remote sensing

MDPI

Article

A Group-Wise Feature Enhancement-and-Fusion Network with Dual-Polarization Feature Enrichment for SAR Ship Detection

Xiaowo Xu [1], Xiaoling Zhang [1], Zikang Shao [1], Jun Shi [1], Shunjun Wei [1], Tianwen Zhang [1] and Tianjiao Zeng [2,*]

[1] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2] School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China
[*] Correspondence: tzeng@uestc.edu.cn

Abstract: Ship detection in synthetic aperture radar (SAR) images is a significant and challenging task. However, most existing deep learning-based SAR ship detection approaches are confined to single-polarization SAR images and fail to leverage dual-polarization characteristics, which increases the difficulty of further improving the detection performance. One problem that requires a solution is how to make full use of the dual-polarization characteristics and how to excavate polarization features using the ship detection network. To tackle the problem, we propose a group-wise feature enhancement-and-fusion network with dual-polarization feature enrichment (GWFEF-Net) for better dual-polarization SAR ship detection. GWFEF-Net offers four contributions: (1) dual-polarization feature enrichment (DFE) for enriching the feature library and suppressing clutter interferences to facilitate feature extraction; (2) group-wise feature enhancement (GFE) for enhancing each polarization semantic feature to highlight each polarization feature region; (3) group-wise feature fusion (GFF) for fusing multi-scale polarization features to realize polarization features' group-wise information interaction; (4) hybrid pooling channel attention (HPCA) for channel modeling to balance each polarization feature's contribution. We conduct sufficient ablation studies to verify the effectiveness of each contribution. Extensive experiments on the Sentinel-1 dual-polarization SAR ship dataset demonstrate the superior performance of GWFEF-Net, with 94.18% in average precision (AP), compared with the other ten competitive methods. Specifically, GWFEF-Net can yield a 2.51% AP improvement compared with the second-best method.

1. Introduction

Synthetic aperture radar (SAR) plays an essential role in the remote sensing (RS) field. Its all-day and all-weather capabilities make it widely applied in both military and civil fields [1–4]. Specifically, ship detection in SAR images attracts increased attention due to its application in marine monitoring, shipping management, shipwreck rescue and illegal vessel control [5–8]. Thus, it is of great significance to obtain accurate ship detection results.

Recently, with the great breakthrough of deep learning (DL) in the computer vision (CV) field, ship detection in SAR images based on convolutional neural networks (CNNs) has attracted an increasing amount of attention. For instance, Mao et al. [9] proposed a lightweight network with an efficient low-cost regression subnetwork for SAR ship detection. Dai et al. [10] proposed a fusion feature extractor network and a refined detection network for the problem of multi-scale ship detection in complex background. Zhang et al. [11] presented a lightweight network for SAR ship detection, where the depth-wise separable convolution was adopted for lightening the model and three other

mechanisms were proposed to compensate for the accuracy. Zhao et al. [12] proposed a two-stage SAR ship detector with a receptive field block (RFB) and a convolutional block attention module (CBAM) for the multi-scale ship detection problem. Pan et al. [13] used a multi-stage RBox detector for arbitrary-oriented ship detection in SAR images. Fu et al. [14] offered an anchor-free CNN composed of a feature-balanced pyramid and a feature refinement network to tackle the multi-scale SAR ship detection problem. Zhang et al. [15] proposed a lightweight SAR ship detector, where a feature fusion module, a feature enhancement module and a scale share feature pyramid module were adopted to guarantee its detection performance. Zhang et al. [16] constructed a hyper-light deep learning network to reach high-accuracy and high-speed ship detection in which five contributions are offered, i.e., a multi-receptive-field module, a dilated convolution module, a channel and spatial attention module, a feature fusion module and a feature pyramid module. Han et al. [17] explored the training of a ship detector from scratch, and they established a CNN-based SAR ship detection model with a multi-size convolution module and a feature-reused module to verify the methodology's effectiveness. Geng et al. [18] proposed a ship detection method where a traditional island filter and a threshold segmentation method were integrated into a CNN model. The above studies have presented promising results in SAR ship detection. However, they all ignore the rich information in the dual-polarization SAR features (i.e., VV polarization features, VH polarization features and polarization coherence features) that have great potential to help achieve better SAR ship detection performance.

There are a few CNN-based researchers focusing on ship detection in multi-polarization SAR images [19–22]. Fan et al. [19] offered a semantic segmentation method for complex scene ship detection. However, in their compact polarimetric (CP) SAR images, only two types of polarization features are used, without considering their polarization coherence features, so their input data do not contain enough SAR polarization information. Jin et al. [20] proposed a pixel-level detector for small-scale ship detection and verified the effectiveness of the network on PolSAR images. Fan et al. [21] aimed to solve the multi-scale ship detection problem and carried out experiments on PolSAR images. They first established a deep convolutional neural network (DCNN)-based sea–coast–ship classifier for ship region extraction, and then proposed a modified Faster-RCNN for ship detection. Hu et al. [22] constructed pseudo-color SAR images composed of rich dual-polarization features and proposed a weakly supervised method for ship detection. However, these works [20–22] all neglect to enhance and fuse different polarization characteristics but directly feed them into the network without any distinguished treatment, which fails to fully mine the respective characteristics from different channels. All in all, the above pixel-level ship detection studies [19,20] and object-level ship detection studies [21,22] all monotonously utilize polarimetric SAR images to verify their respective tasks, without fully excavating polarization SAR features and considering such information when optimizing their network structures.

To tackle the above problems, in this paper, we propose a novel group-wise feature enhancement-and-fusion network with dual-polarization feature enrichment (GWFEF-Net) for better dual-polarization SAR ship detection. We introduce four contributions to guarantee the performance of GWFEF-Net, i.e., (1) a dual-polarization feature enrichment (DFE) proposed for enriching the dual-polarization feature library and suppressing clutter interferences to facilitate the subsequent feature extraction, (2) a group-wise feature enhancement (GFE) designed for autonomously enhancing the various polarization semantic features to highlight various polarization regions, (3) a group-wise feature fusion (GFF) designed for obtaining fused multi-scale polarization features to realize polarization features' group-wise information interaction, (4) a hybrid pooling channel attention (HPCA) proposed for channel modeling to equalize each polarization feature's contribution. We also conduct sufficient ablation experiments to verify the effectiveness of each contribution. Finally, extensive experimental results on the Sentinel-1 dual-polarization SAR ship dataset demonstrate the superior dual-polarization SAR ship

detection performance of GWFEF-Net, with 94.18% in average precision (AP), compared with the other ten competitive methods. Moreover, it can offer a 4% AP improvement over the baseline Faster R-CNN and a 2.51% AP improvement compared with the second-best method.

Our main contributions are as follows:

1. We design a novel two-stage deep learning network named "GWFEF-Net" for better dual-polarization SAR ship detection.
2. To achieve the excellent detection performance of GWFEF-Net, we (1) propose DFE to facilitate subsequent feature extraction; (2) design GFE to highlight each polarization semantic feature region; (3) design GFF to realize polarization features' group-wise information interaction; and (4) propose HPCA to balance each polarization feature's contribution.
3. GWFEF-Net achieves state-of-the-art detection accuracy with AP of up to 94.18% on the Sentinel-1 dual-polarization SAR ship dataset, compared with the other ten competitive methods.

The remaining materials are arranged as follows. Section 2 introduces the materials and methods. Section 3 describes the experimental details. Section 4 shows the quantitative and qualitative experimental results. Section 5 presents ablation studies on four contributions. Section 6 discusses the whole framework. Finally, Section 7 provides the conclusion of the work.

2. Methodology

GWFEF-Net is established based on the mainstream two-stage detector, i.e., Faster R-CNN [23]. Generally speaking, two-stage detectors have superior accuracy performance over one-stage ones [24], so we choose the former as our baseline. Figure 1 shows the network structure of GWFEF-Net. The raw Faster R-CNN contains a backbone network, a region proposal network (RPN) and a detection subnetwork [25]. DFE is proposed as a preprocessing technology. The GFE, GFF and HPCA are inserted into the detection subnetwork for better polarization feature enhancement and fusion.
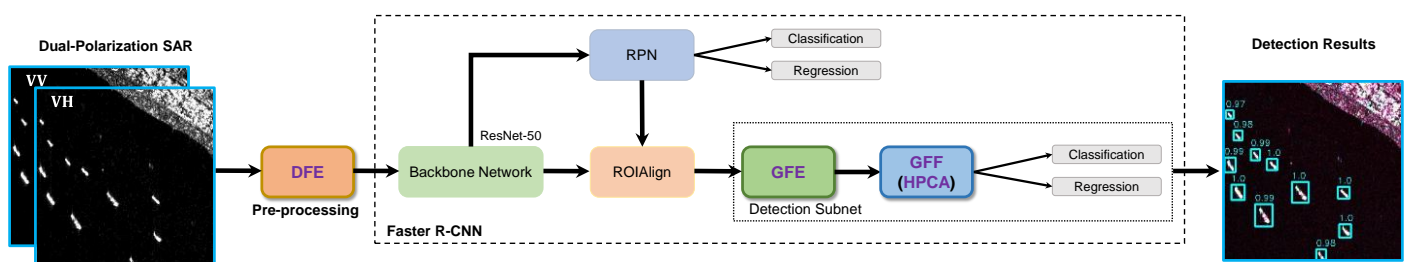
Figure 1. The network structure of GWFEF-Net. DFE denotes the dual-polarization feature enrichment, which is treated as a pre-processing tool. GFE denotes the group-wise feature enhancement. GFF denotes the group-wise feature fusion. HPCA denotes the hybrid pooling channel attention, which is injected into GFF. The mainstream ResNet50 is used as our backbone network. For convenience, the detection results are displayed in the form of pseudo-color images.

Firstly, GWFEF-Net preprocesses the input dual-polarization SAR images by the proposed DFE to enrich the dual-polarization features. The details will be introduced in Section 2.1. Then, GWFEF-Net uses a backbone network to extract ship features from dual-polarization SAR images. Without losing generality, we use the mainstream ResNet50 [26] as our backbone network. Then, an RPN is used for the extraction of regions of interest (i.e., regions containing ships). Afterward, a RoIAlign [27] layer is used to map the generated proposed regions by the RPN to the feature maps of the backbone network for subsequent classification and regression in the detection subnetwork. Finally, feature maps generated by the RoIAlign are input into a detection subnetwork for the final prediction, and the final ship detection results are obtained.

Note that in the detection subnetwork, the proposed GFE, GFF and HPCA are inserted, which are used for better polarization feature enhancement and fusion. Specifically, for better polarization feature enhancement, we insert the GFE into the detection subnetwork. It is used to highlight each polarization semantic feature region by means of enhancing each polarization semantic feature, which will be introduced in detail in Section 2.2. For better polarization feature fusion, we then insert the GFF after the GFE. It is used to realize polarization features' multi-scale information interaction by means of fusing multi-scale polarization features, whose details will be introduced in Section 2.3. In addition, HPCA is inserted in the GFF to balance each polarization feature's contribution by channel modeling. It will be introduced in detail in Section 2.4.

The motivation for the core idea of GWFEF-Net can be summarized as follows.

(1) It has already been demonstrated that dual-polarization features play an important role in improving the detection accuracy of traditional SAR ship detection methods. Inspired by this, employing such features is also likely to improve the performance of DL-based methods in SAR ship detection tasks. However, most of the CNN-based SAR ship detection methods only utilize single-polarization features as the input of networks, ignoring the dual-polarization characteristics with rich structural information of ships. Though a few researchers have tried to utilize polarimetric SAR images to verify their respective ship detection tasks, their networks are not especially designed for the polarimetric characteristics and no special treatments, such as enhancement and fusion, have been applied for different polarization features. Hence, it is of great significance to study how to fully excavate polarization SAR features in a CNN-based network.

(2) To address the above problems, we propose a group-wise feature enhancement-and-fusion network with dual-polarization feature enrichment (GWFEF-Net) to improve the SAR ship detection performance. Specifically, four contributions (i.e., DFE, GFE, GFF and HPCA) are proposed in GWFEF-Net. DFE enables the enrichment of the feature library with more abundant ship polarization information to facilitate the subsequent feature extraction; GFE adopts group-wise features to learn and enhance the semantic representation of various polarization features so as to highlight various target ship regions; GFF performs information interaction between polarization features and multi-scale ship features, which is helpful to extract more abundant information of polarization features and multi-scale ships; HPCA is designed for channel modeling to further balance the contribution of each polarization feature.

Next, we will introduce the DFE, GFE, GFF and HPCA in detail in the following subsections.

2.1. Dual-Polarization Feature Enrichment (DFE)

As for the Sentinel-1 satellite product, it contains two polarization modes of VV polarization and VH polarization. However, the coherence polarization feature is also useful for identifying ships [28]. Inspired by the work [28], we introduce the coherence polarization feature to characterize ship feature relationships in different polarization channels, which can enrich the dual-polarization feature library and suppress clutter interference to further improve the follow-up detection performance. For brevity, we call the above process the dual-polarization feature enrichment (DFE).

We will describe the feature types mentioned above in detail.

(1) VV feature: In the VV polarization image, a ship often has strong backscattering values in the sea background, which means that the outline and texture of the ship are relatively clear [29,30]. Thus, VV features are widely utilized for SAR ship detection.

(2) VH feature: In the VH polarization image, a ship often has lower backscattering values in the sea background, and the sea clutter is lower than the instrument noise level. However, the signal-to-noise-ratio (SNR) of VH is higher than that of VV [31,32]. Thus, VH features are also applicable to SAR ship detection.

(3) $C_{\text{VV-VH}}$ feature: Considering the dual-polarization characteristic in the Sentinel-1 satellite product, a polarization covariance matrix C2 is obtained by the following formula [33]:

$$C_2 = \begin{bmatrix} |S_{vh}|^2 & |S_{vh}S_{vv}{}^*| \\ |S_{vv}S_{vh}{}^*| & |S_{vv}|^2 \end{bmatrix} \tag{1}$$

where S_{vh} denotes the VH polarization complex data, S_{vv} denotes the VV polarization complex data, $|\cdot|$ denotes the function of the absolute value, * denotes the conjugate operation. The polarization coherence feature is defined by

$$C_{\text{VV-VH}} = |S_{vh}S_{vv}{}^*| = |S_{vv}S_{vh}{}^*| \tag{2}$$

$C_{\text{VV-VH}}$ can effectively represent the dual-polarization channel correlation. In the dual-polarization image, the reflection symmetry effect of the sea scene is significant. In other words, the $C_{\text{VV-VH}}$ polarization value can be very low in an image with sea clutter, because the image meets the reflection symmetry; conversely, it can be very high in an image with artificial objects, such as a ship, because the image does not meet the reflection symmetry. In short, the ship-to-clutter-ratio of $C_{\text{VV-VH}}$ features is higher than that of the other two features, that is, the clutter interference can be suppressed with $C_{\text{VV-VH}}$ features. Thus, $C_{\text{VV-VH}}$ features have the potential to improve SAR ship detection.

Briefly speaking, first, we enrich the existing VV features and VH features according to formula (2), and we can obtain the generated $C_{\text{VV-VH}}$ features. Then, the amplitude values of VV, VH and $C_{\text{VV-VH}}$ polarization complex data are integrated into the R, G, B channels of the pseudo-color images, so we can obtain resulting images with less sidelobe and clutter interference. DFE can be described as in Figure 2.

(**a**)

(**b**)
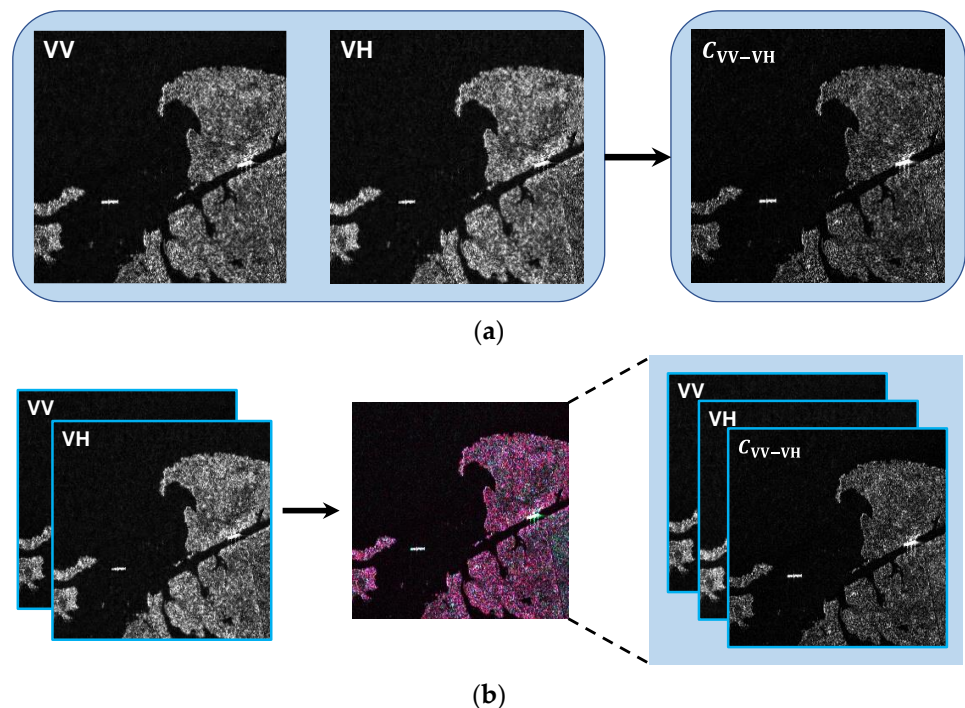
Figure 2. (a) The internal implementation details of the DFE, where VV, VH and $C_{\text{VV-VH}}$, respectively, represent the amplitude value of the VV feature, VH feature and dual-polarization feature. (**b**) The input and output images of DFE, where the left of "→" denotes the input image before DFE, and the right of "→" denotes the output image after DFE. Note that the resulted pseudo-color image has VV, VH and CVV-VH channels. Obviously, the pseudo-color image can reduce sidelobe and noise interference and clarify the skeleton of landing ships.

To summarize, DFE introduces the $C_{\text{VV-VH}}$ feature into the feature library characterizing feature relations in different polarization channels. Moreover, it provides richer polarization information and suppresses clutter interferences, therefore facilitating subsequent feature extraction. In our subsequent implementation, in order to make full use of the polarization features provided by DFE, we insert the proposed GFE, GFF and HPCA into the detection subnetwork for better dual-polarization SAR ship detection.

2.2. Group-Wise Feature Enhancement (GFE)

Group-wise features are widely used in the CV community and can adaptively learn semantic representations of different interested entities. Thus far, a large number of scholars from the SAR ship detection field have devoted themselves to researching single-channel polarization SAR images [9–17]. However, they ignore the exploration of multi-polarization characteristics and further adoption of group-wise features to learn the semantic representation of various polarization features. Thus, different from the former ship detection networks, which are designed for single-polarization SAR images, considering the polarization semantic feature differences of different channels, we adopt group-wise features to autonomously enhance the learned semantic representations of various polarization features.

Specifically, we attempt to conduct feature grouping enhancement along the channel dimension, which is inspired by Sabour et al. [34] and Li et al. [35]. Thus, we propose the GFE to obtain enhanced semantic information for each SAR polarization feature. First, considering that the output data preprocessed by the DFE is composed of three polarization channels, i.e., VV feature, VH feature and $C_{\text{VV-VH}}$ feature, we use a group convolution to enrich the polarization features (i.e., the number of feature channels is tripled). Then, we group-wise enhance the spatial information of the three polarization features in the channel dimension. Finally, we can obtain enhanced semantic features of each SAR polarization feature.

Figure 3 shows the detailed structure of the GFE. In the detection subnetwork, first, the feature map $F_{\text{in}} \in R^{W \times H \times C}$ is input, where W, H and C represent the height, width and channel of the input feature map, respectively. In our implementation, W and H are equal to 7, and C is equal to 256. Its channel is expanded three times to obtain richer SAR polarization features (i.e., VV feature, VH feature and $C_{\text{VV-VH}}$ polarization). Specifically, we obtain three group feature maps along the channel dimension through the following operation:

$$F_i = \text{GC}(F_{in}), \qquad i = 1,2,3 \tag{3}$$

where GC(·) is the group convolution operation and F_i is the i-th group feature maps. Note that i = 1, 2, 3, which keeps the same as the number of polarization features (i.e., VV polarization, VH polarization and $C_{\text{VV-VH}}$ feature).

Because the noise distribution in each polarization image is inconsistent [35], it is necessary to enhance the polarization feature in the group space for highlighting each polarization semantic feature region. Without loss of generality, we first examine a certain group feature map, namely $F_1 = \{f_1, …, f_m\}$, m = 256. The global pooling operation (i.e., global average pooling and global max pooling) is conducted to extract the global semantic feature $\mathbf{g}$ of the group polarization feature map F_1. The operation can be described by

$$\mathbf{g} = \text{GP}(F_1) = \frac{1}{m} \sum_{i=1}^{m} f_i \tag{4}$$

Then, the corresponding importance coefficient c_i is obtained by conducting a dot product between the global semantic feature $\mathbf{g}$ and local feature f_i. The operation formula is defined by

$$c_i = \mathbf{g} \cdot f_i \tag{5}$$

Subsequently, in order to reduce the coefficient deviation caused by the various ship samples (i.e., inshore ones and offshore ones), we conduct the following normalization operations:
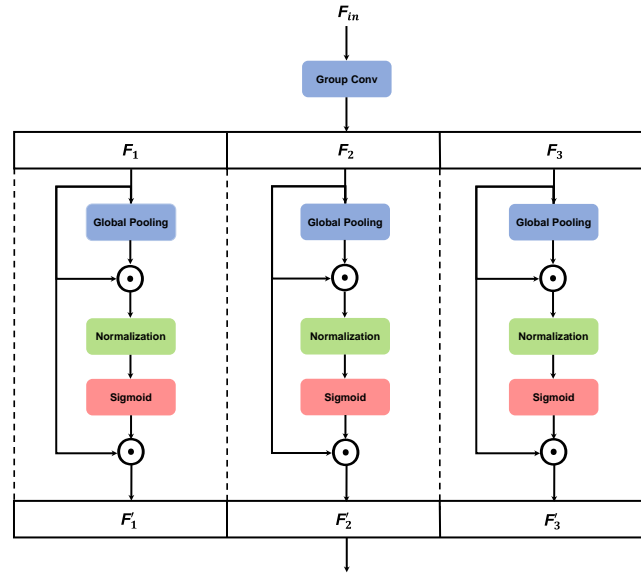
Figure 3. Illustration of the proposed GFE. It expands the original features into three groups of features, and then processes the polarization features of each group in parallel. Finally, it obtains the enhanced polarization semantic feature representation.

$$\mu_c = \frac{1}{m} \sum_{i}^{m} c_i \tag{6}$$

$$\sigma_c^2 = \frac{1}{m} \sum_{i}^{m} (c_i - \mu_c)^2 \tag{7}$$

$$\hat{c}_i = \frac{c_i - \mu_c}{\sigma_c + \varepsilon} \tag{8}$$

where ε (i.e., 1×10^{-5}) is a constant added for numerical stability, which follows the work [35].

Finally, to obtain the final enhanced polarization feature, the original feature f_i is weighted by the corresponding importance coefficient c_i via a sigmoid function $\sigma(\cdot)$:

$$\boldsymbol{f}_i' = \boldsymbol{f}_i \cdot \sigma(c_i) \tag{9}$$

Thus, we can obtain the polarization feature group $\boldsymbol{F}_1'$, i.e., $\boldsymbol{F}_1' = \{\boldsymbol{f}_1', \dots, \boldsymbol{f}_m'\}$, m = 256. In this way, we can obtain all three resulted polarization feature groups $\boldsymbol{F}'$, i.e., $\boldsymbol{F}' = \{\boldsymbol{F}_1', \boldsymbol{F}_2', \boldsymbol{F}_3'\}$. The enhanced features can highlight each polarization semantic feature region so as to enhance the meaningful ship target area and better focus on the interested ship targets. To summarize, the network will detect more meaningful ships and suppress useless clutter interferences with the help of the GFE, which will greatly improve the detection performance of GWFEF-Net.

### 2.3. Group-Wise Feature Fusion (GFF)

It is important for enhancing the represented capability of object detection CNNs to obtain the fused features from different scales. Most existing methods in the SAR ship detection field attempt to fuse the multi-scale features in the layer-wise dimension [36,37]. The former work [36] aims to fuse the high-layer feature maps and low-layer feature maps to achieve enhanced multi-scale semantic features. The latter [37] aims to transmit the

location information of the shallow layer to the deep layer to achieve enhanced multi-scale spatial features. Different from the above works that fuse features from different resolutions, considering various polarization semantic representations of different channels, we aim to obtain the fused multi-scale polarization features in the channel group-wise level. In other words, we tend to achieve multi-scale polarization feature interaction at a channel group-wise dimension besides other existing dimensions, i.e., depth [38], width and cardinality [39]. In addition, our inspiration is derived from the works of Gao et al. [40], Lin et al. [41] and Ezegedy et al. [42], which are recommended to the readers.

We have obtained each enhanced polarization semantic feature group from Section 2.2; to further utilize the advantages of all polarization features, the information interaction between different polarization features should be considered. Thus, a group-wise features fusion (GFF) is proposed. Firstly, this can increase the range of receptive fields of each polarization feature group. Secondly, it can fuse the different polarization feature groups. The above factors all guarantee the extraction capability of multi-scale polarization features and the excellent information interaction capability of the GFF.

Figure 4 shows the detailed structure of the GFF. Note that, after the GFE, we obtain three feature map groups, denoted by F'_i, where $i \in \{1, 2, 3\}$. Each feature group F'_i has the same height, width size and channel amount compared with the original input feature map $F_{in} \in R^{W \times H \times C}$ in Section 2.2. In addition, in order to balance each polarization feature's contribution, a HPCA is inserted in the GFF, which will be described in detail in Section 2.4.
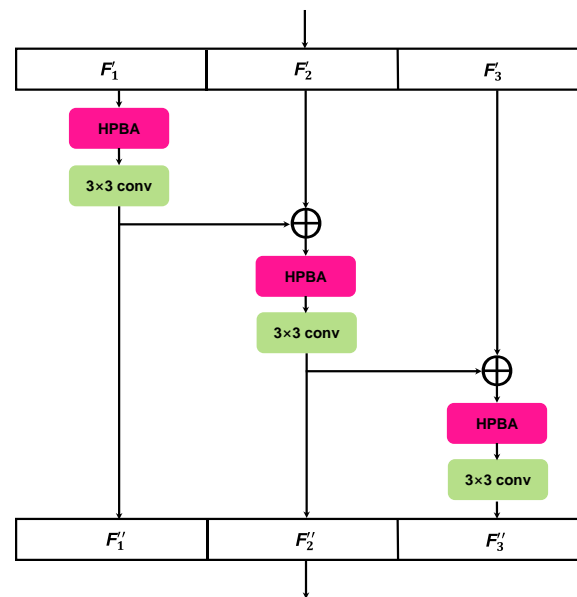
Figure 4. The detailed structure of the GFF. It increases the range of receptive fields of each polarization feature group by 3 × 3 convolution and then fuses the different polarization feature groups. ⊕ denotes the channel-wise summation and 3 × 3 conv denotes the 3 × 3 convolution operation. In addition, in order to balance each polarization feature's contribution, a HPCA is inserted in the GFF, which will be described in detail in Section 2.4.

We conduct a 3 × 3 convolution operation for each feature map group and feed the results into the next group. In this way, we can obtain the fused polarization feature group with a larger range of receptive fields.

In short, the above can be described by

$$F''_i = \begin{cases} Conv_{3 \times 3}(F'_i), i = 1 \\ Conv_{3 \times 3}(F'_i + F''_{i-1}), i = 2,3 \end{cases} \tag{10}$$

Note that each 3 × 3 convolution operation *Conv₃ₓ₃*() could receive polarization information from multi-group features {F_j'', $j \leq i$}. In addition, each instance of conducting a 3 × 3 convolution operation *Conv₃ₓ₃*() on F_j' can provide an output with a larger receptive field than F_j'.

In this way, we can obtain all fused multi-scale polarization features F'', i.e., $F'' = \{F_1'', F_2'', F_3''\}$. In short, GFF can offer excellent information interaction from polarization features and multi-scale ship features, which is helpful to extract more abundant information about ships. Thus, GWFEF-Net can detect more multi-scale ships with the help of the GFE and the final detection performance will be improved.

2.4. Hybrid Pooling Channel Attention (HPCA)

Attention mechanisms have been widely applied in the CV community and can enhance valuable features and improve the expression ability of a CNN through spatial or channel-wise information. Considering the polarization semantic feature differences of different channels, we attempt to balance each polarization feature's contribution at the channel-wise dimension to achieve a better polarization feature fusion. Thus, in our implementation, we choose to use the channel attention mechanism to better obtain reasonable channel modeling during the feature fusion described in Section 2.3. There have been a few attempts [43,44] to incorporate channel attention processing into CNNs to obtain the importance of each channel. However, the above channel attention models are all extracted through the global average pooling operation, which could be suboptimal [45], so we utilize both global average pooling and global max pooling operations to achieve channel attention.

Thus, we propose a hybrid pooling channel attention (HPCA), which is inserted in the GFF of Section 2.3 to obtain the channel importance for balancing each polarization feature's contribution. Figure 5 illustrates the detailed implementation of the HPCA. Then, we will further describe the principle of the HPCA.
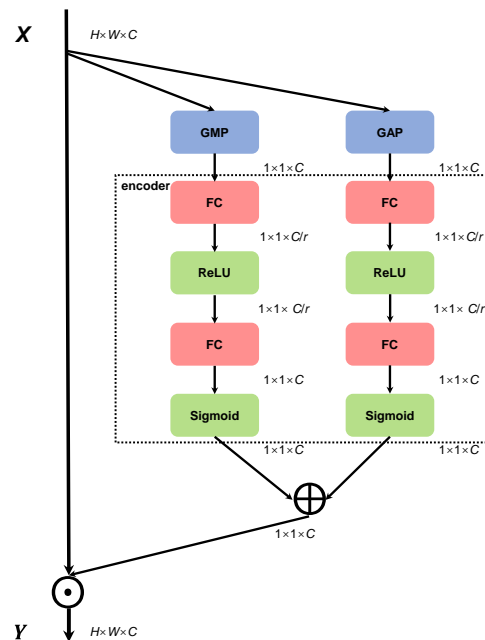
Figure 5. The principle of the HPCA. GAP denotes the global average pooling, GMP denotes the global max pooling, FC denotes the fully connected layer, ReLU denotes the Rectified Linear Unit activation layer, Sigmoid denotes the Sigmoid activation layer, ⊕ denotes the channel-wise summation, and ⊙ denotes the channel-wise multiplication. In our implementation, the reduction ratio *r* is set to 16, which is the same as in the reference [43].

Different from the SE module [43], there are two parallel branches in the HPCA. Specifically, given the input feature map $X \in \mathbb{R}^{W \times H \times C}$, in the first branch, we first conduct the global average pooling operation of each channel to obtain the feature map with the global receptive field; then, two full connection layers with excitation functions are used to predict the channel importance. In the second branch, we first conduct the global max pooling operation of each channel to obtain the feature map with the global receptive field; then, two full connection layers with excitation functions are used to predict the channel importance weight. Next, we add the importance weight of the two branches to obtain the final channel importance coefficient. Finally, the importance coefficient is applied to the corresponding channels to construct the correlation between channels.

The above steps can be described as follows:

$$Y = X \odot W \tag{11}$$

where X denotes the balanced polarization feature map, Y denotes the input feature map, $\odot$ denotes the channel-wise multiplication, and W denotes the channel importance coefficient, i.e.,

$$W = f_{encode}(GAP(X)) \oplus f_{encode}(GMP(X)) \tag{12}$$

where GAP denotes the global average pooling, GMP denotes the global max pooling, $\oplus$ denotes the channel-wise summation, and f_{encode} denotes the channel encoder that can assist in the non-linearity and generality of the model, where two full connection layers with non-linearity excitation functions are adopted.

Finally, we can obtain finer channel information, in which each polarization feature's contribution is more balanced. By inserting HPCA into each polarization feature group, the network can learn the contribution of each polarization feature adaptively in the process of group-wise feature fusion. Therefore, HPCA can equalize the contribution of each group polarization feature, so as to improve the expression ability of the network.

3. Experiments

We use a Personal Computer (PC) with a GPU model of NVIDIA RTX3090, CPU model of i7-10700K and memory size of 32G for the whole experiment. We adopt Pytorch [46] and MMDetection [47] based on the Python 3.8 language as the deep learning framework. We also use CUDA11.1 in our experiments to call the GPU for training acceleration.

3.1. Dataset

The dual-polarimetric SAR ship detection dataset (DSSDD) [22] is used in our work. Moreover, note that DSSDD, with dual-polarization images, is a unique public dataset meeting our research demands. Thus, we utilize DSSDD to verify the effectiveness of GWFEF-Net in dual-polarization SAR ship images. Table 1 shows the details of DSSDD.

Table 1. The details of DSSDD.

Key	Value
Location	Shanghai, the Suez Canal, etc.
Sensors	Sentinel-1
Polarization	VV, VH
Sensor Mode	IW
Incident Angle (°)	29.1~46.0
Resolution (m)	2.3 × 14.0
Swath (km)	250
Number of Images	50

From Table 1, there are 50 large images from different places (Shanghai, the Suez Canal, etc.). Considering the limited computing power of the GPU, we crop the original image into sub-images of 256 pixels × 256 pixels with a 50-pixel overlap, which is in line

with Hu et al. [22]. After the above operation, we obtain 1236 sub-images with 3540 ship targets. Finally, the ratio of the training set to the test set is 7:3, which is the same as in Hu et al. [22].

3.2. Experimental Details

We train GWFEF-Net for 12 epochs using the stochastic gradient descent (SGD) optimizer [48]. The network input size is 256 pixels × 256 pixels. We set the learning rate as 0.008, the momentum as 0.9 and the weight decay as 0.0001. The learning rate is reduced by 10 times at each 8th and 11th epoch. Moreover, due to the limited GPU capability, we set the batch size as 4. To accelerate convergence, we load the pre-training weights of ResNet-50 on ImageNet [49]. Other hyperparameters not mentioned are consistent with Faster R-CNN. During the inference, the Intersection Over Union (IOU) threshold of non-maximum suppression (NMS) [50] is set as 0.30.

3.3. Evaluation Indices

Precision (*P*) is defined by

$$P = \frac{\#TP}{\#TP + \#FP} \times 100\% \tag{13}$$

where # denotes the amount, *TP* denotes the true positives (i.e., the prediction and label are both ships), and *FP* denotes the false positives (i.e., the prediction is a ship but the label is the background).

Recall (*R*) is defined by

$$R = \frac{\#TP}{\#TP + \#FN} \times 100\% \tag{14}$$

where *FN* denotes the false negatives (i.e., the prediction is the background but the label is a ship).

*F*1 can effectively reflect the balance between precision and recall, and is defined by

$$F1 = 2 \times \frac{R \times P}{R + P} \tag{15}$$

The average precision (*AP*) is defined by

$$AP = \int_0^1 P(R) \cdot dR \tag{16}$$

where *P* denotes the precision, and *R* denotes the recall. As *AP* can more comprehensively reflect the detection performance of a detector under different confidence thresholds, we take the *AP* as the core evaluation index in the paper.

4. Results

4.1. Quantitative Results

Table 2 shows the quantitative results of GWFEF-Net on the DSSDD. In Table 2, we present our quantitative results by gradually adding each contribution to the baseline Faster R-CNN. In addition, we also conduct adequate ablation studies to verify the gain in GWFEF-Net of each contribution, which will be introduced in Section 5.

Table 2. The quantitative results of GWFEF-Net on the DSSDD. GT: the number of ground truths; TP: the number of true positives, where the higher the better; FP: the number of false positives, where the lower the better; FN: the number of false negatives, where the lower the better; P: precision, where the higher the better; R: recall, where the higher the better; F1: F1-score, where the higher the better; AP: average precision, the core evaluation index, where the higher the better. The best model is marked in bold.

DFE [1]	GFE [2]	GFF [3]	HPCA [4]	GT	TP	FP	FN	P (%)	R (%)	F1 (%)	AP (%)
-	-	-	-	1203	1102	137	101	88.94	91.60	90.25	90.18
✓				1203	1112	124	91	89.97	92.44	91.18	91.10
✓	✓			1203	1123	108	80	91.23	93.35	92.28	92.30
✓	✓	✓		1203	1130	106	73	91.42	93.93	92.66	93.04
✓	✓	✓	✓	1203	1142	94	61	92.39	94.93	**93.64**	**94.18**

DFE [1] denotes the dual-polarization feature enrichment. GFE [2] denotes the group-wise feature enhancement. GFF [3] denotes the group-wise feature fusion. HPCA [4] denotes the hybrid pooling channel attention.

From Table 2, one can conclude the following.

1. The detection accuracy presents a gradual upward trend by gradually adding each contribution on the baseline Faster R-CNN. Apparently, GWFEF-Net offers a huge AP improvement (i.e., from initial 90.18% to final 94.18%), up to a 4% increment. Moreover, GWFEF-Net offers a perceptible F1 improvement (i.e., from initial 90.25% to final 93.64%), up to a ~3.4% increment. Other accuracy indexes have also been improved more or less, which fully proves the excellent detection performance of the proposed GWFEF-Net.

2. DFE increases both the AP index and the F1 index by ~0.9%. It can improve the recall rate (i.e., from 91.60% to 92.44% with recall) and the detection rate (i.e., from 88.94% to 89.97% with precision), i.e., it can provide more complete and more accurate detection results, owing to its capability of enriching the feature library and suppressing clutter interferences. As mentioned in Section 2.1, DFE can provide richer polarization information and meanwhile suppresses clutter interferences, therefore facilitating subsequent feature extraction. Thus, DFE will greatly improve the detection recall and detection rate of GWFEF-Net.

3. GFE increases the AP index by 1.2% and the F1 index by 1.1%. It is useful for detecting more actual ships (the number of true positives is increased by 11) and decreasing false alarms (the number of false positives is decreased by 16) owing to its capability of group-wise feature enhancement. As mentioned in Section 2.2, the enhanced features can highlight each polarization semantic feature region so as to enhance the meaningful ship target area. In this way, the network will detect more meaningful ships and suppress useless clutter interferences with the help of the GFE, which will greatly improve the detection performance of GWFEF-Net.

4. GFF increases the AP index by ~0.7% and the F1 index marginally. It is still useful because it can provide more true ships (the number of missed detections is decreased by 7) and suppress some false alarms (that is, two false alarms can be suppressed by the GFF), owing to its capability of group-wise feature fusion. As mentioned in Section 2.3, the GFF can offer excellent information interaction from polarization features and multi-scale ship features, which is helpful to extract richer information about ships. Finally, it can obtain a better detection recall capability and detect more multi-scale ships.

5. HPCA increases the AP index by ~1.1% and the F1 index by ~1.0%. It can improve the detection rate (i.e., from 91.42% to 92.39% with precision) and the recall rate (i.e., from 93.93% to 94.93% with recall), owing to its excellent capability of channel modeling. As mentioned in Section 2.4, by inserting HPCA into GWFEF-Net, we can obtain finer channel information, in which each polarization feature's contribution is more balanced. In other words, by inserting HPCA into each polarization feature group, the network can learn the contribution of each polarization feature adaptively

in the process of group-wise feature fusion. Therefore, HPCA can adaptively equalize the contribution of each polarization feature, so as to improve the final detection performance of the network.

6. In short, each contribution has different degrees of gain for the network. In addition, the detection accuracy presents a gradual upward trend by gradually adding each contribution. Finally, GWFEF-Net, equipped with the above contributions, can achieve superior polarimetric SAR ship detection performance.

Table 3 shows the performance comparisons of GWFEF-Net with the other ten competitive models. In Table 3, this paper mainly chooses the Faster R-CNN [23], Cascade R-CNN [51], Dynamic R-CNN [52], Double-Head R-CNN [53], MobileNetv2+SSD [54,55], FreeAnchor [56], ATSS [57], FCOS [58], Fan et al. [21] and Hu et al. [22] for comparison. They all load the pre-training model on the backbone network and are trained on the DSSDD, and their implementation details are consistent with the original paper. It is worth noting that we select the mainstream one-stage detectors (i.e., MobileNetv2+SSD, FreeAnchor, ATSS and FCOS), the mainstream two-stage detectors (i.e., Faster R-CNN, Cascade R-CNN, Dynamic R-CNN and Double-Head R-CNN) in the CV field and typical polarimetric SAR ship detectors (i.e., Fan et al. [21] and Hu et al. [22]) for comparison.

Table 3. The performance comparisons of GWFEF-Net with the other ten competitive models. The best and second-best model are, respectively, marked in bold and underlined.

| Method	GT	TP	FP	FN	P (%)	R (%)	F1 (%)	AP (%)
Faster R-CNN [23]	1203	1102	137	101	88.94	91.60	90.25	90.18
Cascade R-CNN [51]	1203	1102	94	101	92.14	91.60	91.87	90.66
Dynamic R-CNN [52]	1203	1106	93	97	92.24	91.94	92.09	91.02
Double-Head R-CNN [53]	1203	1116	100	87	91.78	92.77	92.27	<u>91.67</u>
MobileNetv2+SSD [54,55]	1203	1119	992	84	53.01	93.02	67.53	88.78
FreeAnchor [56]	1203	1122	433	81	72.15	93.27	81.36	90.72
ATSS [57]	1203	1021	64	182	94.10	84.87	89.25	83.62
FCOS [58]	1203	1006	137	197	88.01	83.62	85.76	80.99
Fan et al. [21]	1203	1092	168	111	86.67	90.77	88.67	89.71
Hu et al. [22]	1203	1126	98	77	91.99	93.60	92.79	91.54
GWFEF-Net	1203	1142	94	61	92.39	94.93	93.64	**94.18**

From Table 3, one can conclude the following.

1. GWFEF-Net achieves the best polarization SAR ship detection performance with the highest accuracy index (whether the AP index or the F1 index). Specifically, GWFEF-Net can achieve a 2.51% improvement in AP compared with the second-best method (i.e., Double-Head R-CNN). It is worth noting that Double-Head R-CNN generates far more missed detections than GWFEF-Net (i.e., from 87 to 61). Therefore, the recall rate of Double-Head R-CNN is lower than that of GWFEF-Net (i.e., from 92.77% to 94.93%). The above reveals the excellent polarization SAR ship detection performance of GWFEF-Net.

2. GWFEF-Net can achieve a 4% improvement in AP (i.e., from 90.18% to 94.18%), a ~3.4% improvement in F1 (i.e., from 90.25% to 93.64%), a ~3.3% improvement in R (i.e., from 91.60% to 94.93%) and a ~3.5% improvement in P (i.e., from 88.94% to 92.39%) compared with the baseline Faster R-CNN. All the above benefit from the four contributions mentioned before.

3. ATSS offers the highest P (i.e., 94.10%), but it has a rather low recall rate (i.e., its 84.87% R << GWFEF-Net's 94.93% R), which leads to a large number of missed detections (i.e., its 182 false negatives >> GWFEF-Net's 61 false negatives). In addition, its detection accuracy is also not ideal—that is, its overall accuracy indexes

are rather low (i.e., its 83.62% AP << GWFEF-Net's 94.18% AP and its 89.25% F1 < GWFEF-Net's 93.64% F1).

4. Due to the failure to mine the respective polarization characteristics of different channels, the accuracies of the typical polarimetric SAR ship detectors [21,22] are undesirable. In particular, Fan et al. [21] only achieve an inferior dual-polarization SAR ship detection result, i.e., its 111 false negatives > GWFEF-Net's 61 false negatives and its 168 false positives > GWFEF-Net's 94 false positives. Similarly, Hu et al. [22] show an observably higher number of false negatives while maintaining similar false positive numbers, i.e., its 77 false negatives > GWFEF-Net's 61 false negatives.

Figure 6 shows the precision–recall (P–R) curves of different models. As shown in Figure 6, compared with other curves, the GWFEF-Net curve is always at the top, which intuitively shows that GWFEF-Net achieves the best detection performance.
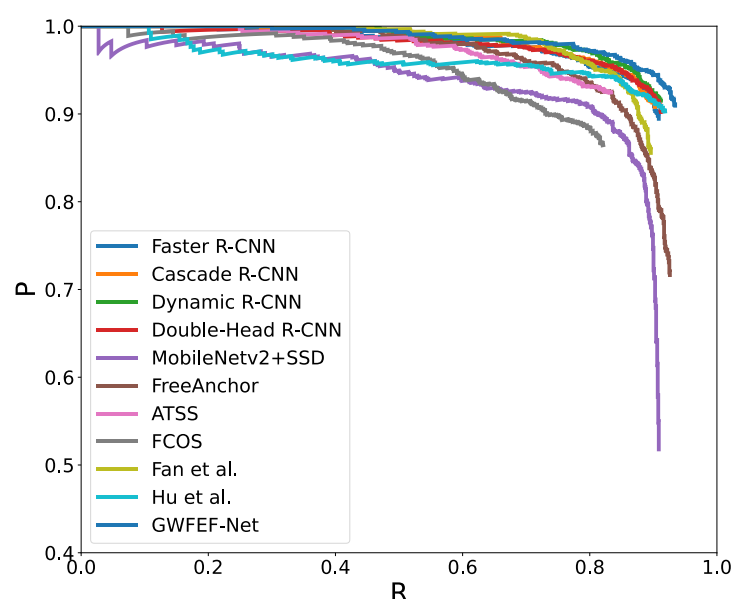
Figure 6. The precision–recall (P–R) curves of eleven state-of-the-art models.

In addition, to avoid the impact of data fluctuation, we also conduct another quantitative experiment. Table 4 shows the comparison of the quantitative evaluation indices with the other ten competitive models, where ten optimal results of each method are used to obtain the mean and standard deviation. From Table 4, compared with other methods, GWFEF-Net is still the best, which further demonstrates the superior detection performance of GWFEF-Net.

Table 4. The comparison of the mean and standard deviation with the other ten competitive models. The best and second-best model are, respectively, marked in bold and underlined.

Method	P (%)	R (%)	F1 (%)	AP (%)
Faster R-CNN [23]	89.55 ± 0.32	91.94 ± 0.15	90.73 ± 0.20	90.53 ± 0.18
Cascade R-CNN [51]	92.13 ± 1.99	91.90 ± 0.80	92.00 ± 0.82	90.91 ± 0.71
Dynamic R-CNN [52]	92.94 ± 0.45	91.95 ± 0.45	92.44 ± 0.24	91.07 ± 0.46
Double-Head R-CNN [53]	91.35 ± 0.76	92.88 ± 0.36	92.10 ± 0.42	<u>91.62 ± 0.38</u>
MobileNetv2+SSD [54,55]	52.86 ± 3.24	91.94 ± 0.62	67.07 ± 2.65	87.30 ± 0.74
FreeAnchor [56]	72.03 ± 1.60	93.26 ± 0.45	81.27 ± 0.98	90.80 ± 0.52
ATSS [57]	94.20 ± 1.49	84.01 ± 1.32	88.80 ± 0.46	82.83 ± 1.07
FCOS [58]	94.08 ± 2.80	82.33 ± 1.31	87.79 ± 1.50	80.85 ± 1.53
Fan et al. [21]	79.61 ± 6.15	91.70 ± 1.25	85.06 ± 3.31	90.09 ± 1.05

|				
Hu et al. [22]	92.09 ± 0.64	93.34 ± 0.36	92.71 ± 0.42	91.30 ± 0.60
GWFEF-Net	91.33 ± 0.55	95.01 ± 0.17	93.13 ± 0.26	**94.20 ± 0.19**

Moreover, we also provide both the training time and testing time of all methods to objectively judge the time efficiency of all models. Table 5 shows the comparison of the training and testing time with the other ten competitive models. From Tables 3 and 5, compared with the second-best model (i.e., Double-Head R-CNN), GWFEF-Net decreases the training time by ~626 s and testing time by ~7 s, but shows a ~2.5% improvement in AP. Thus, although GWFEF-Net is not optimal in terms of time efficiency, it still shows a high cost performance ratio.

Table 5. The comparison of training and testing time with the other ten competitive models.

| Method	Training Time (s)	Testing Time (s)
Faster R-CNN [23]	792.74	7.73
Cascade R-CNN [51]	1163.88	11.14
Dynamic R-CNN [52]	823.70	7.68
Double-Head R-CNN [53]	1891.99	21.19
MobileNetv2+SSD [54,55]	7852.19	4.90
FreeAnchor [56]	683.01	6.63
ATSS [57]	703.60	7.33
FCOS [58]	645.18	7.20
Fan et al. [21]	611.74	5.69
Hu et al. [22]	6614.76	5.14
GWFEF-Net	1265.81	14.16

4.2. Qualitative Results

Figure 7 shows the ground truths of different scenes and the qualitative detection results of different models (i.e., Faster R-CNN, Cascade R-CNN, Dynamic R-CNN, Double-Head R-CNN, MobileNetv2+SSD, FreeAnchor, ATSS, FCOS, Fan et al., Hu et al. and GWFEF-Net). Moreover, their score thresholds used for display are all set as 0.5, except for those of ATSS and FCOS, which are set as 0 due to their poor performance. Here, to sufficiently and intuitively demonstrate the superior dual-polarization ship detection performance of GWFEF-Net, we show the qualitative results by exhibiting some offshore and inshore scenes.
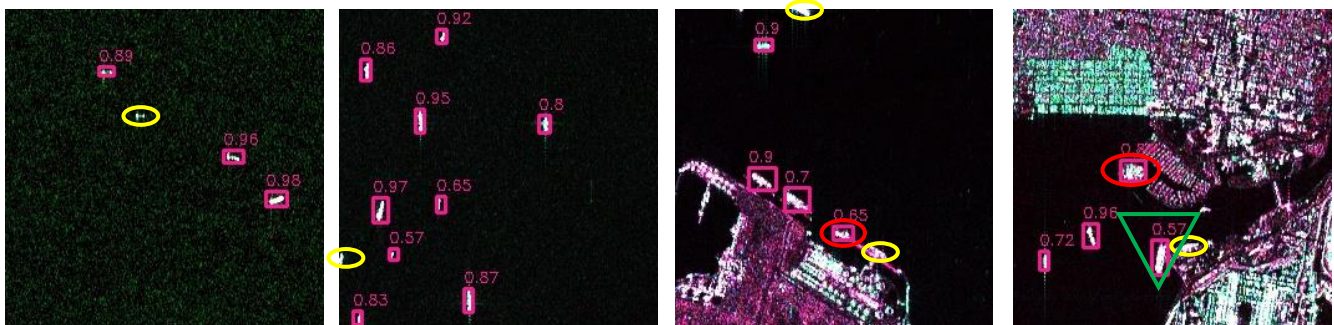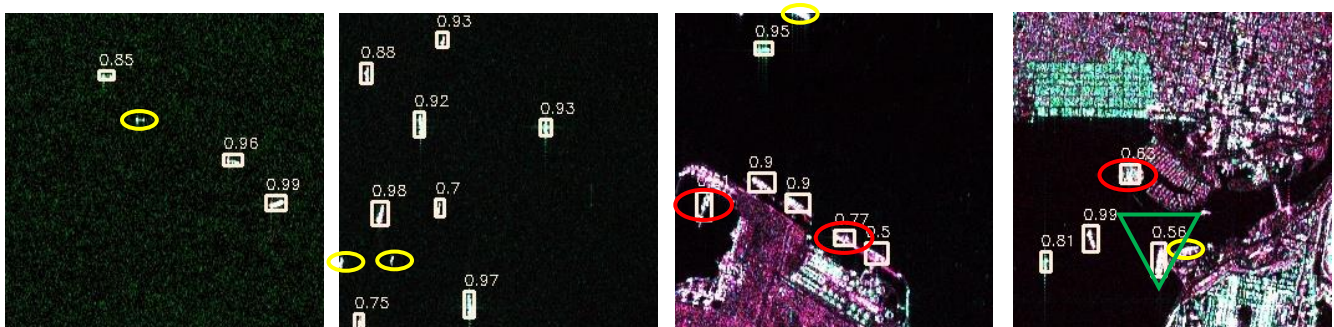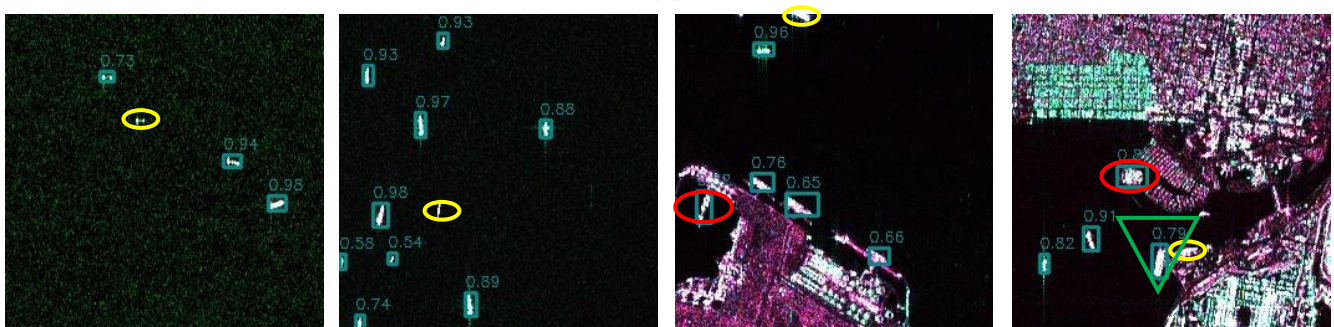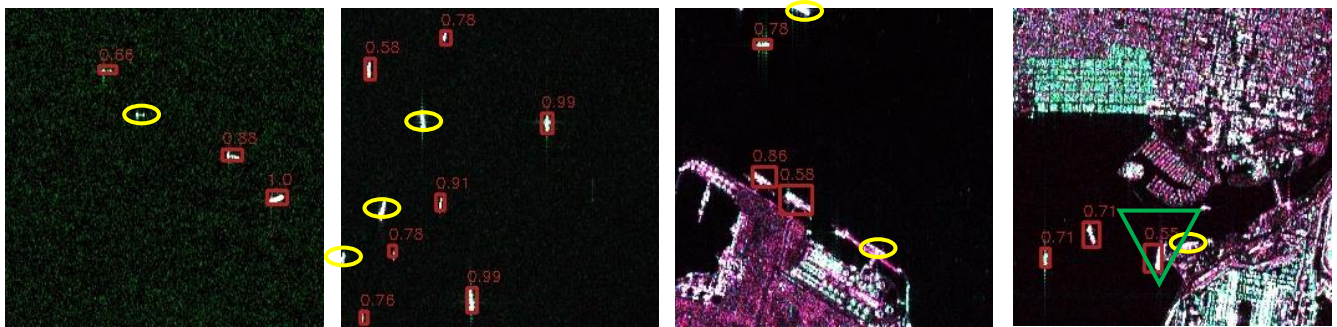

(**a**)

(**b**)

(**c**)

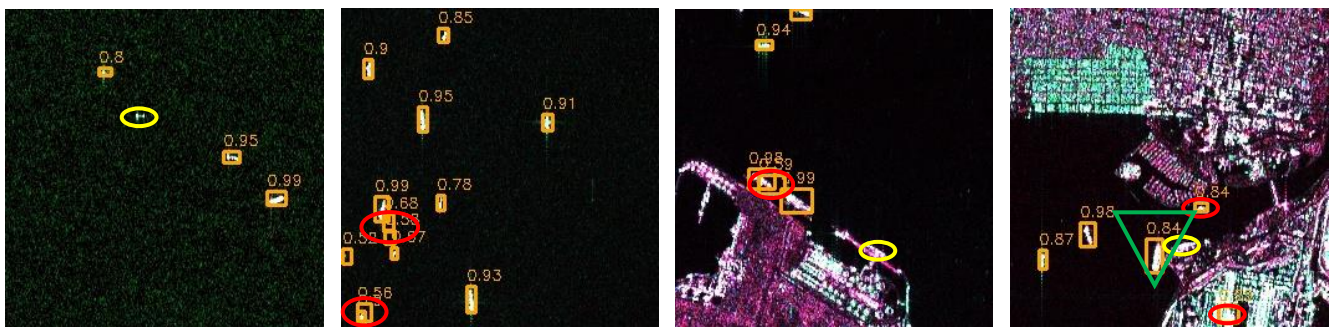

(**d**)

(**e**)

(**f**)

(**g**)

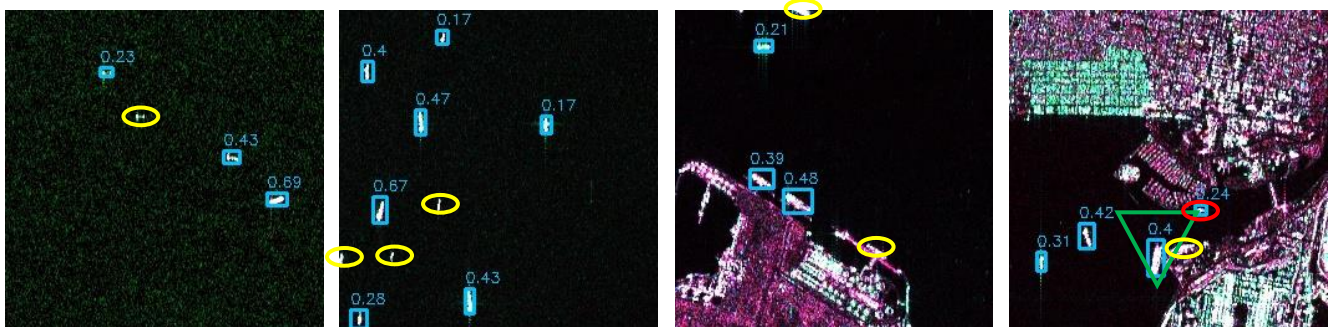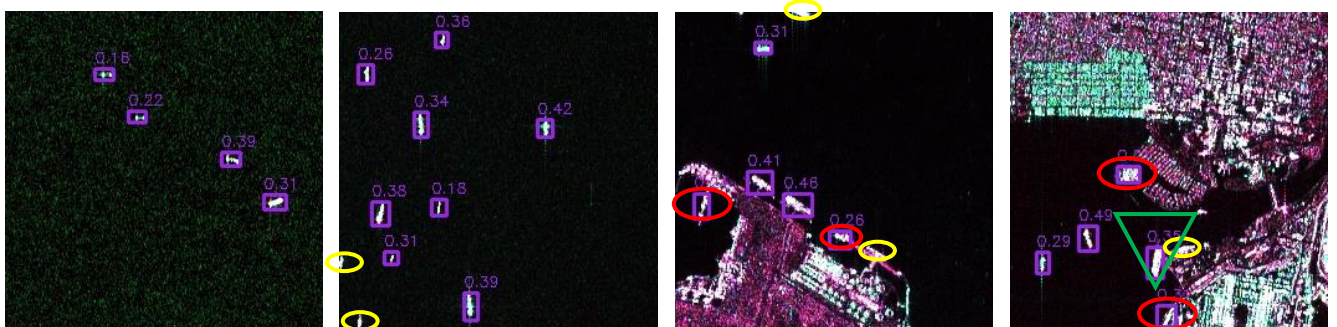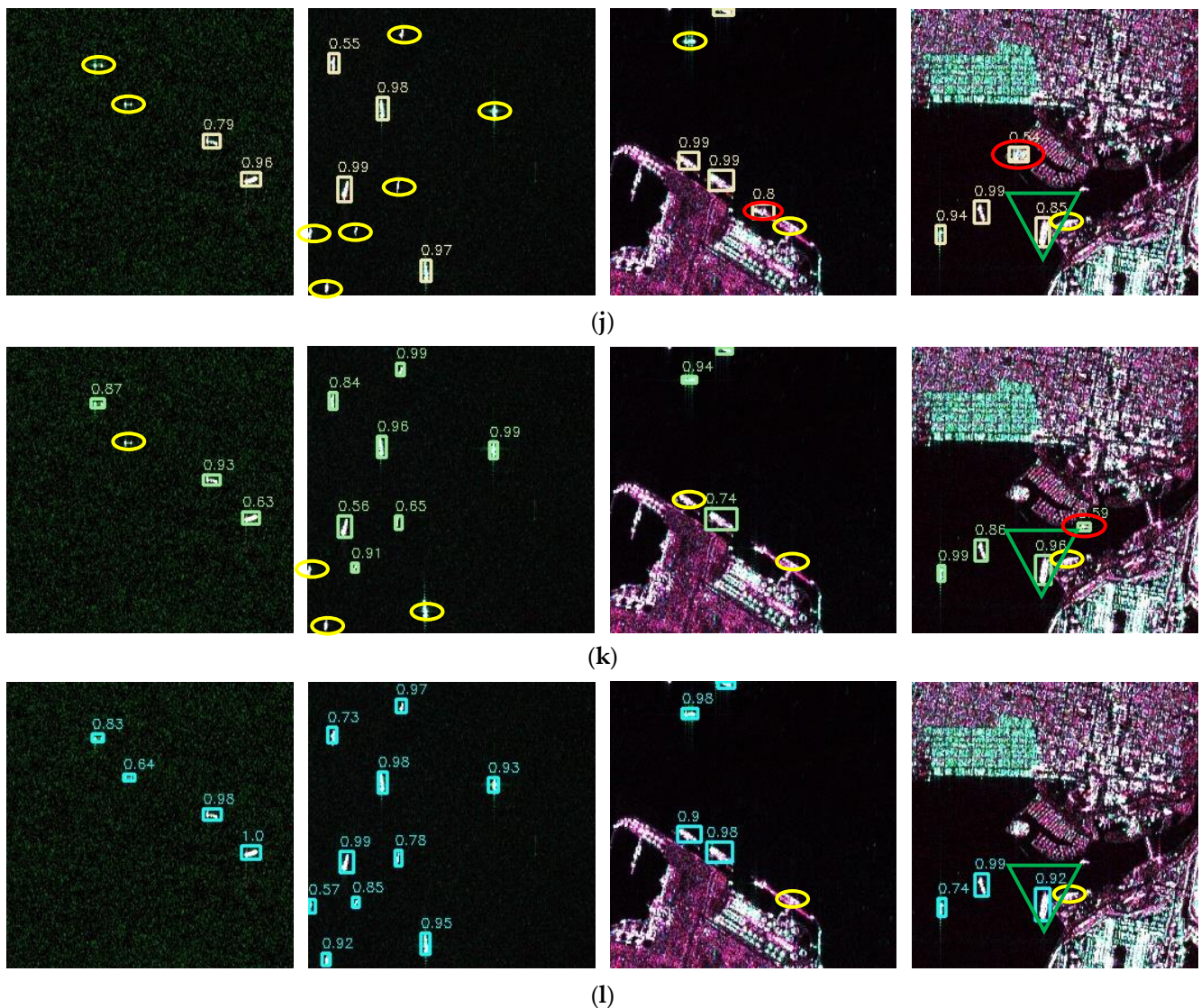
(**h**)

(**i**)

(j)

(k)

(l)
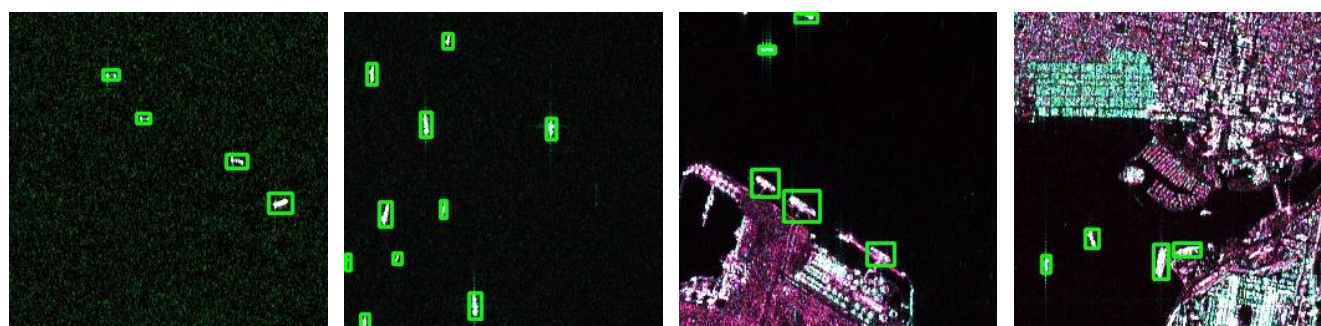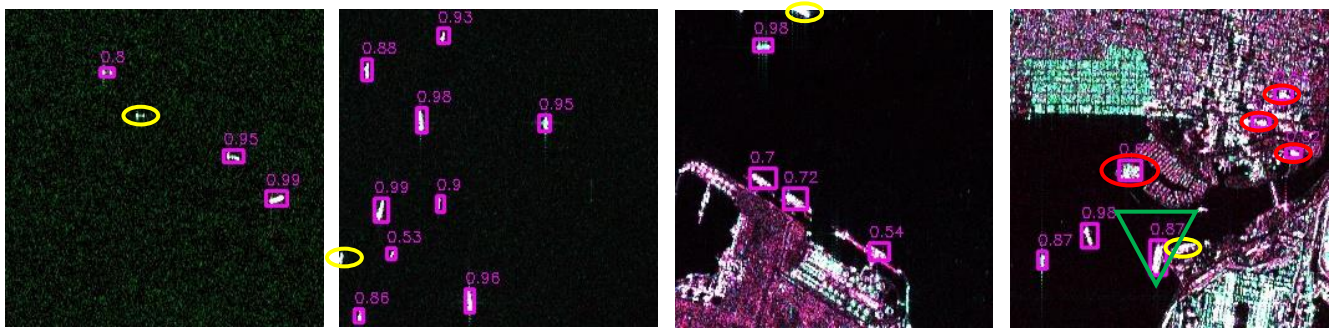
Figure 7. The qualitative detection results of different models. Row (**a**) shows the ground truth; rows (**b**–**l**) represent the detection results of Faster R-CNN, Cascade R-CNN, Dynamic R-CNN, Double-Head R-CNN, MobileNetv2+SSD, FreeAnchor, ATSS, FCOS, Fan et al., Hu et al. and GWFEF-Net, respectively. The detection boxes of different models are marked with different colors, and the number in the upper left corner represents the detection score. In addition, false alarms are marked in the red ellipse and missed detections are marked in the yellow ellipse.

From Figure 7, one can conclude the following.

1. GWFEF-Net shows excellent detection ability even in the case of multi-scale ships. Taking the second column as an example, all models except GWFEF-Net cannot successfully detect the small-scale ship on the image's left edge. Only GWFEF-Net can successfully detect all the multi-scale ships in the image, and provides no unexpected false alarms.

2. GWFEF-Net shows good detection ability even in the case of serious interference from shore buildings. Taking the fourth column as an example, most models tend to mistakenly detect a wharf building as a ship. In fact, in addition to GWFEF-Net, only MobileNetv2+SSD can successfully detect the three ships at the bottom left of the image without generating false alarms, but provides a less tight detection frame (i.e., the IOU of this detection frame and the ground truth is rather small).

3. GWFEF-Net has quite high confidence for dual-polarization ships. Taking the ship in column 4 as an example (marked in a green inverted triangle), the corresponding

confidence of Faster R-CNN is 0.87, the corresponding confidence of Cascade R-CNN is 0.57, the corresponding confidence of Dynamic R-CNN is 0.56, the corresponding confidence of Double-Head R-CNN is 0.79, the corresponding confidence of MobileNetv2+SSD is 0.55, the corresponding confidence of FreeAnchor is 0.84, the corresponding confidence of ATSS is 0.40, the corresponding confidence of FCOS is 0.35, the corresponding confidence of Fan et al. is 0.85, the corresponding confidence of Hu et al. is 0.96, and the corresponding confidence of GWFEF-Net is quite competitive (i.e., 0.92). Moreover, the confidence values of ATSS and FCOS are rather low, so it is unfair to set the score thresholds as 0.50 (if we do so, the two models can hardly detect ships). This situation will be solved in the future.

4. GWFEF-Net achieves the most advanced dual-polarization SAR ship detection performance in both offshore and inshore scenes, compared with the other ten competitive models.

5. Ablation Study

In this section, we describe sufficient ablation experiments on the proposed contributions to verify the effectiveness of each contribution.

5.1. Ablation Study on DFE

Table 6 shows the ablation study of DFE with GWFEF-Net. In Table 6, "VV" denotes the GWFEF-Net detection results on VV polarization data, "VH" denotes the GWFEF-Net detection results on VH polarization data, "C_{VV-VH}" denotes the GWFEF-Net detection results on coherence polarization data, and "VV+VH+C_{VV-VH}" denotes the GWFEF-Net detection results with dual-polarization data (i.e., composed of VV features, VH features and C_{VV-VH} features). From Table 6, it is obvious that using dual-polarization information from three-channel has a greater gain effect on the network than only using polarization information from part of the channels. In other words, the polarization coherence features in DFE are useful to improve GWFEF-Net's detection effect. In particular, using the dual-polarization data increases the AP index (i.e., ~2.3% higher than only using VV and ~1.2% higher than only using VH) and the F1 index (i.e., ~1.3% higher than only using VV and ~1.3% higher than only using VH). As mentioned in Section 2.1, DFE can cooperatively use the information of co-polarization characteristics, cross-polarization characteristics and especially coherence-polarization characteristics. It can compensate for the lack of single-polarization information, finally provide richer dual-polarization features and suppress clutter interferences. In this way, the detection performance of GWFEF-Net can be improved.

Table 6. The ablation study of the DFE.

Input	GT	TP	FP	FN	P (%)	R (%)	F1 (%)	AP (%)
VV	1203	1117	99	86	91.86	92.85	92.35	91.92
VH	1203	1130	114	73	90.84	93.93	92.36	92.98
C_{VV-VH}	1203	1113	114	90	90.71	92.52	91.60	91.48
VV + C_{VV-VH}	1203	1135	111	68	91.09	94.35	92.69	93.34
VH + C_{VV-VH}	1203	1137	106	66	91.47	94.51	92.97	93.58
VV + VH	1203	1131	91	72	92.55	94.01	93.28	93.24
VV + VH + C_{VV-VH}	1203	1142	94	61	92.39	94.93	93.64	94.18

5.2. Ablation Study on GFE

Table 7 shows the ablation study of the GFE by removing and retaining it. In Table 7, "✗" denotes GWFEF-Net with the GFE removed (while retaining the other three contributions), while "✓" denotes GWFEF-Net with the GFE retained (i.e., our proposed model). From Table 7, it can be seen that the GFE increases the AP index by ~1.9% and the

F1 index by ~1.4%, owing to its capability of group-wise feature enhancement. As mentioned in Section 2.2, the enhanced features can highlight each polarization semantic feature region so as to enhance the meaningful ship target area. In this way, the network will detect more meaningful ships and suppress useless clutter interferences with the help of the GFE, which will greatly improve the detection performance of GWFEF-Net.

Table 7. The ablation study of the GFE by removing and retaining it.

GFE	GT	TP	FP	FN	P (%)	R (%)	F1 (%)	AP (%)
✘	1203	1123	110	80	91.08	93.35	92.20	92.27
✓	1203	1142	94	61	92.39	94.93	93.64	94.18

5.3. Ablation Study on GFF

Table 8 shows the ablation study of the GFF by removing and retaining it. In Table 8, "✘" denotes GWFEF-Net with the GFF removed (while retaining the other three contributions), while "✓" denotes GWFEF-Net with the GFF retained (i.e., our proposed model). From Table 8, it can be seen that the GFF increases the AP index by ~2.1% and the F1 index by 1.2%. It is useful for detecting more true ships, i.e., the number of true positives is increased by 23 (that is, GWFEF-Net can effectively decrease missed detections with GFF), owing to its capability of fusing multi-polarization features. As mentioned in Section 2.3, GFF can offer an excellent information interaction from polarization features and multi-scale ship features, which is helpful to extract richer information about ships. Finally, more multi-scale ships can be detected and the final detection performance can be improved.

Table 8. The ablation study of the GFF by removing and retaining it.

GFF	GT	TP	FP	FN	P (%)	R (%)	F1 (%)	AP (%)
✘	1203	1119	99	84	91.87	93.02	92.44	92.06
✓	1203	1142	94	61	92.39	94.93	93.64	94.18

5.4. Ablation Study on HPCA

Table 9 shows the ablation study of the HPCA by removing and retaining it. In Table 9, "✘" denotes GWFEF-Net with the HPCA removed (while retaining the other three contributions), while "✓" denotes GWFEF-Net with the HPCA retained (i.e., our proposed model). From Table 9, it can be seen that the HPCA can increase the AP index by ~1.1% and the F1 index by ~1.0%. It can improve the detection rate by ~1.0% (i.e., from 91.42% to 92.39% with precision) and the recall rate by ~1.0% (i.e., from 93.93% to 94.93% with recall), owing to its capability of channel modeling. As mentioned in Section 2.4, by inserting HPCA into GWFEF-Net, we can obtain finer channel information, in which each polarization feature's contribution is more balanced. In other words, by inserting HPCA into each polarization feature group, the network can learn the contribution of each polarization feature adaptively in the process of group-wise feature fusion. Therefore, HPCA can adaptively equalize the contribution of each polarization feature, and finally ensure the excellent detection performance of GWFEF-Net.

**Table 9.** The ablation study of the HPCA by removing and retaining it.

HPCA	GT	TP	FP	FN	P (%)	R (%)	F1 (%)	AP (%)
✘	1203	1130	106	73	91.42	93.93	92.66	93.04
✓	1203	1142	94	61	92.39	94.93	93.64	94.18

6. Discussion

The above quantitative results, qualitative results and ablation studies fully reveal the superior dual-polarization SAR ship detection performance. The proposed four contributions (i.e., the DFE, GFE, GFF and HPCA) guarantee the excellent ship detection results of GWFEF-Net in dual-polarization SAR images. It can be found that GWFEF-Net can ensure very few missed inspections, which is very applicable to some specific occasions (e.g., the illegal ship monitoring field, where it is essential not to generate missed detections).

We also discuss the generalization ability of GWFEF-Net by conducting an experiment in detecting dual-polarization SAR ships from the Singapore Strait. Note that there is no other public dual-polarimetric SAR ship detection dataset, so we choose to construct some dual-polarization images from the Singapore Strait ourselves. Moreover, these images are not included in the DSSDD dataset used in our paper, and therefore can be used to test the generalization performance of GWFEF-Net. The images are from the Sentinel-1 satellite, with the incident angle of 27.6°~34.8°, resolution of 2.3 m × 14.0 m and swathes of ~250 km. Specifically, in this discussion, the images from Shanghai, the Suez Canal, etc., in DSSDD are selected as our training set, and the dual-polarization images from the Singapore Strait serve as the test set. Figure 8 shows the detection results of GWFEF-Net on the dual-polarization images from the Singapore Strait. From Figure 8, GWFEF-Net can successfully detect many ships in both offshore and inshore scenes. Specifically, GWFEF-Net can correctly detect most ships except for one inshore ship. The above shows the excellent generalization ability of GWFEF-Net.
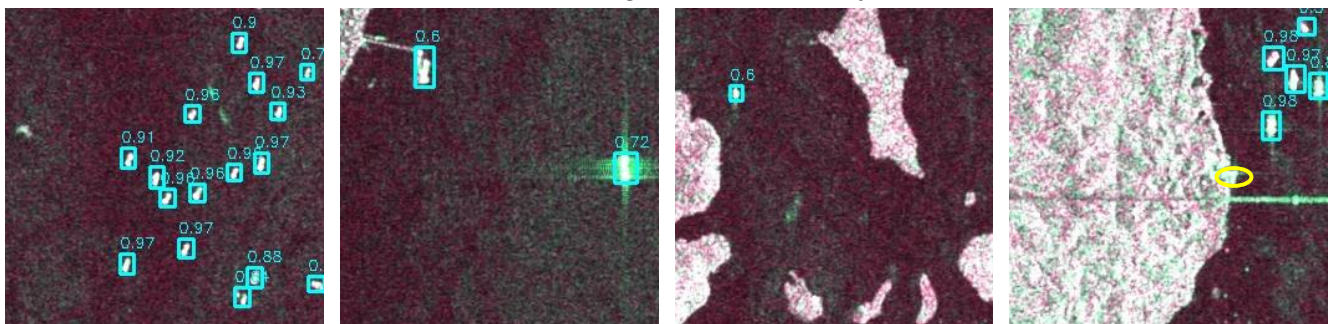
Figure 8. The detection results of GWFEF-Net on the dual-polarization images from the Singapore Strait. The score threshold used for displaying is set as 0.5. The missed detection is marked in the yellow ellipse.

In the future, the typical roll-invariant polarimetric feature will be considered due to its advantage of robustness for rotation ships [59]; the quad-polarization SAR (QP SAR) will be considered because it has the most abundant polarization information [60]; the compact polarimetric SAR (CP SAR) will also be considered because it can reach a balance between swath width and polarization information [61]. In short, we will explore the SAR polarimetric features mentioned above to further improve the ship detection performance. In addition, some traditional artificial features with expert knowledge also reflect the scattering mechanism of ships. Thus, we will also consider integrating the traditional artificial features and polarization features into CNNs to further improve GWFEF-Net's detection performance.

Our future work is as follows:

1. We will explore more SAR polarimetric features to further improve GWFEF-Net's detection performance.
2. We will consider integrating the prior information provided by traditional features into CNNs.

7. Conclusions

In this paper, we present a novel two-stage deep learning network named "GWFEF-Net" for better dual-polarization SAR ship detection. The proposed GWFEF-Net introduces novel contributions on four aspects to achieve better detection performance, i.e., (1) DFE is used to enrich the feature library and suppress clutter interferences to facilitate feature extraction, (2) GFF is used to obtain each enhanced polarization semantic feature to highlight each polarization feature region, (3) GFF is used to obtain fused multi-scale polarization features to realize polarization features' group-wise information interaction, (4) HPCA is used for channel modeling to balance each polarization feature's contribution. Finally, extensive experimental results on the Sentinel-1 dual-polarization SAR ship dataset demonstrate the superior dual-polarization SAR ship detection performance of GWFEF-Net (94.18% in AP), compared with the other ten competitive methods. Specifically, GWFEF-Net can achieve a 4% improvement in AP compared to the baseline Faster R-CNN and a 2.51% improvement in AP compared to the second-best model. In brief, GWFEF-Net can offer high-quality dual-polarization SAR ship detection results, especially ensuring very few missed inspections, which is of great value.

References

1. Bao, J.; Zhang, X.; Zhang, T.; Shi, J.; Wei, S. A Novel Guided Anchor Siamese Network for Arbitrary Target-of-Interest Tracking in Video-SAR. *Remote Sens.* **2021**, *13*, 4504.
2. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. Depthwise Separable Convolution Neural Network for High-Speed SAR Ship Detection. *Remote Sens.* **2019**, *11*, 2483.
3. Chen, S.; Zhang, J.; Zhan, R. R2FA-Det: Delving into High-Quality Rotatable Boxes for Ship Detection in SAR Images. *Remote Sens.* **2020**, *12*, 2031.
4. Zhang, T.; Zhang, X. Insertion of Traditional Hand-Crafted Features into Modern CNN-Based Models for SAR Ship Classification: What, Why, Where, and How. *Remote Sens.* **2021**, *13*, 2091.
5. Cui, Z.; Wang, X.; Liu, N.; Cao, Z.; Yang, J. Ship detection in large-scale SAR images via spatial shuffle-group enhance attention. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 379–391.
6. Xu, X.; Zhang, X.; Zhang, T. Lite-YOLOv5: A Lightweight Deep Learning Detector for On-Board Ship Detection in Large-Scene Sentinel-1 SAR Images. *Remote Sens*. **2022**, *14*, 27.
7. Shao, Z.; Zhang, X.; Zhang, T.; Xu, X.; Zeng, T. RBFA-Net: A Rotated Balanced Feature-Aligned Network for Rotated SAR Ship Detection and Classification. **2022**, *14*, 3345.
8. Zhang, T.; Zhang, X.; Liu, C.; Shi, J.; Wei, S.; Ahmad, I.; Zhan, X.; Zhou, Y.; Pan, D.; Li, J.; et al. Balance Learning for Ship Detection from Synthetic Aperture Radar Remote Sensing Imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *182*, 190–207.
9. Mao, Y.; Yang, Y.; Ma, Z.; Li, M.; Su, H.; Zhang, J. Efficient Low-Cost Ship Detection for SAR Imagery Based on Simplified U-Net. *IEEE Access.* **2020**, *8*, 69742–69753.
10. Dai, W.; Mao, Y.; Yuan, R.; Liu, Y.; Pu, X.; Li, C. A Novel Detector Based on Convolution Neural Networks for Multiscale SAR Ship Detection in Complex Background. *Sensors* **2020**, *20*, 2547.
11. Zhang, T.; Zhang, X. High-Speed Ship Detection in SAR Images Based on a Grid Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 1206.

12. Zhao, Y.; Zhao, L.; Xiong, B.; Kuang, G. Attention Receptive Pyramid Network for Ship Detection in SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2738–2756.

13. Pan, Z.; Yang, R.; Zhang, Z. MSR2N: Multi-Stage Rotational Region Based Network for Arbitrary-Oriented Ship Detection in SAR Images. *Sensors* **2020**, *20*, 2340.

14. Fu, J.; Sun, X.; Wang, Z.; Fu, K. An Anchor-Free Method Based on Feature Balancing and Refinement Network for Multiscale Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1331–1344.

15. Zhang, T.; Zhang, X. ShipDeNet-20: An Only 20 Convolution Layers and 1-MB Lightweight SAR Ship Detector. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1234–1238.

16. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. HyperLi-Net: A hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 123–153.

17. Han, L.; Ran, D.; Ye, W.; Yang, W.; Wu, X. Multi-Size Convolution and Learning Deep Network for SAR Ship Detection from Scratch. *IEEE Access.* **2020**, *8*, 158996–159016.

18. Geng, X.; Shi, L.; Yang, J.; Li, P.X.; Zhao, L.; Sun, W.; Zhao, J. Ship Detection and Feature Visualization Analysis Based on Lightweight CNN in VH and VV Polarization Images. *Remote Sens.* **2021**, *13*, 1184.

19. Fan, Q.; Chen, F.; Cheng, M.; Lou, S.; Xiao, R.; Zhang, B.; Wang, C.; Li, J. Ship Detection Using a Fully Convolutional Network with Compact Polarimetric SAR Images. *Remote Sens.* **2019**, *11*, 2171.

20. Jin, K.; Chen, Y.; Xu, B.; Yin, J.; Wang, X.; Yang, J. A Patch-to-Pixel Convolutional Neural Network for Small Ship Detection with PolSAR Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6623–6638.

21. Fan, W.; Zhou, F.; Bai, X.; Tao, M.; Tian, T. Ship Detection Using Deep Convolutional Neural Networks for PolSAR Images. *Remote Sens.* **2019**, *11*, 2862.

22. Hu, Y.; Li, Y.; Pan, Z. A Dual-Polarimetric SAR Ship Detection Dataset and a Memory-Augmented Autoencoder-Based Detection Method. *Sensors* **2021**, *21*, 8478.

23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149.

24. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. CenterNet: Keypoint Triplets for Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27–28 October 2019; pp. 6568–6577.

25. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 29th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

27. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.

28. Zhang, T.; Zhang, X. Squeeze-and-excitation Laplacian pyramid network with dual-polarization feature fusion for ship classification in sar images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5.

29. Touzi, R.; Hurley, J.; Vachon, P.W. Optimization of the Degree of Polarization for Enhanced Ship Detection Using Polarimetric RADARSAT-2. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5403–5424.

30. Arii, M. Ship detection from full polarimetric SAR data at different incidence angles. In Proceedings of the 2011 3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Seoul, Korea, 26–30 September 2011; pp. 1–4.

31. Touzi, R.; Vachon, P.W.; Wolfe, J. Requirement on Antenna Cross-Polarization Isolation for the Operational Use of C-Band SAR Constellations in Maritime Surveillance. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 861–865.

32. Liu, C.D.; Vachon, P.W.; English, R.A.; Sandirasegaram, N.M. *Ship Detection Using RADARSAT-2 Fine Quad Mode and Simulated Compact Polarimetry Data*; Technical Memorandum; Defence R&D Canada: Ottawa, ON, Canada, 2010.

33. Pelich, R.; Chini, M.; Hostache, R.; Matgen, P.; Lopez-Martinez, C.; Nuevo, M.; Ries, P.; Eiden, G. Large-Scale Automatic Vessel Monitoring Based on Dual-Polarization Sentinel-1 and AIS Data. *Remote Sens.* **2019**, *11*, 1078.

34. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, United states, 4–9 December 2017; pp. 3857–3867.

35. Li, X.; Hu, X.; Yang, J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. *arXiv* **2019**, arXiv:1905.09646.

36. Cui, Z.; Li, Q.; Cao, Z.; Liu, N. Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8983–8997.

37. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection. *Remote Sens.* **2021**, *13*, 2771.

38. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

39. Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995.

40. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2021**, *43*, 652–662.

41. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.

42. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI), San Francisco, CA, USA, 4–10 February 2017; pp. 4278–4284.

43. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

44. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 33rd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online Event, 14–19 June 2020; pp. 11531–11539.

45. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521.

46. Ketkar, N. Introduction to Pytorch. In *Deep Learning with Python: A Hands-On Introduction*; Apress: Berkeley, CA, USA, 2017; pp. 195–208. Available online: https://link.springer.com/chapter/10.1007/978-1-4842-2766-4_12 (accessed on 1 June 2022).

47. Chen, K.;Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Lin, D. MMDetection: Open MMLAB Detection Toolbox and Benchmark. *arXiv* **2019**, arXiv:1906.07155.

48. Sergios, T. Stochastic gradient descent. *Mach. Learn.* **2015**, *5*, 161–231.

49. He, K.; Girshick, R.; Doll´ar, P. Rethinking ImageNet Pre-Training. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27–28 October 2019; pp. 4917–4926.

50. Hosang, J.; Benenson, R.; Schiele, B. Learning Non-Maximum Suppression. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6469–6477.

51. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into High Quality Object Detection. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 6154–6162.

52. Hongkai, Z.; Hong, C.; Bingpeng, M.; Naiyan, W.; Xilin, C. Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training. In Proceedings of the 16th European Conference on Computer Vision (ECCV), Cham, Switzerland, 23–28 August 2020; pp. 260–275.

53. Wu, Y.; Chen, Y.; Yuan, L.; Liu, Z.; Wang, L.; Li, H.; Fu, Y. Rethinking Classification and Localization for Object Detection. In Proceedings of the 33rd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online Event, 14–19 June 2020; pp. 10183-10192.

54. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks *arXiv* **2018**, arXiv:1801.04381.

55. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, Netherlands, 8–16 October 2016; pp. 21–37.

56. Zhang, X.; Wan, F.; Liu, C.; Ji, R.; Ye, Q. FreeAnchor: Learning to match anchors for visual object detection. arXiv **2019**, arXiv:1909.02466.

57. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In Proceedings of the 33rd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online Event, 14–19 June 2020; pp. 9756–9765.

58. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully convolutional one-stage object detection. In Proceedings of the 17th IEEE/CVF International Conference on Computer Vision, （ICCV), Seoul, Korea, 27 October–2 November 2019. pp. 9626–9635.

59. Chen, S.W.; Wang X.S.; Sato M. Uniform polarimetric matrix rotation theory and its applications. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4756–4770.

60. Charbonneau, F.; Brisco, B.; Raney, R. Compact polarimetry overview and applications assessment. *Can. J. Remote Sens.* **2010**, *36*, 298–315.

61. Zhang, B.; Li, X.; Perrie, W.; Garcia-Pineda, O. Compact polarimetric synthetic aperture radar for marine oil platform and slick detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1407–1423.