datasets. Comparative results demonstrate the state-of-the-art performance of our method.

This paper is an extension of our previous conference version [18] in which we proposed an RDN-based SR network tailored with an *SSD* detector for ship detection. Compared to our previous work, we make the following additional contributions in this paper.

- We propose a generic SR-based ship detection method which can be applied to different detectors and backbones to achieve consistent performance improvement.
- We conduct extensive ablation studies and perform feature visualizations to investigate our proposed method. Experimental results validate the effectiveness of using super-resolved feature representation for ship detection.
- We compare our *ShipSR-Faster-RCNN* to several state-of-the-art detectors on three public datasets. Comparative results demonstrate the competitive performance of our method.

2. Related Works

2.1. Ship Detection

With the development of deep learning techniques in object detection [13,14,19,20], ship detection has been deeply investigated in recent years. Different from general object detection, remote sensing ship detection has some special characteristics such as multi-orientation, complex scenarios, large intra-class and small inter-class distance. Most works on remote sensing ship detection aim at handling these challenges to improve the detection accuracy. For example, Ding et al. [21] addressed the arbitrary orientation issue by modifying RPN with RRoI to transform horizontal proposals to rotated ones; Yang et al. [22] added an IoU constant factor to the smooth L1 loss to address the boundary problem for the rotating bounding box; Yang et al. [23] proposed an end-to-end refined single-stage rotation detector using a progressive regression approach to adapt to the dense arrangement scenarios. To handle the large intra-class and small inter-class distance issue, Li et al. [24] proposed a shape-adaptive pooling approach to extract more compact and qualified feature representation for ship classification and localization. To achieve robust ship detection under complex scenarios, Lei et al. [25]

introduced a saliency constraint to the CNN model to enhance the object regions for better detection. Yu et al. [26] developed a pre-processing structure to discriminate whether an image patch contains objects before detection. Using their method, the mount of false positives on background areas can be reduced.

Besides the aforementioned challenges, in recent years, some studies [27,28] addressed the resolution issue in remote sensing images since the low-resolution input images can degrade the detection performance. These methods modified existing networks to extract multi-scale features more effectively, and partially improved the detection performance. However, these methods are relatively complex and have a large computation consumption. Another solution is to perform image SR as a preprocessing step to recover the missing details in input images. The related works in image SR and SR-based ship detection will be briefly reviewed as below.

2.2. Image Super-Resolution

Image super-resolution (SR) aims at reconstructing a high-resolution (HR) image from one or multiple low-resolution (LR) observations. Recently, deep learning has been successfully applied to image SR and has achieved continuously improving performance. Dong et al. [29] proposed the first CNN-based single image SR method to reconstruct HR images by using a 3-layer CNN. Kim et al. [30] proposed a deeper network named VDSR to improve the reconstruction accuracy. Zhang et al. [31] combined residual connection [32] with dense connection [33], and proposed residual dense network (i.e., RDN) to fully exploit hierarchical feature representations for image SR. Li et al. [34] proposed a multiscale residual network to fully exploit the hierarchical feature representation for image super-resolution. Wang et al. [35] explored the sparsity prior in image SR and used sparse convolutions to achieve accurate and efficient image SR. Subsequently, Wang et al. [36] proposed a degradation-aware network and achieve image SR with arbitrary blur kernels and noise levels. Apart from single-image SR methods, several methods [37–40] enhanced SR performance by exploiting the complementary information among multiple input images.

2.3. SR-based Detection

In the field of remote sensing object detection, several methods performed image SR to enhance the detection accuracy. Dong et al. [6] proposed a second-order multi-scale SR network and demonstrated its effectiveness to object detection. Rabbi et al. [9] proposed an edge-enhanced generative adversarial network (GAN), and combined it with an *SSD* [13] detector in an end-to-end manner to improve the detection accuracy. Courtrai, Pham, and Le [10] tailored a GAN-based SR network with a detection network to develop an object-focused detection framework. Wang, Lu, and Zhang [11] modified the loss function to make the SR network more suitable for the detection task. Noh et al. [12] selects relatively small region of interests (RoIs) to perform image SR to improve the detection performance. Note that, although these SR-based detection methods have shown their effectiveness, the benefits introduced by image SR have not been fully utilized since only super-resolved images are fed to the detectors while the informative feature representation generated by SR networks is overlooked.

3. Network Architecture

In this section, we introduce our method in details. As shown in Fig. 2, our method consists of two parts including an SR module and a detection module. Without loss of generality, we use the *Faster-RCNN* detector as our detection module to introduce the details of our method.

3.1. SR Module

As shown in Fig. 2, our SR module takes a mid-low resolution image $\mathcal{I}_{LR} \in \mathbb{R}^{H\times W\times 3}$ as its input to produce an SR image $\mathcal{I}_{SR} \in \mathbb{R}^{\alpha H\times \alpha W\times 3}$ and an intermediate feature

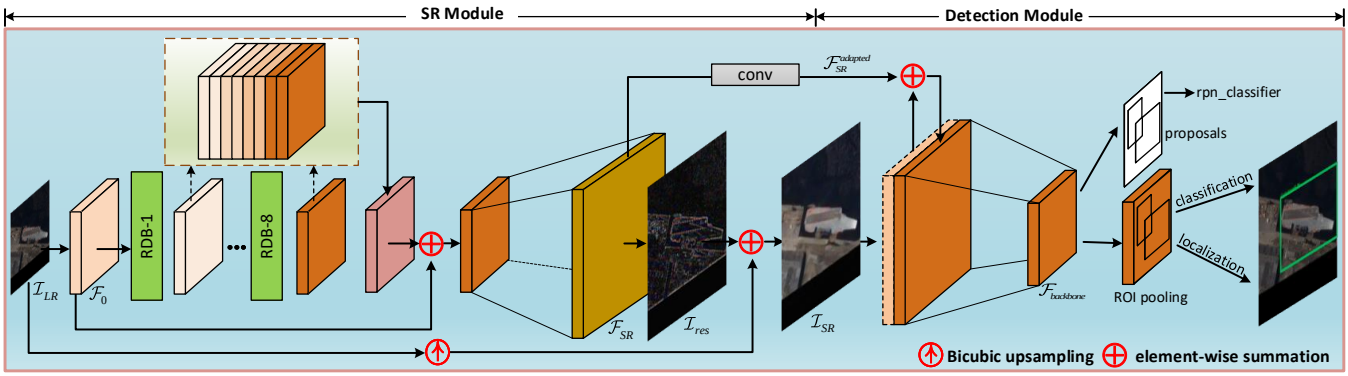

Figure 2. An overview of our *ShipSR-Det*.

$\mathcal{F}_{SR} \in \mathbb{R}^{\alpha H \times \alpha W \times 64}$, where H and W represent the height and width of the input image, and α denotes the upscaling factor. Specifically, the input image $\mathcal{I}_{LR}$ is first fed to a 3×3 convolution to generate initial feature $\mathcal{F}_0 \in \mathbb{R}^{H \times W \times 64}$. Then, $\mathcal{F}_0$ is fed to 8 cascaded residual dense blocks (RDBs) [31] for deep feature extraction. Features from all the RDBs are concatenated for global fusion, and the fused feature is added with the initial feature $\mathcal{F}_0$ and fed to a sub-pixel layer to generate the upsampled feature $\mathcal{F}_{SR}$. Finally, $\mathcal{F}_{SR}$ is fed to a 3×3 convolution to produce the residual prediction which is further added with the bicubicly upsampled input image to generate the final SR image $\mathcal{I}_{SR}$.

3.2. Detection Module

As aforementioned, *Faster-RCNN* is used as our detection module. As a typical object detection method, *Faster-RCNN* first uses a backbone network (e.g., ResNet101 [32]) to extract informative features, and then feeds the extracted features (i.e., $\mathcal{F}_{backbone}$) to a region proposal network (RPN) to generate region proposals. Finally, the generated proposals and $\mathcal{F}_{backbone}$ are fed to the ROI-Pooling layer to extract and resize the features of region proposals, and then fed to a classification layer and a localization layer to produce the final classification and location predictions.

3.3. Feature Connection

When the SR module and the detection module are selected, the SR-based detector can be built by feeding the SR image to the detection module. However, in most SR-based detection methods [6,9–11], the informative features are squeezed to an image, and the backbone of the detection module extracts the features from the SR image for further prediction. The feature squeeze and re-extraction result in the information lost inevitably. In our method, to reduce the information lost and fully use the super-resolved feature representation for ship detection, both the super-resolved image $\mathcal{I}_{SR}$ and the intermediate feature $\mathcal{F}_{SR}$ are fed to backbone of detection module for feature extraction. Note that, since features in the SR module and detection module have different depths and resolutions, we use a 3×3 convolution to adapt $\mathcal{F}_{SR}$ for ship detection. The weights of the 3×3 convolution were initialized as zero values and updated during end-to-end finetuning. The adapted feature $\mathcal{F}_{SR}^{adapted}$ is added to the initial feature extracted from $\mathcal{I}_{SR}$ for deep feature extraction. In this way, the informative feature representation generated by the SR module can be fully used by the detection module for ship detection. Experimental results in Section 3.6 demonstrate the effectiveness of our method. After deep feature extraction, the extracted feature $\mathcal{F}_{backbone}$ is fed to the original neck and head of detection module to produce the final classification and location predictions.

3.4. Losses

The loss of our method can be defined as:

$$L_{overall} = L_{Det} + \lambda L_{SR}, \tag{1}$$

where the L_{Det} and L_{SR} represent the detection loss and SR loss, respectively. λ is a hyper-parameter to balance the SR loss and detection loss.

Specifically, L_{Det} is the detection loss which is identical to that in the *Faster-RCNN*. And L_{SR} is the L_1 distance between the groundtruth image I_{HR} and the super-resolved image I_{SR}. That is,

$$L_{SR} = \| I_{HR} - I_{SR} \|_1 . \tag{2}$$

In this section, we first introduce the datasets and implementation details, then conduct ablation studies and perform feature visualizations to validate the effectiveness of our method. Finally, we compare our method to several state-of-the-art methods on three public datasets.

3.5. Datasets and Implementation Details

We used the HRSC2016 [15], DOTA [16] and NWPU VHR-10 [17] datasets in our experiments.

- **HRSC2016**: HRSC2016 is a public remote sensing dataset for ship detection. It contains 617 images for training and 438 images for validation. We resized these images to 800×512 to generate groundtruth HR images.
- **DOTA**: DOTA is a public dataset with 15 object categories for multi-class object detection in aerial images. In the experiment, we cropped the images in the original datasets into patches of size 512×512, and chose patches containing ship targets to build our training and validation sets. We totally generated 4163 images for training and 1411 images for validation.
- **NWPU VHR-10**: NWPU VHR-10 dataset is a challenging geo-spatial object-detection dataset with 10 categories. We performed the same operations as in the DOTA dataset to generate training and validation samples. Our customized NWPU VHR-10 dataset contains 249 images for training and 52 images for validation.

We used the aforementioned modified images as HR images and performed 8× bicubic downsampling to generate the input mid-low resolution images. We performed a large variety of data augmentations including random horizontal and vertical flipping, random rotation, random color transformation, random brightness and contrast transformation.

Our method was implemented in PyTorch on a PC with an Nvidia RTX 2080Ti GPU. We trained our network progressively following a three-stage pipeline. In the first stage, we trained our SR module using the generated image pairs with an L_1 loss, and used the bicubicly upsampled images as input to train our detection module. When training the SR module, the batch size was set to 4 and the learning rate was initially set to 1×10^{-4} and halved for every 5×10^5 iterations. The training was stopped after 1.2×10^6 iterations. In the second stage, we tailored the SR module with the pretrained detection module via super-resolved image, and performed end-to-end finetuning for 24 epochs. In this stage, the learning rate was initially set to 1×10^{-4} and decreased by a factor of 0.1 for every 10 epochs. In the third stage, we further added the super-resolved feature representation to the detection module (pretrained in the second stage) and performed another round of finetuning. The training settings in this stage were identical to those in the second stage.

For evaluation, we followed [41] to use the mean average precision (mAP) as the quantitative metric with the Intersection over Union (IoU) being set to 0.5 (i.e., mAP50) and 0.75 (i.e., mAP75).

Table 1: Results achieved by different variants on the HRSC2016 dataset. I_{LR}, I_{bic}, I_{SR} and I_{HR} represent the mid-low resolution image, bicubicly upsampled image, super-resolved image and original HR image, which are used as the inputs of the detection module. *FT* represents the fine-tuning operation, *Fea* denotes feeding the super-resolved feature representation to the detection module for ship detection.

Index	I_{LR}	I_{bic}	I_{SR}	*FT*	*Fea*	I_{HR}	*SSD* [13]		*Faster-RCNN* [14]							
							VGG16 [42]		ResNet50 [32]		ResNet101 [32]		HRNet [43]		ResNeXt101 [44]	
							mAP50	mAP75	mAP50	mAP75	mAP50	mAP75	mAP50	mAP75	mAP50	mAP75
0	✓						-	-	0.491	0.074	0.490	0.079	0.515	0.114	0.529	0.086
1		✓					0.597	0.169	0.788	0.495	0.820	0.626	0.848	0.668	0.837	0.647
2			✓				0.688	0.282	0.823	0.611	0.855	0.688	0.857	0.702	0.861	0.689
3			✓	✓			0.711	0.313	0.838	0.663	0.863	0.712	0.863	0.735	0.871	0.717
4			✓	✓	✓		0.725	0.330	0.858	0.706	0.876	0.744	0.881	0.749	0.885	0.742
5						✓	0.773	0.374	0.894	0.728	0.910	0.769	0.915	0.813	0.930	0.808

* Since VGG16 is not fully convolutional, it requires a fixed input image size of 512×512. Consequently, the input images of VGG16 are resized to 512×512 and the results of I_{LR} on VGG16 are unavailable.

3.6. Ablation Study

We compare our method with several variants to investigate the potential benefits introduced by our design choices. Here, we validate the effectiveness of our method by introducing the following variants.

- *Model-0*: We fed the mid-low resolution image to the detection module. We introduce this variant to demonstrate the challenges of mid-low resolution ship detection.
- *Model-1*: We bicubicly upsampled the mid-low resolution image to the target resolution, and fed the upsampled image to the detection module. We introduce this variant to produce baseline results.
- *Model-2*: We use the pretrained SR module to super-resolve the input image, and fed the super-resolved image to the detection module.
- *Model-3*: We finetuned *model-2* to investigate the benefits introduced by end-to-end finetunning.
- *Model-4*: This is our proposed method. Based on *model-3*, we integrated the super-resolved feature representation to the detection module and performed another round of finetuning.
- *Model-5*: We directly fed the original HR images to detection module. We introduce this variant to produce upper-bound results.

Table 1 shows the comparative results achieved by our method and its variants. It can be observed that *model-0* achieves a very poor detection performance. That is because, the detectors cannot exploit enough useful information from mid-low resolution images. Compared to *model-0*, *model-1* uses the bicubicly upsampled versions of the mid-low resolution images as input, and achieves an improved detection performance. Note that, the detection accuracy is significantly improved if image SR is introduced. Taking the ResNet101-based *Faster-RCNN* detector as an example, *model-2* achieves an improvement of 3.5% in mAP50 and an improvement of 6.2% in mAP75 over *model-1*. It demonstrates that the details recovered by the SR module are beneficial to ship detection. Further improvements (0.8% in mAP50 and 2.4% in mAP75) can be achieved if end-to-end finetuning is performed. That is because, by performing end-to-end finetuning, the SR module in *model-3* can learn to super-resolve an image in a detection-driven manner. Compared to *model-3* which only feeds the super-resolved image to the detection module, our proposed method (i.e., *model-4*) can further achieve a performance gain (1.3% in mAP50 and 3.2% in mAP75) by reusing the features in the SR module. It is worth noting that, *model-4* can approximate its upper bound (i.e., 91.0% in mAP50 and 76.9% in mAP75) achieved by *model-5* on the HR image. The above results demonstrate the effectiveness of using super-resolved feature representation for ship detection. Moreover, it can be observed that our method is generic and can introduce consistent performance improvements to different detectors and backbones.

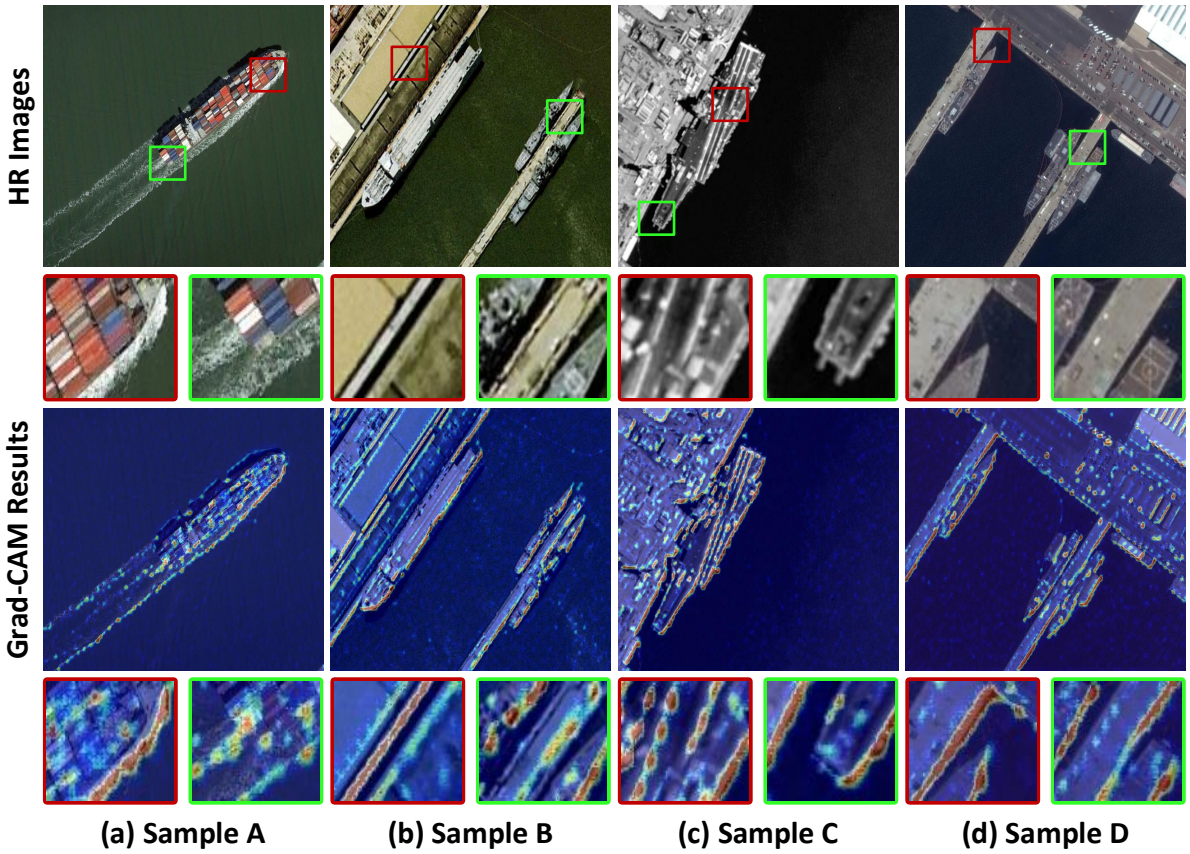

Figure 3. Visualizations of the bicubicly upsampled image (i.e., I_{bic}), super-resolved image (i.e., I_{SR}), the absolute difference of I_{SR} and I_{bic} (i.e., $|I_{SR} - I_{bic}|$), and the heatmaps produced by using the Grad-CAM method.

To qualitatively illustrate how the super-resolved feature representation contributes to ship detection, we use the Gradient-weighted Class Activation Mapping (Grad-CAM) [45] method to perform feature visualizations. As a widely used visual explanation method for CNNs, Grad-CAM can highlight feature regions with a larger influence on the final prediction. Here, we focus on the feature $\mathcal{F}^{SR}$ which is fed to the detection module. Figure 3 shows two example scenes on the HRSC2016 dataset. It can be observed that the major differences between SR images and bicubicly upsampled images are located in the edges. That is, SR images can provide much more edge information than bicubicly upsampled images. As shown in the heatmaps produced by the Grad-CAM method, edges in $\mathcal{F}^{SR}$ are highlighted and thus make more contributions to the final detection results. The above results demonstrate that the super-resolved feature representation contributes to ship detection by providing abundant edge information.

Table 2: Results achieved by our method with different settings of λ.

λ	0.001	0.01	0.1	1	10	100
mAP50	0.880	0.876	0.869	0.880	0.878	0.869
mAP75	0.728	0.732	0.740	0.751	0.744	0.739

Moreover, we investigate the influence of the hyper-parameter λ to the detection performance. As shown in Table 2, it can be observed that mAP75 changes more obviously than mAP50 with the increasing of λ, and when $\lambda = 1$, our *ShipSR-Faster-RCNN* can achieve the best performance in terms of both mAP50 and mAP75. Based on the comparative results, we set $\lambda = 1$ in our experiments.

Table 3: Quantitative results (i.e., mAP75) achieved by different methods (based on ResNet101) on the HRSC2016, DOTA and NWPU VHR-10 datasets. Here, we adopt the *Faster-RCNN* detector as our detection module. Our *ShipSR-Faster-RCNN* achieves state-of-the-art detection performance.

Method	Datasets			Parameters (MB)	Inference time (ms)
	HRSC2016 [15]	DOTA [16]	NWPU VHR-10 [17]		
GFL [49]	0.632	0.181	0.488	411.2	46.95
Reppoints [48]	0.453	0.161	0.563	331.8	51.28
HTC [46]	0.679	0.296	0.568	791.3	104.17
DetectoRS [47]	0.735	0.311	0.580	1336.9	158.73
Faster-RCNN [14]	0.626	0.233	0.549	485.6	48.31
ShipSR-Faster-RCNN	**0.744**	**0.342**	**0.608**	503.3	56.50

⋆ Inference time is averaged on the HRSC2016 dataset with an input mid-low resolution image of size 100×64.

3.7. Comparison to the State-of-the-art Methods

We apply our method to *Faster-RCNN* and compare our *ShipSR-Faster-RCNN* with four state-of-the-art detection methods including HTC [46], DetectoRS [47], Reppoints [48] and GFL [49]. We use the bicubicly upsampled images as the inputs of the compared methods to ensure the input size of different detectors is identical to our detection module.

3.7.1. Quantitative results

Comparative results are shown in Table 3. It can be observed that our *ShipSR-Faster-RCNN* achieves significant improvements over the original *Faster-RCNN* with only 17.7 MB increase of model size and 8.19 ms/image increase in inference time. Moreover, our *ShipSR-Faster-RCNN* outperforms Reppoints, GFL, HTC and DetectoRS on all the three datasets. Compared with DetectoRS, our method achieves a better performance with much fewer parameters and less inference time. Note that, although HTC and DetectoRS are also developed on *Faster-RCNN*, these two methods are less competitive due to the missing details in the input bicubicly upsampled images. In contrast, by using the super-resolved images and features, our method can well handle this problem and achieves state-of-the-art detection accuracy.

3.7.2. Qualitative results

Figure 4 shows the detection results achieved by different methods on three typical scenes, and these scenes indicate the following three challenges in ship detection: packed, multi-scaled and with complex background. It can be observed that the ships in scene A are closely packed, thus most detectors can not detect them accurately, and produce miss detection or error detection. That is because, the insufficient detail information makes the boundaries of these ships blurring, and thus only the most salient target can be recognized. By using the super-resolved images and feature representation, our *ShipSR-Faster-RCNN* can detect all the targets. In scene B, the background similar to the target (i.e., the docks marked by the red circle and ellipse) cannot be recognized due to the missing details of input images. Although the ships in scene C are salient for human vision, most detectors cannot detect them due to the insufficient appearance information. Note that, the small ships in both scene A and scene C cannot be detected by most detectors. In contrast, our *ShipSR-Faster-RCNN* can detect them accurately by using the beneficial detail information provided by the super-resolved images and feature representations.

4. Conclusion

In this paper, we propose a novel SR-based method *ShipSR-Det* for mid-low resolution ship detection. In our method, both the output image and the intermediate feature representation from the SR module are fed to the detection module to better

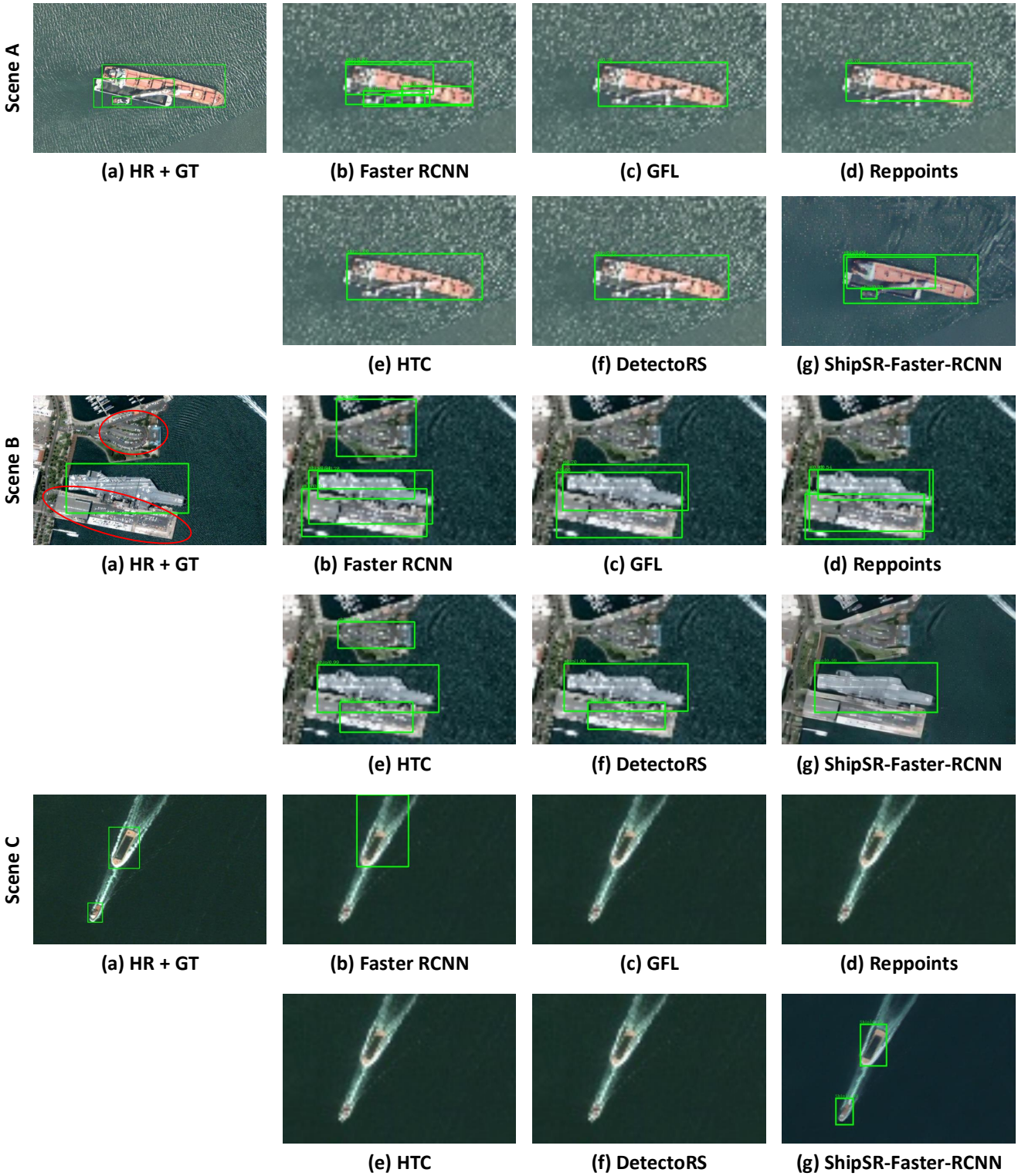

Figure 4. Qualitative results achieved by different methods on four example scenes. "HR" denotes high-resolution images and "GT" represents groundtruth labels. We use green bounding boxes to mark the detection results.

utilize the super-resolved information. Extensive ablation studies and visualizations have demonstrated the effectiveness of our method with different detection modules and backbones. Comparative results on three public datasets have demonstrated that our method can well recover the missing details in the mid-low resolution images, and achieves higher detection accuracy as compared to several state-of-the-art methods.

Author Contributions: H.Z. determined the research direction and modified the article expression; S.H. and Y.W. conceived the innovative ideas, and H.Z. helped to modify the conception. S.H. designed the ShipSR-Det framework, conducted the experiment and completed the first paper. Y.W. revised the manuscript and provided suggestions in expression. R.L., F.C. and X.C checked out the article's writing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 62071474.

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

References

1. Li, Q.; Mou, L.; Liu, Q.; Wang, Y.; Zhu, X.X. HSF-Net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery. *IEEE T-GRS* **2018**, *56*, 7147–7161.
2. Zhang, F.; Wang, X.; Zhou, S.; Wang, Y.; Hou, Y. Arbitrary-Oriented Ship Detection through Center-Head Point Extraction. *arXiv preprint arXiv:2101.11189* **2021**.
3. Wang, P.; Sun, X.; Diao, W.; Fu, K. FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery. *IEEE T-GRS* **2019**, *58*, 3377–3390.
4. Shermeyer, J.; Etten, A. The effects of super-resolution on object detection performance in satellite imagery. CVPR Workshop, 2019.
5. Dong, X.; Xi, Z.; Sun, X.; Gao, L. Transferred multi-perception attention networks for remote sensing image super-resolution. *Remote Sensing* **2019**, *11*, 2857.
6. Dong, X.; Wang, L.; Sun, X.; Jia, X.; Gao, L.; Zhang, B. Remote Sensing Image Super-Resolution Using Second-Order Multi-Scale Networks. *IEEE T-GRS* **2020**.
7. Pan, Z.; Yu, J.; Huang, H.; Hu, S.; Zhang, A.; Ma, H.; Sun, W. Super-resolution based on compressive sensing and structural self-similarity for remote sensing images. *IEEE T-GRS* **2013**, *51*, 4864–4876.
8. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-enhanced GAN for remote sensing image superresolution. *IEEE T-GRS* **2019**, *57*, 5799–5812.
9. Rabbi, J.; Ray, N.; Schubert, M.; Chowdhury, S.; Chao, D. Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network. *Remote Sensing* **2020**, *12*, 1432.
10. Courtrai, L.; Pham, M.; Lefèvre, S. Small Object Detection in Remote Sensing Images Based on Super-Resolution with Auxiliary Generative Adversarial Networks. *Remote Sensing* **2020**, *12*, 3152.
11. Wang, B.; Lu, T.; Zhang, Y. Feature-Driven Super-Resolution for Object Detection. IEEE CRC. IEEE, 2020, pp. 211–215.
12. Noh, J.; Bae, W.; Lee et al., W. Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection. ICCV, 2019, pp. 9725–9734.
13. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. ECCV. Springer, 2016, pp. 21–37.
14. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE T-PAMI* **2016**, *39*, 1137–1149.
15. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. A High Resolution Optical Satellite Image Dataset for Ship Recognition and Some New Baselines. ICPR, 2017, Vol. 2, pp. 324–331.
16. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. IEEE CVPR, 2018, pp. 3974–3983.
17. Cheng, G.; Han, J.; Zhou, P.; Guo, L. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS* **2014**, *98*, 119–132.
18. He, S.; Zou, H.; Wang, Y.; Li, R.; Cheng, F. ShipSRDet: An End-to-End Remote Sensing Ship Detector Using Super-Resolved Feature Representation. *arXiv preprint arXiv:2103.09699* **2021**.
19. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. IEEE CVPR, 2017, pp. 2117–2125.
20. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. IEEE ICCV, 2017, pp. 2961–2969.
21. Ding, J.; Xue, N.; Long, Y.; Xia, G.S.; Lu, Q. Learning roi transformer for oriented object detection in aerial images. IEEE CVPR, 2019, pp. 2849–2858.
22. Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Sun, X.; Fu, K. Scrdet: Towards more robust detection for small, cluttered and rotated objects. IEEE ICCV, 2019, pp. 8232–8241.

23. Yang, X.; Liu, Q.; Yan, J.; Li, A.; Zhang, Z.; Yu, G. R3det: Refined single-stage detector with feature refinement for rotating object. *arXiv preprint arXiv:1908.05612* **2019**, *2*.
24. Li, L.; Zhou, Z.; Wang, B.; Miao, L.; Zong, H. A Novel CNN-Based Method for Accurate Ship Detection in HR Optical Remote Sensing Images via Rotated Bounding Box. *IEEE T-GRS* **2020**, *59*, 686–699.
25. Lei, J.; Luo, X.; Fang, L.; Wang, M.; Gu, Y. Region-enhanced convolutional neural network for object detection in remote sensing images. *IEEE T-GRS* **2020**, *58*, 5693–5702.
26. Yu, Y.; Yang, X.; Li, J.; Gao, X. A Cascade Rotated Anchor-Aided Detector for Ship Detection in Remote Sensing Images. *IEEE T-GRS* **2020**.
27. Cao, G.; Xie, X.; Yang, W.; Liao, Q.; Shi, G.; Wu, J. Feature-fused SSD: Fast detection for small objects. ICGIP 2017. International Society for Optics and Photonics, 2018, Vol. 10615, p. 106151E.
28. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single-shot refinement neural network for object detection. IEEE CVPR, 2018, pp. 4203–4212.
29. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. ECCV. Springer, 2014, pp. 184–199.
30. Kim, J.; Kwon Lee, J.; Mu Lee, K. Accurate image super-resolution using very deep convolutional networks. IEEE CVPR, 2016, pp. 1646–1654.
31. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. CVPR, 2018, pp. 2472–2481.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. IEEE CVPR, 2016, pp. 770–778.
33. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. IEEE CVPR, 2017, pp. 4700–4708.
34. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale residual network for image super-resolution. ECCV, 2018, pp. 517–532.
35. Wang, L.; Dong, X.; Wang, Y.; Ying, X.; Lin, Z.; An, W.; Guo, Y. Exploring Sparsity in Image Super-Resolution for Efficient Inference. IEEE CVPR, 2021, pp. 4917–4926.
36. Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; Guo, Y. Unsupervised Degradation Representation Learning for Blind Super-Resolution. IEEE CVPR, 2021, pp. 10581–10590.
37. Wang, Y.; Wang, L.; Yang, J.; An, W.; Yu, J.; Guo, Y. Spatial-Angular Interaction for Light Field Image Super-Resolution. ECCV, 2020, pp. 290–308.
38. Wang, Y.; Yang, J.; Wang, L.; Ying, X.; Wu, T.; An, W.; Guo, Y. Light field image super-resolution using deformable convolution. *IEEE T-IP* **2021**, *30*, 1057–1071.
39. Wang, L.; Guo, Y.; Wang, Y.; Liang, Z.; Lin, Z.; Yang, J.; An, W. Parallax attention for unsupervised stereo correspondence learning. *IEEE T-PAMI* **2020**.
40. Ying, X.; Wang, L.; Wang, Y.; Sheng, W.; An, W.; Guo, Y. Deformable 3D convolution for video super-resolution. *IEEE SPL* **2020**, *27*, 1500–1504.
41. Everingham, M.; Winn, J. The pascal visual object classes challenge 2012 (VOC2012) development kit. *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep* **2011**, *8*.
42. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* **2014**.
43. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. IEEE CVPR, 2019, pp. 5693–5703.
44. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. IEEE CVPR, 2017, pp. 1492–1500.
45. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. IEEE ICCV, 2017, pp. 618–626.
46. Chen, K.; Pang, J.; Wang, J.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Shi, J.; Ouyang, W.; others. Hybrid task cascade for instance segmentation. IEEE CVPR, 2019, pp. 4974–4983.
47. Qiao, S.; Chen, L.C.; Yuille, A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. IEEE CVPR, 2021, pp. 10213–10224.
48. Yang, Z.; Liu, S.; Hu, H.; Wang, L.; Lin, S. Reppoints: Point set representation for object detection. IEEE ICCV, 2019, pp. 9657–9666.
49. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *arXiv preprint arXiv:2006.04388* **2020**.