AMHC-DTENet: Attention-Based Multipath Hybrid Convolutional Distribution Target Extraction Network for Polarimetric Channel Imbalance Assessment

Haoyang Li , Mingjie Zheng , *Member, IEEE*, Yonghui Han , and Xingjie Zhao

Abstract—Manual selection of the distribution target and determination of its spatial location is not only time consuming but also has a serious impact on the polarimetric channel imbalances assessment if the area is incorrect. However, the existing automatic extraction methods are affected by the morphology of the features, and their local feature extraction capability is limited. On the other hand, only local information is considered, while the feature channels and global information are ignored, which in turn leads to the limited application scenarios, especially in some urban areas. The polarimetric channel imbalance estimation accuracy is so low that it exceeds the system tolerance limit. Therefore, in order to more effectively mine the polarimetric features of distributed targets in polarimetric synthetic aperture radar images, we propose an attention-based multipath hybrid convolutional distribution target extraction network for polarimetric channel imbalance assessment. First, in order to develop the local feature extraction capability of the hierarchical network, we design a hybrid convolutional module with adaptive adjustment of the receptive field size. Second, a polarimetric feature channel reconstruction module is constructed in order to utilize the spatial information of the polarimetric feature channels. Then, considering that the polarimetric information of the ground feature is sensitive to the relative geometry of the target attitude to the radar line of sight, the vision transformer architecture is used to capture polarimetric global information and extend it to multipath. Finally, Gaofen-3(GF-3) full-polarimetric data are used for experimental verification. Experimental results demonstrate the effectiveness and reliability of the proposed method.

Index Terms—Distributed target (DT), Gaofen-3 (GF-3), hybrid convolutional, polarimetric assessment, polarimetric feature channel reconstruction (PFCR), vision transformer (ViT).

I. INTRODUCTION

P OLARIMETRIC synthetic aperture radar (PolSAR) has all-day and all-weather remote sensing capabilities. By

Haoyang Li, Yonghui Han, and Xingjie Zhao are with the Department of Space Microwave Remote Sensing System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: lihaoyang21@mails.ucas.ac.cn; hanyonghui18@mails.ucas.ac.cn; zhaoxingjie19@mails.ucas.ac.cn).

Mingjie Zheng is with the Department of Space Microwave Remote Sensing System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhengmj@mail.ie.ac.cn).

transmitting and receiving different combinations of electromagnetic waves, the system records the backscattering characteristics of different ground targets, which are widely used in land cover classification [1], [2], [3], [4], [5], [6], [7], building area extraction [8], [9], ship detection [10], [11], and so on.

As the polarimetric antenna is affected by the environment and the real scattering information of the target cannot be reflected, polarimetric distortion is mainly reflected in the polarimetric crosstalk and polarimetric channel imbalance. At the same time, according to the current polarimetric crosstalk calibration level of the satellite-borne synthetic aperture radar (SAR) system [12], [13], [14], the polarimetric crosstalks are all better than 35 dB. And polarimetric crosstalk accuracy avoids the effect of its distortion parameters on polarimetric applications [15]. Therefore, it is necessary to use polarimetric calibration technology to calibrate the polarimetric channel imbalance and improve the quality of polarimetric data. Polarimetric calibration methods are generally divided into three types: the calibration method only depends on artificial calibrators [16], [17] or distributed targets [18], [19], [20], [21], and the calibration method depends on the combination of them [22], [23], [24]. Although the calibration accuracy of calibrators is high, it is limited by the layout scene, such as mountains, moon, and so on. In addition, the artificial calibrators are expensive.

In order to meet the need of rapid polarimetric calibration and reduce the dependence on artificial corner reflectors, polarimetric calibration use distribution targets that meet the requirements of reciprocity and reflection symmetry, where a point target that exhibits scattering randomness due to environmental influences and eliminates the random scattering component after multilook processing is called the distribution target [25]. Research of the polarimetric calibration [26], [27], [28], [29], [30], [31], [32], [33], [34] based on distributed targets are carried out at home and abroad. In 2018, Jiang et al. [27] proposed a statistical method for evaluating polarimetric data, which can estimate channel imbalance values within a scene relatively quickly. However, in Jiang's method, the distribution target areas need to be manually selected from the PolSAR images to be evaluated, which not only requires a priori knowledge of the PolSAR images, but also results in the manually selected distribution target areas artificially influencing the evaluation results. Moreover, in some scenarios with more discrete distribution targets, it is difficult to manually select the distribution targets, and thus, the PolSAR system cannot automatically process the scene batches in

current methods. In 2020, Shangguan et al. [35] first proposed a deep learning method that adopts an improved convolutional neural network (CNN) to extract distributed targets and automatically evaluate the quality of GF-3 full-polarimetric data. The CNN methods instead of manual selection method are utilized in extracting distributed targets within the scene, alleviating the burden of processing massive amounts of polarimetric data.

However, the automatic extraction method of distributed targets proposed by Shangguan has scene limitations. Although the accuracy of polarimetric distortion parameter estimation performs well in a large number of forest distribution scenes, it performs so poorly in a large number of urban distribution scenes that it exceeds the system tolerance limit. The following reasons may exist for this problem. First, in PolSAR images, the distribution of targets is irregular, which requires a large enough receptive field for feature extraction to ensure the extraction of more spatial information of targets. The traditional convolutional kernel, however, will result in too little spatial information sampling of the targets due to the fixed size of the sampling window when extracting local features, which in turn affects the distinguishability of the polarimetric features of the targets. Not only that, the same problem occurs in applications such as hyperspectral image classification and remote sensing image cloud detection [36], [37], [38]. Second, the selected distribution targets are mainly volume scattering dominated regions, which include multiple types of targets. And the traditional network methods have difficulty in extracting their effective homogeneity features without imposing any attention mechanism, which leads to serious misclassification of the distribution targets and thus affects the estimation accuracy of the imbalance of the polarimetric channels. And Wang et al. [39] used SE channel attention for feature preference of polarimetric feature channels of the hierarchical network, but there are some limitations in dealing with channel space relations. In addition, according to the uniform polarization rotation matrix theory, the scattering characteristics of the target are influenced by the relative geometric relationship between the target attitude and the radar line of sight. That is to say, in PolSAR images, there is heterogeneity in the scattering characteristics of distributed targets in different regions but the based-network methods by Shangguan cannot effectively use the global information.

To solve the aforementioned problems, we propose a new distributed target extraction model that integrates the receptive field theory, attention mechanism, and polarimetric rotation matrix theory into a distributed target extraction network. The main contributions are as follows.

1) To improve the local and global feature extraction of distribution targets and ensure the effectiveness of distribution target extraction in some special scenarios, we propose an attention-based multipath hybrid convolutional distribution target extraction network (AMHC-DTENet) for polarimetric channel imbalance assessment. The single branch network is composed of backbone network—ResNet18, hybrid convolution module, polarimetric feature channel reconstruction (PFCR) module and global information capture module (GIC).

2) In order to exploit the homogeneity features of the distribution targets and the spatial information of the polarimetric channels, we design a PFCR module based on effective channel attention (ECA) [39]. The polarimetric channel weights are automatically generated without dimension reduction. The use of PFCR module meets automatic processing of multiple polarimetric characteristic channel information at the same level.

3) In order to solve the problem of poor extraction of spatial information from irregularly shaped targets due to the restricted reception field of conventional convolutional kernels, a hybrid deformable convolutional module is added to the deep-layer hierarchical network to adaptively adjust the receptive field size by a learnable offset.

4) To better exploit the heterogeneity characteristics of distributed targets in different regions of PolSAR images, we design a GIC module. Global information can be used effectively under the vision transformer (ViT) architecture [40] using a self-attentive mechanism. And a contrast learning strategy is used to form a multibranch feature extraction network.

The rest of this article is organized as follows. Section II introduces the basic theory of PolSAR and the polarimetric distortion estimation method. Section III constructs the AMHC-DTENet. Section IV gives the results of the experiments. Section V provides discussion analysis. Finally, Section VI concludes this article.

II. BASIS OF THE IMAGES REPRESENTATION AND THE ASSESSMENT METHOD

A. Representation of PolSAR Images

The PolSAR images can be synthesized from the full-polarimetric single look complex data, and the polarimetric scattering matrix of the target in the image can be expressed as

$$\boldsymbol{S} = \begin{bmatrix} S_{\mathrm{HH}} & S_{\mathrm{HV}} \\ S_{\mathrm{VH}} & S_{\mathrm{VV}} \end{bmatrix} \tag{1}$$

where the subscript HV stands for receiving horizontal polarimetric waves and transmitting vertical polarimetric waves. And the other elements of the scattering matrix can be defined similarly. Based on the Pauli decomposition with the assumption of scattering reciprocity, the target scattering characteristics can be represented by the scattering vector

$$\boldsymbol{k}_p = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{\mathrm{HH}} + S_{\mathrm{VV}} & S_{\mathrm{HH}} - S_{\mathrm{VV}} & 2S_{\mathrm{HV}} \end{bmatrix}^T \tag{2}$$

where the superscript T denotes the transpose. Finally, the coherent decomposition of the target using the Pauli component yields the polarimetric coherency matrix $\boldsymbol{T}$

$$\boldsymbol{T} = \langle \boldsymbol{k}_p \cdot \boldsymbol{k}_p^H \rangle = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \tag{3}$$

where $< \cdot >$ and superscript H denote spatial average and conjugate transpose, respectively.

Considering that the backward scattering coefficient of the target is sensitive to the relative geometry of the target attitude to the radar line of sight, Chen et al. [41] proposed a uniform polarimetric matrix rotation theory, which establishes the connection between polarimetric rotation matrix and polarimetric coherence matrix

$$\boldsymbol{T}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta \\ 0 & -\sin 2\theta & \cos 2\theta \end{bmatrix} \boldsymbol{T} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta \\ 0 & -\sin 2\theta & \cos 2\theta \end{bmatrix}^{-1}$$

(4)

where θ represents the polarimetric rotation angle. It is solved by the relative geometric relationship between the target attitude and the radar line of sight.

B. Polarimetric Data Quality Assessment

Due to the influence of error sources such as channel imbalance and polarimetric crosstalk, the polarimetric scattering matrix $\boldsymbol{M}$ measured in practice cannot represents the real scattering matrix $\boldsymbol{S}$ of ground targets. The distortion matrix relationship between $\boldsymbol{M}$ and $\boldsymbol{S}$ can be expressed as follows:

$$\begin{bmatrix} M_{\text{HH}} & M_{\text{HV}} \\ M_{\text{VH}} & M_{\text{VV}} \end{bmatrix} = \text{K} \cdot \boldsymbol{R} \cdot \boldsymbol{S} \cdot \boldsymbol{T} + \boldsymbol{N_s}$$

$$= \text{K} \cdot \begin{bmatrix} 1 & \delta_2 \\ \delta_1 & f_r \end{bmatrix} \cdot \begin{bmatrix} S_{\text{HH}} & S_{\text{HV}} \\ S_{\text{VH}} & S_{\text{VV}} \end{bmatrix} \cdot \begin{bmatrix} 1 & \delta_3 \\ \delta_4 & f_t \end{bmatrix}$$

$$+ \boldsymbol{N_s}$$

(5)

where K is the absolute radiation amplitude phase constant and this article does not deal with K, f_r and f_t are channel imbalances of receiver and transmitter respectively, and subscript H/V represents horizontal/vertical polarimetric modes, δ represents polarimetric crosstalk, and $\boldsymbol{N_s}$ represents noise.

Performance metrics for polarimetric crosstalk of current PolSAR systems are all better than 35 dB, such as RADARSAT-2 [14], Gaofen-3 [12], and ALOS-2 [13], where the polarimetric crosstalk of them is better than 42, 35, 40 dB, respectively. Channel imbalances are only estimated in this article. In the assessment method [27], the following formula can be derived:

$$\begin{cases} |f_r|_L = \frac{1}{2}(\Delta f_\alpha - \Delta f_\beta) \\ \quad + \frac{1}{2}(|M_{\text{VV}}|_L - |M_{\text{HH}}|_L + |M_{\text{HV}}|_L - |M_{\text{VH}}|_L) \\ |f_t|_L = \frac{1}{2}(\Delta f_\alpha + \Delta f_\beta) \\ \quad + \frac{1}{2}(|M_{\text{VV}}|_L - |M_{\text{HH}}|_L + |M_{\text{VH}}|_L - |M_{\text{HV}}|_L) \end{cases}$$

(6)

$$\begin{cases} \Delta f_\alpha = |S_{\text{HH}}|_L - |S_{\text{VV}}|_L \\ \Delta f_\beta = |S_{\text{HV}}|_L - |S_{\text{VH}}|_L. \end{cases}$$

(7)

Here, $|\cdot|_L$ represents the logarithm of the spatial average of the absolute value squared. Δf_α and Δf_β are the residuals of copolarization and cross polarization, respectively. Phase

imbalance can be written as

$$\begin{cases} |\theta_r|_L = \frac{1}{2}(\Delta\theta_\alpha - \Delta\theta_\beta) \\ \quad + \frac{1}{2}(P(< M_{\text{HV}}M_{\text{VH}}^* >) - P(< M_{\text{HH}}M_{\text{VV}}^* >)) \\ |\theta_t|_L = \frac{1}{2}(\Delta\theta_\alpha + \Delta\theta_\beta) \\ \quad - \frac{1}{2}(P(< M_{\text{HV}}M_{\text{VH}}^* >) + P(< M_{\text{HH}}M_{\text{VV}}^* >)) \end{cases}$$

(8)

$$\begin{cases} \Delta\theta_\alpha = P(< S_{\text{HH}}S_{\text{VV}}^* >) \\ \Delta\theta_\beta = P(< S_{\text{HV}}S_{\text{VH}}^* >) \end{cases}$$

(9)

where $P(\cdot)$ represents the phase of complex numbers. The aforementioned is the channel imbalance formula derived from the polarimetric distortion matrix model and the parameter basis for extracting the distributed targets.

III. METHOD

In this section, the estimation of polarimetric channel imbalance within the scene is first presented. Then, Fig. 2 shows the architecture of the proposed distributed target extraction network (AMHC-DTENet) and all of its components, including the hierarchical polarimetric feature extraction network, PFCR module, and GIC module.

A. Estimation of Polarimetric Channel Imbalance Within the Scene

The workflow consists of data acquisition and preprocessing, slicing and training, testing and polarimetric channel imbalance estimation, as shown in Fig. 1. The first part includes reading the full-polarimetric data, generating PolSAR images, and extracting the polarimetric channel residual terms. The single look complex data of the full-polarimetric channels are first read from the metafile of the GF-3 product, and the data of each polarimetric channel are quantified by the quantitative constants in the file. The data used in this article are in Quad-Pol StripmapI (QPSI) imaging mode with a resolution of 8 m. Second, to eliminate the effect of coherent spots, the full-polarimetric data are subjected to multilook processing. Then, PolSAR images are generated using Pauli basis. The polarimetric channel residual terms are then extracted directly from the PolSAR images to the ground targets and the dataset labels are generated by setting a threshold condition on the polarimetric channel imbalance residual terms in (7) and (9). Based on the current PolSAR design metrics for the polarimetric distortion parameters [12], [13], [14], the amplitude imbalance is better than 0.5 dB and the phase imbalance is better than 10°. Therefore, we set the threshold conditions for the amplitude and phase imbalance residual terms to 0.5 dB and 10°, respectively. In addition the condition for phase imbalance is loose [35], the targets selected by the condition of amplitude imbalance is almost the intersection of the two items. We use the condition of amplitude imbalance and the contrast of the optical images for labeling all SAR image slices.

The second part first slices the PolSAR images along the range and azimuthal directions, with each slice having a size of 50 × 50 pixels. Next, the slice labels are generated based on the
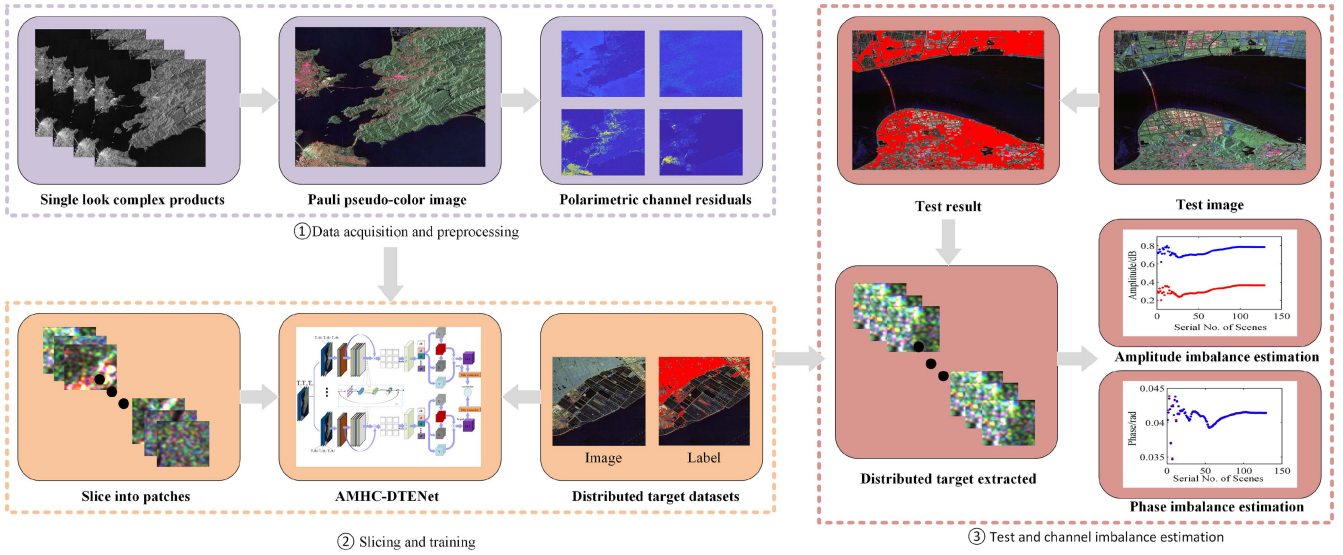
Fig. 1. Whole workflow of the distribution target-based polarimetric channel imbalance estimation of the scene.
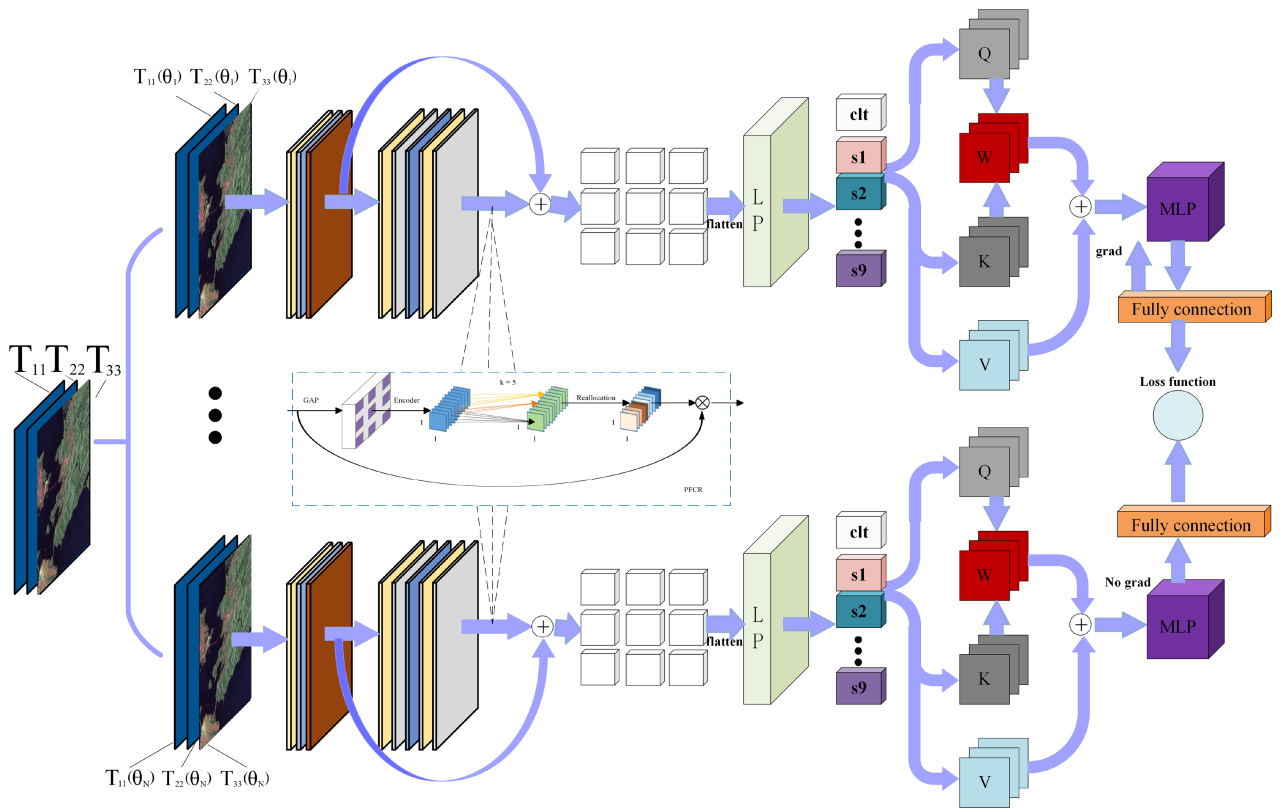

Fig. 2. Aarchitecture of the AMHC-DTENet.

adaptive threshold condition. Then, the datasets are fed to the proposed AMHC-DTENet for training and learning.

The third part first keeps the training parameters of the model, then inputs the test images into the network, automatically extracts the distribution targets in the images, and finally, the channel imbalance estimation of the receiver and the transmitter is obtained by the distributed target extracted from each scene.

B. Proposed AMHC-DTENet Architecture

Traditional network-based methods can extract effective distribution targets well in PolSAR images with uniformly distributed volume scattering targets, however, they perform poorly in areas with irregular target distribution and large urban areas because of their limited local feature and global information
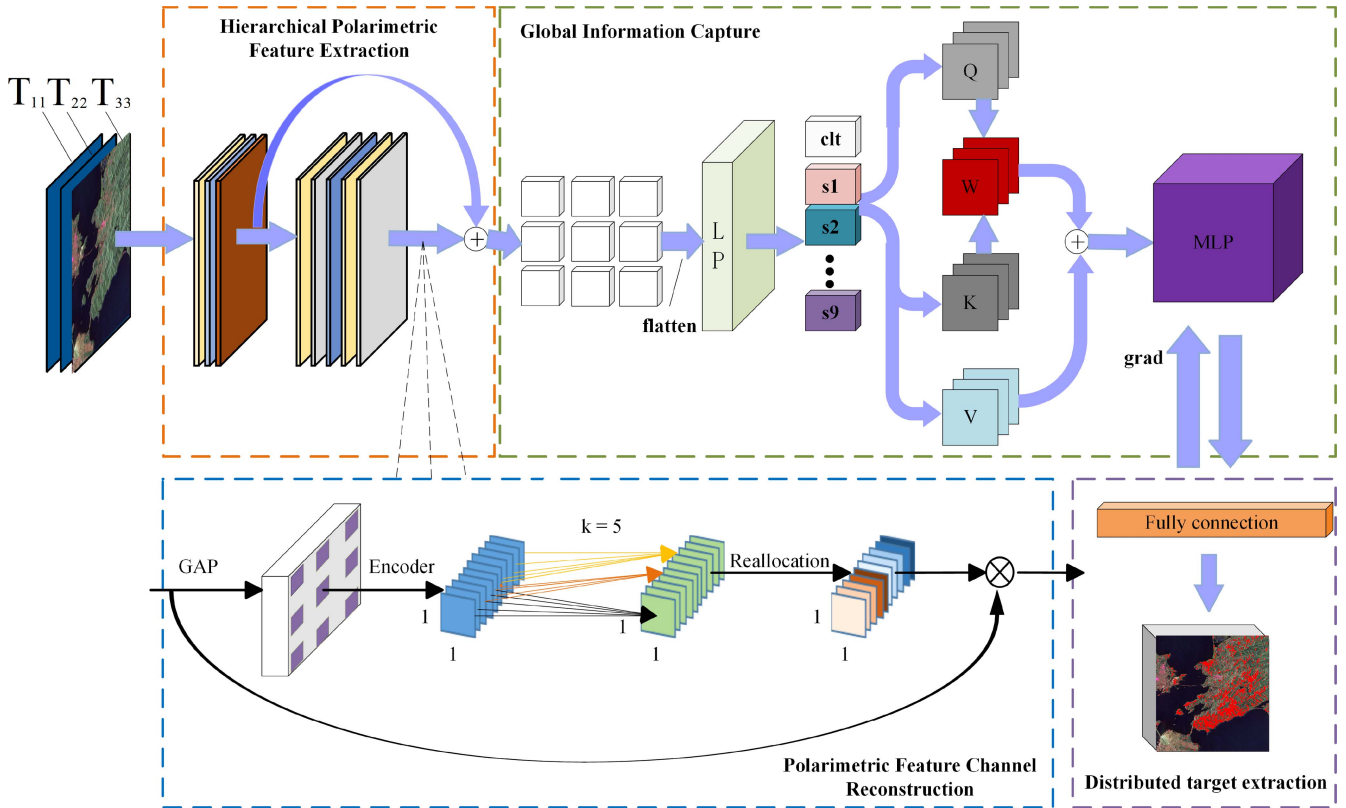
Fig. 3. Structure of the base network in this article. It consists of three components, a hierarchical polarimetric feature extraction network, PFCR module, and GIC module.

extraction capabilities. In this article, in order to solve the problems of difficulty in sampling local spatial information due to the irregular shape of features and the difficulty in using feature channels and global information, we propose the AMHC-DTENet for polarimetric channel imbalance estimation. Fig. 3 shows the architecture of its base network. In the AMHC-DTENet, the polarimetric feature channel weights of multiple branches are shared, and the training errors of all branches are only gradient updated for the backbone during backpropagation, reducing unnecessary gradient iterations.

1) Hierarchical Polarimetric Feature Extraction Network: In the local feature extraction of some irregularly distributed targets, the conventional convolution kernel is limited in its field of receptive due to its fixed size, which makes it impossible to extract more spatial information of the targets. In order to solve the aforementioned problems, we design the hierarchical polarimetric feature extraction network based on a residual network with hybrid deformable convolutional layers, as shown in Fig. 4. Deformable convolution is used to improve the polarimetric feature receptive field by additionally learning the convolution kernel offsets, and both convolution kernels update the weights simultaneously while performing backpropagation.

The hierarchical polarimetric feature extraction network is divided into shallow hierarchical polarimetric feature extraction based on traditional residual blocks and deep hierarchical polarimetric feature extraction based on hybrid deformable convolution. First, the residual blocks in the shallow hierarchical
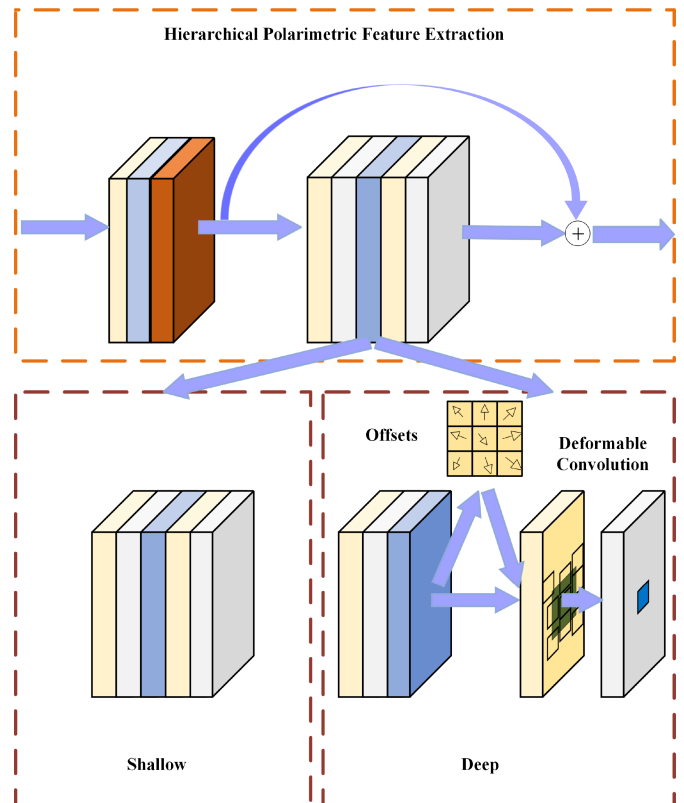

Fig. 4. Structure of the hierarchical polarimetric feature extraction network.
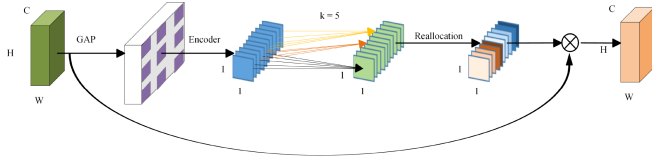
Fig. 5. Structure of the PFCR.

polarimetric feature extraction are composed of convolutional, batchnormalization, activation, convolutional, and batchnormalization layers in order. The output can be expressed as

$$z = \sigma_{\text{relu}} \left(\left(\text{BN} \left(w * \sigma_{\text{relu}} \left(\text{BN}(y_1) \right) + b_2 \right) + x \right) \right) \tag{10}$$

where σ_{relu} represents relu activation function, w represents the input weights for the layer, $\text{BN}(\cdot)$ represents the layer of BatchNormalization, y_1 represents convolutional layer features output, and x represents polarimetric feature input.

After the shallow hierarchical network downsampling, the offset needs to be learned in parallel in the deep hierarchical polarimetric feature extraction, where the offsets are changed by bilinear differences along the range direction and azimuth direction, respectively, so that the polarimetric feature output after the deformable convolution layer can be expressed as

$$y(p_0) = \sum_{p_n \in R} \omega(p_n) * x(p_0 + p_n + \Delta p_n) + b \tag{11}$$

where $*$ represents convolution operation. $x, y \in \mathbb{R}^{m \times m \times 1}$ represents polarimetric feature input and output. p_n represents the location of the neighboring pixel point at position p_0. Δp_n represents the range and azimuthal offsets that can be learned. $\omega(p_n) \in \mathbb{R}^{n \times n \times c \times 1}$ represents the polarimetric feature weights in the convolution layer, and b represents the convolution layer bias parameter.

2) PFCR Module: CNNs are widely used in PolSAR images. However, the difficulty of high-dimensional data processing caused by multiple polarimetric features as input has always been the research topic of polarimetric feature selection. Among them, Dong et al. [1] uses an improved squeeze excitation (SE) block to reduce the dimension of data to achieve the polarimetric image scene classification, but SE block has the problem of low efficiency of capturing feature channel spatial information and the loss of detailed information of polarimetric channel characteristics due to direct use of fully connected layers. To solve these problems, we propose the PFCR module based on ECA [39], which can select effective features by capturing spatial information between different polarimetric features in the extraction of polarimetric hierarchical features. The PFCR are outlined in Fig. 5, which consists of global averaging pooling (GAP) layer, channel coprocessor (CP), and recalibration operation. Relevant formulas are given to explain the principle of this module.

For any hierarchical feature input $x \in \mathbb{R}^{m \times m \times c}$, first by a GAP layer aggregate into channel information, and can be written

$$e = f^{\text{GAP}}(x) = \frac{1}{m \times m} \sum_{h=1}^{m} \sum_{w=1}^{m} x_{h,w} \tag{12}$$
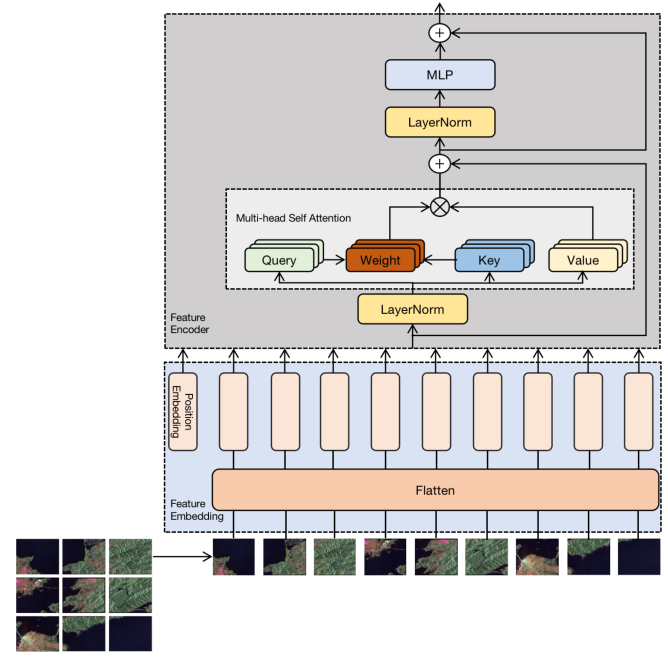

Fig. 6. Structure of GIC.

where $e = (e_1, e_2, e_3, \ldots, e_c)$ represents the channel descriptor corresponding to each polarimetric feature of the current level. $f^{\text{GAP}}(\cdot)$ denotes the GAP. m represents the space size of the polarimetric feature, and c represents the channel size. Then, learnable polarimetric feature channel weights are then generated by a 1-D convolutional kernel of size k, i.e.,

$$\omega = \sigma \left(\text{C1D}_k(e) \right) \tag{13}$$

where C1D stands for 1-D convolution, and the hyperparameter only k. And there is a mapping relationship between k and channel size c, namely

$$k = \psi(c) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{\text{odd}} \tag{14}$$

where $| \cdot |_{\text{odd}}$ represents the nearest odd number, γ and b are the hyperparameter. Finally, the function of input and output before and after PFCR processing is obtained, i.e.,

$$x^{\text{PFCR}} = \sigma \left(\text{C1D}_k \left(f^{\text{GAP}}(x) \right) \right) \cdot x \tag{15}$$

where x^{PFCR} represents the output after the processing of the PFCR module.

3) GIC Module: CNN with PFCR already have good extraction of polarimetric hierarchical features. However, the direct use of fully connected layers to process global information can lead to ineffective utilization of high-level semantic features. Considering the heterogeneity of global information of targets from different regions, under the ViT architecture, we design the GIC module to capture the long-range interrelationship of global information using a self-attentive function. The overview of GIC module is shown in Fig. 6. It is composed of feature embedding layer and feature encoder. In order to better explain

the principle of this module, we combine the text description and related formulas to explain.

In order to match the token embeddings sequence input of the feature encoder, first slice the 3-D hierarchical polarimetric feature map $S_1 \in \mathbb{R}^{m \times m \times c}$ into a series of flat 2-D patches $S_1^p \in \mathbb{R}^{N \times (p^2 \cdot c)}$ in the feature embedding layer. Assuming that there is a constant mapping vector size D, each patch is linearly mapped to a 1-D vector with length D according to (16). The learnable class tag embedding ($s_{\text{class\_token}}$) is preinserted before the embedded patches sequence, and the learnable 1-D position embedding $E_{\text{pos}} \in \mathbb{R}^{(N+1) \times D}$ is embedded in the output block of the feature embedding. It is used to reserve the position and relative position information before each embedded patch is flattened. Finally, ϵ is input as the feature encoder.

$$\epsilon = \left[s_{\text{class\_token}}; (s_1^p)^1 E; (s_1^p)^2 E; \cdots ; (s_1^p)^N E \right] + E_{\text{pos}} \quad (16)$$

where $N = (m/p)^2$ represents the total number of patches, and p represents the size of each patch. E represents the linear projection operator.

Then, the image output $s_1' \in \mathbb{R}^{(N+1) \times D}$ is obtained after the feature encoder, where the feature encoder includes layer norm (LN), multihead attention (MSA), and multilayer perceptron (MLP) blocks

$$s_1' = \text{LN} \left(\text{MLP} \left(\text{LN} \left(\epsilon_{\text{MSA}} \right) \right) + \epsilon_{\text{MSA}} \right) \quad (17)$$

where ϵ_{MSA} stands for connecting the output of the MSA layer to the input jumper. The MSA layer constructs h self-attention header functions for all ϵ to capture the correlation between sequences. And each attention since the head function for a given query Q_i, key K_i, and value V_i, respectively, through training the learning weight matrix W_i^Q, W_i^K, and W_i^V, namely

$$Q_i = \epsilon W_i^Q, K_i = \epsilon W_i^K, V_i = \epsilon W_i^V. \quad (18)$$

The dot product of query and key is scaled and activated by softmax function to get the weight of the self-attention. And value is weighted to get the self-attention head function score $head_i$, that is,

$$\begin{aligned} head_i &= \text{Attention} \left(Q_i, K_i, V_i \right) \\ &= \text{softmax} \left(\frac{Q_i K_i^T}{\sqrt{d_k}} \right) \cdot V_i \end{aligned} \quad (19)$$

where d_k represents the dimension of K_i.

IV. EXPERIMENTAL RESULTS

To verify the effectiveness of our AMHC-DTENet, four types of experiments are carried out on the full-polarimetric SAR datasets of GF-3 with different strips in the QPSI mode. All experiments are conducted on a PC with PyCharm software configured with expansion packs such as Python3.9 and Pytorch, 2.1 GHz i7-12700F CPU, NVIDIA GTX 3090Ti 24 G GPU, and 64-GB RAM.

A. Experimental Setup

1) Datasets: In order to verify the validity and reliability of the model, the datasets used are from the C-band GF-3 full-polarimetric data in QSPI imaging mode, as described in Section III-A. Table I shows a single look complex data that use or test about 80 scenes, including multiple strips. The imaging time of these scenes is between 2019 and 2023. These scenes contain rich types of land targets, such as urban areas, farmland, forest areas, water bodies, bare land, and so on. Fig. 7 shows the scenarios for validation, testing, and robustness analysis described as follows:

1) The two scenarios shown in Fig. 7(a) and (b) are used for verification. The first verification scenario is mainly water, forest, bare land, and urban areas. The second verification scenario is mainly forest area and water area. And these two verification scenarios are used to verify the performance of the distributed target extraction of the module.
2) The two scenarios shown in Fig. 7(c) and (d) are used for test. The first test scenario is in Xi'an. The second test scenario is in Nanning. In these two scenarios, urban areas are the main distribution, and the validity of the distribution targets extracted by the methods in urban areas is tested.
3) The last scenario shown in Fig. 7(e) is used for robustness analysis due to its rich feature types, mainly including forest areas, towns, waters, and farmland. According to the adaptive threshold condition of polarimetric channel imbalance, most of the forest areas in PolSAR images are directly labeled as positive classes. Among them, the existence of unmarked bands at the bottom of the distribution target extraction map is because it does not meet the slice size.

2) Evaluation Metrics: In order to measure the effectiveness of the proposed model, the accuracy of distributed target extraction and polarimetric channel imbalance estimation are used to evaluate the model. Among them, the number of predicted correct distribution targets is expressed as true positive (TP), while the number of targets incorrectly predicted as distribution targets is expressed as false positive (FP), but the number of targets incorrectly predicted in the actual distribution targets is expressed as false negative (FN). Then precision, recall, and F1 score are used to compare the accuracy of distributed target extraction. These indicators are commonly used in the field of target detection.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (20)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (21)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (22)$$

The precision represents the proportion of correctly predicted distributed targets in all predicted distributed targets. The recall represents the proportion of all actual distribution targets that are correctly predicted as distribution targets. The F1 score provides a measure of the accuracy of distributed target extraction for balancing precision and recall.

TABLE I
INFORMATION OF GF-3 SCENES USED IN QPSI MODE

| Wave code	Product ID	Direction	Imaging time of center scene	Incidence angle/° (near range – far range)	Center location	Number of scenes used
190	00000000xx	ASC	2022.2.28	22.24−25.00	45.4N 117.3E ∼ 31.0N 121.5E	9
	00000000xx	DEC	2022.11.18	22.24−25.00	27.0N 107.4E ∼ 49.6N 125.0E	11
191	00000000xx	DEC	2023.2.6	23.67−26.27	23.9N 118.3E ∼ 27.9N 119E	9
194	00000000xx	DEC	2022.11.6	28.32−30.73	29.4N 117E ∼ 37.6N 118.1E	8
195	00000000xx	ASC	2022.10.29	29.37−31.46	45.3N 132.3E ∼ 49.1N 131.1E	10
197	00000000xx	DEC	2022.10.25	32.26−34.24	31.7N 117.1E ∼ 40.0N 119.7E	7
199	000xxxxxxx	DEC	2019.3.27	35.32−37.15	37.9N 122.5W	1
	00000000xx	ASC	2022.2.26	35.29−37.05	42.5N 98.3E ∼ 36.9N 101.6E	8
201	00000000xx	ASC	2022.11.23	37.77−39.01	39.2N 119.4E ∼ 30.2N 120.1E	9
216	0000000xxx	DEC	2022.12.3	48.98−49.91	28.3N 117.7E ∼ 37.9N 118.3E	8
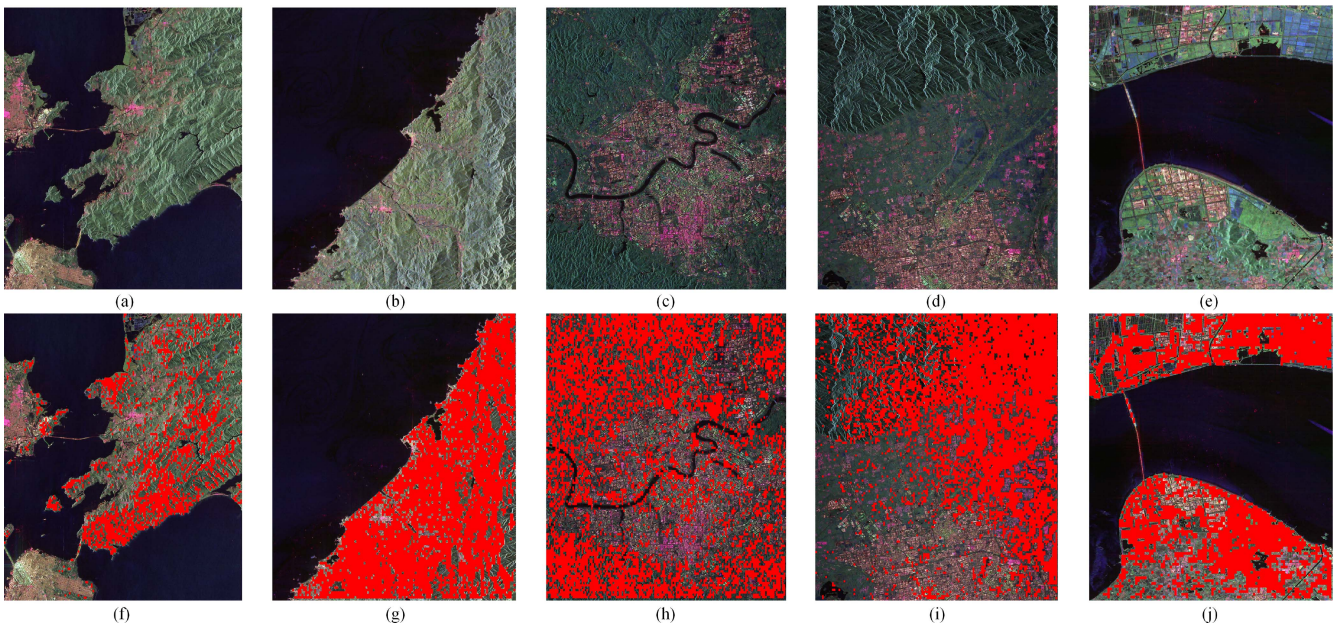
Fig. 7. Pseudocolor drawings and distributed target extraction diagram in these scenes. (a)–(e) Pseudocolor drawings of the these set scenes. (f)–(j) Corresponding distributed target extraction diagrams.

In order to measure the estimation accuracy of the polarimetric channel imbalance of the model, the estimation value of the polarimetric channel imbalance is obtained by the distributed target extracted in the scene, and then, the model is compared by the difference with the real value. The difference is mainly divided into the amplitude and phase imbalance difference at the receiving end ($|\Delta f_r|$ and $|\Delta\theta_r|$) and the amplitude and phase imbalance difference at the transmitting end ($|\Delta f_t|$ and $|\Delta\theta_t|$).

The polarimetric channel imbalance value of the extracted distributed target is calculated according to (6) and (8). The copolarization and cross polarization of the extracted distributed targets should satisfy weak reciprocity [27]. According to (7) and (9), the channel imbalance residual terms are set to zero.

3) Implementation Details: In the experiment, in addition to the aforementioned three PolSAR images used for verification and testing, one scene is randomly selected from each wave code image sequence for generalization analysis. Most of the remaining PolSAR images are used for training. In order to extract the polarimetric features of distributed targets under different incidence angles, we select PolSAR images for each wave code, and a total of ten PolSAR images are selected as training set, which contain rich feature scenes. These scenes are sliced and mixed to form the training set. We train the fine-tuning network with the strategy of step-by-step migration learning and the training set needs to be updated at each training stage.

At the same time, each PolSAR image needs to be sliced, and the slice size is $50 \times 50 \times 3$. When calculating the polarimetric channel imbalance value of each slice, the statistical value of the polarimetric channel imbalance value of all scattering points in the slice is taken as the estimated value of the slice.

There are some experimental details as follows. We adopt 0.0001 initial learning rate, 0.65 drop rate, cross entropy loss function [42], and Adam optimizer [43]. The mini batch size was set to 128. The training iteration with random selection is over 20 000. Based on the aforementioned experimental conditions, the models used in this article are trained and tested under the same experimental conditions.
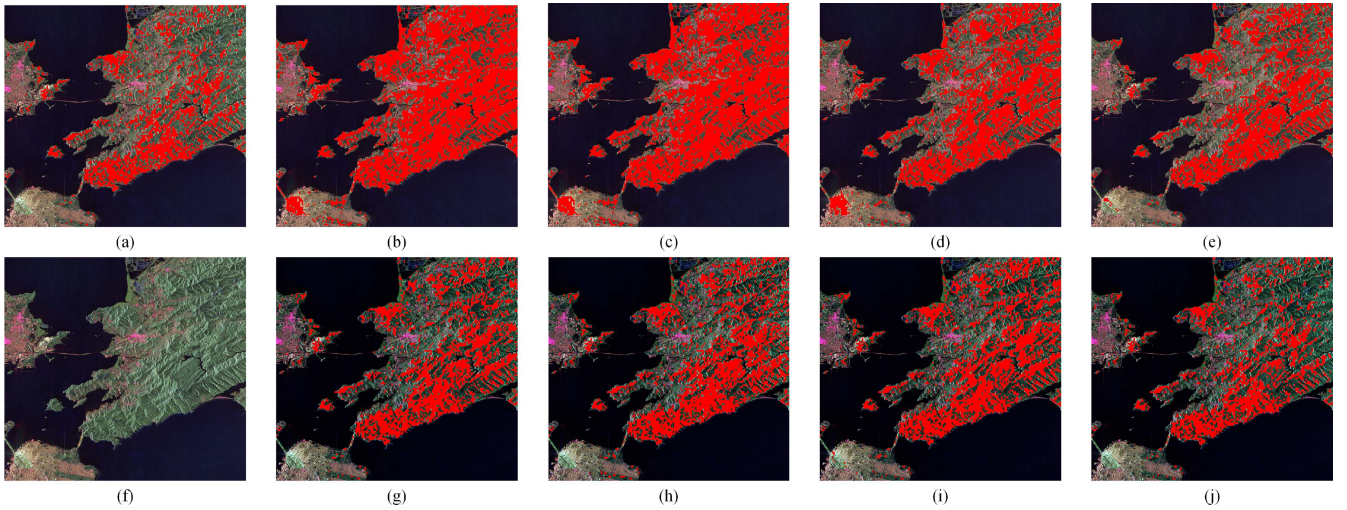
Fig. 8.   Distributed target extraction diagrams in the No.1 verification scenario. (a) Label image. (b) ResNet18. (c) PFCR. (d) GIC. (e) HDC. (f) Pauli pseudocolor image. (g) PFCR + GIC. (h) PFCR + HDC. (i) HDC + GIC. (j) AMHC-DTENet.
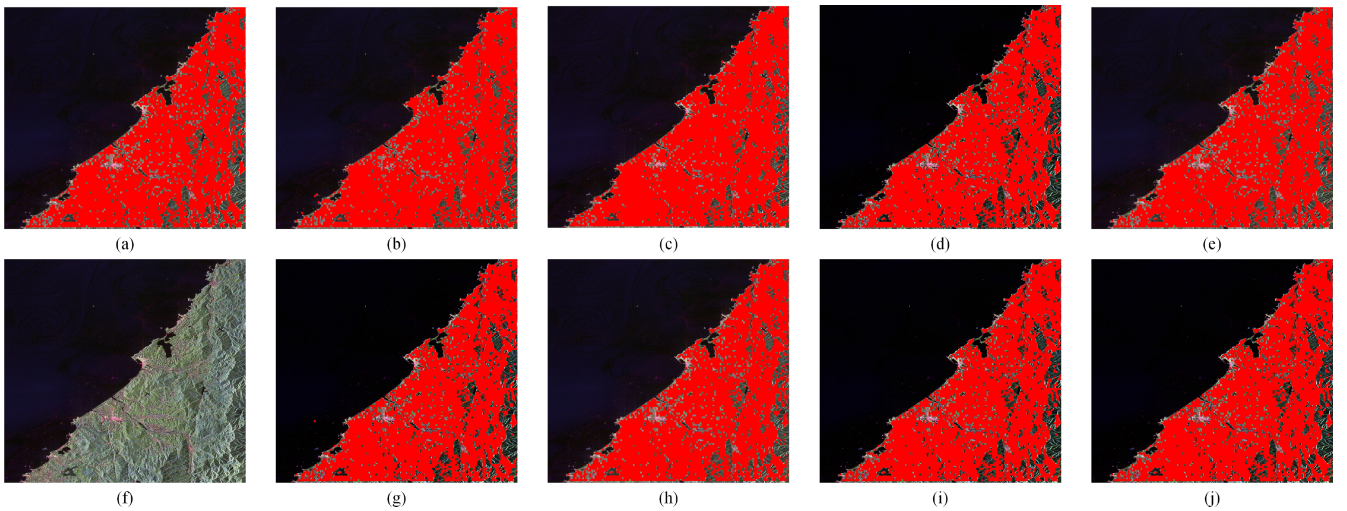

Fig. 9.   Distributed target extraction diagrams in the No. 2 verification scenario. (a) Label image. (b) ResNet18. (c) PFCR. (d) GIC. (e) HDC. (f) Pauli pseudocolor image. (g) PFCR + GIC. (h) PFCR + HDC. (i) HDC + GIC. (j) AMHC-DTENet.

TABLE II
DISTRIBUTED TARGET EXTRACTION ACCURACY METRICS FOR DIFFERENT METHODS IN THE NO.1 VERIFICATION SCENARIO

Method	ResNet18	PFCR	GIC	HDC	PFCR+GIC	PFCR+HDC	GIC+HDC	AMHC-DTENet
Precision	0.4713	0.4707	0.4889	0.492	0.5179	0.5218	0.5004	**0.5565**
Recall	0.5144	0.5591	0.5904	0.5924	0.5963	0.6405	0.6713	**0.6892**
F1	0.4919	0.5111	0.5349	0.5376	0.5544	0.5751	0.5734	**0.6158**

The bold values represent the higher extraction accuracy of distributed targets under this method.

B. Experimental Results

1) Ablation Experiments: In order to verify the effectiveness of our proposed method, we used ResNet18 as the baseline model and performed ablation experiments with different modules. PFCR represents the PFCR module, GIC represents the global information capture module, and HDC represents the hybrid deformable convolution module.

We verify the models in the two scene verification set mentioned previously, where Figs. 8 and 9 shows the distribution target extraction of these methods. Tables II and III compare the accuracy of these methods in extracting distributed targets.

TABLE III
DISTRIBUTED TARGET EXTRACTION ACCURACY METRICS FOR DIFFERENT METHODS IN THE NO.2 VERIFICATION SCENARIO

Method	ResNet18	PFCR	GIC	HDC	PFCR+GIC	PFCR+HDC	GIC+HDC	AMHC-DTENet
Precision	0.8375	0.8359	0.8535	0.8445	0.8457	0.8579	0.8523	**0.8584**
Recall	0.8371	0.8462	0.8488	0.8683	0.8819	0.8991	0.8948	**0.9507**
F1	0.8373	0.8410	0.8511	0.8563	0.8634	0.8780	0.8730	**0.9022**

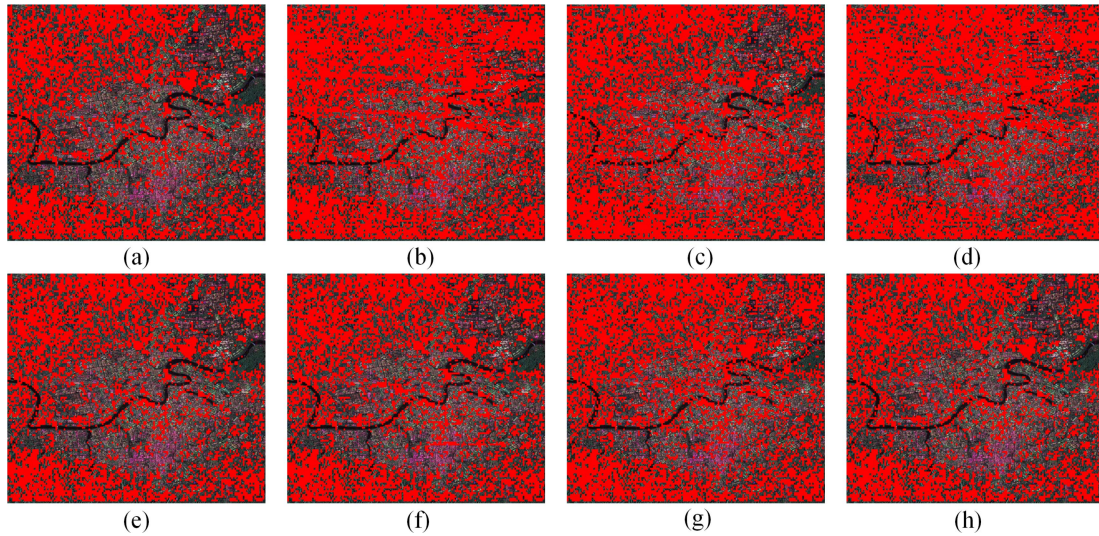The bold values represent the higher extraction accuracy of distributed targets under this method.

Fig. 10.   Distributed target extraction diagrams in the No.1 test scenario. (a) Label image. (b) VGGNet16. (c)SENet. (d) CBAM. (e) ResNet50. (f) MobileNet. (g) ViT. (h) AMHC-DTENet.

In No.1 verification scenario, ResNet18 with one of the PFCR or GIC modules and the baseline network inevitably extract some urban areas, such as the urban area in the lower left part of the whole scene. The network with more than two modules can extract the distributed target area well. And compared with other methods, the distribution of the distributed targets extracted by the proposed method is closer to the label image. In addition, from the comparison of the extraction accuracy of distributed targets, it can be seen that the network performance with more than two modules is higher than that with a single module. Compared with the baseline network, the network Precision with more than two modules increased by 2.91%–8.52%, the Recall increased by 8.19%–17.48%, and the F1 score increased by 6.25%–12.39%. Our AMHC-DTENet increased by 8.52%, 17.48%, and 12.39%, respectively. In the second verification scenario, the distribution target extraction accuracy of the network with a single module is improved but not significantly. Compared with the baseline network, the distribution target extraction accuracy of the network with two or more modules improved significantly, where the Precision improved by 0.82%–2.09%, the Recall improved by 4.48%–11.36%, and the F1 score improved by 2.61%– 6.49%, compared with the aforementioned network, the improvement of the distributed targets extraction performance of our AMHC-DTENet is the largest. They were 2.09%, 11.36%, and 6.49%, respectively.

The aforementioned experimental results show that the AMHC-DTENet with PFCR, GIC, and HDC modules exerts the advantages of each module and improves the extraction performance of distributed targets.

2) Comparative Experiments: In order to further verify the performance of our AMHC-DTENet, in addition to comparing with the distributed target extraction network (VGGNet16) mentioned in [35], we also selected the target recognition network methods, 50-layer residual network (ResNet50) [44], convolutional block attention module (CBAM) [45], and squeeze excitation network (SENet) [46]. In addition, we added two network-based method in the comparative experiments. One is MobileNet [47], a lightweight network commonly used on mobile, and the other is ViT [40], which learns the information between feature maps using the self-attention mechanism.

After the aforementioned network is trained according to the strategy of segmented transfer learning, Fig. 10 shows the extraction results of Nanning as the first test scene, in which the urban area is mainly distributed in the center of the image, and the surrounding area is mainly forest area. And the polarimetric channel imbalance estimation error of the scene is obtained by the distribution target extracted in the image, as shown in Table IV.

Compared with the label image, all methods can extract most of the distribution targets of the label image. However, the distribution targets extracted by VGGNet16, SENet, CBAM, and ViT have a large number of false detections in the central city area. At the same time, VGGNet16, CBAM, and ViT incorrectly extract distribution targets in the township area on the upper right of the image. Compared to ResNet50, MobileNet still has few misdetected targets in the center and upper right urban areas. The distribution target distribution extracted by the method proposed in this article and ResNet50 is close to

TABLE IV
POLARIMETRIC CHANNEL IMBALANCE ESTIMATION ACCURACY METRICS FOR DIFFERENT METHODS IN THE NO.1 TEST SCENARIO

Method	VGGNet16	SENet	CBAM	ResNet50	MobileNet	ViT	AMHC-DTENET
$|\Delta f_r|/$dB	0.5715	0.4894	0.5918	0.2537	0.2841	0.3770	**0.2301**
$|\Delta f_t|/$dB	0.6187	0.5024	0.6521	0.2855	0.3157	0.4207	**0.2584**
$|\Delta \theta_r|/$rad	0.2281	0.2381	0.2444	0.1016	0.1167	0.1527	**0.0927**
$|\Delta \theta_t|/$rad	0.2276	0.2400	0.2442	0.1028	0.1182	0.1540	**0.0940**

The bold values represent lower scene channel imbalance estimation errors obtained by the method through the extracted distribution targets.
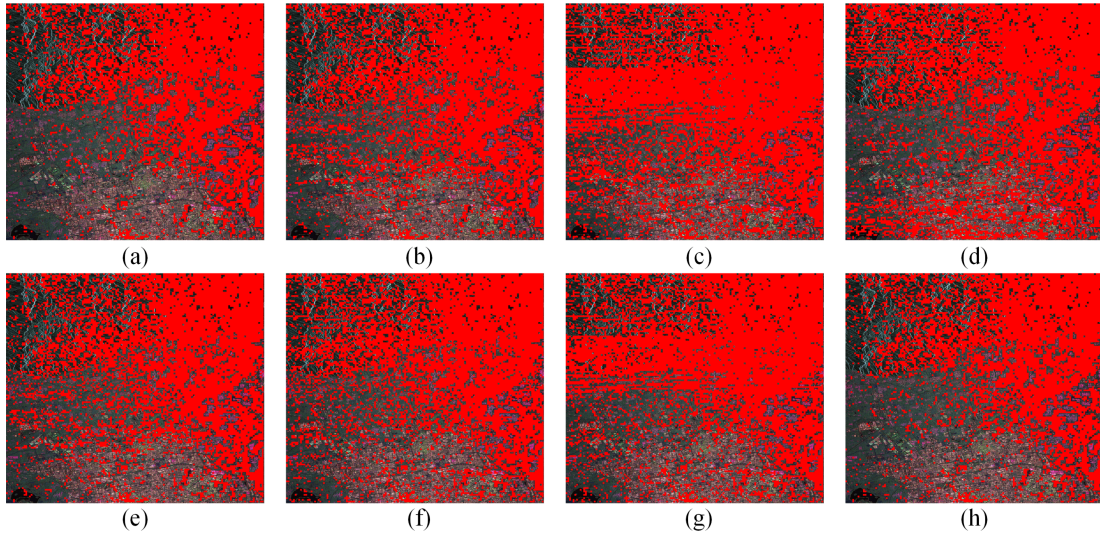

Fig. 11. Distributed target extraction diagrams in the No. 2 test scenario. (a) Label image. (b) VGGNet16. (c) SENet. (d) CBAM. (e) ResNet50. (f) MobileNet. (g) ViT. (h) AMHC-DTENet.

the label image distribution. This shows that when the PolSAR images have a large number of urban areas, the distribution target extraction process of VGGNet16 will be affected, resulting in a serious deviation in the distribution of the extracted distribution target, which may affect the estimation accuracy of the polarimetric channel imbalance. The method proposed in this article can extract the distribution targets in urban areas more accurately.

In addition, the distributed targets extracted by all methods are used to estimate the polarimetric channel imbalance of the scene. The results show that the amplitude imbalance errors of the receiver and transmitter obtained by the proposed method are 0.0236 to 0.3617 dB and 0.0271 to 0.3937 dB lower than other methods, respectively, and 0.3414 and 0.3603 dB lower than VGGNet16. In terms of phase imbalance estimation, the phase imbalance errors of the receiver and transmitter obtained by the proposed method are 0.0089 to 0.1517 rad and 0.0088 to 0.1502 rad lower than other methods, respectively, which are 0.1354 and 0.1336 rad lower than VGGNet16. Meanwhile, the distribution targets extracted by the lightweight network MobileNet are close to ResNet50 in terms of polarimetric channel imbalance estimation performance, where the amplitude imbalance estimation accuracies at the receiver and transmitter ends differ by 0.0304 and 0.0302 dB, respectively, whereas, the phase imbalance estimation accuracies differ by 0.0151 and 0.0154 rad, respectively.

The second test scene shown in Fig. 11 is located in Xi'an. The mountain forest area is mainly located in the upper part of the image, while the urban area is located in the lower part of the image, and the middle part is mainly some rural areas. From the results of extraction, SENet and CBAM have a large number of misdetected targets in the lower urban region, while VGGNet16, ResNet50, and MobileNet extracted distribution targets with similar distribution, but misdetected targets in the middle township region. The misdetected targets of the ViT are mostly found in the upper forest region, and the impact of bias in the polarimetric channel imbalance estimation obtained may be small. In contrast, the distribution of distribution targets extracted by the method proposed in this article is closer to the label image.

In addition, the distribution targets extracted by all methods are estimated for polarimetric channel imbalance, and the estimated bias results are shown in Table V. The amplitude imbalance deviation of receiver and transmitter obtained by CBAM is the largest, which is 0.5866 and 0.6759 dB, respectively. Second, the amplitude imbalance deviation of the receiver and transmitter obtained by the proposed method in this article is the smallest, which is 0.0917 and 0.1018 dB lower than that of VGGNet16, respectively. From the phase imbalance estimation results, the phase imbalance error obtained by the CBAM is the highest. The phase imbalance deviation obtained by the proposed method in this article is the smallest, which is 0.0495
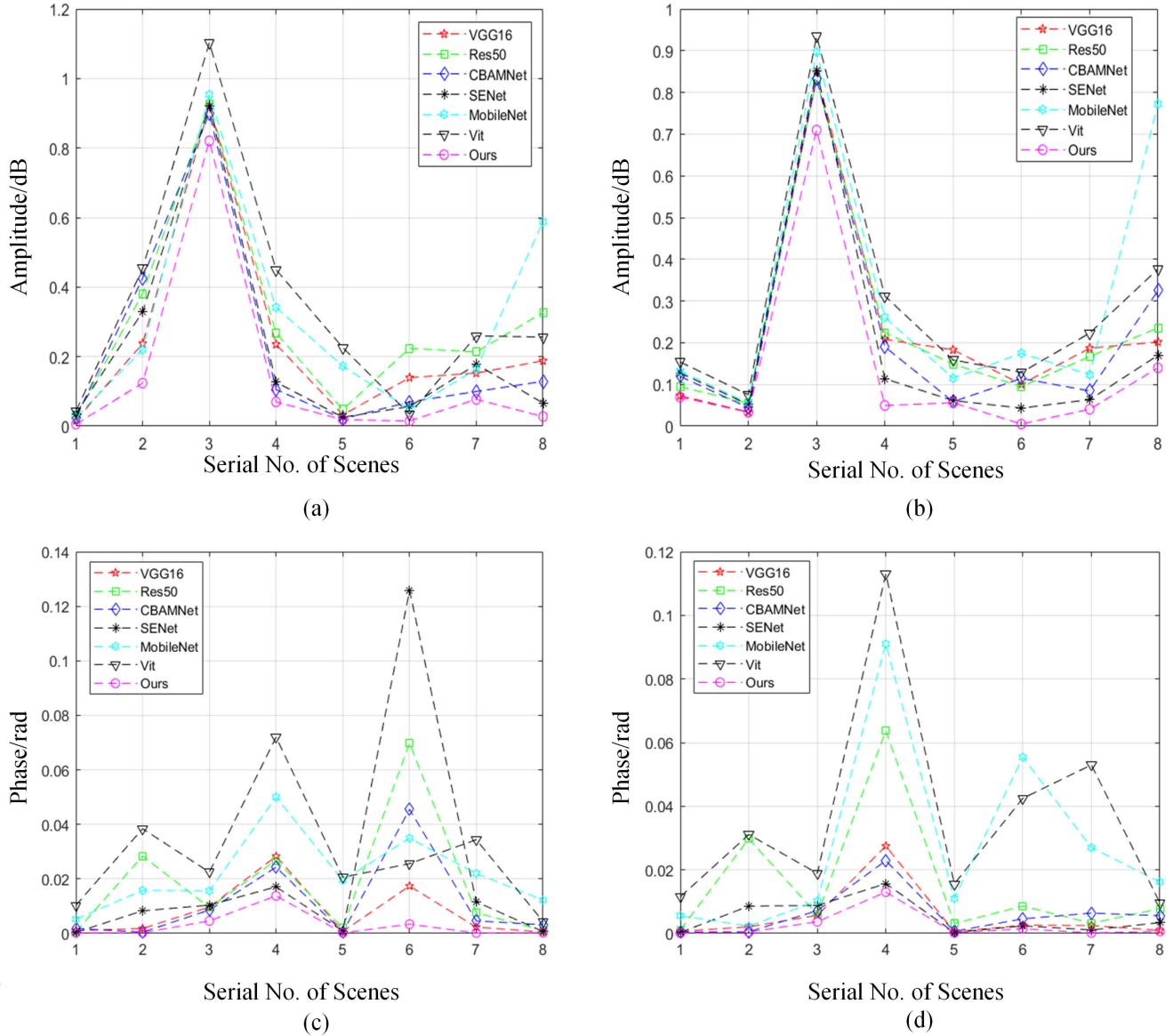
Fig. 12. Generalization analysis of the polarimetric channel imbalance estimation for all methods. (a) Amplitude imbalance of the receiver. (b) Amplitude imbalance of the transmitter. (c) Phase imbalance of the receiver. (d) Phase imbalance of the transmitter.

TABLE V
POLARIMETRIC CHANNEL IMBALANCE ESTIMATION ACCURACY METRICS FOR DIFFERENT METHODS IN THE NO.2 TEST SCENARIO

Method	VGGNet16	SENet	CBAM	ResNet50	MobileNet	ViT	AMHC-DTENET
$|\Delta f_r|$/dB	0.3480	0.3767	0.5866	0.3438	0.3338	0.2898	**0.2563**
$|\Delta f_t|$/dB	0.4052	0.4351	0.6759	0.3996	0.3884	0.3393	**0.3034**
$|\Delta \theta_r|$/rad	0.0779	0.0457	0.1573	0.0449	0.0431	0.0337	**0.0284**
$|\Delta \theta_t|$/rad	0.0805	0.0477	0.1612	0.0471	0.0452	0.0356	**0.0304**

The bold values represent lower scene channel imbalance estimation errors obtained by the method through the extracted distribution targets.

and 0.0501 rad lower than that of VGGNet16 at the receiver and transmitter, respectively. Compared to other networks, the accuracy of the polarimetric channel imbalance estimation obtained by the distribution target extracted by the ViT is close to the results obtained by the method proposed in this article, whereas the amplitude imbalance accuracies obtained in this article are

accurate by 0.0335 dB at the receiver side and 0.0359 dB at the transmitter side, respectively.

Combined with the results of distributed target extraction, the wrong extraction of urban targets will seriously affect the amplitude–phase imbalance estimation. In addition, from the test results of these two test scenes, the proposed method is
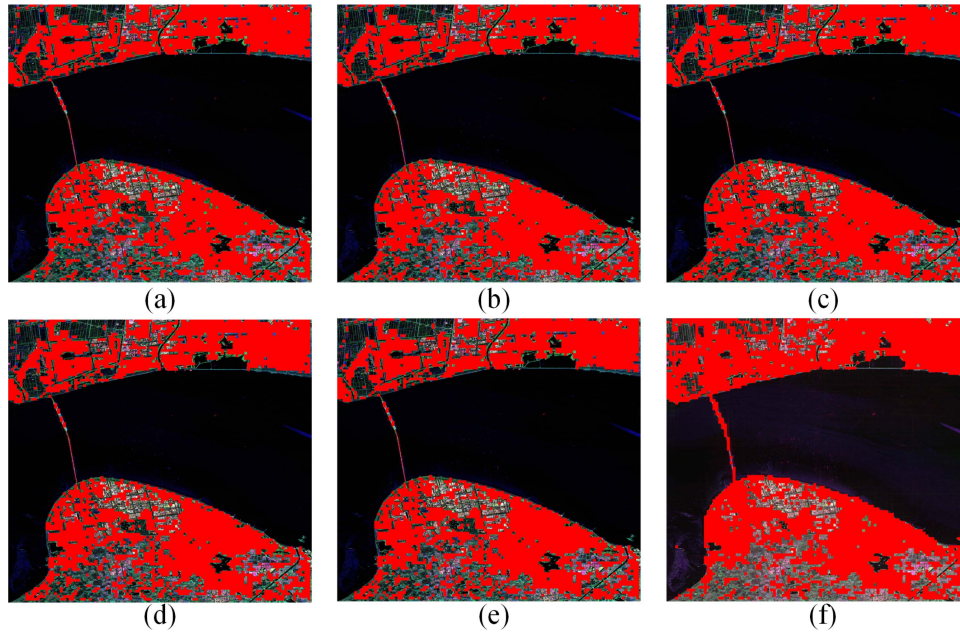
Fig. 13.   Robustness analysis of distributed target extraction of the AMHC-DTENet. (a) Label image. (b) Add 0.2-dB amplitude imbalance. (c) Add 0.3-dB amplitude imbalance. (d) With 3-dB SNR noise. (e) With 1.5 dB/30° imbalance, −10-dB crosstalk. (f) With −1 dB/−20° imbalance, −30-dB crosstalk and 10-dB SNR noise.

more effective in extracting distributed targets, and can obtain better estimation accuracy of polarimetric channel imbalance in PolSAR images with urban as the main distribution.

V. DISCUSSION

In order to verify the generalization ability of the AMHC-DTENet proposed in this article, we randomly select a PolSAR image from each wave code image sequence in Table I as the test image. The evaluation results of all methods are shown in Fig. 12. In this experiment, each scene is performed four times, and the average value is taken as the polarimetric channel imbalance estimation result of the scene. It can be seen from Fig. 12(a) and (b) that all methods have large deviations in the amplitude imbalance estimation when testing the third scene, which may be due to the fact that the distribution targets of the scene are less. This leads to a high false detection rate of distributed targets. Therefore, the amplitude imbalance estimate obtained from the extracted distribution target has a large deviation. In addition, among all the network-based methods, the ViT has a poor generalization ability and performs poorly in most scenarios, except for individual scenarios where it yields high accuracy in estimating polarimetric channel imbalance.

Compared with other methods, the estimation bias of the polarimetric channel imbalance obtained by the AMHC-DTENet proposed in this article is at a lower level in each wave code test scenario. Therefore, it can illustrate the reliability of the polarimetric channel imbalance estimation results obtained by the AMHC-DTENet.

In order to verify the stability of the AMHC-DTENet, we add different combinations of polarimetric distortion and noise with different signal-to-noise ratios (SNRs) in the test scene, as

shown in Fig. 7(e), where the polarimetric distortion is added into the real scattering matrix according to (5) to construct the actual observation matrix, while the noise of the same variance is added to each of the four polarimetric channels according to the same SNR. Here, amplitude imbalance (−1.5 to 1.5 dB), phase imbalance (−30° to 30°), polarimetric crosstalk (−50 to −10 dB), SNR of 3 dB, or −10-dB noise.

To this end, we selected several typical combinations of polarimetric distortion parameters for robustness experiments. The experimental results are shown in Fig. 13. Fig. 13(b)–(d) shows that after adding a certain degree of amplitude imbalance or noise, the extracted distribution target has a trace of false detection target. It mainly appears in the lower part of the image. Second, Fig. 13(e) and (f) shows the impact in extreme cases. Although there were misdetections in the urban areas and water areas, a large number of distributed target areas could be extracted, indicating that the network performed well.

VI. CONCLUSION

In this article, an AMHC-DTENet is proposed for the polarimetric channel imbalance assessment. In this network, the PFCR module can adaptively learn the polarimetric channel weight through 1-D convolution kernel without dimensionality reduction, which utilizes the spatial information of the polarimetric feature channels and avoids the loss of polarimetric feature information caused by dimensionality reduction processing of the polarimetric channel information. At the same time, the hybrid deformable convolution module is embedded into the deep hierarchical polarimetric feature extraction, replacing the residual block with traditional convolution. The receptive field size is adaptively changed by the learnable convolution offset,

and the multiscale features of the spatial information from irregularly shaped targets are extracted. Second, under the ViT framework, the polarimetric high-level semantic features are encoded and different behavior information is learned through the self-attention mechanism, replacing the fully connected layer, which can effectively capture global information and exploit the heterogeneity characteristics of distributed targets in different regions of PolSAR images. And a contrast learning strategy is used to form a multibranch feature extraction network.

We conduct ablation experiments on two verification sets with rich ground target types. Through the extraction performance of distributed targets, the network module designed by us can improve the extraction accuracy of distributed targets. Compared with the distributed target extraction network proposed by previous scholars and the network used in remote sensing image applications, the polarimetric channel imbalance estimation deviation obtained from the extracted distributed targets by the method proposed in this article is lower, which shows the effectiveness of the AMHC-DTENet proposed in this article. At the same time, in the generalization experiment, the polarimetric channel imbalance estimation bias obtained by the AMHC-DTENet proposed in this article is at a low level, indicating the reliability of the AMHC-DTENet proposed in this article. In addition, in the robustness experiment, the AMHC-DTENet proposed in this article performs well under different combinations of polarimetric distortion. And through the test experiment of two scenes of urban area as the main distribution, it can be seen that the estimation accuracy of the polarimetric channel imbalance obtained by the method proposed in this article is higher, and the distributed target can be effectively extracted in the urban area.

REFERENCES

[1] H. Dong, L. Zhang, D. Lu, and B. Zou, "Attention-based polarimetric feature selection convolutional network for PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4001705, doi: 10.1109/LGRS.2020.3021373.

[2] P. Zhang et al., "PolSAR image classification using hybrid conditional random fields model based on complex-valued 3-D CNN," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 3, pp. 1713–1730, Jun. 2021.

[3] W. Nie, K. Huang, J. Yang, and P. Li, "A deep reinforcement learning-based framework for PolSAR imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4403615, doi: 10.1109/TGRS.2021.3093474.

[4] Q. Wu, B. Hou, Z. Wen, Z. Ren, and L. Jiao, "Cost-sensitive latent space learning for imbalanced PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4802–4817, Jun. 2021.

[5] L. Zhang, S. Zhang, B. Zou, and H. Dong, "Unsupervised deep representation learning and few-shot classification of PolSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5100316, doi: 10.1109/TGRS.2020.3043191.

[6] C. Yang, B. Hou, B. Ren, Y. Hu, and L. Jiao, "CNN-Based polarimetric decomposition feature selection for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8796–8812, Nov. 2019.

[7] H. Dong, B. Zou, L. Zhang, and S. Zhang, "Automatic design of CNNs via differentiable neural architecture search for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6362–6375, Sep. 2020.

[8] U. Balss, H. Runge, S. Suchandt, and X. Y. Cong, "Automated extraction of 3-D ground control points from SAR images—An upcoming novel data product," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 5023–5026.

[9] Y. Tan, F. Ren, and S. Xiong, "Automatic extraction of built-up area based on deep convolution neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3333–3336.

[10] Q. Bai, G. Gao, X. Zhang, L. Yao, and C. Zhang, "LSDNet: Lightweight CNN model driven by PNF for PolSAR image ship detection," *IEEE J. Miniaturization Air Space Syst.*, vol. 3, no. 3, pp. 135–142, Sep. 2022.

[11] K. Jin, Y. Chen, B. Xu, J. Yin, X. Wang, and J. Yang, "A patch-to-pixel convolutional neural network for small ship detection with PolSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6623–6638, Sep. 2020.

[12] Y. Chang, J. Yang, P. Li, L. Shi, and L. Zhao, "Polarimetric calibration and assessment of GF-3 images in steppe," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-3, pp. 143–146, Apr. 2018. [Online]. Available: https://isprs-archives.copernicus.org/articles/XLII-3/143/2018/

[13] T. Motohka, O. Isoguchi, M. Sakashita, and M. Shimada, "Results of ALOS-2 PALSAR-2 calibration and validation after 3 years of operation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4169–4170.

[14] R. Caves, "RADARSAT-2 polarimetric calibration performance over five years of operation," in *Proc. 10th Eur. Conf. Synthetic Aperture Radar*, 2014, pp. 1–4.

[15] Y. Chang et al., "Polarimetric calibration of SAR images using reflection symmetric targets with low helix scattering," *Int. J. Appl. Earth Observ. Geoinform.*, vol. 104, Dec. 2021, Art. no. 102559. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030324342100266X

[16] M. Whitt, F. Ulaby, P. Polatin, and V. Liepa, "A general polarimetric radar calibration technique," *IEEE Trans. Antennas Propag.*, vol. 39, no. 1, pp. 62–67, Jan. 1991.

[17] A. Freeman, Y. Shen, and C. Werner, "Polarimetric SAR calibration experiment using active radar calibrators," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 2, pp. 224–240, Mar. 1990.

[18] X. Zhao, Y. Deng, H. Guo, X. Liu, C. Wang, and G. Song, "A bragg-like point extraction method for co-polarization channel imbalance calibration," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3850–3866, Mar. 2023, doi: 10.1109/JSTARS.2023.3252014.

[19] H. Guo, X. Zhao, X. Liu, and W. Yu, "Channel imbalance calibration based on the zero helix of Bragg-like targets," *Remote Sens.*, vol. 15, no. 7, pp. 2072–4292, 2023.

[20] X. Zhao, Y. Deng, H. Zhang, and X. Liu, "A channel imbalance calibration scheme with distributed targets for circular quad-polarization SAR with reciprocal crosstalk," *Remote Sens.*, vol. 15, no. 5, pp. 2072–4292, 2023.

[21] Y. Han et al., "On the method of circular polarimetric SAR calibration using distributed targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5203216, doi: 10.1109/TGRS.2023.3244584.

[22] J. van Zyl, "Calibration of polarimetric radar images using only image parameters and trihedral corner reflector responses," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 3, pp. 337–348, May 1990.

[23] S. Quegan, "A unified algorithm for phase and cross-talk calibration of polarimetric data-theory and observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 1, pp. 89–99, Jan. 1994.

[24] T. Ainsworth, L. Ferro-Famil, and J. -S. Lee, "Orientation angle preserving a posteriori polarimetric SAR calibration," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 4, pp. 994–1003, Apr. 2006.

[25] L. Liang, Y. Zhang, and D. Li, "Fast Huynen–Euler decomposition and its application in disaster monitoring," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4231–4243, Apr. 2021, doi: 10.1109/JSTARS.2021.3070897.

[26] L. Shi, P. Li, J. Yang, L. Zang, X. Ding, and L. Zhao, "Research process of full-polarimetric SAR calibration without using corner reflectors," *Nat. Remote Sens. Bull.*, vol. 25, no. 11, pp. 2211–2219, 2021.

[27] S. Jiang, X. Qiu, B. Han, H. U. Wenlong, and L. U. Xiaojun, "A fast evaluation method of channel imbalance of polarimetric SAR based on common distributed targets," *J. Univ. Chin. Acad. Sci.*, vol. 36, no. 2, pp. 280–288, 2019, doi: 10.7523/j.issn.2095-6134.2019.02.017.

[28] M. Shimada, "Model-based polarimetric SAR calibration method using forest and surface-scattering targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1712–1733, May 2011.

[29] H. W. Zhang Jingjing and Y. Qiang, "Robust distributed-target-based calibration method for polarimetric SAR using spherically truncated covariance matrix," *J. Radars*, vol. 5, no. 6, pp. 701–710, 2016, doi: 10.12000/JR16138.

[30] J. Chen, M. Sato, and J. Yang, "Polarimetric calibration using distributed odd-bounce targets," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 1079–1082.

[31] H. Kimura, "Calibration of polarimetric PALSAR imagery affected by faraday rotation using polarization orientation," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 3943–3950, Dec. 2009.

[32] L. Shi, P. Li, J. Yang, L. Zhang, X. Ding, and L. Zhao, "Co-polarization channel imbalance phase estimation by corner-reflector-like targets," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 255–266, 2019.

[33] L. Shi, J. Yang, and P. Li, "Co-polarization channel imbalance determination by the use of bare soil," *ISPRS J. Photogrammetry Remote Sens.*, vol. 95, no. 9, pp. 53–67, 2014.

[34] A. Villa, L. Iannini, D. Giudici, A. Monti-Guarnieri, and S. Tebaldini, "Calibration of SAR polarimetric images by means of a covariance matching approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 674–686, Feb. 2015.

[35] S. Shangguan, X. Qiu, K. Fu, B. Lei, and W. Hong, "GF-3 polarimetric data quality assessment based on automatic extraction of distributed targets," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4282–4294, Jul. 2020, doi: 10.1109/JSTARS.2020.3012151.

[36] C. Zhao, W. Zhu, and S. Feng, "Hyperspectral image classification based on kernel-guided deformable convolution and double-window joint bilateral filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5506505, doi: 10.1109/LGRS.2021.3084203.

[37] Y. Liu, W. Wang, Q. Li, M. Min, and Z. Yao, "DCNet: A deformable convolutional cloud detection network for remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8013305, doi: 10.1109/LGRS.2021.3086584.

[38] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018.

[39] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542. [Online]. Available: https://doi.org/10.48550/arXiv.1910.03151

[40] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[41] S.-W. Chen, X.-S. Wang, and M. Sato, "Uniform polarimetric matrix rotation theory and its applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4756–4770, Aug. 2014.

[42] P. T. D. Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Operations Res.*, vol. 134, no. 1, pp. 19–67, 2005.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *J. Comput. Sci.*, 2014.

[44] F. Jia, "Fire flame detection in tunnel based on Halcon and ResNet50," in *Proc. 3rd Int. Conf. Neural Netw., Inform. Commun. Eng.*, 2023, pp. 256–259.

[45] Y. Guo, S. Chen, R. Zhan, W. Wang, and J. Zhang, "SAR ship detection based on YOLOV5 using CBAM and BiFPN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 2147–2150.

[46] S. Mei, X. Wei, B. Jin, and J. Guo, "SAR image target recognition based on senet depth separable convolutional neural network," in *Proc. 2nd China Int. SAR Symp.*, 2021, pp. 1–6.

[47] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for Mobile Vision Applications," 2017. [Online]. Available: http://arxiv.org/abs/1704.04861

Haoyang Li received the B.S. degree in aerospace science and technology from Xidian University, Xi an, China, in 2021. He is currently working toward the master's degree in electronic and information engineering with the Department of Space Microwave Remote Sensing System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

He is currently with the University of Chinese Academy of Sciences, Beijing. His research interests include polarimetric synthetic aperture radar image calibration and deep learning.

Mingjie Zheng (Member, IEEE) received the Ph.D. degree in engineering from the Institute of Electronics, Chinese Academy of Sciences (IECAS), Beijing, China, in 2003.

In 2003, she joined the Department of Space Microwave Remote Sensing System, IECAS, where she was involved in spaceborne synthetic aperture radar (SAR) system design and simulation analysis, during which she presided some research projects about spaceborne SAR technology. She is currently a Research Fellow and a Doctoral Supervisor with the Aerospace Information Research Institute, Chinese Academy of Sciences. Her research interests include spaceborne SAR ground moving target indicator technology, spaceborne SAR full-link system design, and spaceborne SAR dual-frequency interferometric technology and polarimetric technology.

Yonghui Han received the B.S. degree in aerospace science and technology from Xidian University, Xi an, China, in 2017. He is currently working toward the Ph.D. degree in communication and information systems engineering with the National Key Laboratory of Microwave Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

He is also with the University of Chinese Academy of Sciences, Beijing. His research interests include synthetic aperture radar imaging and signal processing.

Xingjie Zhao was born in Xingtai, China, in 1998. He received the B.S. degree in electronic and information engineering from Northwestern Polytechnical University, Xi'an, China, in 2019. He is currently working toward the Ph.D. degree in PolSAR calibration with the Department of Space Microwave Remote Sensing System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

He is currently with the University of Chinese Academy of Sciences, Beijing. His research interests include polarimetric synthetic aperture radar information extraction and calibration.