remote sensing

Article

A Dual-Polarization Information-Guided Network for SAR Ship Classification

Zikang Shao †, Tianwen Zhang *,† and Xiao Ke

School of Information and Communication Engineering,
University of Electronic Science and Technology of China, Chengdu 611731, China
* Correspondence: twzhang@std.uestc.edu.cn
† These authors contributed equally to this work.

Abstract: Synthetic aperture radar (SAR) is an advanced active microwave sensor widely used in marine surveillance. As part of typical marine surveillance missions, ship classification in synthetic aperture radar (SAR) images is a significant task for the remote sensing community. However, fully utilizing polarization information to enhance SAR ship classification remains an unresolved issue. Thus, we proposed a dual-polarization information-guided network (DPIG-Net) to solve it. DPIG-Net utilizes available dual-polarization information from the Sentinel-1 SAR satellite to adaptively guide feature extraction and feature fusion. We first designed a novel polarization channel cross-attention framework (PCCAF) to model the correlations of different polarization information for feature extraction. Then, we established a novel dilated residual dense learning framework (DRDLF) to refine the polarization characteristics for feature fusion. The results on the open OpenSARShip dataset indicated DPIG-Net's state-of-the-art classification accuracy compared with eleven other competitive models, which showed the potential of DPIG-Net to promote effective and sufficient utilization of SAR polarization data in the future.

Keywords: synthetic aperture radar; ship classification; polarization-guided

check for updates

Citation: Shao, Z.; Zhang, T.; Ke, X. A Dual-Polarization Information-Guided Network for SAR Ship Classification. *Remote Sens.* **2023**, *15*, 2138. https://doi.org/10.3390/rs15082138

Academic Editor: John Trinder

Received: 23 February 2023
Revised: 3 April 2023
Accepted: 6 April 2023
Published: 18 April 2023

1. Introduction

Ship classification plays an important role in ocean surveillance [1–6]. It can distinguish the specific types of ships and provide more comprehensive and extensive marine surveillance information, which is conducive to trade management, marine traffic and transportation monitoring, fishery management, etc. The specific types of ships are related to their functions, such as bulk carriers carrying industrial and commercial resource, container ships carrying important goods in international trade, oil tankers carrying industrial petroleum, ore carriers carrying materials from coal mines, fishing vessels carrying out marine fishing, cruise ships carrying passengers for sightseeing, law enforcement vessels carrying out marine river traffic management, etc. Ship classification belongs to a typical image classification task, that is, a two-dimensional ship image is input into an image classification model and the specific category label of the ship in the image is generated as output.

Imaging radiometers, optical sensors, and SAR are commonly used systems in the related area [7,8]. Optical sensors can provide high-quality image information, but they are easily disturbed by clouds and the revisit period is long [9,10]. Visible infrared imaging radiometers can obtain a wider field of view and have a shorter revisit period, but they are still susceptible to cloud interference [11]. In comparison with the above systems, SAR is able to obtain relatively clear images, the operation of which is usually unhindered by both light and weather [12]. Therefore, compared with other technologies, SAR has unique advantages that makes it very suitable for marine ship classification [13]. Nowadays, SAR ship classification is receiving much attention.

Similar to SAR automatic target recognition (ATR) methods [14–18] designed for vehicle targets, traditional ship classification methods [1,19–23] focus on feature extraction based on handcraft features according to expert experience, such as geometric features, texture features, scattering intensity features, directional gradient histogram features, etc. Subsequently, the extracted features are input into classic machine learning classifiers such as SVM, decision tree, and KNN to complete the classification of ships. In the last ten years, most of the subsequent scholars [19–27] have followed this research approach to design SAR ship classification models, such as improving the representation of ship manual features and selecting machine learning classifiers with better performance to process ship manual features. However, these traditional methods still have many defects, such as being time-consuming and having laborious manual feature extraction processes, complex mathematical theory, limited migration ability, and so on [28]. Nowadays, it is difficult to adapt to the needs of more intelligent SAR ship classification.

With the boom in deep learning in recent years [29,30], SAR ship classification methods based on deep learning [31–41] are receiving more attention. Deep convolutional neural networks have been used by researchers to achieve SAR ship classification with higher accuracy and faster speeds than traditional classification models. Moreover, deep convolutional neural networks are able to conduct end-to-end training and testing, greatly simplifying the design process and reducing the burden of manual feature extraction. For example, Hou et al. [31] designed a simple convolutional neural network to classify ships in Gaofen-3 images. Huang et al. [32] proposed a group squeeze excitation sparsely connected convolutional network to extract robust ship features. Wang et al. [33] studied transfer learning to solve the few-shot ship classification problem. Wang et al. [34] proposed a semi-supervised learning method via self-consistent augmentation to boost classification accuracy. He et al. [35] designed a densely connected triplet CNN and integrated Fisher discrimination regularized metric learning for ship classification in medium-resolution SAR images. Zhang et al. [36] fused HOG features into CNNs to reduce model risk. However, these works did not consider ship polarization information, which is acquired through specific polarized antenna transmitting one polarization and simultaneously receiving multiple polarizations. Ships imaged by such sensors have different backscattering characteristics in different polarization channels. Therefore, utilization of polarization information is helpful to improve the performance of SAR ship classification, especially for low-resolution SAR images where the spatial features of ships are limited and more useful information is needed to guide the classification task.

Several works [38–41] tried to utilize polarization information for better classification performance. For example, Zeng et al. [38] proposed a loss function for better dual-polarization feature training, but their network ignored feature interaction that might lead to local optimization. Zhang et al. [39] designed a squeeze-and-excitation Laplacian pyramid network for multi-resolution feature extraction, but their network did not highlight salient features and yielded limited accuracy gains. Xiong et al. [40] established a mini hourglass region extraction network for dual-channel feature fusion, but they did not consider channel correlation, resulting in insufficient utilization of polarization information. Zhang et al. [41] established a polarization fusion and geometric feature embedded network to increase feature richness, but their network treated each polarization branch equally and resulted in difficult training and incomplete feature extraction.

The above information demonstrates that it remains a challenging and unresolved issue to make full use of polarization information to further boost SAR ship classification performance. Previous works have not provided a simple and effective implementation so far. Therefore, we proposed a dual-polarization information-guided network (DPIG-Net) to address the problem. DPIG-Net utilizes available dual-polarization information from the Sentinel-1 satellite to guide SAR ship classification from two aspects—feature extraction and feature fusion. In the feature extraction process, we designed a novel polarization channel cross-attention framework (PCCAF) to model feature correlations of different polarization information, which was used to guide the network to extract more representative features. In the feature fusion process, we designed a novel dilated residual dense learning framework (DRDLF) to refine the features, which enabled better feature fusion benefits. The results on the open three-and six-category OpenSARShip datasets [42] revealed the state-of-the-art classification accuracy of DPIG-Net compared with eleven other competitive models.

The main contributions of this paper are as follows:

(1) DPIG-Net is proposed for the sufficient utilization of polarization information to boost classification accuracy. It is a brand new architecture for achieving dual-polarization SAR ship classification. Compared with other state-of-the-art methods, DPIG-Net can make full use of ship polarization information and has the potential to implicitly mine useful dual-polarization feature patterns for better classification accuracy.

(2) PCCAF is proposed for representative polarization feature extraction. It is a brand new framework for dual-polarization feature extraction. Compared with other state-of-the-art methods, PCCAF can model the correlations between different polarization channels by the proposed cross-attention subnetwork so as to serve for better feature extraction.

(3) DRDLF is proposed for refined polarization feature fusion. It is a brand new framework for achieving dual-polarization feature fusion. Compared with other state-of-the-art methods, DRDLF can maintain a large receptive field in network depth and its idea of feature reuse is conducive to the deep supervision of feature learning, thus reducing overfitting risk.

(4) For the community of SAR ship detection and classification, we provide the idea of using polarization information to guide the intelligent interpretation of SAR images, and we contribute a network framework (PCCAF-DRDLF) that makes it possible to make full use of dual-polarization information.

The rest of the paper is organized as follows: Section 2 introduces DPIG-Net. Section 3 introduces the experiments and the results. The discussion is described in Section 4. Section 5 sums up this paper.

2. Method

Figure 1 shows the network architecture of DPIG-Net. It is similar to [43], but it is closely related to ship polarization. The data used in this work were the open OpenSARShip dataset, samples of which were from the Sentinel-1 [44] SAR satellite. Sentinel-1 works in dual-polarization mode, i.e., vertical–vertical (VV) and vertical–horizontal (VH). The offered data were denoted by S_{VV} and S_{VH} which were in the form of complex numbers. Since S_{VV} has higher scattering energy of ships [7], it was selected as the source of the middle main branch guiding other branches for feature extraction. Additionally, the input of the middle main branch was denoted by $I_2 = |S_{VV}|$. We selected S_{VH} as the source of the upper branch since S_{VH} reflects less scattering energy of ships than S_{VV} [7], and the input of the upper branch was denoted by $I_1 = |S_{VH}|$. See [7] for more details.

Moreover, to fully leverage the polarization information, the lower branch in PCCAF was constructed to measure the polarization channel difference for a more comprehensive description of ship characteristics, and its input was given by:

$$I_3 = |S_{VV} \cdot S_{VH}^*| \quad (1)$$

where * denotes a complex conjugate operation. Significantly, S_{VV} and S_{VH} used in our work must be complex data, rather than the previous commonly used amplitude-based real data. To the best of our knowledge, OpenSARShip might be the only data that can meet this requirement. Notably, FUSAR-Ship only offers amplitude real data, so I_3 could not be obtained by Equation (1). Moreover, images in FUSAR-Ship are not paired in the form of VV–VH or HH–HV, which prevents the application of our network.

In particular, our current work only considered the dual-polarization case due to the limitation of available data. If full-polarization data is available in the future, one can expand DPIG-Net into four parallel branches to receive four different polarization inputs (or more branches for the cross-channel model).

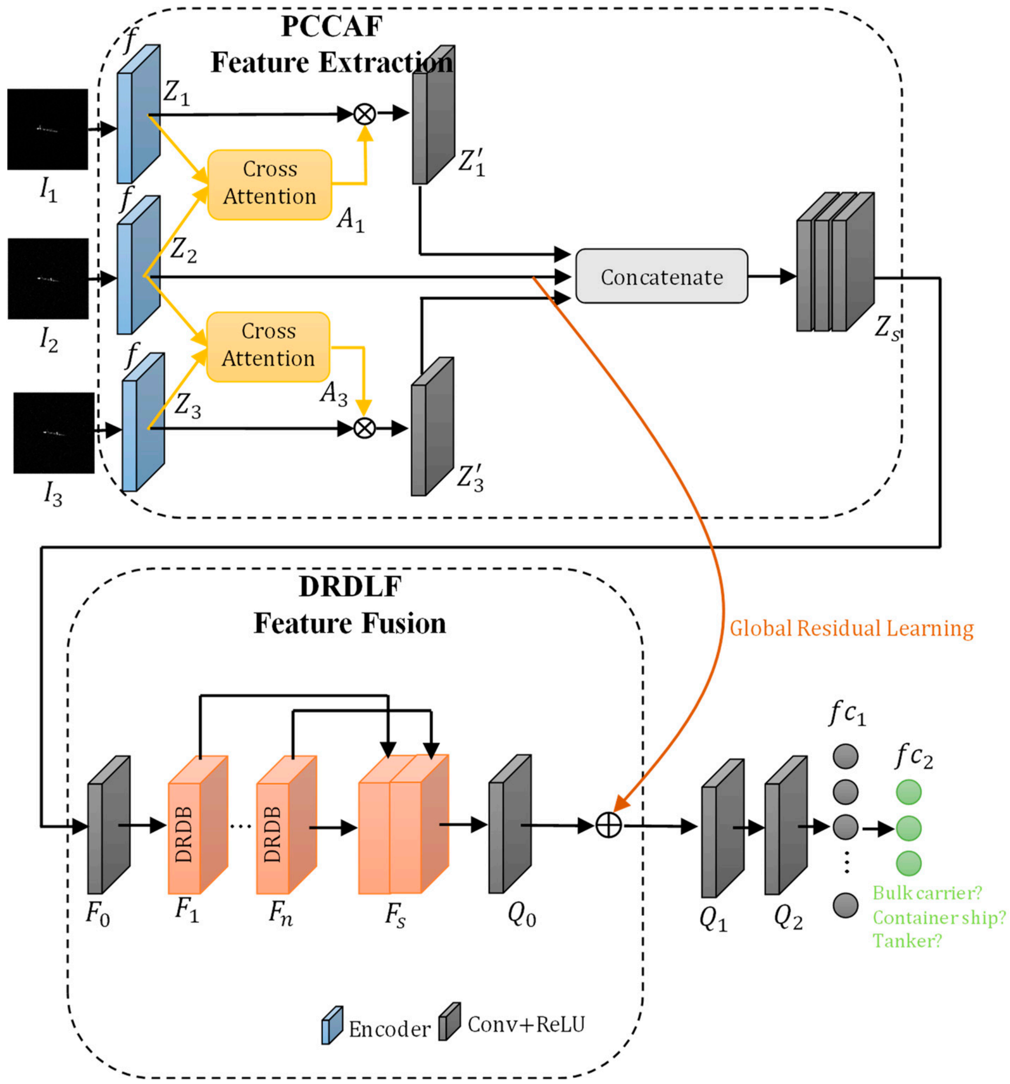

Figure 1. Network architecture of DPIG-Net. PCCAF denotes the polarization channel cross-attention framework. DRDLF denotes the dilated residual dense learning framework. DRDB denotes the dilated residual dense block.

PCCAF received three types of data (I_1, I_2, and I_3) for feature extraction. Its output was denoted by Z_S, which contained the high-level semantic features [45] of the three types of data. DRDLF received Z_S for feature fusion through several cascaded dilated residual dense blocks and global residual learning from the main branch I_2. Finally, 2D feature maps were flattened into 1D feature vectors to transmit into a fully-connected (fc_1) layer. The terminal fc_2 was responsible for category prediction with the soft-max function. Significantly, the reason that we set two fully connected layers was to gradually aggregate the flattened feature, which was conducive to keeping important semantic features and training the network. More fully connected layers may provide benefits, but the amount of calculations and number of parameters will increase sharply. Therefore, we only kept two fully connected layers in DRDLF.

DPIG-Net showed a tendency of feature aggregation from the three input branches to the terminal feature integration. Most previous works only adopted I_2 to predict ship categories, i.e., the middle main branch of PCCAF. In contrast, we made full use of the polarization information (I_1 and I_3) to guide the classification prediction of I_2. We named the above paradigm the dual-polarization information-guided SAR ship classification.

2.1. Polarization Channel Cross-Attention Framework (PCCAF)

PCCAF established a simple encoder f to preliminarily extract features from the three types of data. The encoder structure is shown in Table 1. The encoder f used standard convs to extract features, batch normalization (BN) [46] to ensure training, and ReLU to activate neurons. The max-pooling operation was used to reduce the size of feature maps. With network deepening, the channel width increased by a multiple of 2. Significantly, the number of channels is known to increase as the resolution decreases in order to prevent the loss of discriminative features [47]. Moreover, our feature encoder f only had four stages, rather than the usual five stages [36]. This was to avoid the loss of spatial features [48] due to the small size [49] of SAR ships. Their outputs were denoted by Z_1, Z_2, and Z_3 for the subsequent processing. A more advanced encoder might achieve better performance, but it was not within the scope of this research.

Table 1. Encoder Structure in PCCAF.

Stage	Layer	Input Shape	Output Shape	Kernel@Stride
S_1	Conv + BN + ReLU	224 × 224 × 1	224 × 224 × 8	3 × 3 × 8@1
	Max-pooling	224 × 224 × 8	128 × 128 × 8	@2
S_2	Conv + BN + ReLU	128 × 128 × 8	128 × 128 × 16	3 × 3 × 16@1
	Max-pooling	128 × 128 × 16	64 × 64 × 16	@2
S_3	Conv + BN + ReLU	64 × 64 × 16	64 × 64 × 32	3 × 3 × 32@1
	Max-pooling	64 × 64 × 32	32 × 32 × 32	@2
S_4	Conv + BN + ReLU	32 × 32 × 32	32 × 32 × 64	3 × 3 × 64@1
	Max-pooling	32 × 32 × 64	16 × 16 × 64	@2

To better exploit the benefit of polarization information, we designed a cross-attention subnetwork to model the correlations between different polarization branches. The design concept of the cross-attention subnetwork was that the middle main branch generated referenced feature maps to guide the other two auxiliary branches. Most existing attention networks merely refine their own feature maps in the uncrossed mode, which cannot solve the multi-branch dual-polarization-guided case. That is, their module input has only one entry, but our proposed cross-attention subnetwork was specially designed for dual-polarization ship missions, i.e., our module input had two entries. The cross-attention subnetwork could be summarized as:

$$A_i = a_i(Z_i, Z_r) \tag{2}$$

where Z_r denotes the referenced feature maps (in this paper, $Z_r = Z_2$, i.e., the main VV branch), Z_i denotes the feature maps to be corrected (in this paper, Z_i means the VH branch Z_1 or the polarization difference branch Z_3), a_i denotes the learned mapping, and A_i denotes the cross-attention map.

Figure 2a shows the network implementation. Taking Z_1 and Z_2 as an example, the same procedure was applied to Z_3 and Z_2. We first concatenated the two input feature maps directly, and then, three convs with a skip connection were employed to learn the inputs' interrelations. Finally, the learning knowledge was activated by a sigmoid to obtain the final cross-attention map A_1. Significantly, the reason that we selected a sigmoid as the activation function was that a sigmoid is easily differentiated for backpropagation and can narrow the range of attention weights in the cross-attention map for stable network training. Moreover, in comparison with other activation functions, such as Tanh and ReLU, a sigmoid is able to map any real number to output from 0 to 1, which is suitable for measuring the attention level of one position in a feature map [50]. Specifically, the closer the attention weight in the cross-attention map is to 0, the less important the feature of the corresponding position in the feature map, and vice versa.

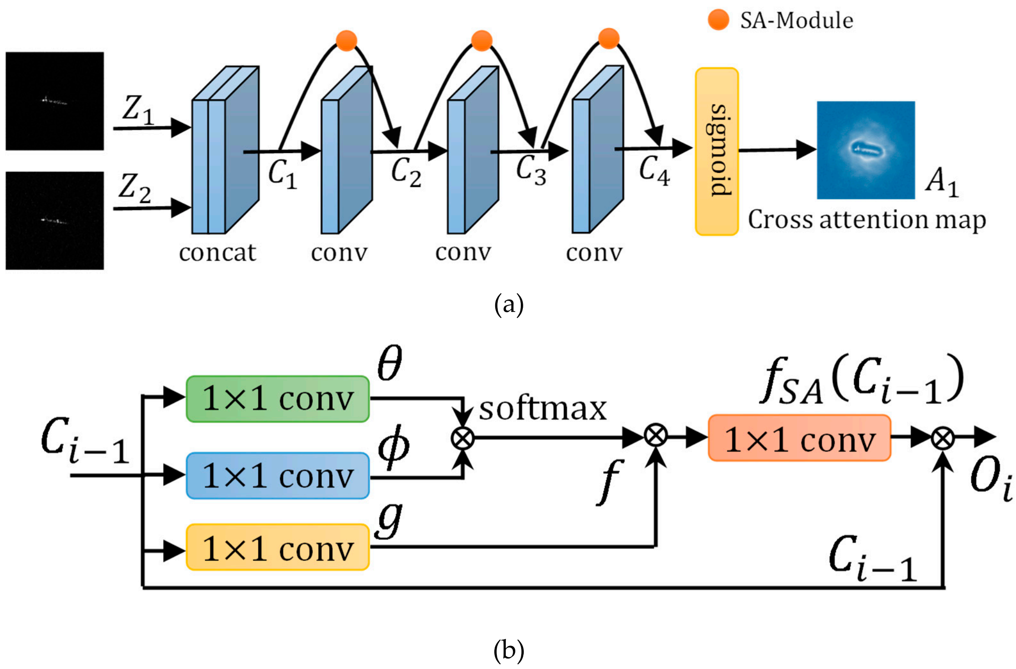

Figure 2. Implementation of the cross-attention subnetwork. (**a**) Cross-attention subnetwork. (**b**) SA-module.

Furthermore, for better skip connection fusion between shallow low-level features and deep high-level features, we designed a self-attention module (SA-module) to refine the previous features. The motivation for the SA-module was also related to SAR image characteristics, e.g., speckle noise and sea clutter. It can relieve their related interferences to enhance ship saliency, as shown in Figure 2a. The SA-module could highlight important global information in space [51], suppress low-value information, and promote network information flow. Ablation studies in Section 4.1 indicated that it could offer a ~2% accuracy improvement on the six-category task. The SA-module generated a self-attention map to modify input and then the result was added to the raw conv branch. The above was described as:

$$C_i = C_{i-1} \cdot f_{SA}(C_i - 1) + f_{3\times3}(C_{i-1}) \tag{3}$$

where C_i denotes the i-th conv feature map, f_{SA} denotes the SA-Module operation, and $f_{3\times3}$ denotes the 3×3 conv. Figure 2b shows the implementation process of the SA-module. The representation of the input at j-position was embedded by g, which was instantiated by a1 $\times$ 1 conv. The spatial features of the i-position were embedded by θ. The spatial features of the j-position were embedded by ϕ. The relationship between i-position and j-position was calculated through the relationship function f, which was defined as:

$$f = \frac{e^{(W_\theta x_i)^\top (W_\phi x_j)}}{\sum\limits_{\forall j} e^{(W_\theta x_i)^\top (W_\phi x_j)}} \tag{4}$$

where W_θ and W_ϕ serve as learnable weights. $\sum\limits_{\forall j} e^{(W_\theta x_i)^\top (W_\phi x_j)}$ serves as a normalization factor to normalize the relationship between two positions for stable training of the network. In practice, we instantiated $W_\theta x_i$ and $W_\phi x_j$ through a 1×1 conv, respectively. $\frac{e^{y_j}}{\sum\limits_{\forall j} e^{y_j}}$ was instantiated by soft-max along dimension j, where $y_j = (W_\theta x_i)^\top (W_\phi x_j)$ was instantiated by matrix multiplication after 1×1 conv was completed. The response at i-position was obtained by a matrix element-wise multiplication between input C_{i-1} and self-attention map $f_{SA}(C_i - 1)$. Significantly, the reason that soft-max was selected for normalization was derived from concerns about the definition of the relationship function f. On the one hand, f needs a normalization factor as the denominator for normalization in case network training is unstable [52]. On the other hand, f should be conveniently instantiated in consideration of efficiency and operability. Using existing operators such as convolution and soft-max is suitable for instantiating f while designing a network. Therefore, using soft-max along dimension j as the instantiation of $\frac{e^{y_j}}{\sum\limits_{\forall j} e^{y_j}}$ was a convenient method for normalization [51].

The final resulting cross-attention map was acted on the other two branches by matrix-element multiplication to obtain the refined polarization-guided features:

$$Z_i' = Z_i \otimes A_i \tag{5}$$

where Z' denotes polarization-guided features that will be used to guide the main polarization branch.

Finally, the output of the main polarization branch was the concatenation of three types of features:

$$Z_s = \text{Concat}(Z_1', Z_2, Z_3') \tag{6}$$

where Z_s denotes the output of PCCAF. We found that feature concatenation performed better than feature adding because the former could avoid the resistance effects between different polarization features with our subsequent feature fusion operations.

2.2. Dilated Residual Dense Learning Framework (DRDLF)

DRDLF used some dilated residual dense blocks (DRDBs) to fuse the extracted polarization features coming from the previous PCCAF stage. The input of DRDLF was denoted as Z_s, which was associated with the dual-polarization information using the concatenation operation of Equation (5) where Z_1' denotes the feature maps of I_1 VH information, Z_2 denotes that of I_2 VV information, and Z_3' denotes that of VV-VH correlation information. Z_s was refined by a 3×3 conv for feature concentration and channel dimensionality reduction. The result was denoted by F_0. Then, several cascaded DRDBs were used for feature aggregation. DRDB was motivated by RDB [53], which was designed for image super-resolution tasks. However, there are many speckle noises around SAR ship images [54,55], so we inserted a dilated rate of 2 to the standard conv for larger receptive fields.

Figure 3 shows the DRDB's implementation. Its input was the previous output F_i, and its output was denoted by F_{i+1}. DRDB contained three 3×3 conv layers with a dilated rate of 2, and their results were denoted by D_1, D_2, and D_3 respectively. They were concatenated directly as D_S. To meet the requirement of residual connection in the entire DRDB, a 1×1 conv was used for channel reduction. Finally, the sum between F_i and D_S was its output. In DRDLF, we arranged n DRDBs for feature fusion where n was empirically set to the optimal value 3. The results of n DRDBs from F_1 to F_n were concatenated and then processed by a 1×1 conv for overall channel reduction. The result was denoted by Q_0. Significantly, we did not select dilated convs with a higher dilated rate or more dilated convs for feature extraction. Even though a higher dilated rate and more dilated convs can obtain a larger receptive field, which is helpful to extract contextual information and discriminate between the foreground and background [56], this will deteriorate the spatial details of ships, especially in the case of low-resolution SAR images. Therefore, the chosen dilated rate and number of dilated convs was more like a trade-off in the design of the network.

Significantly, we observed that after a series of DRDB processing with multiple dense connections, the details of the main VV branch might be gradually diluted, causing unstable training and deteriorating performance. Thus, inspired by [57], we proposed a global residual learning to solve this problem. As shown in Figure 1, the global residual learning connected PCCAF and DRDLF, thus maintaining the dominant position of the main branch and making the other two branches smoothly play an auxiliary guiding role. This was an important design aspect of our dual-polarization-guided network. The global residual learning was described by:

$$Q_1 = Q_0 + Z_2 \tag{7}$$

where Q_1 denotes the final output of DRDLF. From Figure 1, we set another two 3×3 convs to process Q_1 for more semantic features Q_2, which was helpful for balancing spatial and semantic information.

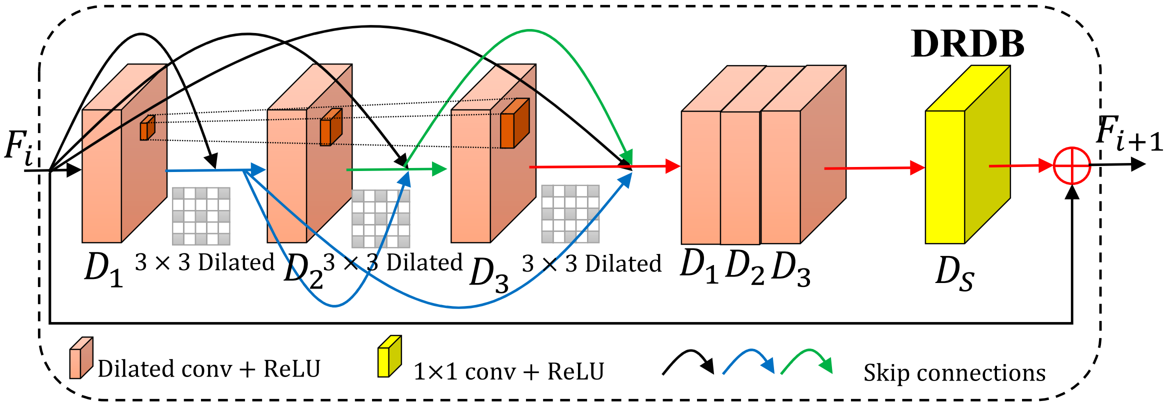

Figure 3. Implementation of the dilated residual dense block (DRDB).

To sum up, combined with the above designed PCCAF and DRDLF, our proposed DPIG-Net could make full use of the polarization information ignored in previous works. The other two types of polarization data were well refined to assist in the feature extraction and feature fusion of the main branch. Finally, an effective dual-polarization information-guided SAR ship classification paradigm was realized. DPIG-Net successfully handled the problems of how to conduct polarization guidance and how to carry out more effective polarization guidance, which are of great value.

3. Result

3.1. Dataset

The open OpenSARShip dataset [42] was used to evaluate the effectiveness of DPIG-Net. It offers VV–VH dual-polarization SAR ship data from Sentinel-1 with different environmental conditions. The labels of SAR ships are annotated through automatic identification system (AIS) messages corrected for position shifts, which ensures the high reliability of labeling. The raw data covered five typical ports, including Shanghai Port (China), Shenzhen Port (China), Tianjin Port (China), Yokohama Port (Japan), and Singapore Port (Singapore), with the form of single look complex (SLC) type. Same as [39], two subsets of the data were used for experiments, i.e., a three-category subset and a six-category one. As previously mentioned, OpenSARShip is the only dataset that could satisfy our experimental requirements, i.e., paired dual-polarization complex data with corresponding ground truth labels. Tables 2 and 3 show descriptions of the data. Figures 4 and 5 show some samples of different ship categories.

Table 2. Three-Category Data.

Category	Training	Test
Bulk carrier	169	164
Container ship	169	404
Tanker	169	73

Table 3. Six-Category Data.

Category	Training	Test
Bulk carrier	100	233
Cargo	100	571
Container ship	100	473
Fishing	100	25
General cargo	100	42
Tanker	100	142

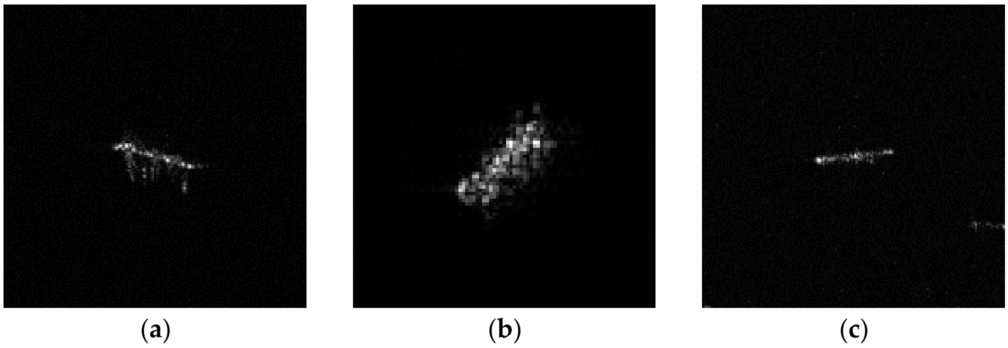

Figure 4. Three-category data. (**a**) Bulk carrier; (**b**) container ship; (**c**) tanker.

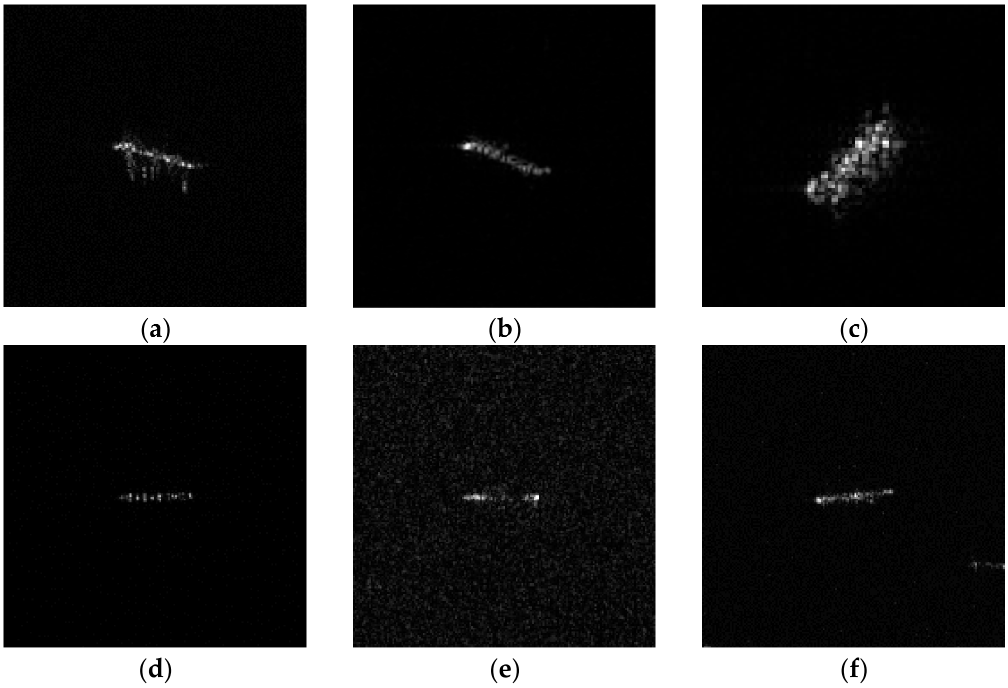

Figure 5. Six-category data. (**a**) Bulk carrier; (**b**) cargo ship; (**c**) container ship; (**d**) fishing vessel; (**e**) general cargo ship; (**f**) tanker.

3.2. Training Details

We trained DPIG-Net by 100 epochs from scratch using Adam with a learning rate of 0.0001. The network parameters were initialized by [58,59]. Samples were resized to 224 × 224 by bilinear interpolation. It is worth noting that there are a lot of other classic interpolation methods, such nearest neighbor interpolation, bicubic interpolation, and Lanczos interpolation [60]. In comparison with these methods, bilinear interpolation is able to balance interpolation performance and computational burden and hence is widely used in the computer vision community. Moreover, since this paper focused on the SAR ship classification method, resampling was not within the scope of this paper. Therefore, we selected bilinear interpolation for resampling, which was the same as many SAR ship classification methods [36,39,41,56,61–63]. The batch size was set to 16 in consideration of the theoretical guidance and hardware limitations. Specifically, in theory, the batch size should not be too small or too large [64–67]. When the batch size is too large, optimization of the network tends to be trapped at a local optimum and generalization of the trained network is weak due to the lack of randomness in gradient descent. When the batch size is too small, the speed of convergence is restricted due to excessive noise resulting from the small batch size. Therefore, the batch size is usually set as 16, 32, or 64. However, a batch size of 32 or 64 was not available due to our limited GPU memory. Hence, 16 was set as the batch size in the experiment. The multi-category cross entropy [34] served as the loss function of the network, which was defined as:

$$Loss = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{K-1} y_{i,k} \ln p_{i,k} \tag{8}$$

where $y_{i,k}$ is the ground truth of kth category of ith sample and $p_{i,k}$ is the predicted result of kth category of ith sample. N denotes the number of samples in this batch, while K denotes the number of categories. We chose multi-category cross entropy for two reasons. On the one hand, multi-category cross entropy is sensitive to wrong predictions. Specifically, when $y_{i,k} = 0$ and $p_{i,k}$ is close to 0, the loss will be much closer to positive infinity, which guides the network towards wrong predictions. On the other hand, multi-category cross entropy is more likely to avoid the vanishing gradient in classification tasks, which is suitable for network training. Specifically, the derivation of multi-category cross entropy with respect to weight in a network was suitable, whereas the value of other loss functions, such as mean squared error, tends to be extremely small in the case of classification tasks where a sigmoid or soft-max are used before loss functions. Therefore, multi-category cross entropy was selected as the loss function of the network. We reproduced other models that were basically consistent with their raw reports. The experiments were run on a personal computer (PC) with the Intel i9-9900K CPU, NVIDIA RTX2080Ti GPU, and 32G memory. We use PyTorch based on the CUDA10.1 and CUDNN7.4 framework for network training and evaluation.

3.3. Evaluation Criteria

Accuracy (Acc) was calculated to evaluate the network's ability to classify ships, which was described by:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

where TP denotes the true positives, TN denotes the true negatives, FP denotes the false positives, and FN denotes the false negatives. Significantly, we did not adapt class-related global measures such as completeness, correctness, or F1 score since class-related global measures would be affected by imbalance in sample categories. In the OpenSARShip dataset, the imbalance in ship categories was quite severe. As shown in Table 2, in the test set of three-category data, the number of tankers in the sample was 73, while that of container ships was 404. As shown in Table 3, in the test set of six-category data, the number of fishing vessels in the sample was 25, while that of cargo was 571. This severe imbalance in ship categories would force the class-related measures to pay more attention to categories with small sample numbers, which may result in a fairish indicator of the class-related global measure but poor performance of the model in reality. Therefore, we selected accuracy as the global measure. Moreover, to evaluate the ship classification ability more specifically, we selected a confusion matrix as the class-wise measure to evaluate the classification ability of each category, which was also performed in previous SAR ship classification research [36,39,41,61,62].

3.4. Classification Performance

Table 4 shows the quantitative evaluation of different models. The top-10 best results among 20 trainings were used to calculate the average and standard deviation, except for DenseNet-LRCS [62].

Table 4. Classification Performance of Different Models.

Methods	Three-Category Acc (%)	Six-Category Acc (%)	Time (ms)
Hou et al. [22]	67.41 ± 1.13	47.44 ± 2.01	4.30
GSESCNN [23]	74.98 ± 1.46	54.78 ± 2.08	4.28
Wang et al. [24]	69.27 ± 0.27	48.43 ± 3.71	4.33
HOG-ShipCLSNet [27]	78.15 ± 0.57	53.77 ± 3.63	4.52
Zeng et al. [28]	77.41 ± 1.74	55.26 ± 2.36	4.47
SE-LPN-DPFF [29]	79.25 ± 0.83	56.66 ± 1.54	5.05
Mini Hourglass Net [30]	75.44 ± 2.68	54.93 ± 2.61	4.52
PFGFE-Net [31]	<u>79.84 ± 0.53</u>	<u>56.83 ± 2.68</u>	4.95
VGGNet-Grey [49]	78.51 ± 0.93	55.80 ± 2.05	4.63
GBCNN [51]	78.84 ± 0.26	56.48 ± 1.94	4.85
DenseNet-LRCS [50]	78.00 ± 0.00	56.29 ± 0.00	5.36
DPIG-Net (Ours)	**81.28 ± 0.65**	**58.68 ± 2.02**	5.12

The best result is in bold and the second best is underlined.

From Table 4, it can be noted that the accuracy of the three-category task was obviously higher than that of the six-category task. The reason is that more categories caused more misclassification, especially in low-resolution SAR images where the characteristics of ships in different categories tended to be similar [31,42,68,69].

Moreover, the accuracy of models using polarization information is usually higher than that of models [31–33,36] that do not consider polarization information and directly input SAR images into the ship classification model. The reason is that the characteristics of ships in different polarization modes are different, which may be complemented by combining different polarization information together. However, there is one exception: HOG-ShipCLSNet adds tradition HOG features to guide the classification and hence surpasses some ship classification models that utilize polarization information. In fact, the exception of HOG-ShipCLSNet makes sense from a general point of view. That is, more information will lead to smarter decisions. Additionally, fully utilizing the polarization information may further improve the performance of SAR ship classification. As can be seen from Table 4, DPIG-Net obviously outperformed the other eleven comparative models. The second-best model offered 79.84% accuracy on the three-category task, which was still lower than our network by 1.44%, and 56.83% accuracy on the six-category task, which was still lower than our network by 1.85%. This revealed the state-of-the-art classification performance of DPIG-Net. Note that such accuracy improvement was already huge progress for the SAR ship classification community. Compared with the other methods, DPIG-Net could make full use of ship polarization information with the potential to implicitly mine useful dual polarization feature patterns for better classification accuracy.

Figure 6 shows the computational efficiency comparison of different methods. From Figure 6, it can be noted that the speeds of models without considering polarization information were usually faster than those of models using polarization information. The reason is that the utilization of polarization information usually needs more prepossessing and merging of different polarization information, which will lead to more computations and hence slower speeds. However, it is worth sacrificing a little speed for higher classification accuracy in consideration of relatively long SAR imaging processing, which usually takes several hours or days and makes the speed of SAR ship classification less important in a way than accuracy. Moreover, it can be noted in Figure 6 that DPIG-Net consumed more time (5.12 ms) to classify ships than most other methods, but it was still faster than DenseNet-LRCS [62]. Furthermore, the speed gap between DPIG-Net and other methods was relatively small (within 1 ms), so DPIG-Net might still meet practical applications.

According to our theoretical statistics of the network parameters, DPIG-Net had about 17,961,536 (~18M) parameters. This indicated that DPIG-Net might be a little heavy, which led to its longer running time in our experiments, as shown in Figure 6. Thus, speed optimization will be studied in the future.

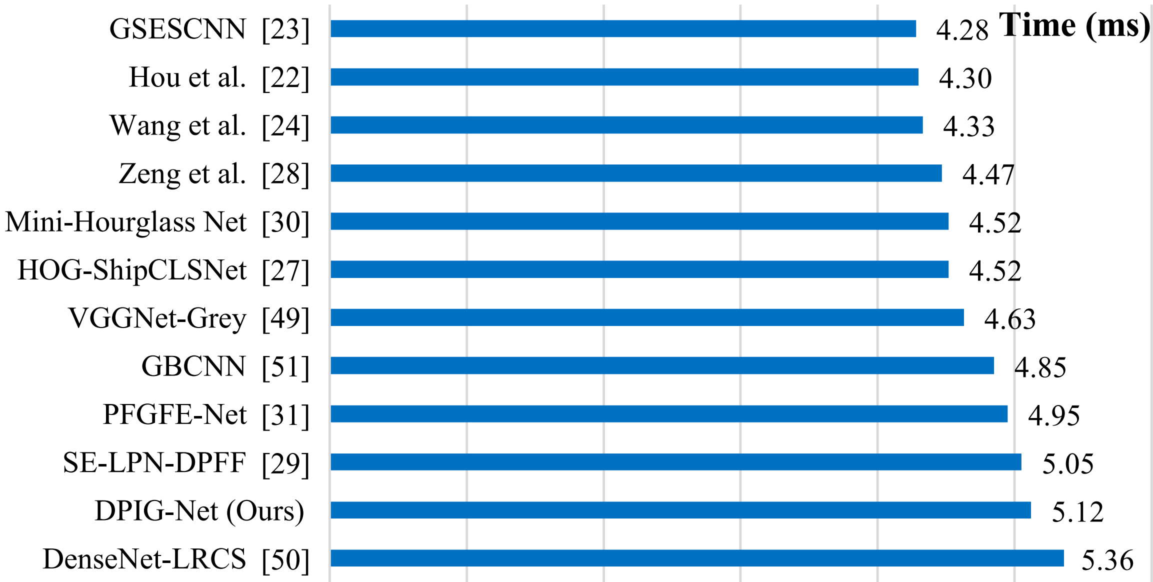

Figure 6. Classification time comparison. Time is sorted from short to long.

3.5. Confusion Matrix

Tables 5 and 6 show the confusion matrix of DPIG-Net. From Tables 5 and 6, it can be observed that DPIG-Net could successfully identify most ships, i.e., the diagonal value was greater than others at the same line in most cases, which revealed the superior ship classification ability of DPIG-Net.

Table 5. Confusion Matrix of the Three-Category Task.

True \ Predicted	**Bulk Carrier**	**Container Ship**	**Tanker**
Bulk carrier	125	21	8
Container ship	48	342	14
Tanker	11	8	54

Table 6. Confusion Matrix of the Six-Category Task.

True \ Prdicted	**Bulk Carrier**	**Cargo**	**Container Ship**	**Fishing**	**General Cargo**	**Tanker**
Bulk carrier	143	23	43	0	22	2
Cargo	69	325	19	27	83	48
Container ship	67	16	359	0	29	2
Fishing	0	2	0	22	0	1
General cargo	6	20	2	0	9	5
Tanker	13	91	1	4	19	14

Moreover, as can be seen from Table 5, container ships had the highest class-wise classification accuracy (i.e. 342/(48 + 342 + 14) = 84.65%) in the three-category task. The reason could be that container ships have a duplicate texture derived from the grid structure of the cabin and strong scattering characteristics [70], which makes container ships relatively easy to classify. Similarly, as seen in Table 6, container ships had the second highest class-wise classification accuracy (i.e. s359/(67 + 16 + 359 + 29 + 2) = 75.90%) in the six-category task, while fishing vessels had the highest class-wise classification accuracy (i.e. 22/(2 + 22 + 1) = 88.00%). Meanwhile, tankers had the lowest class-wise classification accuracy in both tasks (i.e. 54/(11 + 8 + 54) = 73.97% in the three-category task and 14/(13 + 91 + 1 + 4 + 19 + 14) = 9.86% in the six-three category task). This was a typical case in that the class-wise classification accuracy of a certain category will decrease as the number of categories increases in the dataset.

4. Discussion

4.1. Discussion on PCCAF

To confirm the effectiveness of PCCAF, we conducted ablation studies including the polarization-guided paradigm and the proposed cross-attention module. The results are shown in Table 7. As shown in Table 7, the polarization-guided paradigm offered obvious accuracy gains. Taking the six-category task as an example, I_1 (the VH polarization channel) boosted the accuracy by 1.47%, and I_3 (the polarization channel difference) boosted the accuracy by 2.67%. The combination of two inputs was better than a single input; the combination of three inputs was better than the combination of two inputs. The above showed the effectiveness of utilizing polarization information. Moreover, the offered accuracy gain was greater than some previous works [39,41]. This showed that PCCAF could make full use of the polarization information. Finally, the proposed cross-attention module could further improve the classification accuracy (~2% improvement on the six-category task), which was in line with the subjective analysis in Section 2.1. This was because the network could establish correlations between channels to extract features with more mutual recognition. As a result, the information flow between channels was promoted for better feature extraction.

Table 7. Discussion Results on PCCAF.

I_2	Polarization-Guided		Cross Attention	Three-Category Acc (%)	Six-Category Acc (%)
	I_1	I_3			
✓			–	78.45 ± 1.78	52.69 ± 2.98
	✓		–	77.86 ± 1.94	50.82 ± 1.75
		✓	–	75.28 ± 1.50	50.32 ± 2.03
✓	✓		–	79.89 ± 2.01	54.16 ± 2.57
✓	✓	✓	×	80.86 ± 1.08	56.83 ± 1.93
✓	✓	✓	✓	**81.28 ± 0.65**	**58.68 ± 2.02**

The best result is in bold.

We discussed the effect of different inputs in the main branch on the results, as shown in Table 8. It can be observed that the VV of I_2 offered better results than the others since it contained more ship scattering energy. Additionally, I_3 had the worst results, which indicated that improper utilization of the merged polarization information may backfire and the original data (i.e. VH of I_1 and VV of I_2 in OpenSARShip dataset) should serve as the foundation for SAR ship classification.

Table 8. Results of Different Main Branches in PCCAF.

Main Branch	Three-Category Acc (%)	Six-Category Acc (%)
I_1	80.02 ± 0.84	57.45 ± 1.85
$\boldsymbol{I_2}$	**81.28 ± 0.65**	**58.68 ± 2.02**
I_3	75.38 ± 1.64	51.52 ± 2.36

The best result is in bold.

We conducted another experiment to verify the advantage of feature concatenation over feature adding. The results are presented in Table 9. The former performed better than the latter, indicating that features between different polarization channels should not be added directly or they might cause feature resistance effects.

Table 9. Results of Feature Concatenation or Feature Adding in PCCAF.

Type	Three-Category Acc (%)	Six-Category Acc (%)
Feature Adding	80.65 ± 1.26	57.66 ± 2.14
Feature Concatenation	$\mathbf{81.28 \pm 0.65}$	$\mathbf{58.68 \pm 2.02}$

The best result is in bold.

Finally, we performed experiments to confirm the effectiveness of the SA-module in the cross-attention subnetwork, as shown in Table 10. The SA-module further improved the accuracy since it could enable more prominent features for multi-stage residual fusion. Furthermore, the SA-module could ease the negative effects of the SAR image characteristics of speckle noise and sea clutter in order to enhance ship saliency, as shown in Figure 2a. This was in line with the experimental results in Table 10.

Table 10. Results on Effectiveness of SA-Module.

SA-Module	Three-Category Acc (%)	Six-Category Acc (%)
×	80.98 ± 0.87	57.48 ± 2.35
✓	$\mathbf{81.28 \pm 0.65}$	$\mathbf{58.68 \pm 2.02}$

The best result is in bold.

4.2. Discussion on DRDLF

To verify the effectiveness of DRDLF, we conducted ablation studies. The results are shown in Table 11. DRDB improved the accuracy by 1.54% on the three-category task and by 2.08% on the six-category task. It could learn context information more effectively to achieve more concentrated feature fusion effects. Furthermore, the global residual learning further boosted the accuracy because it could effectively restore original feature details from the main branch I_2, which avoided possible feature loss from multiple convs and pooling operations. Significantly, the basic operation behind global residual learning is actually feature adding, which was different from the feature concatenation in PCCAF. The difference was derived from the fact that we regarded the feature of global residual learning as a residual correction [71] to the original output in DRDB, while the feature of different polarization channels in PCCAF was regarded as three complementary features extracted by different feature extraction subnetworks, just as we concatenated the output of different convolution kernels in the same conv layer together instead of adding them together [72].

Table 11. Discussion Results on DRDLF.

DRDB	Global Residual Learning	Three-Category Acc (%)	Six-Category Acc (%)
–	–	79.44 ± 0.82	55.38 ± 1.98
✓		80.98 ± 0.63	57.46 ± 2.25
✓	✓	$\mathbf{81.28 \pm 0.65}$	$\mathbf{58.68 \pm 2.02}$

The best result is in bold.

We determined the number of DRDBs empirically via experiments, as shown in Table 12. Table 12 indicates that the accuracy first increased and then decreased as the number of DRDBs increased. One possible reason is that excessive DRDBs may lead to overfitting for its large number of network parameters. Another possible reason is that a more dilated convolution brought by DRDB may cause adverse effects on the network. Specifically, although a dilated convolution can broaden the receptive field and extract the contextual information of features, which is helpful to suppress the effects of speckle noise in SAR images [73], it may dilute the spatial details of ships. All in all, it is a trade-off when it comes to the number of DRDBs. For this study, we set the number of DRDBs to the optimal value of 3.

Table 12. Results on Different Numbers of DRDBs.

Number	Three-Category Acc (%)	Six-Category Acc (%)
1	80.69 ± 0.48	57.05 ± 2.26
2	80.99 ± 0.32	57.87 ± 2.18
3	**81.28 ± 0.65**	**58.68 ± 2.02**
4	81.02 ± 0.17	58.27 ± 3.01
5	80.78 ± 0.84	58.02 ± 3.18

The best result is in bold.

5. Conclusions

In this paper, DPIG-Net was proposed for dual-polarization-guided SAR ship classification. DPIG-Net exploits available dual-polarization information to adaptively model the correlations of different polarization channels, implicitly mining useful dual-polarization feature patterns for feature extraction from Sentinel-1 to guide better ship classification performance. PCCAF was designed for better dual-polarization feature extraction through a cross-attention network. DRDLF was designed for fine dual-polarization feature fusion through multiple dilated convolutions and residual dense connections. We performed extensive experiments on the public OpenSARShip dataset to confirm the effectiveness of DPIG-Net. The results showed that DPIG-Net achieved 81.28% accuracy in the three-category task and 58.68% accuracy in the six-category task, surpassing the second-best model PFGFE-Net by 1.44% in the three-category task and 1.85% in the six-category task. These findings indicated the state-of-the-art ship classification ability of DPIG-Net and the effectiveness of exploiting SAR polarization data.

Our future work will be as follows:

1. Strive to improve the speed of DPIG-Net without sacrificing the classification accuracy of ships.
2. Study the generalization of DPIG-Net for more polarization information.
3. Study how to combine traditional handcraft features and different polarization information together for higher classification accuracy.
4. Explore a transformer-related feature extraction subnetwork for better modeling of long-range dependencies among different parts of ships, such as prows and sterns, to improve the performance of ship classification.
5. Strive to improve the accuracy of tanker classification in the OpenSARShip dataset.

Author Contributions: Conceptualization, X.K., T.Z. and Z.S.; methodology, Z.S., X.K.; software, Z.S., X.K.; validation, Z.S., T.Z. and X.K.; formal analysis, T.Z.; investigation, Z.S., X.K.; resources, Z.S., X.K.; data curation, Z.S.; writing—original draft preparation, X.K. and Z.S.; writing—review and editing, X.K. and Z.S.; visualization, Z.S.; supervision, T.Z.; project administration, T.Z.; funding acquisition, T.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the editors and the four anonymous reviewers for their valuable comments that greatly improved our manuscript. They would also like to thank Liu from Shanghai Jiao Tong University [1] and Xu from Fudan University [14], respectively, for providing the OpenSARShip-1.0 and FUSAR-Ship datasets. Moreover, the authors would like to thank Wang [35] from the University of Electronic Science and Technology of China for directly providing the three-category OpenSARShip-1.0 dataset. The authors would also like to thank Xiaoling Zhang and Jun Shi from the University of Electronic Science and Technology of China, for providing much writing guidance and code guidance.

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

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic Aperture Radar
DPIG-Net	Dual-Polarization Information-Guided Network
PCCAF	Polarization Channel Cross-Attention Framework
DRDLF	Dilated Residual Dense Learning Framework
ATR	Automatic Target Recognition
VV	Vertical–Vertical
VH	Vertical–Horizontal
BN	Batch Normalization
SA-Module	Self-Attention Module
DRDB	Dilated Residual Dense Block
AIS	Automatic Identification System
SLC	Single Look Complex

References

1. Jiang, M.; Yang, X.; Dong, Z.; Fang, S.; Meng, J. Ship classification based on superstructure scattering features in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 616–620. [CrossRef]
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. [CrossRef]
3. 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. *Remote Sens.* **2022**, *14*, 3345. [CrossRef]
4. Di, Y.; Jiang, Z.; Zhang, H. A public dataset for fine-grained ship classification in optical remote sensing images. *Remote Sens.* **2021**, *13*, 747. [CrossRef]
5. Zhang, X.; Lv, Y.; Yao, L.; Xiong, W.; Fu, C. A new benchmark and an attribute-guided multilevel feature representation network for fine-grained ship classification in optical remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1271–1285. [CrossRef]
6. 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. [CrossRef]
7. Pelich, R.; Chini, M.; Hostache, R.; Matgen, P.; Lopez-Martinez, C.; Nuevo, M.; Eiden, G. Large-scale automatic vessel monitoring based on dual-polarization sentinel-1 and AIS data. *Remote Sens.* **2019**, *11*, 1078. [CrossRef]
8. 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*, 1018. [CrossRef]
9. Zhang, T.; Zhang, X. ShipDeNet-20: An only 20 convolution layers and <1-MB lightweight SAR ship detector. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1234–1238.
10. Xu, X.; Zhang, X.; Zhang, T.; Yang, Z.; Shi, J.; Zhan, X. Shadow-Background-Noise 3D Spatial Decomposition Using Sparse Low-Rank Gaussian Properties for Video-SAR Moving Target Shadow Enhancement. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
11. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. High-Speed Ship Detection in SAR Images by Improved Yolov3. In Proceedings of the 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, Chengdu, China, 14–15 December 2019; IEEE: Piscataway, NJ, USA, 2019.
12. 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. [CrossRef]
13. Zhang, T.; Zhang, X.; Shi, J.; Wei, S.; Wang, J.; Li, J. Balanced Feature Pyramid Network for Ship Detection in Synthetic Aperture Radar Images. In Proceedings of the 2020 IEEE Radar Conference (RadarConf20), Florence, Italy, 21–25 September 2020; IEEE: Piscataway, NJ, USA, 2020.
14. Amrani, M.; Bey, A.; Amamra, A. New SAR target recognition based on YOLO and very deep multi-canonical correlation analysis. *Int. J. Remote Sens.* **2022**, *43*, 5800–5819. [CrossRef]
15. Amrani, M.; Chaib, S.; Omara, I.; Jiang, F. Bag-of-visual-words based feature extraction for SAR target classification. In Proceedings of the Ninth International Conference on Digital Image Processing (ICDIP 2017), Hong Kong, China, 19–22 May 2017; SPIE: Washington, DC, USA, 2017.
16. Amrani, M.; Jiang, F. Deep feature extraction and combination for synthetic aperture radar target classification. *J. Appl. Remote Sens.* **2017**, *11*, 042616. [CrossRef]
17. Amrani, M.; Jiang, F.; Xu, Y.; Liu, S.; Zhang, S. SAR-oriented visual saliency model and directed acyclic graph support vector metric based target classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3794–3810. [CrossRef]
18. Amrani, M.; Yang, K.; Zhao, D.; Fan, X.; Jiang, F. An efficient feature selection for SAR target classification. In Proceedings of the Advances in Multimedia Information Processing—PCM 2017: 18th Pacific-Rim Conference on Multimedia, Harbin, China, 28–29 September 2017; Revised Selected Papers, Part II 18. Springer: Berlin/Heidelberg, Germany, 2018.

19. Lang, H.; Zhang, J.; Zhang, X.; Meng, J. Ship classification in SAR image by joint feature and classifier selection. *IEEE Geosci. Remote Sens. Lett.* **2015**, *13*, 212–216. [CrossRef]
20. Margarit, G.; Mallorqui, J.J.; Rius, J.M.; Sanz-Marcos, J. On the usage of GRECOSAR, an orbital polarimetric SAR simulator of complex targets, to vessel classification studies. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3517–3526. [CrossRef]
21. Margarit, G.; Tabasco, A. Ship classification in single-pol SAR images based on fuzzy logic. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3129–3138. [CrossRef]
22. Xing, X.; Ji, K.; Zou, H.; Chen, W.; Sun, J. Ship classification in TerraSAR-X images with feature space based sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1562–1566. [CrossRef]
23. Xu, Y.; Lang, H.; Niu, L.; Ge, C. Discriminative adaptation regularization framework-based transfer learning for ship classification in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1786–1790. [CrossRef]
24. Chaturvedi, S.K.; Yang, C.S.; Ouchi, K.; Shanmugam, P. Ship recognition by integration of SAR and AIS. *J. Navig.* **2012**, *65*, 323–337. [CrossRef]
25. Zhang, H.; Tian, X.; Wang, C.; Wu, F.; Zhang, B. Merchant vessel classification based on scattering component analysis for COSMO-SkyMed SAR images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1275–1279. [CrossRef]
26. Zhao, Z.; Ji, K.; Xing, X.; Chen, W.; Zou, H. Ship classification with high resolution TerraSAR-X imagery based on analytic hierarchy process. *Int. J. Antennas Propag.* **2013**, *2013*, 698370. [CrossRef]
27. Wu, F.; Wang, C.; Jiang, S.; Zhang, H.; Zhang, B. Classification of vessels in single-pol COSMO-SkyMed images based on statistical and structural features. *Remote Sens.* **2015**, *7*, 5511–5533. [CrossRef]
28. Zhang, T.; Zhang, X. HTC+ for SAR ship instance segmentation. *Remote Sens.* **2022**, *14*, 2395. [CrossRef]
29. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
30. Zhang, T.; Zhang, X. A full-level context squeeze-and-excitation ROI extractor for SAR ship instance segmentation. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4506705. [CrossRef]
31. Hou, X.; Ao, W.; Song, Q.; Lai, J.; Wang, H.; Xu, F. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Sci. China Inf. Sci.* **2020**, *63*, 140303. [CrossRef]
32. Huang, G.; Liu, X.; Hui, J.; Wang, Z.; Zhang, Z. A novel group squeeze excitation sparsely connected convolutional networks for SAR target classification. *Int. J. Remote Sens.* **2019**, *40*, 4346–4360. [CrossRef]
33. Wang, Y.; Wang, C.; Zhang, H. Ship classification in high-resolution SAR images using deep learning of small datasets. *Sensors* **2018**, *18*, 2929. [CrossRef]
34. Wang, C.; Shi, J.; Zhou, Y.; Yang, X.; Zhou, Z.; Wei, S.; Zhang, X. Semisupervised learning-based SAR ATR via self-consistent augmentation. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4862–4873. [CrossRef]
35. He, J.; Wang, Y.; Liu, H. Ship classification in medium-resolution SAR images via densely connected triplet CNNs integrating Fisher discrimination regularized metric learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3022–3039. [CrossRef]
36. Zhang, T.; Zhang, X.; Ke, X.; Liu, C.; Xu, X.; Zhan, X.; Wang, C.; Ahmad, I.; Zhou, Y.; Pan, D.; et al. HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5210322. [CrossRef]
37. Xu, X.; Zhang, X.; Shao, Z.; Shi, J.; Wei, S.; Zhang, T.; Zeng, T. A Group-Wise Feature Enhancement-and-Fusion Network with Dual-Polarization Feature Enrichment for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 5276. [CrossRef]
38. Zeng, L.; Zhu, Q.; Lu, D.; Zhang, T.; Wang, H.; Yin, J.; Yang, J. Dual-polarized SAR ship grained classification based on CNN with hybrid channel feature loss. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 4011905. [CrossRef]
39. 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*, 4019905. [CrossRef]
40. Xiong, G.; Xi, Y.; Chen, D.; Yu, W. Dual-polarization SAR ship target recognition based on mini hourglass region extraction and dual-channel efficient fusion network. *IEEE Access* **2021**, *9*, 29078–29089. [CrossRef]
41. Zhang, T.; Zhang, X. A polarization fusion network with geometric feature embedding for SAR ship classification. *Pattern Recognition* **2022**, *123*, 108365. [CrossRef]
42. Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; Yu, W. OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *11*, 195–208. [CrossRef]
43. Wu, S.; Xu, J.; Tai, Y.W.; Tang, C.K. Deep high dynamic range imaging with large foreground motions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
44. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [CrossRef]
45. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
46. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015.
47. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
48. Ke, X.; Zhang, X.; Zhang, T. GCBANET: A global context boundary-aware network for SAR ship instance segmentation. *Remote Sens.* **2022**, *14*, 2165. [CrossRef]

49. Xie, X.; Li, L.; An, Z.; Lu, G.; Zhou, Z. Small Ship Detection Based on Hybrid Anchor Structure and Feature Super-Resolution. *Remote Sens.* **2022**, *14*, 3530. [CrossRef]
50. Sharma, S.; Sharma, S.; Athaiya, A. Activation functions in neural networks. *Towards Data Sci.* **2017**, *6*, 310–316. [CrossRef]
51. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 23 June 2018.
52. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; Volume 30.
53. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 23 June 2018.
54. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sens.* **2021**, *13*, 2771. [CrossRef]
55. Shao, Z.; Zhang, X.; Wei, S.; Shi, J.; Ke, X.; Xu, X.; Zhan, X.; Zhang, T.; Zeng, T. Scale in Scale for SAR Ship Instance Segmentation. *Remote Sens.* **2023**, *15*, 629. [CrossRef]
56. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
57. Ke, X.; Zhang, X.; Zhang, T.; Shi, J.; Wei, S. Sar Ship Detection Based on Swin Transformer and Feature Enhancement Feature Pyramid Network. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022.
58. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Las Condes, Chile, 7–13 December 2015.
59. Zhang, T.; Zhang, X. High-speed ship detection in SAR images based on a grid convolutional neural network. *Remote Sens.* **2019**, *11*, 1206. [CrossRef]
60. Han, D. Comparison of commonly used image interpolation methods. In Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), Hangzhou, China, 22–23 March 2013; Atlantis Press: Amsterdam, The Netherlands, 2013.
61. Zhang, T.; Zhang, X. Integrate Traditional Hand-Crafted Features into Modern CNN-based Models to Further Improve SAR Ship Classification Accuracy. In Proceedings of the 2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Bali, Indonesia, 1–3 November 2021.
62. Zhang, T.; Zhang, X. Injection of traditional hand-crafted features into modern CNN-based models for SAR ship classification: What, why, where, and how. *Remote Sens.* **2021**, *13*, 2091. [CrossRef]
63. He, J.; Chang, W.; Wang, F.; Liu, Y.; Wang, Y.; Liu, H.; Li, Y.; Liu, L. Group Bilinear CNNs for Dual-Polarized SAR Ship Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4508405. [CrossRef]
64. Zhang, T.; Zhang, X.; Shi, J. Balance scene learning mechanism for offshore and inshore ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4004905. [CrossRef]
65. Kandel, I.; Castelli, M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express* **2020**, *6*, 312–315. [CrossRef]
66. Zhang, T.; Zhang, X. A mask attention interaction and scale enhancement network for SAR ship instance segmentation. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
67. Radiuk, P.M. Impact of training set batch size on the performance of convolutional neural networks for diverse datasets. *Inf. Technol. Manag. Sci.* **2017**, *20*, 20–24. [CrossRef]
68. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]
69. Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1. 0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images. *Remote Sens.* **2020**, *12*, 2997. [CrossRef]
70. Lin, H.; Song, S.; Yang, J. Ship classification based on MSHOG feature and task-driven dictionary learning with structured incoherent constraints in SAR images. *Remote Sens.* **2018**, *10*, 190. [CrossRef]
71. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
72. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Türkiye, 21–23 August 2017.
73. Ke, X.; Zhang, X.; Zhang, T.; Shi, J.; Wei, S. SAR ship detection based on an improved faster R-CNN using deformable convolution. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021.