Group Bilinear CNNs for Dual-Polarized SAR Ship Classification

Jinglu He, Wenlong Chang, Fuping Wang, Ying Liu, *Senior Member, IEEE*, Yinghua Wang, *Senior Member, IEEE*, Hongwei Liu, *Member, IEEE*, Yinghua Li, and Lei Liu, *Member, IEEE*

***Abstract*— Ship classification from synthetic aperture radar (SAR) images tends to be a hotspot in the remote sensing community. Currently, more efforts have been made to the single-polarization (single-pol) SAR ship classification with limited performance. This letter proposes to explore the dual-polarization (dual-pol) SAR images for better ship classification. To be specific, a novel group bilinear convolutional neural network (GBCNN) model is developed to deeply extract discriminative second-order representations of ship targets from the pairwise vertical-horizontal polarization (VH) and vertical-vertical polarization (VV) SAR images. In particular, the deep bilinear features are efficiently acquired by performing the bilinear pooling on subgroups of deep feature maps derived from the single-pol SAR images (self-bilinear pooling) and dual-pol SAR images (cross-bilinear pooling). To fully explore the polarization information, the multipolarization fusion loss (MPFL) is constructed to train the proposed model for superior SAR ship representation learning. Through extensive experiments, the proposed method can achieve an overall accuracy (OA) of 88.80% and 66.90% on the three- and five-category dual-pol OpenSARShip datasets, which outperforms the state-of-the-art methods by at least 2.00% and 2.37%, respectively.**

***Index Terms*— Convolutional neural network (CNN), group bilinear pooling, multipolarization fusion loss (MPFL), ship classification, synthetic aperture radar (SAR).**

I. Introduction

OVER the last couple of decades, the successful launches of several advanced synthetic aperture radar (SAR) satellites have greatly boosted the development of SAR remote sensing [1]. In the maritime surveillance area, ship classification using SAR images is a more advanced and challenging task, which has received increasing attention in recent years [2]–[20]. With the rapid development of data volume and interpretation algorithms, SAR ship classification has progressed, to some extent, from the traditional handcrafted methods to the automatic data-driven deep learning (DL) paradigm [2]–[20], where the ship signature representation is of core importance, which will be specifically reviewed in the following.

To acquire discriminative SAR ship representation, on the one hand, most early works made efforts to design traditional handcrafted features, derived from the geometric properties and scattering characteristics. These features, such as the geometric structures [2], [3] and backscattering intensity density [4], [5], were mostly evaluated on the limited high-resolution (HR) SAR ship samples, whose performance degraded significantly if applied to the medium-resolution (MR) SAR images [6]. Moreover, large-scale SAR ship data impose imperative challenges on the aforementioned features. To address the above issues, lots of researchers in the SAR remote sensing regime strive to apply the impressive DL techniques to acquire more discriminative SAR ship representations, on which solid foundations have been built for superior classification performance. For example, in [7] and [8], two domain-specific convolutional neural network (CNN) models were proposed to perform ship classification in HR and MR SAR images. Bentes *et al.* [9] proposed a multiple-input resolution CNN model to classify maritime targets in HR SAR images. Wang *et al.* [10] applied the DL methods for small sample SAR ship classification.

On the other hand, seeking for boost mechanisms for available SAR ship features is also an effective approach to promote the classification accuracy. He *et al.* [8] proposed to employ the metric learning (ML) scheme on SAR ship deep features to enhance the discriminability of different ship types. Similarly, Yan *et al.* [11] also verified the effectiveness of ML-based learning framework for small sample SAR ship classification. As a counterpart work, Xu and Lang [12], [13] proved that the traditional distance ML methods can boost the naive geometric features [6] for better SAR ship classification. In addition, Zhang and Zhang [14], [15] and Zhang *et al.* [16] proposed that it is also important to inject the handcrafted features into modern CNN models to further improve the SAR ship classification accuracy.

To summarize, most of the existing classification methods lay more attention on the intensity information of single-polarization (single-pol) SAR images for ship representation while ignoring the intrinsic information mining of the imaging process such as the polarization property. As pioneer works, Xi *et al.* [17], [18] proposed a novel feature-loss double fusion Siamese network (DFSN) with a fusion loss appended for dual-polarized SAR ship classification. Zeng *et al.* [19] also developed a hybrid channel feature loss (HCFL) for

Manuscript received December 25, 2021; revised March 21, 2022; accepted May 17, 2022. Date of publication May 26, 2022; date of current version June 8, 2022. This work was supported in part by the Scientific Research Program Funded by Shaanxi Provincial Education Department under Grant 21JK0911, in part by the Natural Science Basic Research Program of Shaanxi Province under Grant 2022JQ-687, and in part by the Ph.D. Startup Fund of the Xi'an University of Posts and Telecommunications under Grant 101-315020049. *(Corresponding author: Jinglu He.)*

Jinglu He, Wenlong Chang, Fuping Wang, Ying Liu, Yinghua Li, and Lei Liu are with the School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China (e-mail: jlhe20@xupt.edu.cn).

Yinghua Wang and Hongwei Liu are with the National Laboratory of Radar Signal Processing and Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, Xi'an 710071, China.

Digital Object Identifier 10.1109/LGRS.2022.3178080

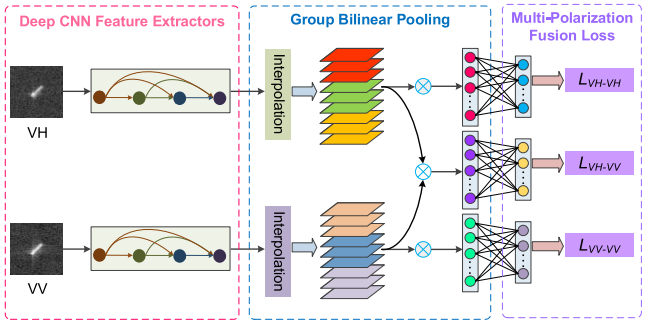


Fig. 1. Overall framework of the proposed method. ⊗ means the group bilinear pooling operation among the subgroups of the input feature maps.

deep CNN features to jointly utilize the information contained in the vertical-horizontal polarization (VH) and vertical-vertical polarization (VV) SAR images. In addition, Zhang and Zhang [20] proposed a squeeze-and-excitation Laplacian pyramid network with dual-polarization (dual-pol) feature fusion (SE-LPN-DPFF) for SAR ship classification. These preliminary works made a good starting point to exploit the imaging mechanism for prominent SAR ship classification, and we firmly believe that there is still room for performance improvement by deeply exploring the polarization information.

In this letter, we propose a group bilinear CNN (GBCNN) model to extensively extract the ship representations from dual-pol SAR images. Different from [17]–[20], we apply the bilinear pooling [21] to fuse the deep CNN features of a paired single-pol SAR images from the feature level, which can capture the second-order interactions between different feature channels to further improve the classification performance. Compared to [21], the input layer in GBCNN takes two different single-pol SAR images as input, and we propose to perform the bilinear pooling on subgroups of the feature maps. Hence, the number of parameters in the following embedding layer is effectively reduced, and thus, the computational efficiency and classification performance are enhanced. To summarize, our main contributions are given as follows.

1) Our method is the first framework applying the bilinear pooling for SAR ship classification. The group bilinear pooling is further proposed to achieve high model efficiency and more discriminative ship representation.
2) As decision-level fusion strategy, a multipolarization fusion loss (MPFL) is proposed to fully enhance the utilization of dual-pol information for prominent model learning.
3) Extensive experiments demonstrate the superiority of the proposed method over the state-of-the-art methods.

The remainder of this letter is organized as follows. Section II presents the proposed methodology in detail. Section III gives the experiments and analyses. Conclusion and future work are finally summarized in Section IV.

II. Methodology

A. Framework of the Proposed Methodology

As shown in Fig. 1, there are mainly three modules in the proposed dual-polarized SAR ship classification method. First, the VH and VV polarization SAR images of a ship target are separately sent to two structure-equal CNNs, from which the deep feature maps are extracted. Second, to fully mine the information contained in the deep features of SAR images and to reduce the network complexity, the bilinear pooling is applied to subgroups of the feature maps of two single-pol SAR ship images in individual and mutual ways, and thus, three bilinear vectors are obtained. Note that the input feature maps are first interpolated by a 1×1 convolution to ensure that the number of feature maps in each group is an integer. Third, the three bilinear vectors are inputted to a parameter-shared embedding layer to further extract high-level semantic representations, which are then processed by the Softmax classifier [8] for fully supervised network training.

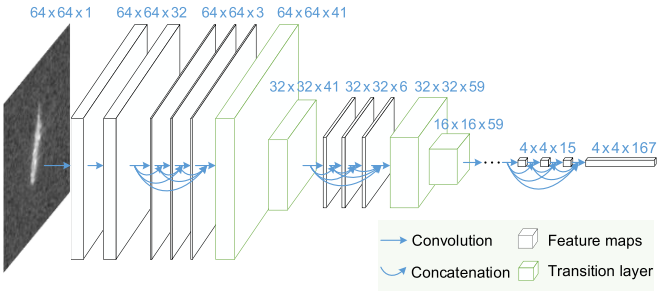


Fig. 2. Schematic architecture of the backbone network.

B. Backbone Network Architecture

The densely connected CNN elaborately modified for MR SAR images (SAR-DenseNet) proposed in our previous work [8] is favorably utilized as the backbone network for deep CNN feature extraction. Note that, considering the limited amount of dual-polarized SAR ship data and network training efficiency, we take a simplified network version that is referred to as SAR-DenseNet-v1 (see Fig. 2). To be specific, there are mainly five dense blocks (DBs) with each containing three densely connected convolution layers. Following each of the first four DBs, there is a transition layer, which consists of a convolution layer and an average pooling layer. The growth rates in the five DBs are sequentially set to 3, 6, 9, 12, and 15. The filter size of all the convolution layers, except for specification, is set to 3×3. The feature maps of the last DB are then passed to the group bilinear pooling module.

C. Group Bilinear Pooling

Given the input feature maps $\boldsymbol{F} \in \mathbb{R}^{H \times W \times D}$, where H, W, and D, respectively, represent the height, width, and the number of channels, the bilinear pooling on $\boldsymbol{F}$ itself to extract the second-order statistics is defined by the following outer product [21]:

$$\boldsymbol{B}(\boldsymbol{F}, \boldsymbol{F}) = \frac{1}{HW} \sum_{i=1}^{HW} \boldsymbol{f}_i \boldsymbol{f}_i^T \tag{1}$$

where $\boldsymbol{f}_i \in \mathbb{R}^D, i = 1, 2, \ldots, HW$ is a feature vector along the channel dimension at the spatial location i of the input feature maps and $\boldsymbol{B} \in \mathbb{R}^{D \times D}$ is the global feature descriptor. As in [21], the vectorized global feature descriptor $\boldsymbol{b} = \text{vec}(\boldsymbol{B})$ is then successively processed by the elementwise signed square root $\boldsymbol{y} = \text{sign}(\boldsymbol{b})(|\boldsymbol{b}|)^{1/2}$ and l_2 normalization $\boldsymbol{z} = \boldsymbol{y}/\|\boldsymbol{y}\|_2$, where $\boldsymbol{z}$ is the resulting global bilinear vector.

One can indeed find that the size of the bilinear vector can go huge if the channel number D of the input feature

maps is very large, which can induce great parameters in the following fully connected (FC) layer. Moreover, due to the symmetric property of the global feature descriptor [22], there is lot of redundant information in the resulting bilinear vectors. Hence, we propose to apply the bilinear pooling on subgroups of the feature maps and then concatenate the sub-bilinear vectors to acquire compact bilinear vectors. Suppose that the input feature maps F are divided into G groups with each containing d subfeature maps, i.e., $\boldsymbol{F} = \{\boldsymbol{F}^i | \boldsymbol{F}^i \in \mathbb{R}^{H\times W\times d}, i = 1, 2, \ldots, G\}$. Then, the compact bilinear vector can be obtained as follows:

$$z^G = \left[\boldsymbol{B}(\boldsymbol{F}^i_{\text{pol1}}, \boldsymbol{F}^j_{\text{pol2}})\right]_{\substack{1 \le i \le G \\ i \le j \le G}} \tag{2}$$

where $[\cdot]$ refers to the concatenation of the vectorized sub-bilinear vectors between either two subgroups of the interpolated feature maps resulting from two single-pol SAR images (pol1 and pol2 indicate the combination of the VH and VV polarization modes, respectively).

Considering the methodologies of [23] and [24], the proposed group bilinear pooling can implicitly map the class information into the subgroups of the feature maps. Thus, we set the group number G to the number of ship types reasonably in the experiments. In addition, the compact bilinear vector is approximately equivalent to extract the upper triangular part of the global feature descriptor computed by the full feature maps (denoted as FBCNN). The size of the compact bilinear vector is equal to $(D\ /\ G)^2\ (G\ (G\ -\ 1)/2\ +\ G) = (1\ +\ 1/G)D^2/2$. Thus, if the group number is large enough, the size of the compact bilinear vector is about half that of the global bilinear vector, which is beneficial to the network efficiency.

D. Multipolarization Fusion Loss

To sufficiently employ the polarization information, we propose to develop a novel MPFL to supervise the training of the network. Specifically, inspired by [8], the three compact bilinear vectors are passed through a weight-shared embedding layer to extract high-level semantic features. Then, the softmax log loss [8] for the deep embeddings is computed as follows:

$$L = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i^T \log(\boldsymbol{S}(\boldsymbol{x}_i^{\text{pol1}}, \boldsymbol{x}_i^{\text{pol2}})) \tag{3}$$

where $\boldsymbol{y}_i$ is a one-hot vector representing the ground-truth label of the corresponding ship target, N is the total number of training samples, and $\boldsymbol{S}(\boldsymbol{x}_i^{\text{pol1}}, \boldsymbol{x}_i^{\text{pol2}})$ is the output of the softmax function for two/one single-pol SAR images. Finally, the MPFL loss is constructed as

$$L_{\text{MPFL}} = \alpha(L_{\text{VH-VH}} + L_{\text{VV-VV}}) + (1-\alpha)L_{\text{VH-VV}} \tag{4}$$

where the subscript of L on the right-hand side indicates the combination way of the single-pol SAR image for bilinear pooling operation and α is a hyperparameter to balance the importance between the self- and cross-bilinear pooling terms. By backpropagating the MPFL values, the network parameters can learn better from the dual-pol SAR images for excellent ship classification. For inference, the predictions from the three MPFL branches are averaged for the final classification of different ship types, which is proven effective both in our experiments and by those in [16].

TABLE I
STATISTICS OF THE PAIRWISE EXPERIMENTAL DATASET

Category	Tanker	Container ship	Bulk carrier	Cargo	General cargo
Training	280	167	632	707	91
Test	80	50	170	270	30
Total	360	217	802	977	121

The first three categories are used for the three-category classification task.

III. EXPERIMENTS AND ANALYSIS

This section performs extensive experiments to evaluate the effectiveness of the proposed method. The network training and inference were conducted by the DL library of TensorFlow with the NVIDIA GeForce RTX 2080 Ti GPU for acceleration.

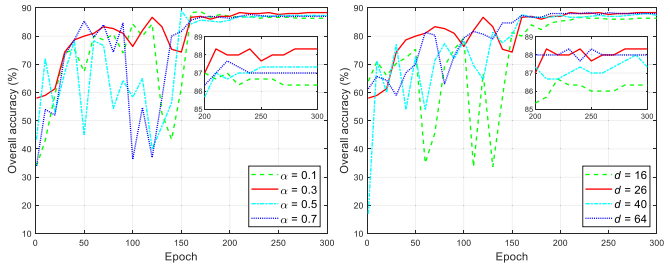
A. Data Preparation and Training Details

As most of the recent works [8], [11], [14]–[20], the three- and five-category SAR ship datasets in [8] that are curated from the OpenSARShip database [25] are readily applied to validate our method, in which each ship target is represented by pairwise VH and VV ground range detected (GRD) amplitude data samples. Similar to [8], the total number of ship targets in each classification task is randomly divided into five groups of training and test sets. The results are reported by the mean and standard deviation of the five splits. The numbers of pairwise ship samples for each category are listed in Table I. The data augmentation and normalization in [8] are applied here, and the numbers of the augmented pairwise training samples for three and five categories are 4008 and 3640, respectively. The training details keep the same as that in [8] except that the batch size is set to 32 for better dual-pol information mining.

B. Evaluation Metrics and Parameter Setting

To evaluate the effectiveness of the proposed method, we select the most commonly used metrics, such as the overall accuracy (OA) being defined as the ratio of the number of the correctly classified ship samples to that of the total samples in the test set. The precision (P), recall (R), and F_1 measure are also adopted for comprehensive evaluation [26]. Note that the P, R, and F_1 metrics are calculated using the one-versus-the-rest approach [27]. To implement, we train the networks three times for each of the randomly selected train–test datasets, and the median value is recorded for each metric.

We first validate the selection of the hyperparameters. Specifically, α in (4) and the number of feature maps in each group (i.e., d) are chosen from the sets of $\{0.1, 0.3, 0.5, 0.7\}$ and $\{16, 26, 40, 64\}$, respectively. For the sake of relative performance evaluation, we only perform this experiment on the first train–test dataset. Fig. 3 shows the OA curves for different values of the two parameters. We can conclude that the two hyperparameters should choose moderate values even though the network performance is less sensitive to the parameter change. Specifically, the parameter α should be less



(a) (b)

Fig. 3. Parameter setting evaluation on (a) α and (b) d. Different curves are plotted by sampling the OA values every ten epochs.

TABLE II

OVERALL ACCURACIES AND STANDARD DEVIATIONS (%) OF THE THREE-CATEGORY TEST SET FOR ABLATION STUDY

Model	Loss	Polarization	OA
SAR-DenseNet-v1	softmax log-loss	VH	85.47 ± 1.95
SAR-DenseNet-v1	softmax log-loss	VV	83.53 ± 2.14
SAR-DenseNet-v1	softmax log-loss	VH + VV	82.40 ± 2.15
GBCNN	$L_{\text{VH-VV}}$	VH + VV	86.27 ± 1.36
FBCNN	L_{MPFL}	VH + VV	88.13 ± 1.35
GBCNN	L_{MPFL}	VH + VV	**88.80 ± 1.15**

than 0.5 to leverage more mutual information between the dual-pol SAR images. The number of subfeature maps should be chosen to balance the model efficacy and complexity. Thus, we appropriately set the parameters of α and d to 0.3 and 26, respectively. Note that the adaptive selection of the hyperparameters will be considered in our future work.

C. Ablation Study for the Proposed Method

This section will perform an ablation study to extensively verify the effectiveness of the proposed method. Table II shows the OA values of the SAR-DenseNet-v1 trained by the softmax log loss on the separate and combined single-pol SAR images, the GBCNN trained by the $L_{\text{VH}-\text{VV}}$ and the L_{MPFL} losses, as well as the bilinear model operated on the full feature maps trained by the L_{MPFL} loss. All the training details are determined as mentioned above.

From Table II, we can draw the following conclusions. First, the proposed SAR-DenseNet-v1 has superior performance for the VH images compared with the VV ones, which indicates that the densely connected property is conducive to capture the context information as the VH images possess high signal-to-noise ratio [20]. Second, jointly exploiting the dual-pol SAR images by the GBCNN model with $L_{\text{VH}-\text{VV}}$ is better than that use the single-pol SAR images in the separate and directly combined manners. Third, the proposed GBCNN model trained by the MPFL achieves the optimal OA of as high as 88.80% for the three-category ship classification, being better than that of the FBCNN model by 0.67%. To summarize, the proposed method could sufficiently explore the multipolarization information for outstanding SAR ship classification.

D. Comparison With the State-of-the-Art

To thoroughly evaluate the efficacy of the proposed method, we present the experimental results by comparison with the state-of-the-art methods, including the DFSN [17], [18], the HCFL [19], and the SE-LPN-DPFF [20], for task-specific consideration. To make a fair comparison, we reimplement the above methods. In particular, the training schemes in DFSN and HCFL are directly applied to the SAR-DenseNet-v1 backbone network. Unlike the original implementation, the DFSN and HCFL utilized two separate CNNs to extract features as we found that the initial weight-shared scheme cannot make the network convergent well. The experimental results of the comparing methods on the three- and five-category classification tasks are shown in Tables III and IV, respectively.

TABLE III

COMPARISON WITH STATE-OF-THE-ART METHODS ON THE THREE-CATEGORY TEST SET (%)

Method	Metrics (mean ± std.)			
	R	P	F_1	OA
DFSN	82.73 ± 1.35	86.41 ± 2.86	84.51 ± 1.89	86.80 ± 1.32
HCFL	83.08 ± 1.70	85.08 ± 0.79	84.06 ± 1.02	86.60 ± 0.86
SE-LPN-DPFF	69.51 ± 1.84	62.38 ± 1.33	65.76 ± 1.56	63.87 ± 1.30
GBCNN	**85.25 ± 2.04**	**88.04 ± 2.04**	**86.61 ± 1.77**	**88.80 ± 1.15**

TABLE IV

COMPARISON WITH STATE-OF-THE-ART METHODS ON THE FIVE-CATEGORY TEST SET (%)

Method	Metrics (mean ± std.)			
	R	P	F_1	OA
DFSN	55.07 ± 2.51	56.11 ± 2.22	55.57 ± 2.07	64.53± 1.88
HCFL	56.05 ± 1.86	54.51 ± 2.69	55.25 ± 2.02	64.00 ± 2.26
SE-LPN-DPFF	38.56 ± 1.49	32.99 ± 1.12	35.55 ± 1.15	33.20 ± 1.30
GBCNN	**57.79 ± 2.03**	**57.33 ± 1.93**	**57.54 ± 1.66**	**66.90 ± 1.20**

From the experimental results, we can find that the proposed method can obtain the optimal OA of 88.80% and 66.90% for the three- and five-category ship classification, which outperforms the second optimal method DFSN by 2.00% and 2.37%, respectively. Due to the MR property of the SAR data used and the initial unbalanced samples of the five-category ship targets, it is more challenging for the proposed method to classify as many as ship types effectively. As for other metrics, the proposed method can obtain a recall and precision both higher than 85% for the three-category classification task. However, the recall and precision of the five-category ship classification drop largely with regard to the OA, which may be caused by the small and unbalanced number of different ship types. This can be clearly observed in Fig. 4. Note that the SE-LPN-DPFF [20] behaves much worse than other compared methods. This may be greatly due to the lower data resolution in this letter than that used in [20], and the direct multiplication of the dual-pol GRD SAR images can also decrease the performance as some coherent information has been lost. Therefore, the full-resolution SAR images will be used to evaluate the robustness and generalization ability of different methods in the future.

E. Extensive Evaluation for the Proposed Method

To further evaluate the proposed method, the confusion matrices of the best three- and five-category classification results are shown in Fig. 4. From the confusion matrices, one can find that most of the ship types with fewer samples are generally misclassified to that with lots of samples, which deeply validates the results in Tables III and IV. In addition,

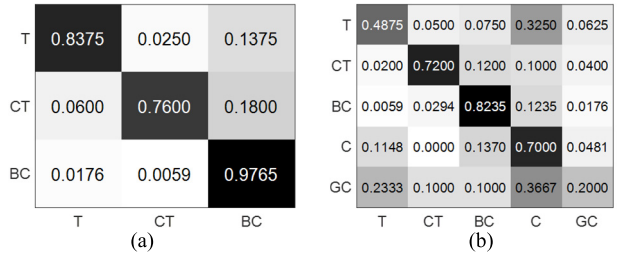


Fig. 4. Confusion matrices for the (a) three-and (b) five-category classification tasks, respectively. The five ship types of the tanker, container ship, bulk carrier, cargo, and general cargo are, respectively, denoted as T, CT, BC, C, and GC.

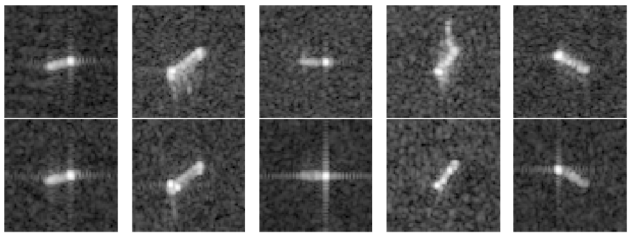


Fig. 5. Misclassified cases for the (first row) VH and (second row) VV images. (a) Tanker to cargo. (b) Container ship to bulk carrier. (c) Bulk carrier to cargo. (d) General cargo to cargo. (e) Cargo to bulk carrier.

some samples that are the most misclassified from each other are shown in Fig. 5. It seems that some interferences present more challenges for the proposed method to achieve better performance.

IV. Conclusion and Future Work

This letter proposes a GBCNN model to fully exploit the dual-polarized SAR images for promising fine-grained ship classification. Specifically, the group bilinear pooling is proposed to perform bilinear pooling between subfeature maps to efficiently acquire compact and discriminative bilinear features. Moreover, the MPFL loss is developed to ensure better learning of the proposed model for dual-pol SAR images. Extensive experiments on the three- and five-category OpenSARShip datasets demonstrate the effectiveness and superiority of the proposed method for promising SAR ship classification. Future work and challenging problems are given as follows.

1) To adaptively perform the group bilinear pooling and explicitly embed the semantic information into different groups of feature maps for further classification performance improvement.
2) Taking account of the imaging interferences to design an advanced network framework and to further improve the network efficiency.
3) The unbalanced data classification and the adaptive selection of hyperparameters will be particularly considered in the future.

References

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[2] G. Margarit and A. Tabasco, "Ship classification in single-Pol SAR images based on fuzzy logic," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 3129–3138, Aug. 2011.

[3] J. Zhu, X. Qiu, Z. Pan, Y. Zhang, and B. Lei, "Projection shape template-based ship target recognition in TerraSAR-X images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 222–226, Feb. 2017.

[4] H. Zhang, X. Tian, C. Wang, F. Wu, and B. Zhang, "Merchant vessel classification based on scattering component analysis for COSMO-SkyMed SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1275–1279, Nov. 2013.

[5] X. Leng, K. Ji, S. Zhou, X. Xing, and H. Zou, "2D comb feature for analysis of ship classification in high-resolution SAR imagery," *Electron. Lett.*, vol. 53, no. 7, pp. 500–502, 2017.

[6] H. Lang and S. Wu, "Ship classification in moderate-resolution SAR image by naive geometric features-combined multiple kernel learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1765–1769, Sep. 2017.

[7] Y. Dong, H. Zhang, C. Wang, and Y. Wang, "Fine-grained ship classification based on deep residual learning for high-resolution SAR images," *Remote Sens. Lett.*, vol. 10, no. 11, pp. 1095–1104, Nov. 2019.

[8] J. He, Y. Wang, and H. Liu, "Ship classification in medium-resolution SAR images via densely connected triplet CNNs integrating Fisher discrimination regularized metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3022–3039, Apr. 2021.

[9] C. Bentes, D. Velotto, and B. Tings, "Ship classification in TerraSAR-X images with convolutional neural networks," *IEEE J. Ocean. Eng.*, vol. 43, no. 1, pp. 258–266, Jan. 2018.

[10] Y. Wang, C. Wang, and H. Zhang, "Ship classification in high-resolution SAR images using deep learning of small datasets," *Sensors*, vol. 18, no. 9, p. 2929, Sep. 2018.

[11] Y. Yan, J. Sun, J. Yu, and J. Sun, "Small sample radar target recognition based on metric learning," in *Proc. IEEE 4th Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Jun. 2020, pp. 441–445.

[12] Y. Xu and H. Lang, "Distribution shift metric learning for fine-grained ship classification in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2276–2285, 2020.

[13] Y. Xu and H. Lang, "Ship classification in SAR images with geometric transfer metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6799–6813, Aug. 2021.

[14] T. Zhang and X. Zhang, "Injection of traditional hand-crafted features into modern CNN-based models for SAR ship classification: What, why, where, and how," *Remote Sens.*, vol. 13, no. 11, p. 2091, May 2021.

[15] T. Zhang and X. Zhang, "A polarization fusion network with geometric feature embedding for SAR ship classification," *Pattern Recognit.*, vol. 123, Mar. 2022, Art. no. 108365.

[16] T. Zhang *et al.*, "HOG-ShipCLSNet: A novel deep learning network with HOG feature fusion for SAR ship classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–22, 2022.

[17] Y. Xi, G. Xiong, and W. Yu, "Feature-loss double fusion Siamese network for dual-polarized SAR ship classification," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP)*, Dec. 2019, pp. 1–5.

[18] G. Xiong *et al.*, "Dual-polarization SAR ship target recognition based on mini hourglass region extraction and dual-channel efficient fusion network," *IEEE Access*, vol. 9, pp. 29078–29089, 2021.

[19] L. Zeng *et al.*, "Dual-polarized SAR ship grained classification based on CNN with hybrid channel feature loss," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[20] T. Zhang and X. Zhang, "Squeeze-and-excitation Laplacian pyramid network with dual-polarization feature fusion for ship classification in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[21] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1449–1457.

[22] T.-Y. Lin and S. Maji, "Improved bilinear pooling with CNNs," 2017, *arXiv:1707.06772*.

[23] D. Chang *et al.*, "The devil is in the channels: Mutual-channel loss for fine-grained image classification," *IEEE Trans. Image Process.*, vol. 29, pp. 4683–4695, 2020.

[24] L. Liu, C. Shen, and A. van den Hengel, "The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4749–4757.

[25] L. Huang *et al.*, "OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018.

[26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[27] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.