

Dual Attention-Based Global-Local Feature Extraction Network for Unsupervised Change Detection in PolSAR Images

Dazhi Xu¹, Graduate Student Member, IEEE, Ming Li¹, Member, IEEE, Yan Wu¹, Member, IEEE, Peng Zhang¹, Member, IEEE, and Xinyue Xin¹

Abstract—Due to the interference of multiplicative speckles, it is challenging to accurately detect changes in polarimetric synthetic aperture radar (PolSAR) images. Convolutional neural network has been proven to learn rich local features from PolSAR data. However, convolution kernels with limited receptive fields have difficulty in exploring global information. Here, a dual attention-based global-local feature extraction network (DA-GLN) is developed for unsupervised PolSAR image change detection (CD). First, we use fuzzy C-means clustering on the enhanced Shannon entropy difference image to automatically generate pseudolabeled samples required for unsupervised CD. Subsequently, our DA-GLN utilizes a deep residual shrinkage network that incorporates channel attention mechanisms and soft-thresholding to weaken the influence of speckle noise and capture local features. Meanwhile, a pooling-based vision transformer is adopted in DA-GLN to extract global features, which introduces pooling layers to complete self-attention spatial information interaction with higher efficiency than the visual transformer. Furthermore, a global-local constraint feature fusion strategy is designed to effectively fuse local and global features. Finally, we employ a feature constraint-focal loss function including feature constraint loss and focal loss as the objective function of DA-GLN. Specifically, the feature constraint loss function is constructed to eliminate feature redundancy and fully exploit the complementarity between features, while the focal loss function is introduced to balance the impact of the inequality between changed and unchanged samples on the network. Numerical experiments on five real spaceborne PolSAR datasets demonstrate that our DA-GLN is more competitive than other state-of-the-art methods.

Index Terms—Dual attention-based global-local feature extraction network (DA-GLN), focal loss function, polarimetric synthetic aperture radar (PolSAR) image, pooling-based vision transformer (PiT), unsupervised change detection.

I. INTRODUCTION

CHANGE detection (CD) aims to identify differences of the same surface in different periods and has been widely

utilized in remote sensing (RS) applications [1], [2], [3], [4], [5], [6], [7]. Benefiting from the multipolarization transceiver operation mode [8], polarimetric synthetic aperture radar (PolSAR) can provide comprehensive polarization information for object recognition. Hence, it has made advanced development in various fields [9], [10], [11], [12], [13], [14]. Over the last two decades, PolSAR has also been gradually applied to CD tasks and plays an important role in disaster monitoring [15], vegetation assessments [16], and glacier dynamics [17]. However, PolSAR image CD is challenging and rarely researched owing to the complex semantic scenes as well as the inherent influence of speckles. Therefore, it is urgent to propose a novel method for PolSAR image CD.

Since manually labeling surface changes of PolSAR images is time consuming and labor-intensive, unsupervised CD methods become mainstream. Several studies have gone into developing traditional approaches for unsupervised PolSAR image CD, such as those based on polarization distance [18], statistical distribution [19], [20], and polarization scattering features [15]. For example, Liu et al. [18] used the Wishart distance [21] to detect specific changes. Akbari et al. [19] adopted the Hotelling-Lawley trace (HLT) statistic to measure the difference between bitemporal PolSAR images. A Shannon entropy (SE)-based approach [20] is developed to obtain difference image (DI) and further detect changes using the empirical test size (ETS). Mahdavi et al. [15] introduced neighborhood variation coefficients to guide the measurement of differences between the scattering matrices. The above methods generate DI based on the similarity measures and further detect changes by analyzing the DI [22]. Hence, the rationality of the similarity measurement and the validity of the DI analysis greatly affect the final CD results. Nevertheless, the change information may be missed during the process of generating the DI because of the interference of speckle noise, resulting in incomplete detection. Furthermore, changes cannot be fully analyzed from DI utilizing global thresholds, simple clustering, or weak semantic hand-crafted features.

Over the past few years, deep learning theory has been widely applied in computer vision (CV) [23], [24], [25]. As one of the representative models of deep learning, the convolutional neural network (CNN) performs well in feature extraction and is gradually being applied to SAR [26], [27], [28], [29], [30] and PolSAR image CD tasks [31], [32], [33], [34]. For example, a principal component analysis network (PCANet) [26] replaces

Manuscript received 2 February 2024; revised 7 May 2024; accepted 24 May 2024. Date of publication 28 May 2024; date of current version 14 June 2024. This work was supported in part by the Natural Science Foundation of China under Grant 62172321, and in part by the Civil Space Thirteen Five Years Pre-Research Project under Grant D040114. (Corresponding authors: Ming Li; Yan Wu.)

Dazhi Xu, Ming Li, Peng Zhang, and Xinyue Xin are with the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: liming@xidian.edu.cn).

Yan Wu is with the Remote Sensing Image Processing and Fusion Group, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: ywu@mail.xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3406421

the convolutional kernels with PCA filters to capture noise-robust pixel neighborhood features. Campos et al. [27] used CNN to reduce the false alarm rate of CD by distinguishing change targets from irrelevant strong scatterers. A dual-domain network (DDNet) [28] utilizes both frequency and spatial domain feature representations to mitigate scattering noise and complete the change feature modeling. Zhang et al. [29] introduced statistical texture information as auxiliary features to effectively suppress the effect of scattering noise. Robust unsupervised small area CD (RUSACD) [30] uses the two-stage center-constrained fuzzy C-means (FCM) algorithm to obtain the pseudolabeled set and mitigate the adverse effects of sample imbalance. As for the research of PolSAR image CD, a local restricted CNN [31] effectively detects changes from the discriminative DI. Based on the DI generated by transfer learning, a novel end-to-end three-channel deep neural network (TCD-Net) [32] uses adaptive multiscale shallow blocks and residual blocks to detect changes in an unsupervised manner. Seydi et al. [33] proposed an end-to-end multidimensional CNN (EEMCNN) for PolSAR image CD, which adopts multidimensional dilated convolution to simultaneously extract change-aware features. A novel joint CD network (NJCDN) [34] processes the amplitude PolSAR data and covariance matrix using metric learning. Those aforementioned CNN-based methods implicitly learn rich local feature representations from the raw data and perform better than traditional methods. However, CNN with limited receptive fields has difficulty exploring global spatial correlations between change-aware features. Besides, some change irrelevant features from the PolSAR images may be unexpectedly captured.

In addition to purely convolution-based methods, methods based on attention mechanisms, such as channel attention and spatial attention, have also been applied in SAR image CD [35], [36], [37]. For example, a deep cascade network [35] introduces residual learning to mitigate gradient explosion and designs a channel weight-based model to enhance the saliency of features. Gao et al. [36] integrated multiscale feature representation using the attention-based fusion mechanism and improved feature discrimination through metric learning. Zhao et al. [37] integrated the spatial and frequency domain features of SAR images to suppress noise interference. The aforementioned methods adopt attention mechanisms to reweight change-aware features in the channel or spatial dimension, thereby noting changed areas of interest and improving feature discriminability [38]. In addition, these methods can capture global information to some extent [39]. Nevertheless, CNN using attention mechanisms is still difficult to effectively correlate distant concepts in space-time [40] owing to the local property of convolution.

With the ability to establish long-range correlation by interacting with self-attention spatial information, transformer [41] has been widely applied in natural language processing and CV. As a classic application of transformer in CV, a new transformer-based model named visual transformer (ViT) [42] is proposed for the image classification tasks, which can model long-range dependencies of tokens and learn image dependencies at different locations. Different from CNN, ViT can capture global context information without stacking a large number of convolutional layers. Currently, several studies have successfully applied ViT

to SAR image CD [43], [44]. For example, a differential attention metric-based network [43] employs the modified attention module in ViT to enhance differential image features. Du et al. [44] encoded contextual features to suppress noise and described the boundary structures of the changed areas. The aforementioned methods indicate that the ViT-based models perform well in SAR image CD. Noteworthy, the features captured only by ViT contain insufficient local semantic information and are not conducive to detecting local detail information of the changed areas. Moreover, the applicability of the ViT in PolSAR image CD has not been validated.

To avoid the limitations of using CNN- and transformer-based models alone for unsupervised PolSAR image CD, we develop a novel dual attention-based global-local feature extraction network (DA-GLN). It combines the advantages of channel attention, self-attention, CNN, transformer, constraint feature fusion, and focal loss function. First, the pseudolabeled samples can be constructed from the enhanced SE DI using the FCM clustering, avoiding the manual labeling of training samples. Second, the developed DA-GLN can simultaneously capture global and local change-aware features from bitemporal PolSAR images. Furthermore, a global-local constraint feature fusion strategy (GLCFF) is designed in DA-GLN to efficiently integrate the above features with different attributes. Finally, a feature constraint-focal loss (FC-F loss) function including feature constraint loss and focal loss function is designed as the objective function of DA-GLN, which considers the consistency and difference constraints between features with different attributes and the sample imbalance in CD. To the best of our knowledge, we first apply the transformer-based structures to the PolSAR image CD tasks. In this article, the main contributions are as follows.

- 1) The DA-GLN is developed for the final CD, which combines the advantages of CNN for learning local spatial context and transformers for modeling global long-range correlation. It employs the DRSN based on channel attention to adaptively filter irrelevant features affected by speckle noise and focus on capturing local features. Moreover, self attention-based pooling-based vision transformer (PiT) is also adopted in the network for extracting global features with efficient spatial information interaction efficiency.
- 2) The GLCFF strategy is designed in DA-GLN to obtain compact and nonredundant fusion features. Based on the consistency and the difference between global and local features, GLCFF constructs a feature constraint loss function to constrain the learning process of features, thus removing feature redundancy and fully exploiting the complementarity between features with different attributes.
- 3) The FC-F loss function is employed to supervise the training process of DA-GLN. In addition to the feature constraint loss constructed in GLCFF, the focal loss function is also introduced as part of the FC-F loss function to balance the impact of the inequality between changed and unchanged areas on the network.

The rest of this article is organized as follows. Section II presents related works. Section III details the methodology.

Section IV describes the datasets, experiment settings, evaluation indicators, ablation study, discussion, and experimental comparisons in detail. Finally, Section V concludes this article.

II. RELATED WORK

A. ViT

As a successful exploration of applying the pure Transformer structure to CV tasks, ViT [42] has received widespread attention in image processing and provided new ideas for subsequent research. It contains only the Transformer encoder and utilizes stacked MSA to explore the global relationships between tokens. As a result, ViT can fully explore the global contextual information as well as the long-range dependencies of images [45].

Currently, ViT has been extensively applied to RS image processing, such as image classification [46], scene classification [47], image segmentation [48], target detection [49], and CD [50], [51]. For example, Xue et al. [46] designed a transformer-based structure to capture the long-range dependencies and hierarchical spatial features. On this basis, multimodal features are effectively fused for the accurate classification of multimodal RS data. Considering the geometric information as well as the channel information of the image, a spatial-channel feature preserving ViT [47] effectively enhances the classification ability of ViT. After innovatively converting the 2-D features extracted by ViT into 3-D features. Wang et al. [48] fused the generated multiscale features with rich context information to improve the segmentation accuracy. Zhou et al. [49] designed a Transformer-based detector to effectively represent dense objects and alleviate the semantic gap between multiple scales. Wang et al. [50] proposed a joint spectral, spectral, and temporal transformer for hyperspectral image CD. Dai et al. [51] designed a MobileViT-based network to capture the spatial features of planetary images and improve the distinguishability of change-aware features.

B. Global-Local Feature Extraction Networks for RS Image Processing

Due to the limited receptive field of the convolution kernels, CNN-based models focus on capturing local features, thus ignoring global contextual information. To address this issue, several studies have combined CNN with advanced models capable of capturing global structural information for RS image downstream tasks, such as image classification [52], [53], [54], [55], [56], object detection [57], [58], semantic segmentation [59], and CD [60], [61]. For example, Zhuo et al. [52] simultaneously utilized multiscale CNN and multihop GCN to capture multiscale features containing local-global structural relationships. A novel global-local transformer network [53] learned local spatial features using multiscale aggregated CNN and extracts global spectral sequence properties using ViT. Taking global spatial context into account, [54] learned discriminative spatial features by overcoming the limitation of the receptive field and develops a dual-view spectral aggregation model to capture short- and long-view spectral features. Liu et al. [55] adopted CNN to

learn the local features and ViT for the extraction of the global context information. Duan et al. [56] designed two different branches to learn the deep aggregation features and extracted the global structural features through grafting regular spatial convolution. Teng et al. [57] mined global contextual information by encoding the image from a global perspective and designed a clip-long short-term memory to learn local correlations of image features. ATC-Net [58] captured global-local context features from Transformer and CNN and enhanced the fused features using attention mechanisms. A dual-branch backbone network of CNN-transformer [59] adopts self-attention and cross-fusion mechanisms to fuse extracted global-local features. As for the CD tasks, [60] employed the transformer to capture local-global semantic features by encoding the patches from the CNN feature maps. Li et al. [61] used the parallel-branch ConvTrans block as the basic component to fully capture multiscale global-local information.

III. PROPOSED METHOD

The architecture of the proposed method is shown in Fig. 1. A SE DI is first generated and then refined to acquire an enhanced DI (EDI). Then, pseudolabeled samples for training DA-GLN are constructed based on the EDI. Finally, uncertain samples are classified by the trained DA-GLN to obtain the CD result.

In our implementations, the DA-GLN can simultaneously capture global-local features and efficiently fuse them using the GLCFF strategy, as shown in Fig. 2. Moreover, we design the FC-F loss to supervise the training process of our DA-GLN. In Sections III-D and III-E, we will introduce the proposed DA-GLN and the FC-F loss function in detail.

A. Preprocessing

In the acquisition of PolSAR data, external uncertainties may make the PolSAR data inaccurate, resulting in the loss of credibility of CD results. Therefore, before extracting difference information and detecting changes from bitemporal PolSAR images, the original Gaofen3 PolSAR data should be preprocessed, including the following steps.

- 1) Geometric correction is first employed to maximize the alignment of the acquired PolSAR images with the real geographic coordinates.
- 2) Radiometric correction with DEM is applied to correct the radiation in different regions to the same energy level.
- 3) The filtering method proposed by Cui et al. [62] was used to effectively suppress speckle noise and increase the signal-to-noise ratio of the PolSAR images.
- 4) As an extended application of SAR-SIFT [63] on PolSAR, PolSAR-SIFT is developed to register bitemporal PolSAR images into the same coordinate system.
- 5) The Z-score normalization is used to normalize the nine elements of each polarimetric covariance matrix, thus obtaining PolSAR data with the same metric scales and improving the convergence speed of the model.

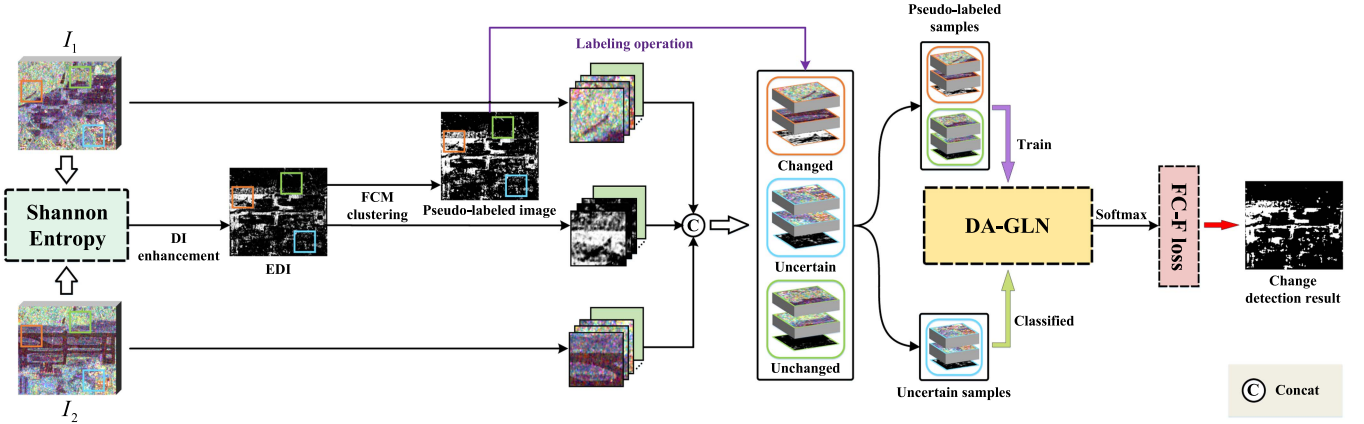


Fig. 1. Architecture of the proposed method.

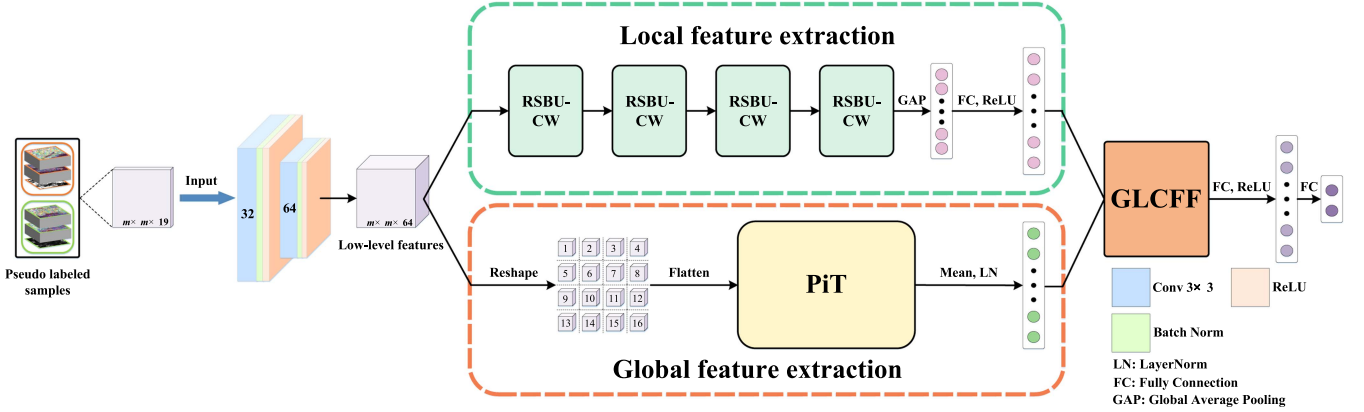


Fig. 2. Illustration of the proposed DA-GLN.

B. Generation of DI

Since the proposed method detects changes in an unsupervised manner, DIs need to be first generated by measuring the degree of difference between the bitemporal PolSAR images. Based on the DI, the high-confidence pseudolabeled set required for unsupervised CD can be easily constructed. In a multilook processed PolSAR image, each pixel is expressed by an average polarimetric covariance matrix \mathbf{C} , which is defined as

$$\mathbf{C} = \frac{1}{\mathcal{L}} \sum_{l=1}^{\mathcal{L}} \Psi_l \Psi_l^H = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \quad (1)$$

where \mathcal{L} is the nominal number of looks used for averaging, $\Psi \in \mathbb{C}^3$ is the complex scattering vector expressed in SLC format, and $(\cdot)^H$ is the Hermitian transposition operator.

On the assumption that \mathbf{C} obeys the complex Wishart distribution, [20] employed the SE operator to measure the difference between bitemporal PolSAR images and generate the SE DI DI_S , which is defined as

$$DI_S = \frac{\mathcal{N} \left[H_S(\hat{\phi}_A) - H_S(\hat{\phi}_B) \right]^2}{\sigma^2(\hat{\phi}_A) + \sigma^2(\hat{\phi}_B)} \quad (2a)$$

$$H_S(\hat{\phi}_A) = E \{ -\log p_{\mathbf{C}}(\mathbf{C}_1) \} \quad (2b)$$

$$H_S(\hat{\phi}_B) = E \{ -\log p_{\mathbf{C}}(\mathbf{C}_2) \} \quad (2c)$$

Where \mathbf{C}_1 and \mathbf{C}_2 represent the pre- and postchanged covariance matrices, $p_{\mathbf{C}}(\mathbf{C})$ is the probability density function of \mathbf{C} , $E\{\cdot\}$ denotes the statistical expectation operator, $\hat{\phi}$ represents the maximum likelihood estimation of ϕ , ϕ denotes the parameter vector of $p_{\mathbf{C}}(\mathbf{C})$, $H_S(\hat{\phi})$ represents the SE of $\hat{\phi}$, $\mathcal{N}[\cdot]$ is the sampling size, and $\sigma^2(\hat{\phi})$ represents the variance of $\hat{\phi}$.

Since entropy is an effective measure of the concentration trends in data, the SE method can suppress the detrimental effects of noise by capturing the average information of the pixels in bitemporal images. However, the generated DI_S cannot preserve the critical change edge information and its overall discriminability needs to be further enhanced. In [64], an EDI DI_E improved by the DI_S was designed to enhance boundary localization and difference discrimination. It combines the gamma correction principle with the edge information contained in the bitemporal PolSAR images. In this article, we utilize the generated DI_E to obtain the pseudolabeled set and provide

critical difference information. The EDI DI_E is expressed as

$$DI_E = DI_S^\gamma \quad (3a)$$

$$\gamma = \begin{cases} \beta(1 + \bar{E}), & \beta \geq 1 \\ \beta(1 - \bar{E}), & \beta < 1 \end{cases} \quad (3b)$$

$$\beta = (1 - \bar{E})10^{T_{DI_S} - DI_S} \quad (3c)$$

where T_{DI_S} is the average of the thresholds generated by binary segmentation of the DI_S using the conventional Ostu and Kittler-Illingworth threshold methods, and \bar{E} is the edge information extracted from the bitemporal PolSAR images.

C. Construction of Pseudolabeled Samples

Since the proposed method is unsupervised, we perform FCM clustering to divide DI_E into a pseudolabeled set $\Omega = \{\Omega_{ch}, \Omega_{uch}, \Omega_{uc}\}$ containing the changed class Ω_{ch} , the unchanged class Ω_{uch} , and the uncertain class Ω_{uc} . In Ω , Ω_{ch} and Ω_{uch} can provide high-confidence labels for training samples, whereas the classes of the pixels belonging to Ω_{uc} need to be further determined by classification. Notably, the number of Ω_{ch} and Ω_{uch} in the training samples is imbalanced, with the former usually having a smaller number than the latter. This is mainly attributed to two factors: 1) changed areas are more difficult to be detected and thus have fewer labels than unchanged areas; and 2) the changed areas are often much smaller than the unchanged areas in actual ground object scenes. Hence, considering the sample imbalance issue is necessary when designing the network for CD.

After that, considering that both covariance matrices C_1 and C_2 in (1) are symmetric matrices and the diagonal elements are real, we use a 9-D real vector $[C_{11}, \Re(C_{12}), \Re(C_{13}), C_{22}, \Re(C_{23}), C_{33}, \Im(C_{12}), \Im(C_{13}), \Im(C_{23})]$ as the original feature of the monotemporal PolSAR image. $\Re(\cdot)$ and $\Im(\cdot)$ represent the real and imaginary parts of the complex number. For CD tasks, the bitemporal images $I_1, I_2 \in \mathbb{R}^{W \times H \times 9}$ are usually stacked as the input information of the network, where W and H represent the width and height of the bitemporal images. However, speckle noise in PolSAR images may interfere with the ability of the network to capture change-aware features. Consequently, $DI_E \in \mathbb{R}^{W \times H \times 1}$ containing rich difference information is concatenated with I_1 and I_2 as the input information $\mathbf{I}^{Input} \in \mathbb{R}^{W \times H \times 19}$. The introduction of difference information can enhance feature discrimination and thus accelerate the convergence of the network. After that, we extract multichannel stacked patches with size m at the same position of $\mathbf{X}^{Input} \in \mathbb{R}^{m \times m \times 19}$.

D. DA-GLN for CD

In order to generate the final CD results, the DA-GLN is proposed as a classifier to accurately predict the pixel classes in Ω_{uc} . As shown in Fig. 2, DA-GLN simultaneously utilizes the CNN- and transformer-based feature extractors to capture both local and global features. In addition, the GLCFF strategy is also designed to fuse global-local features and obtain nonredundant and compact fusion features. Before extracting

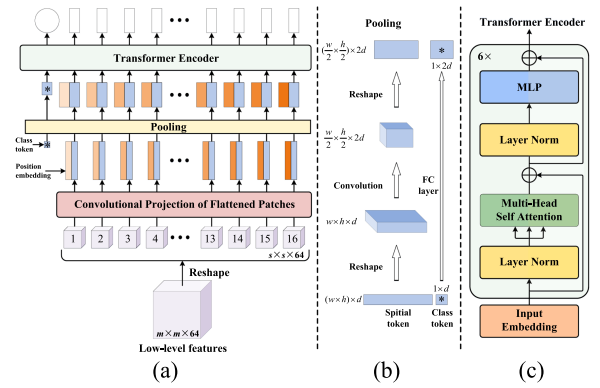


Fig. 3. (a) Structure of the PiT. (b) Pooling layer. (c) Transformer encoder.

global and local features, the DA-GLN first feeds pseudolabeled samples into the initial convolution module to obtain shallow features \mathbf{X}^{Base} . The change discrimination of shallow features is more salient compared to the raw input information. Two alternating 3×3 convolutional layers, batch normalization (BN) layers, and ReLU activation layers are included in the initial convolution module. The amounts of filters in the convolution layer are 32 and 64. The generated \mathbf{X}^{Base} is further fed into the two subsequent branches for extracting global-local features.

1) *Global Feature Extraction Network*: To address the shortcoming that the CNN-based structure cannot effectively explore the global spatial correlation of features, we first design a global feature extraction network as a branch of the DA-GLN for extracting global features. Based on the self-attention mechanism, ViT can adaptively model the semantic relationships between any pixel pairs in space-time to capture global contextual information. Inspired by the design principles of CNN, Heo et al. [65] proposed a novel PiT based on the ViT. The PiT embeds pooling layers into transformer-based structures and transforms features in the spatial dimension, which is similar to CNN. The introduction of the pooling layer effectively improves the efficiency of self-attention spatial information interaction (similar to the receptive field of CNN), thereby improving the performance and generalization of ViT. Besides, PiT computes faster than ViT. In summary, PiT can capture more discriminative global features from bitemporal PolSAR images with complex feature components and chaotic region features. The structure of the whole PiT is shown in Fig. 3(a). Since the transformer takes the feature sequence as input, we further divide \mathbf{X}^{Base} into N small pieces of $s \times s$ and vectorize these pieces as $\mathbf{X}_t \in \mathbb{R}^{N \times (s^2 \cdot C)}$, where $N = m \times m / s^2$ and C denotes the number of feature channels. Then, patch flatten is performed on \mathbf{X}_t and mapped to a d D patch embedding $\mathbf{X}_0 \in \mathbb{R}^{N \times d}$ using a trainable linear projection layer. Furthermore, the learnable location features can be obtained from the added patch embedding. The above process can be described as

$$\mathbf{X}_0 = [\mathbf{X}_t^1 \mathbf{P}; \mathbf{X}_t^2 \mathbf{P}; \dots; \mathbf{X}_t^N \mathbf{P}] + \mathbf{P}_{pos} \quad (4)$$

where $\mathbf{X}_t^1, \mathbf{X}_t^2, \dots, \mathbf{X}_t^N$ represent the vectorized pieces, $\mathbf{P} \in \mathbb{R}^{(s^2 \cdot C) \times d}$ denotes the trainable projection transform, $\mathbf{P}_{pos} \in \mathbb{R}^{N \times d}$.

In contrast to ViT which directly feeds the embedded features into the transformer encoder, PiT first transforms the embedded features in the spatial dimension using the pooling layer. As shown in Fig. 3(b), the pooling layer first reshapes the spatial 2-D token features $\mathbf{X}_0 \in \mathbb{R}^{(w \cdot h) \times d}$ into 3-D token features $\hat{\mathbf{X}}_0 \in \mathbb{R}^{w \times h \times d}$ with spatial structure. Then, a 3×3 convolution layer is used to obtain the feature $\hat{\mathbf{X}}_0^p \in \mathbb{R}^{\frac{w}{2} \times \frac{h}{2} \times 2d}$ that the spatial dimension is halved and the channel dimension is doubled. Finally, reshape $\hat{\mathbf{X}}_0^p$ into a 2-D token feature $\mathbf{X}_0^p \in \mathbb{R}^{(\frac{w}{2} \cdot \frac{h}{2}) \times 2d}$. Since the cls token used for classification is independent of spatial features, the mapping is performed using a fully connected (FC) layer to expand the channel dimensions.

Furthermore, the features \mathbf{X}_0^p obtained from the pooling layer are fed into the transformer encoder with L layers. The structure of a transformer encoder layer is shown in Fig. 3(c), which consists of MSA, multilayer perceptron (MLP), and LayerNorm (LN). The output representation of the l th transformer \mathbf{X}_l^p is defined as

$$\mathbf{X}_l^{p'} = \text{MSA}(\text{LayerNorm}(\mathbf{X}_{l-1}^p)) + \mathbf{X}_{l-1}^p \quad (5a)$$

$$\mathbf{X}_l^p = \text{MLP}(\text{LayerNorm}(\mathbf{X}_l^{p'})) + \mathbf{X}_l^{p'} \quad (5b)$$

where $l = 1, 2, \dots, L$. As the core of the transformer, MSA provides an effective modeling approach for capturing global context information. Multiple headers allow the transformer to learn multiple dependencies from various representation subspaces in different locations. The process can be described as

$$\text{MSA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_u) \mathbf{W}^O \quad (6a)$$

$$\text{head}_i = \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d_k}}\right) \mathbf{V}_i \quad (6b)$$

where $\mathbf{Q}_i = \mathbf{X}_i \mathbf{W}_i^q$, $\mathbf{K}_i = \mathbf{X}_i \mathbf{W}_i^k$, and $\mathbf{V}_i = \mathbf{W}_i^v$ denote key, query, and value. \mathbf{W}^O , \mathbf{W}^q , \mathbf{W}^k , and \mathbf{W}^v represent the learnable parameter matrices corresponding to output, query, key, and value. \mathbf{X}_i is the i th head of the feature matrix, u is the number of multiheads, and d_k is the dimensions of \mathbf{Q} and \mathbf{K} .

In addition, the transformer includes a MLP that can enhance its nonlinear transform capabilities. The MLP is composed of two linear transform layers, with a Gaussian error linear unit in the middle, which can be expressed as

$$\text{MLP}(\mathbf{x}) = \max(0, \mathbf{x} \mathbf{W}_1 + b_1) \mathbf{W}_2 + b_2 \quad (7)$$

where \mathbf{x} is the input features of the MLP, \mathbf{W}_1 and \mathbf{W}_2 are the learnable weight matrices of the two linear transform layers, and b_1 and b_2 are the biases of the two linear transform layers.

After the above process, we can obtain the global features $\mathbf{X}^{\text{Global}}$ corresponding to the shallow features \mathbf{X}^{Base} .

2) *Local Feature Extraction Network*: Affected by the inherent speckle, purely convolution-based structures cannot completely extract local change-aware features and eliminate irrelevant features affected by speckle noise. Hence, we utilize the deep residual shrinkage network (DRSN) [66] to focus on capturing local features. The DRSN has been applied to image denoising with good performance [67], [68]. It embeds soft thresholds as a trainable shrinkage function into the CNN

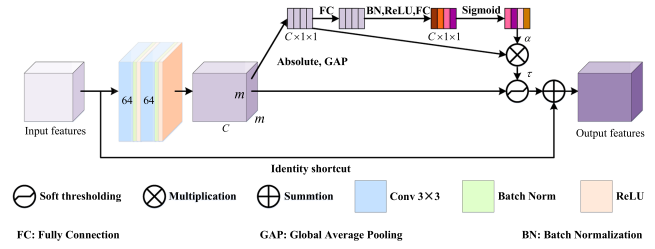


Fig. 4. Structure of the RSBU-CW.

to force unimportant features to zero, thereby weakening the adverse effect of noise and learning more discriminative local features. The core of the DRSN is the residual shrinkage building unit with channel-wise thresholds (RSBU-CW) that can adaptively estimate thresholds in soft-thresholding. Noteworthy, RSBU-CW is an extended application of the squeeze-and-excitation networks (SENet) [69], and both obtain channel weights through the channel attention mechanism. Unlike SENet, RSBU-CW further sets thresholds for each channel of the learnable feature mapping based on the channel weights. As shown in Fig. 4, the RSBU-CW with channel thresholding first simplifies the input features into 1-D vectors through absolute value operation and global average pooling (GAP) layer, and then further propagates them to two FC layers. The number of neurons in FC layers is equal to the number of channels of the input feature, which is set to 64 in this article. The output value of the latter FC layer is converted to a scale parameter $\alpha \in (0, 1)$ using the Sigmoid function. Based on this, the thresholds can be calculated by

$$\tau_c = \alpha_c \cdot \text{average}_{m,m} |\mathbf{x}_{m,m,c}| \quad (8)$$

where τ_c is the threshold for the c th channel corresponding to the feature map, and m and c are the indexes of the side length and channel of the features \mathbf{x} , respectively. Finally, the optimized features can be obtained by summing \mathbf{x} with the soft threshold shrunken features using the identity shortcut.

As shown in Fig. 2, the proposed local feature extraction branch consists of four stacked RSBU-CW modules, a GAP layer, and a FC layer with $2d$ neurons. In this way, the shallow features \mathbf{X}^{Base} can be converted into the local features $\mathbf{X}^{\text{Local}}$ by the aforementioned local feature extraction network.

3) *Global-Local Constraint Feature Fusion (GLCFF)*: The above two branches can learn rich local space contexts and remote dependencies. However, it is challenging to integrate global features $\mathbf{X}^{\text{Global}}$ and local features $\mathbf{X}^{\text{Local}}$ with high quality. Although the local feature extraction branch focuses on extracting local features using the local connection attributes of convolution, the stacked multiple convolution layers and channel attention mechanism have been proved to improve the receptive field and thus model the global information to a certain extent [39]. As a result, there may be redundancy in the features captured by the two branches receiving the same input. To address this issue, we design the GLCFF strategy to obtain nonredundant and compact fusion features. Based on the consistency between shared features and the difference between

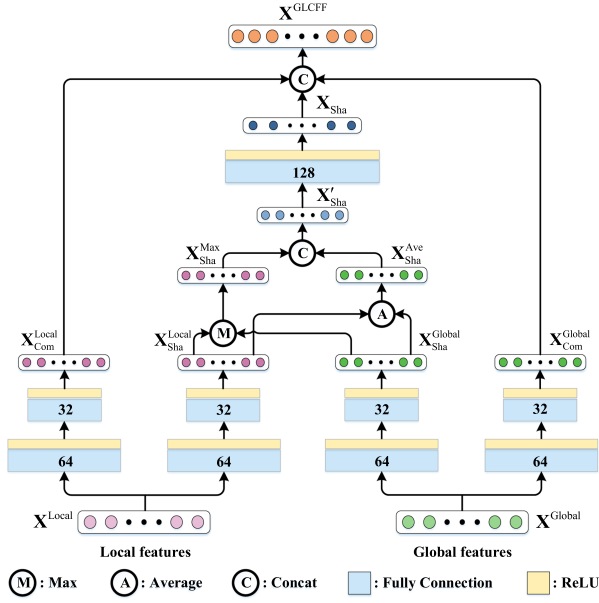


Fig. 5. Illustration of the proposed GLCFF.

complementary features under different attributes, as well as the difference between shared and complementary features under specific attributes, the GLCFF strategy effectively eliminates feature redundancy and fully exploits the complementarity between different attribute features.

As shown in Fig. 5, the proposed GLCFF strategy employs four feature learning branches to learn the shared feature representation $\mathbf{X}_{\text{Sha}}^n$ and the complementary feature representation $\mathbf{X}_{\text{Com}}^n$ between the global and local attributes $n = \{\text{Global}, \text{Local}\}$. Each branch is composed of two alternating FC layers with 64 and 32 neurons and a ReLU activation function to facilitate training. Considering the consistency and difference between features, two loss functions named the consistency constrain loss function \mathcal{L}_{Con} and the difference constrain loss function \mathcal{L}_{Dif} are constructed to constrain the feature learning process of the above branches. Among them, \mathcal{L}_{Con} represents the consistency between the shared features under different attributes. As for \mathcal{L}_{Dif} , it measures the difference between the shared features and complementary features under specific attributes. In addition, the correlation between $\mathbf{X}_{\text{Com}}^{\text{Global}}$ and $\mathbf{X}_{\text{Com}}^{\text{Local}}$ is also constrained in \mathcal{L}_{Dif} , thus further eliminating feature redundancy more thoroughly. Concretely, to ensure that the acquired shared features are adequately similar, the normalized Pearson distance is adopted to express \mathcal{L}_{Con} . Moreover, since the shared features and complementary features as well as the complementary features under different attributes are different and cannot be separated automatically, we use the normalized Pearson correlation coefficients to represent \mathcal{L}_{Dif} , which ensures sufficient discrimination between features and avoids mutual contamination. The Pearson correlation coefficient centralizes the vectors to better measure the correlation between two random variables. On this basis, the feature constraint loss function can be expressed as

$$\mathcal{L}_{\text{FC}} = \mathcal{L}_{\text{Con}} + \mathcal{L}_{\text{Dif}} \quad (9a)$$

$$\mathcal{L}_{\text{Dif}} = \left| \rho(\mathbf{X}_{\text{Sha}}^{\text{Global}}, \mathbf{X}_{\text{Com}}^{\text{Global}}) \right| + \left| \rho(\mathbf{X}_{\text{Sha}}^{\text{Local}}, \mathbf{X}_{\text{Com}}^{\text{Local}}) \right| + \left| \rho(\mathbf{X}_{\text{Com}}^{\text{Global}}, \mathbf{X}_{\text{Com}}^{\text{Local}}) \right| \quad (9b)$$

$$\mathcal{L}_{\text{Con}} = 1 - \left| \rho(\mathbf{X}_{\text{Sha}}^{\text{Global}}, \mathbf{X}_{\text{Sha}}^{\text{Local}}) \right| \quad (9c)$$

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{E[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})]}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}} \quad (9d)$$

where \mathcal{L}_{FC} denotes the feature constraint loss function, $\rho(\cdot)$ represents the calculation of the Pearson correlation coefficient, $|\cdot|$ denotes the absolute value, μ denotes the mean, and σ denotes the standard deviation.

For the $\mathbf{X}_{\text{Sha}}^{\text{Global}}$ and $\mathbf{X}_{\text{Sha}}^{\text{Local}}$, we first calculate the average value $\mathbf{X}_{\text{Sha}}^{\text{Ave}}$ and maximum value $\mathbf{X}_{\text{Sha}}^{\text{Max}}$ of both and concatenate them to obtain the initial shared feature representation \mathbf{X}'_{Sha} . Moreover, a FC layer with 128 neurons as well as a ReLU activation function are used to further eliminate the redundancy of the \mathbf{X}'_{Sha} and generate the final shared feature representation \mathbf{X}_{Sha} . After that, the $\mathbf{X}_{\text{Com}}^{\text{Global}}$ and $\mathbf{X}_{\text{Com}}^{\text{Local}}$ are concatenated with \mathbf{X}_{Sha} to obtain the global-local constraint fusion features. The above process can be described as

$$\mathbf{X}^{\text{GLCFF}} = \text{Concat}(\mathbf{X}_{\text{Sha}}, \mathbf{X}_{\text{Com}}^{\text{Global}}, \mathbf{X}_{\text{Com}}^{\text{Local}}) \quad (10a)$$

$$\mathbf{X}_{\text{Sha}} = f^{\text{ReLU}}(f^{\text{FC}}(\mathbf{X}'_{\text{Sha}})) \quad (10b)$$

$$\mathbf{X}'_{\text{Sha}} = \text{Concat}(\mathbf{X}_{\text{Sha}}^{\text{Ave}}, \mathbf{X}_{\text{Sha}}^{\text{Max}}) \quad (10c)$$

$$\mathbf{X}_{\text{Sha}}^{\text{Ave}} = \text{average}(\mathbf{X}_{\text{Sha}}^{\text{Global}}, \mathbf{X}_{\text{Sha}}^{\text{Local}}) \quad (10d)$$

$$\mathbf{X}_{\text{Sha}}^{\text{Max}} = \text{max}(\mathbf{X}_{\text{Sha}}^{\text{Global}}, \mathbf{X}_{\text{Sha}}^{\text{Local}}) \quad (10e)$$

where $\mathbf{X}^{\text{GLCFF}}$ represents the global-local constraint fusion features, $\text{Concat}(\cdot)$ denotes the connection operation, $f^{\text{FC}}(\cdot)$ denotes the FC layer, $f^{\text{ReLU}}(\cdot)$ denotes the ReLU activation function, $\text{average}(\cdot)$ denotes the average value, and $\text{max}(\cdot)$ denotes the maximum value.

To obtain the classification results of changed and unchanged samples, the $\mathbf{X}^{\text{GLCFF}}$ is sequentially fed into two FC layers containing 64 and 2 neurons as well as the ReLU activation function.

E. Feature Constraint-Focal (FC-F) Loss Function

GLCFF strategy constructs the \mathcal{L}_{FC} to eliminate feature redundancy and fully exploit the complementarity between global and local features, thereby obtaining nonredundant and compact fused features. However, as mentioned in Section III-C, the sample imbalance issue prevents the training process from adequately learning critical change features. Some studies [30], [70] generate a smaller number of changed virtual samples to make the samples balanced. Admittedly, such approaches may not guarantee the quality of the virtual samples and incur high computational costs. To address this issue, we introduce the focal loss function [71] to form a novel FC-F loss function together with the \mathcal{L}_{FC} for supervising the training process of the DA-GLN. From the perspective of sample difficulty classification, the focal loss function evolved from the cross-entropy loss function and focuses on improving the classification accuracy of a small number of changed samples. The binary classification

focal loss function is applicable to the binary CD tasks studied in this article, which is defined as

$$\mathcal{L}_F = -(1 - p_t)^\gamma \log(p_t) \quad (11)$$

where \mathcal{L}_F represents the focal loss function, p_t denotes the output value of the corresponding change class, $-\log(p_t)$ is the initial cross-entropy function value, and $\gamma \in [0, +\infty)$ is the focusing coefficient.

On this basis, the proposed FC-F loss function includes \mathcal{L}_F and \mathcal{L}_{FC} , which can be expressed as

$$\mathcal{L}_{FC-F} = \mathcal{L}_F + \lambda \times \mathcal{L}_{FC} \quad (12)$$

where \mathcal{L}_{FC-F} represents the designed FC-F loss function, and λ is the balance parameter.

IV. EXPERIMENTS

In this section, the measured PolSAR data, experimental settings, evaluation indicators, ablation study, discussion, and experimental comparison are reported in Sections IV-A–IV-E.

A. Experimental Datasets

Five multiscenario PolSAR datasets are tested to evaluate the performance of the proposed method. Each dataset is multilook processed and corresponds to a specific scene of the ground, including the airport, Jinsha River, and the city of Los Angeles. Multiscenarios help to verify the wide suitability of the proposed method. The details of the above datasets are described as follows.

1) *Airport Dataset*: The first dataset used in our experiment corresponds to an airport scene in China. It was taken from the multipolarization SAR imaging satellite Gaofen3. The Pauli-RGB images of this dataset are shown in Fig. 6. Several earth surfaces are contained in this dataset, such as airport runways, vegetation, and buildings. Since the size of the original image is too large and most areas remain unchanged, it is challenging to show the CD results in the actual experiment [31]. As a result, two densely changed subareas are cropped from the original images as testing datasets, i.e., Datasets R1–R2, as shown in the areas circled by the red block in Fig. 6.

2) *Jinsha River Dataset*: The second dataset was taken in Jinsha River, China. Like the first dataset, it was also captured from the Gaofen3 satellite. The Pauli-RGB images of this dataset are shown in Fig. 6. Rivers, mountains, and geological landslides constitute the earth surfaces of this dataset. The main changes are caused by landslides around the river. Similar to the airport dataset, we select two subareas with dense changes (i.e., Datasets R3–R4) from this dataset to conduct our experiments.

3) *Los Angeles Dataset*: The third dataset (i.e., Dataset R5) was taken in the city of Los Angeles. It was acquired by the UAVSAR satellite. The Pauli-RGB images of this dataset are shown in Fig. 7. Changes caused by the transformation of the city are recorded in this open-source dataset [32].

In summary, the whole dataset consisted of five image pairs, totaling ten images. The five datasets R1–R5 and their ground-truth images are shown in Fig. 7. In the ground-truth images of Datasets R1–R5, the white areas represent changed, and the

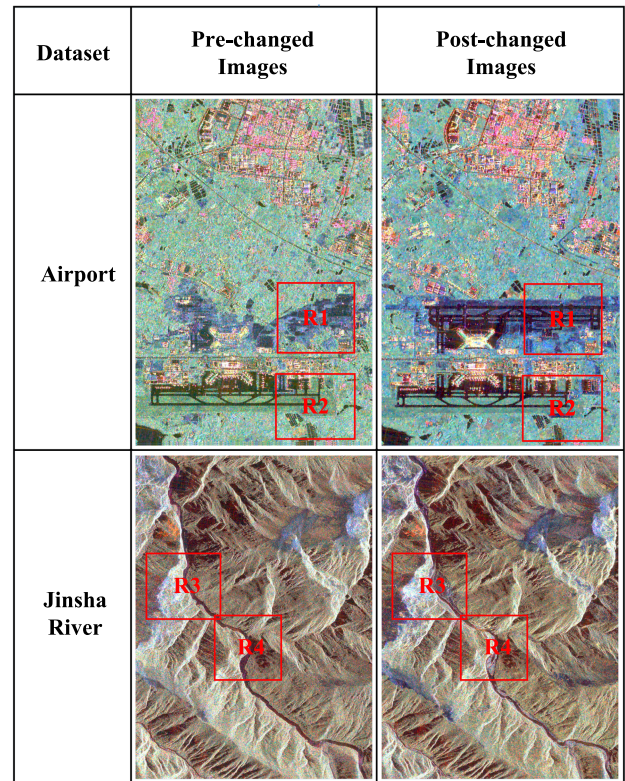


Fig. 6. Pauli images of the experimental datasets in large scenes.

TABLE I
DETAILS OF THE FIVE DATASETS

Dataset	Band	Resolution	Temporal 1	Temporal 2	Size
R1	C	5 m	2017/06	2019/01	360×324
R2	C	5 m	2017/06	2019/01	367×300
R3	C	5 m	2018/09	2019/12	488×410
R4	C	5 m	2018/09	2019/12	412×404
R5	L	1.6 m	2009/04	2015/05	786×300

black areas represent unchanged. Table I records the detailed information of each dataset, including band, resolution, acquisition time, and size.

B. Experimental Settings and Evaluation Indicators

All experiments are conducted on a PC with a 2.10 GHz Intel(R) Core(TM) i7-12700 CPU and 24 GB of RAM. The generation of EDI and the acquisition of pseudolabeled samples are implemented on the CPU and MATLAB R2020b platform. The training and prediction process of the DA-GLN is implemented on NVIDIA GeForce RTX 2060 s GPU with 8 GB memory and PyTorch 1.12.0 software.

When training the DA-GLN, an Adam optimizer with a constant learning rate of 0.001 is employed, the batch size is set to 150, and the number of epochs is empirically set as 20 for all datasets. In the global feature extraction branch of the DA-GLN, we set the amounts of pieces N to 16, the dimension of the generated patch embedding d is set to 384, the number of

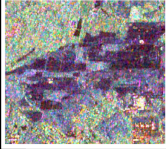



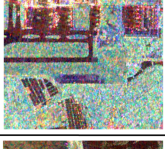

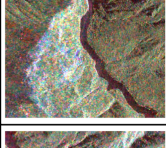
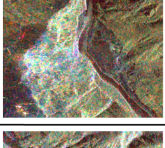
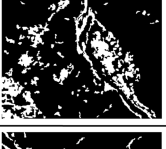
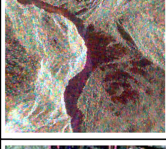
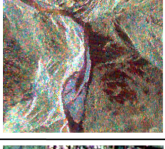




Dataset	Pre-changed Images	Post-changed Images	Ground-truth
R1			
R2			
R3			
R4			
R5			

Fig. 7. Experimental datasets.

layers in the transformer encoder L is set to 6, and the number of multiheads u is set to 8. Besides, the focusing parameter γ is set to 2 in the focal loss function. We select 80% of the entire samples as the training set and the remaining samples for network validation. The entire samples are constructed with 7 pixels as the sampling interval. The patch size m is set to 14 for the airport dataset and 10 for the Jinsha River and Los Angeles datasets. For the balance parameter λ in the proposed FC-F loss function, we set $\lambda = 0.7$ for Dataset R1, $\lambda = 0.8$ for Datasets R2 and R3, $\lambda = 0.1$ for Dataset R4, and λ is set to 0.4 for Dataset R5. We will discuss and choose the suitable patch sizes m and λ values for Datasets R1–R5 in Sections IV-D2 and IV-D3.

To evaluate the proposed CD method quantitatively, six evaluation indicators are utilized to evaluate the CD performance: Precision (Pre), recall (Rec), overall errors (OE), F1 score (F1), percentage correct classification (PCC), and Kappa coefficient ($\kappa \times 100$).

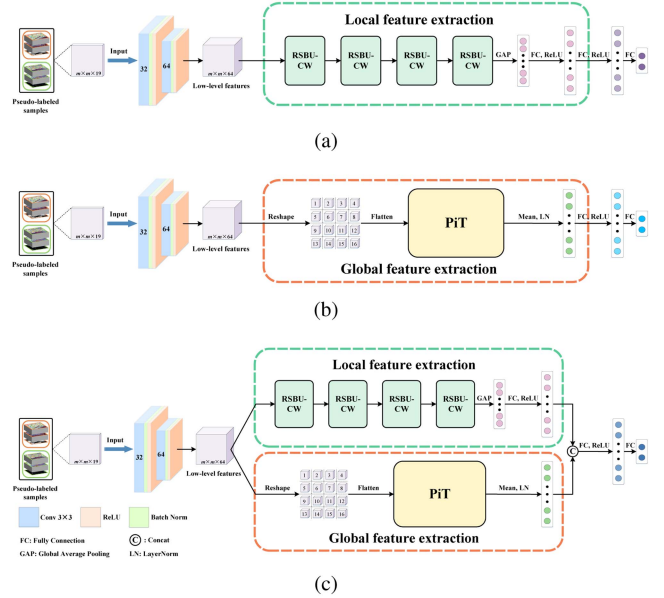


Fig. 8. Comparison models for ablation analysis. (a) Module 1. (b) Module 2. (c) Module 3.

C. Ablation Study

The proposed DA-GLN includes the following four modules:

- 1) the CNN-based branch is adopted to extract local features;
- 2) the PiT is introduced for global feature extraction;
- 3) the GLCFF strategy is designed to remove redundancy and efficiently integrate global-local features;
- 4) the focal loss is adopted to alleviate the adverse effects of the sample imbalance.

To analyze the effectiveness of the above modules more clearly, we conduct the following ablation experiments, including Local, Global, GLCFF, and Focal loss. Concretely, five experiments are designed by adding modules incrementally and trained according to the parameters in Section IV-B.

In the ablation experiments, Model 1 only employs the local feature extraction branch for CD. The PiT is introduced in Model 2 to focus on modeling long-range contextual patterns and extracting global features. Based on the structure of Models 1 and 2, Model 3 integrates dual branches based on CNN and transformer to capture both global and local features simultaneously. Noteworthy, a simple concatenation fusion strategy is adopted to fuse the global and local features in Model 3. Models 1–3 use the cross-entropy loss function as the objective function and their structures are shown in Fig. 8(a)–(c). Unlike Model 3, the GLCFF strategy is designed for feature fusion in Model 4. The structure of Model 4 is the same as shown in Fig. 2. In addition to the cross-entropy loss function, the \mathcal{L}_{FC} in GLCFF is also used to supervise the training process of Model 4. The only difference between Model 5 (i.e., the proposed DA-GLN) and Model 4 is the introduction of focal loss instead of cross-entropy loss as part of the proposed FC-F loss function. Table II presents the experimental results.

The results on five datasets indicate that each module raises the detection accuracy of the models and the complete model has

TABLE II
COMPARISON ABLATION OF EACH MODULE ON FIVE DATASETS

	Model	1	2	3	4	5
Dataset	Local	✓	✗	✓	✓	✓
	Global	✗	✓	✓	✓	✓
	GLCFF	✗	✗	✗	✓	✓
	Focal loss	✗	✗	✗	✗	✓
Dataset R1	F1(%)	85.11	83.65	86.20	87.28	87.74
	PCC(%)	94.69	94.17	94.96	95.17	95.34
	$\kappa \times 100$	81.91	80.15	83.13	84.30	84.86
Dataset R2	F1(%)	85.23	84.59	86.45	87.02	87.39
	PCC(%)	97.42	97.28	97.58	97.65	97.67
	$\kappa \times 100$	83.84	83.11	85.13	85.74	86.12
Dataset R3	F1(%)	81.30	80.91	82.24	83.49	84.48
	PCC(%)	94.18	94.06	94.34	94.53	94.76
	$\kappa \times 100$	77.89	77.43	78.89	80.21	81.33
Dataset R4	F1(%)	87.31	85.77	87.82	88.11	88.26
	PCC(%)	98.41	98.27	98.47	98.48	98.51
	$\kappa \times 100$	86.46	84.85	87.01	87.30	87.46
Dataset R5	F1(%)	77.90	77.64	78.19	78.59	78.70
	PCC(%)	93.31	93.34	93.35	93.37	93.49
	$\kappa \times 100$	73.96	73.74	74.27	74.66	74.86

The bolded values have the highest precision compared to other values.

the highest accuracy. Specifically, by comparing Models 1 and 2, the branch that focuses on extracting local features performs better than the branch that extracts global features. This is because the local branch consisting of stacked RSBU-CWs forces change-independent features to zero and takes advantage of the induction bias. As can be seen from Models 1–3, the combination of the two branches outperforms the performance of a single branch, which proves that the fusion of global-local features helps to capture more complete features related to the changes in interest. Noteworthy, employing simple concatenation to fuse the two branches may cause feature redundancy and restrict the maximization of global-local feature dominance. Compared with the concatenation strategy used in Model 3, Model 4 employs the more efficient GLCFF strategy to remove feature redundancy and mine the complementarity between features, thus performing better than the concatenation fusion strategy as well as a single branch on all datasets. The improvement of Model 5 relative to Model 4 illustrates that the introduction of focal loss can attenuate the effect of unbalanced samples and further improve the ability of the network.

D. Discussion

1) *Discussion of the Computational Cost*: In this subsection, we count the computational cost of the proposed method and the other relevant methods in recent years. Concretely, the time cost of comparison methods on Datasets R1–R5 is listed in the last column of Tables VIII–XII. Moreover, Fig. 9 presents the number of training parameters (Params.) and floating-point operations per second (FLOPs) of the end-to-end deep learning methods. As can be seen from Fig. 9, the lightweight DDNet consisting of a small number of convolutional layers, linear layers, and discrete cosine transforms has the lowest parameters and FLOPs (i.e., 12.1 K and 0.02 G). This allows it to generate

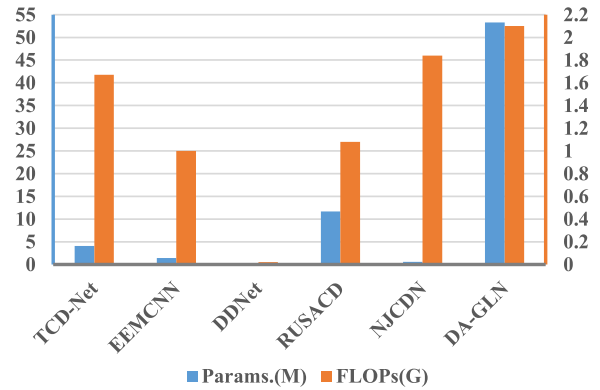


Fig. 9. Parameters and FLOPs of different methods.

CD results in a relatively short time among deep learning methods. The proposed DA-GLN has higher parameters and FLOPs (i.e., 53.3 M and 2.10 G) than other CNN-based methods owing to the introduction of transformer-based structures. However, our network can converge quickly in fewer training epochs and thus has the lowest run time compared with the aforementioned approaches, which makes the effect of the added parameters negligible. In conclusion, our method can accurately detect changes in a few minutes and is practicable.

2) *Discussion of the Patch Size m* : The proposed DA-GLN can simultaneously capture global-local features efficiently by employing CNN- and transformer-based branches, thus improving the accuracy of CD. Nevertheless, choosing the suitable patch size to generate satisfactory CD results is challenging. Generally, small patch sizes allow for better retention of change details. In addition, the training time of the network will increase with the increase of the patch size [72]. As a result, small sizes can complete network training at a low time cost. However, too small a size may constrain the network from capturing more comprehensive global information and is even ineffective in establishing remote dependencies. From this perspective, it seems that larger patch sizes should be selected as much as possible to explore the global spatial correlation of change-aware features. Unfortunately, too large size inevitably contains more interfering and redundant information, leading to performance degradation. Meanwhile, the higher cost of training time may also limit the practical applicability of the network. Therefore, we conduct experiments to test and analyze the CD performance of each dataset at different patch sizes, thus ensuring that the patches are semantically rich and contain low-redundancy information. Specifically, the patch sizes vary in a range between 8 and 18. Fig. 10 presents the CD accuracy for different patch sizes on the five datasets.

As seen from Fig. 14, the CD performance gets progressively better as the size becomes larger. This indicates that the suitable patch size already contains enough spatial context for optimal CD performance. Concretely, the optimal accuracy for Datasets R1 and R2 is achieved when patch size $m = 14$. Simultaneously, from Tables VIII and IX we can see the time cost (i.e., 150.37 and 159.22 s) is appropriate. For the other Datasets R3–R5, the optimal CD indicators can be obtained when $m = 10$. As shown in Tables X–XII, the time cost for these three larger datasets has

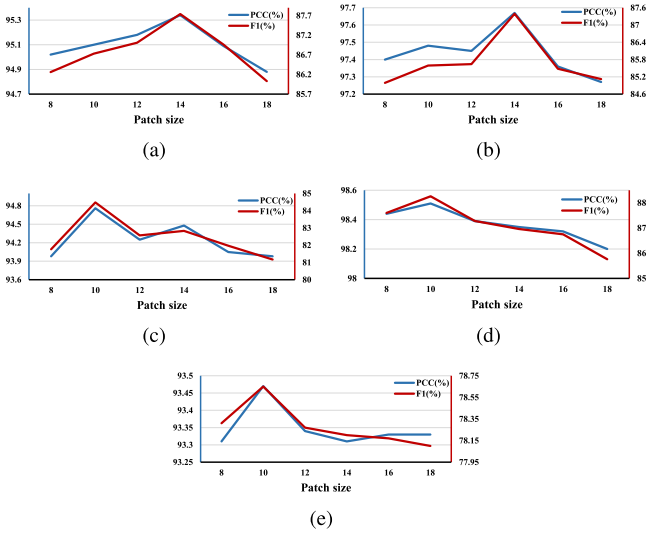


Fig. 10. Experimental results of different patch sizes m on five datasets. (a) R1. (b) R2. (c) R3. (d) R4. (e) R5.

TABLE III

EXPERIMENTAL RESULTS OF DIFFERENT BALANCE PARAMETERS λ ON FIVE DATASETS

Dataset	λ	0.1	0.3	0.4	0.7	0.8	0.9	1.0
R1	F1(%)	87.05	87.02	87.26	87.74	86.69	86.42	85.87
	PCC(%)	95.14	95.12	95.20	95.34	95.07	95.01	94.81
	$\kappa \times 100$	84.07	84.02	84.30	84.86	83.67	83.37	82.70
R2	F1(%)	82.85	84.10	86.79	87.22	87.39	86.67	85.52
	PCC(%)	97.06	97.26	97.50	97.63	97.67	97.58	97.36
	$\kappa \times 100$	81.28	82.62	85.41	85.92	86.12	85.35	84.08
R3	F1(%)	82.46	83.21	83.34	83.56	84.48	82.76	80.52
	PCC(%)	94.43	94.39	94.24	94.52	94.76	94.53	94.07
	$\kappa \times 100$	79.17	79.84	79.86	80.28	81.33	79.53	77.07
R4	F1(%)	88.26	86.78	87.79	87.79	87.79	88.20	86.78
	PCC(%)	98.51	98.46	98.47	98.45	98.44	98.47	98.32
	$\kappa \times 100$	87.46	87.15	86.97	87.06	86.96	87.38	85.88
R5	F1(%)	78.56	78.57	78.70	78.60	78.56	78.63	78.51
	PCC(%)	93.41	93.39	93.49	93.40	93.42	93.38	93.45
	$\kappa \times 100$	74.67	74.67	74.86	74.70	74.67	74.71	74.64

The bolded values have the highest precision compared to other values.

increased compared to Datasets R1 and R2, but is still within the acceptable range (i.e., 318.32, 297.27, and 426.80 s). Based on this, we conduct our experiments with patch size $m = 14$ for Datasets R1 and R2 and $m = 10$ for Datasets R3–R5.

3) *Discussion of the Balance Parameter λ* : The balance parameter λ of the FC-F loss function \mathcal{L}_{FC-F} impacts the CD performance. Hence, we design experiments to discuss its effect on the CD results and set the suitable values for five datasets. Specifically, λ varies in the range from 0 to 1. Table III summarizes the numerical results on five datasets. Apparently, different λ produces the best performance on different datasets. The proposed method achieves optimal performance when λ is set to 0.7 on Dataset R1. For Datasets R2 and R3, the three comprehensive evaluation indicators are highest when $\lambda = 0.8$. When λ is set to 0.1 and 0.4, the F1, PCC, and $\kappa \times 100$ are

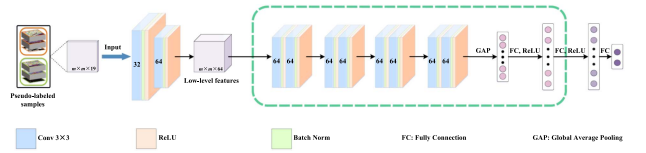


Fig. 11. Structure of the PCNN.

TABLE IV

COMPARISONS BETWEEN PCNN AND RSBU-CW ON FIVE DATASETS

Dataset	Attention	F1(%)	PCC(%)	$\kappa \times 100$
Dataset R1	PCNN	84.84	94.58	81.57
	RSBU-CW	85.11	94.69	81.91
Dataset R2	PCNN	83.77	97.09	82.19
	RSBU-CW	85.23	97.42	83.84
Dataset R3	PCNN	80.90	93.90	77.30
	RSBU-CW	81.30	94.18	77.89
Dataset R4	PCNN	86.21	98.26	85.28
	RSBU-CW	87.31	98.41	86.46
Dataset R5	PCNN	77.69	93.23	73.71
	RSBU-CW	77.90	93.31	73.96

The bolded values have the highest precision compared to other values.

the highest on Datasets R4 and R5, respectively. On this basis, we set $\lambda = 0.7$ for Dataset R1, $\lambda = 0.8$ for Datasets R2 and R3, $\lambda = 0.1$ for Dataset R4, and $\lambda = 0.4$ for Dataset R5 in our experiments.

4) *Application of the RSBU-CW*: In our DA-GLN, the designed local feature extraction branch employs the DRSN to adaptively filter irrelevant features affected by speckle noise and extract more discriminative local features. As shown in Fig. 2, the local feature extraction branch consists of four stacked RSBU-CWs, which are the core components of the DRSN. To verify the effectiveness of combining the channel attention mechanism with soft thresholding to eliminate irrelevant features affected by speckle noise, we design experiments to compare RSBU-CW with conventional convolutional layers. Concretely, channel attention, soft thresholding, and identity shortcut contained in RSBU-CWs are removed. As a result, each RSBU-CW of the local feature extraction branch is replaced by two alternating 3×3 convolutional layers with 64 filters, BN layers, and ReLU activation layers. Fig. 11 and Table IV present the structure of the pure CNN (PCNN) and the accuracy indicators of the comparison experiments. As shown in Table IV, the adopted RSBU-CW produces the highest accuracy on all datasets than the PCNN and is more suitable for extracting local change-aware features in PolSAR images. The RSBU-CW uses the channel weight generated by SENet to adaptively obtain the shrinkage thresholds. On this basis, the soft threshold is employed to force the unimportant feature to be zero. In this way, the noise in the PolSAR images can be effectively suppressed and the features related to the changes can be adequately captured.

5) *Application of the GLCFF*: DA-GLN adopts the GLCFF strategy to integrate global and local features with different attributes in high quality, which can effectively remove redundancy and make the fusion features more compact. Several verification experiments are conducted to discuss the validity

TABLE V
CD PRECISION INDICATORS OF DIFFERENT FUSION STRATEGIES ON FIVE DATASETS

Dataset	Network	F1(%)	PCC(%)	$\kappa \times 100$
Dataset R1	max	86.10	94.79	82.90
	concat	86.20	94.96	83.13
	add	86.76	95.07	83.74
	GLCFF	87.28	95.17	84.30
Dataset R2	max	85.84	97.44	84.43
	concat	86.45	97.58	85.13
	add	86.06	97.43	84.66
	GLCFF	87.02	97.65	85.74
Dataset R3	max	82.31	94.30	78.92
	concat	82.24	94.34	78.89
	add	82.41	94.28	79.01
	GLCFF	83.49	94.53	80.21
Dataset R4	max	87.75	98.41	86.90
	concat	87.82	98.47	87.01
	add	87.77	98.43	86.93
	GLCFF	88.11	98.48	87.30
Dataset R5	max	78.08	93.36	74.17
	concat	78.19	93.35	74.27
	add	78.16	93.36	74.24
	GLCFF	78.59	93.37	74.66

The bolded values have the highest precision compared to other values.

and superiority of GLCFF. Specifically, we compare the CD results obtained by the proposed GLCFF strategy with some common fusion algorithms, such as maximum, concatenation, and addition in the same network structure. To avoid the influence of focal loss function, the networks adopting the common fusion algorithms are supervised by cross-entropy loss function, while the network employing the GLCFF strategy uses both cross-entropy loss and \mathcal{L}_{FC} to supervise its training process. As listed in Table V, when applying traditional fusion algorithms for the PolSAR image CD tasks, concatenation achieves the highest $\kappa \times 100$ on Datasets R2, R4, and R5, whereas data addition achieves better $\kappa \times 100$ for Datasets R1 and R3. However, simple data concatenation or addition may cause feature redundancy and weaken feature discrimination, thus limiting the further improvement of CD accuracy. Compared with the three fusion algorithms mentioned above, the highest CD accuracy is obtained on all datasets when using the GLCFF strategy. The above experiments demonstrate that GLCFF can fully exploit the global and local information in bitemporal PolSAR images and achieve better CD performance than traditional fusion algorithms.

6) *Application of the EDI*: The unsupervised CD method utilizes distance measurement to generate DIs and further thresholds them to obtain pseudolabeled samples required for unsupervised CD. Therefore, it is crucial to choose a well-performing DI generation method. In this article, we adopt the information theoretic-based SE difference operator to measure the difference between bitemporal PolSAR images and enhance them to obtain the EDI. Hence, we conduct experiments to analyze the performance of EDI.

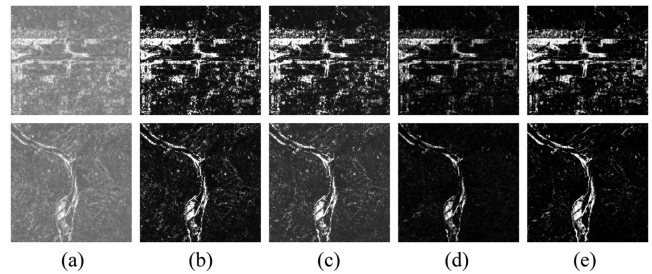


Fig. 12. DIs of the five methods on Datasets R1 and R4. (a) LRT. (b) WD. (c) HLT. (d) SE. (e) EDI.

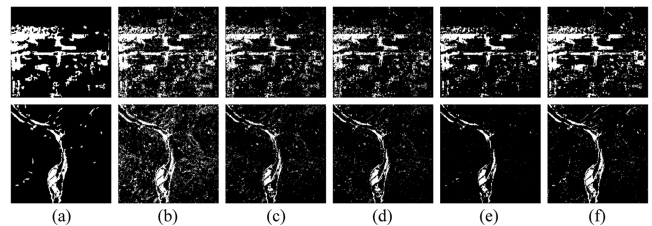


Fig. 13. CD results of the five DI methods using Otsu thresholding on Datasets R1 and R4. (a) Ground-truth image. (b) LRT. (c) WD. (d) HLT. (e) SE. (f) EDI.

Five DI generation methods are compared in the experiment: HLT [19], WD [73], LRT [74], SE [20], and EDI [64]. The visual results for Datasets R1 and R4 are shown in Fig. 12. Compared with other DIs, EDI suppresses outliers caused by speckle noise and enhances edge information. This is due to the fact that the entropy value measures the central tendency of the data and the edge information of the PolSAR images is also introduced to enhance the DI.

For the sake of fairness and to eliminate the influence of the network, we utilize the classical OSTU method to segment the above DIs and obtain CD results, as shown in Fig. 13 and Table VI. The EDI outperforms other methods in CD and has the highest accuracy. The other methods perform unsatisfactorily on the five datasets. Concretely, HLT, WD, and LRT measure the similarity between isolated pixels, which makes it possible to describe local detail information well but sensitive to noise, especially on datasets with high-level noise. SE uses entropy to measure data concentration and capture important spatial information, thus effectively suppressing speckle noise. However, it is ineffective in accurately describing the details of changes, and some nonnegligible areas are also missed. As an improvement of SE, EDI utilizes the principle of gamma correction and introduces edge information to suppress noise, enhancing the edge information and separability between the changed areas and the background.

E. Performance Comparison

To demonstrate the superiority of the proposed DA-GLN, we compare it with eight relevant methods in recent years. They are the traditional methods 1)–2) and deep learning-based methods 3)–8). Comparison of our method with traditional methods allows for exploring the advancement and efficacy of deep learning

TABLE VI
CD PRECISION INDICATORS OF FIVE DI METHODS USING OTSU
THRESHOLDING ON FIVE DATASETS

Dataset	Methods	F1(%)	PCC(%)	$\kappa \times 100$
Dataset R1	LRT	76.33	89.95	70.01
	WD	74.37	90.99	68.97
	HLT	75.69	91.44	70.55
	SE	76.26	92.29	71.85
	EDI	82.99	93.59	79.04
Dataset R2	LRT	64.01	90.26	58.91
	WD	70.86	93.49	67.28
	HLT	73.89	94.28	70.73
	SE	72.78	95.76	70.62
	EDI	79.30	96.22	77.23
Dataset R3	LRT	65.64	85.86	57.09
	WD	64.39	88.98	57.94
	HLT	68.57	90.26	62.86
	SE	71.68	91.91	67.14
	EDI	77.89	92.53	73.39
Dataset R4	LRT	53.90	89.91	49.21
	WD	67.67	95.73	65.38
	HLT	70.98	96.31	69.02
	SE	72.68	97.11	71.25
	EDI	80.70	97.42	79.32
Dataset R5	LRT	72.97	90.84	67.50
	WD	74.62	92.61	70.32
	HLT	75.17	92.19	70.54
	SE	75.29	93.07	71.32
	EDI	76.70	93.21	72.75

The bolded values have the highest precision compared to other values.

theory, while comparison with deep learning-based methods can verify the effectiveness of extracting local-global features by combining CNN and transformer. Among them, 5)–7) are proposed to detect changes in SAR images. In this article, polarization data is uniformly used as the input of these networks to detect changes from the EDI.

The added part with a detailed description of the above comparison experiments is as follows.

- 1) HLT [19] used the HLT statistic to measure the difference between covariance matrices and estimated the sampling of HLT from the FisherSnedecor distribution. On this basis, the constant false alarm rate is used to generate the CD results.
- 2) SE [20] obtained the DI using the information entropy and further detected changes using the ETS based on the chi-square distribution.
- 3) TCD-Net [32] used adaptive multiscale convolutional blocks and residual blocks to capture the change-aware features of objects with different sizes.
- 4) EEMCNN [33] uniformly utilized multidimensional dilated convolution layers to extract deep features from bitemporal images and change deep features.
- 5) DDNet [28] took the characteristics of the discrete cosine transform domain into account, and integrated the

TABLE VII
EXPERIMENTAL PARAMETER SETTINGS OF THE DEEP LEARNING-BASED
COMPARISON METHODS

Methods	Patch	Batch	Epoch	Lr
TCD-Net	11	150	250	10^{-3}
EEMCNN	11	100	750	10^{-4}
DDNet	7	128	50	10^{-3}
RUSACD	28	50	50	10^{-4}
NJCDN	13	128	250	10^{-4}

reshaped DCT coefficients as the frequency domain branches into the proposed model.

- 6) PCANet [26] used PCA filters as convolutional filters and exploited the representative neighborhood features of each pixel.
- 7) RUSACD [30] designed a two-stage center-constrained FCM algorithm to obtain the pseudolabeled set from DIs. Moreover, a deep convolution generative adversarial network and a convolutional wavelet neural network are adopted to deal with the sample imbalance issue and detect changes.
- 8) NJCDN [34] used metric learning and incorporates low-, mid-, and high-level features to capture high-resolution change features in both the covariance matrix and amplitude PolSAR data.

For the two traditional methods, both are analyzed directly on the generated DIs to obtain CD results, making them simple in structure and have low computational complexity. However, due to the interference of speckle noise, the rationality of the similarity measurement and the validity of the DI analysis significantly affect the final CD results. Unlike our DA-GLN which combines transformer with CNN to capture both global and local features, the deep learning-based methods all attempt to improve CD performance by optimizing conventional CNN for capturing multidimensional or multilevel local features.

The aforementioned deep learning-based methods are implemented using the default parameters provided in their original papers. Table VII records the experimental parameter values for these comparison methods, including patch size (Patch), batch size (Batch), training epochs (Epoch), and learning rate (LR). Noteworthy, all of the above methods employ an Adam optimizer for training. As for PCANet with more parameters, we set the patch size as 5. In Gabor feature extraction, the orientation and scale of the Gabor kernel are set to 8 and 5, respectively. The maximum frequency $k_{\max} = 2\pi$ is utilized. For fairness, we have performed the same preprocessing as described in Section III-A on all Gaofen3 PolSAR images. The CD performance comparison of the aforementioned methods on Datasets R1–R5 is as follows.

1) *Performance on Dataset R1:* In Dataset R1, the major change regions include newly built airport runways and a small number of irregular buildings. The detection accuracy evaluation of different methods on Dataset R1 is listed in Table VIII. The proposed DA-GLN achieves the highest detection accuracy. The comprehensive indicators F1, PCC, and $\kappa \times 100$ increased by

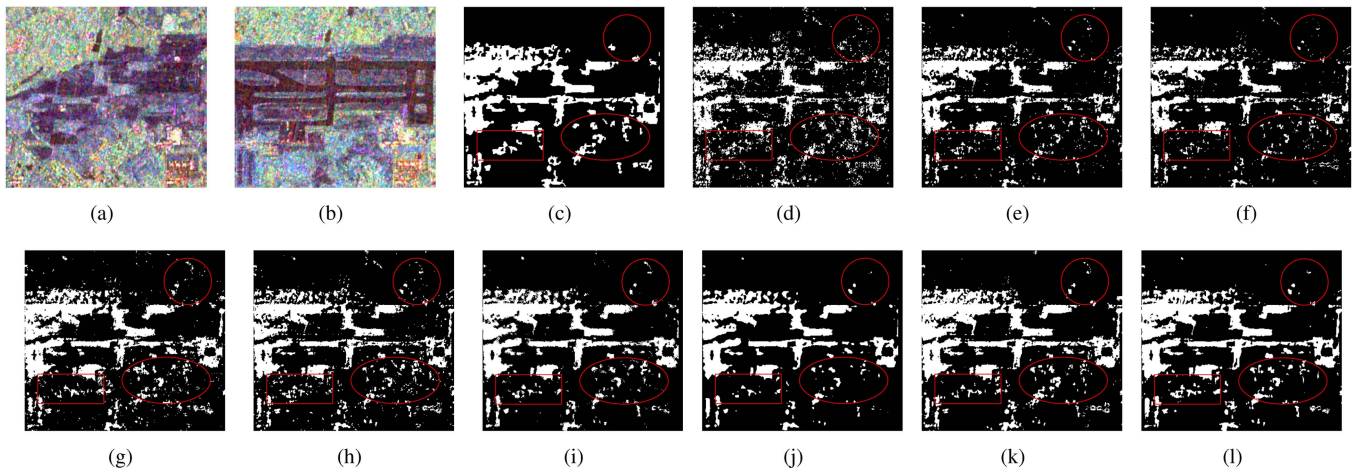


Fig. 14. CD results of different methods on Dataset R1. (a) Prechanged image. (b) Postchanged image. (c) Ground-truth image. (d) HLT. (e) SE. (f) TCD-Net. (g) EEMCNN. (h) DDNet. (i) PCANet. (j) RUSACD. (k) NJCDN. (l) DA-GLN.

TABLE VIII
QUANTITATIVE COMPARISON PERFORMANCE OF DIFFERENT METHODS ON DATASET R1

Methods	Pre(%)	Rec(%)	OE	F1(%)	PCC(%)	$\kappa \times 100$	Time(s)
HLT [19]	82.05	73.67	9548	77.63	91.81	72.64	21.85
SE [20]	92.83	67.19	8548	77.95	92.67	73.69	1.92
TCD-Net [32]	94.22	61.63	9481	74.52	91.87	69.93	1036.84
EEMCNN [33]	82.55	80.90	8149	81.70	93.01	77.39	1349.07
DDNet [28]	88.24	75.87	7702	81.59	93.40	77.59	652.68
PCANet [26]	90.77	78.38	6658	84.12	94.29	80.66	1388.27
RUSACD [30]	95.18	77.01	6050	85.13	94.81	82.03	340.73
NJCDN [34]	87.83	83.69	6278	85.71	94.62	82.39	1264.41
DA-GLN	88.98	86.55	5440	87.74	95.34	84.86	150.37

The bolded values have the highest precision compared to other values.

TABLE IX
QUANTITATIVE COMPARISON PERFORMANCE OF DIFFERENT METHODS ON DATASET R2

Methods	Pre(%)	Rec(%)	OE	F1(%)	PCC(%)	$\kappa \times 100$	Time(s)
HLT [19]	62.63	88.19	6848	73.24	93.78	69.84	20.82
SE [20]	93.54	60.59	4633	73.54	95.79	71.37	1.79
TCD-Net [32]	94.07	58.53	4799	72.16	95.64	69.94	469.76
EEMCNN [33]	86.38	63.68	4962	73.14	95.49	70.75	982.00
DDNet [28]	88.88	70.68	4060	78.72	96.31	76.73	296.07
PCANet [26]	92.62	76.49	3147	83.78	97.14	82.23	1019.38
RUSACD [30]	96.22	74.72	2999	84.11	97.28	82.65	334.43
NJCDN [34]	89.49	83.46	2800	86.37	97.46	84.97	602.49
DA-GLN	91.57	83.84	2563	87.39	97.67	86.12	159.22

The bolded values have the highest precision compared to other values.

more than 2.03%, 0.53%, and 2.47 over other state-of-the-art methods.

The visual results of different methods on Dataset R1 are shown in Fig. 14. Among the traditional methods, HLT adopts the HLT statistic to measure the similarity between isolated pixels, which makes it sensitive to speckle noise. As shown in Fig. 14(d), HLT recognizes most of the unchanged pixels in the background as noise (areas farmed in red). However, it can describe the local detail information well. SE suppresses noise more effectively than HLT (as shown in the red boxes) by measuring the concentration of data and capturing important spatial information. Unfortunately, SE cannot accurately describe the change details and misses some changed areas. As for the deep learning-based methods, the visual results of TCD-Net, EEMCNN, and DDNet cover the major changed areas and retain the change details; however, the background noise widely distributed on them is still nonnegligible, as shown by the regions framed in red. Although PCANet, RUSACD, and NJCDN perform better than the aforementioned methods, their ability to describe the change details still needs to be improved. The visual results of the proposed DA-GLN have better regional homogeneity and can suppress background noise (highlighted by

the red boxes), validating the effectiveness of fusing global and local features. Fig. 14 and Table VIII demonstrate that our dual attention-based method performs better than other traditional and purely CNN-based methods on Dataset R1.

2) *Performance on Dataset R2*: Dataset R2, which belongs to the same overall dataset as Dataset R1, is also utilized to validate the validity of the DA-GLN. The major changes in Dataset R2 are the result of airport expansion, including red runways in the horizontal direction and irregular discontinuous aprons in the vertical direction. The changed areas below the runway are also included. The CD evaluation indicators on Dataset R2 are listed in Table IX. The F1, PCC, and $\kappa \times 100$ increased by more than 1.02%, 0.21%, and 1.15, respectively, which validates the superiority of our DA-GLN.

Fig. 15 shows the CD results of each method on Dataset R2. The effect of speckle noise in Dataset R2 is not as severe as in Dataset R1 owing to the relatively concentrated changed areas. However, HLT still cannot effectively suppress speckle noise (areas farmed in red). SE, TCD-Net, EEMCNN, DDNet, and PCANet attenuate the impact of noise, but miss many areas that cannot be ignored. RUSACD and NJCDN describe the subject changed areas and deal with isolated pixels efficiently; however,

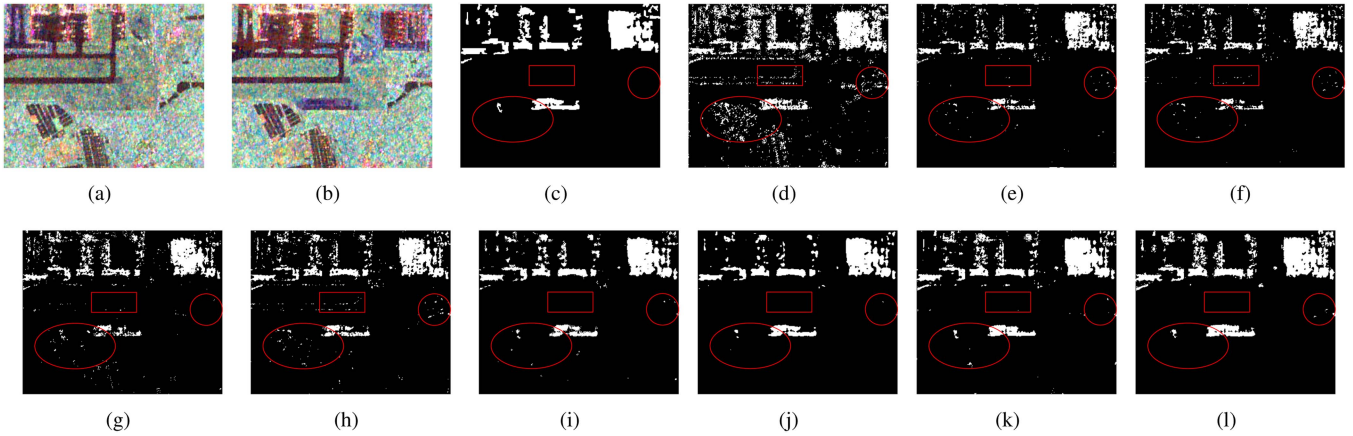


Fig. 15. CD results of different methods on Dataset R2. (a) Prechanged image. (b) Postchanged image. (c) Ground-truth image. (d) HLT. (e) SE. (f) TCD-Net. (g) EEMCNN. (h) DDNet. (i) PCANet. (j) RUSACD. (k) NJCDN. (l) DA-GLN.

TABLE X
QUANTITATIVE COMPARISON PERFORMANCE OF DIFFERENT METHODS ON DATASET R3

Methods	Pre(%)	Rec(%)	OE	F1(%)	PCC(%)	$\kappa \times 100$	Time(s)
HLT [19]	74.34	65.22	19361	69.48	90.32	63.76	37.45
SE [20]	88.37	58.40	16652	70.33	91.68	65.72	3.22
TCD-Net [32]	86.36	64.69	15385	73.97	92.31	69.57	1778.57
EEMCNN [33]	78.28	75.77	15323	76.96	92.34	72.37	1768.11
DDNet [28]	75.10	81.40	15409	78.12	92.30	73.45	1187.00
PCANet [26]	84.44	76.95	12667	80.39	93.67	76.63	2553.47
RUSACD [30]	90.06	74.35	11442	81.45	94.28	78.11	384.36
NJCDN [34]	79.31	82.56	13211	80.86	93.40	76.87	1728.30
DA-GLN	84.61	84.52	10493	84.48	94.76	81.33	318.32

The bolded values have the highest precision compared to other values.

TABLE XI
QUANTITATIVE COMPARISON PERFORMANCE OF DIFFERENT METHODS ON DATASET R4

Methods	Pre(%)	Rec(%)	OE	F1(%)	PCC(%)	$\kappa \times 100$	Time(s)
HLT [19]	63.49	79.76	7272	70.70	95.63	68.37	30.94
SE [20]	93.53	63.25	4524	75.46	97.28	74.08	2.75
TCD-Net [32]	87.49	75.03	3927	80.78	97.64	79.53	670.40
EEMCNN [33]	82.40	81.36	3968	81.85	97.62	80.58	695.44
DDNet [28]	72.13	88.58	5021	79.51	96.98	77.90	444.16
PCANet [26]	86.10	83.01	3343	84.53	97.99	83.45	1766.58
RUSACD [30]	91.37	84.16	2616	87.62	98.43	86.78	359.90
NJCDN [34]	87.31	87.51	2777	87.39	98.33	86.50	664.99
DA-GLN	91.82	85.03	2487	88.26	98.51	87.46	297.27

The bolded values have the highest precision compared to other values.

RUSACD omits some change details and NJCDN overdetects some changed areas outside the newly built runway. Our proposed method suppresses noise (highlighted by the red boxes) and accurately locates edges by fully capturing spatial context information, resulting in the best visual effect. The results of the evaluation indicators are consistent with the visualization results.

3) *Performance on Dataset R3*: To verify the broad applicability of the proposed method, another landslide scenario dataset is used to validate the broad applicability and superiority of our method. Dataset R3 primarily records the changes caused by landslides. The CD evaluation indicators on Dataset R3 are presented in Table X. The proposed DA-GLN has achieved the highest accuracy. The F1, PCC, and $\kappa \times 100$ increased by more than 3.03%, 0.48%, and 3.22, respectively.

Fig. 16 shows the visual results of each method on Dataset R3. Since Dataset R3 suffers from a high level of noise, we can observe that HLT fails to suppress the speckle noise. Similar situations occur in the visual results of TCD-Net and DDNet (as shown in the red boxes). Nevertheless, the boundary positioning ability of the above methods is commendable. SE can describe the subject changed areas well, but misses some changes related to the edge structure. EEMCNN and PCANet further attenuate

the detrimental effects of background noise highlighted by red boxes, whereas the visual results of both suffer from over-detection. RUSACD and NJCDN perform better than the above methods. However, RUSACD lost some nonnegligible areas and NJCDN overdetects some unchanged areas. Our DA-GLN can effectively suppress noise (areas farmed in red) and has less over-detection. The CD evaluation indicators on Dataset R3 are listed in Table X. The results in Table X also validate the conclusions drawn in Fig. 16.

4) *Performance on Dataset R4*: Similarly, we select Dataset R4 from the Jinsha River Dataset to validate the wide applicability of our DA-GLN. Table XI records the CD evaluation indicators on Dataset R4. The F1, PCC, and $\kappa \times 100$ increased by more than 0.64%, 0.08%, and 0.68, respectively.

As shown in Fig. 17, the visual result of HLT describes the information in detail but with much background noise highlighted by red boxes. SE attempts to further attenuate the noise (as shown in the red boxes) but misses some of the nonnegligible edge detail information. As for the deep learning-based methods, TCD-Net, EEMCNN, and DDNet can accurately localize the main areas of change and maintain edge structure information; however, these methods still generate numerous false alarms in the background (highlighted by the red boxes). PCANet, RUSACD, and NJCDN

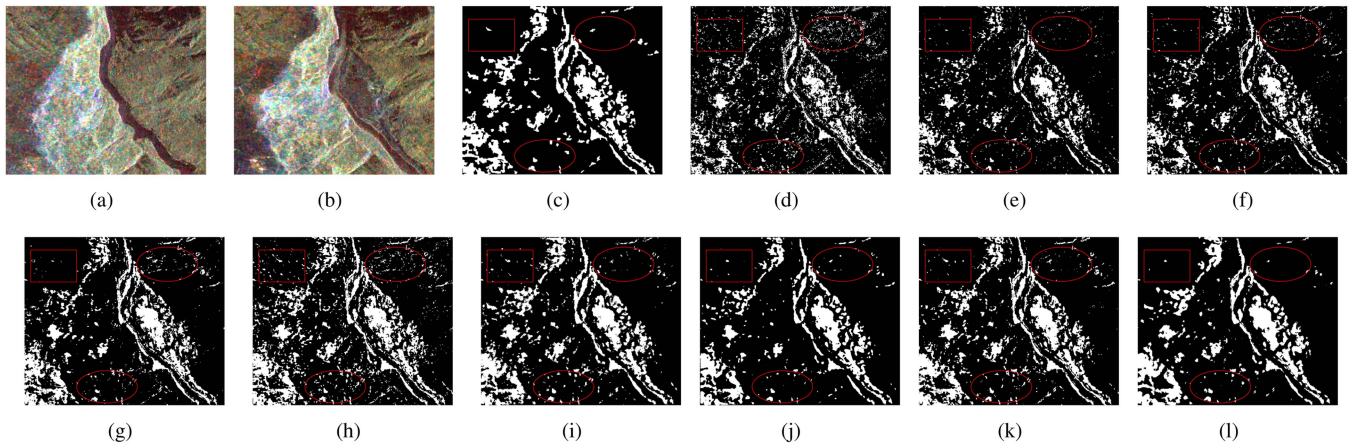


Fig. 16. CD results of different methods on Dataset R3. (a) Prechanged image. (b) Postchanged image. (c) Ground-truth image. (d) HLT. (e) SE. (f) TCD-Net. (g) EEMCNN. (h) DDNet. (i) PCANet. (j) RUSACD. (k) NJCDN. (l) DA-GLN.

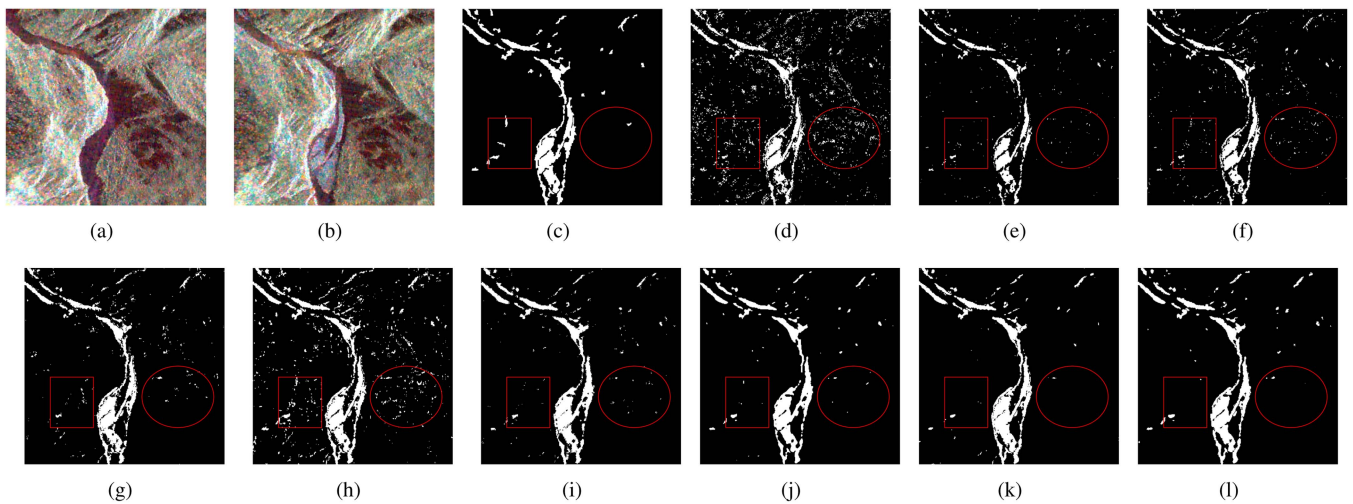


Fig. 17. CD results of different methods on Dataset R4. (a) Prechanged image. (b) Postchanged image. (c) Ground-truth image. (d) HLT. (e) SE. (f) TCD-Net. (g) EEMCNN. (h) DDNet. (i) PCANet. (j) RUSACD. (k) NJCDN. (l) DA-GLN.

perform well and validate the superiority of the CNN-based structures. Nevertheless, none of the aforementioned approaches show the same performance as our DA-GLN combining CNN with transformer. The results of Table XI are consistent with Fig. 17, indicating that our dual attention-based method has the optimal performance and highest accuracy.

5) *Performance on Dataset R5*: In addition to the aforementioned datasets describing airport and landslide scenarios, we further use the dataset describing city scenarios to validate the broad applicability and superiority of our method. Dataset R5 primarily records changes caused by urban transformation. The CD evaluation indicators on Dataset R5 are listed in Table XII. According to the observation, the proposed method has achieved the best performance. The F1, PCC, and $\kappa \times 100$ increased by more than 0.58%, 0.28%, and 0.76, respectively.

Fig. 18 shows the visual results for each method on Dataset R5. HLT fails to suppress the background noise highlighted in red boxes. SE, TCD-Net, and EEMCNN attempt to mitigate

TABLE XII
QUANTITATIVE COMPARISON PERFORMANCE OF DIFFERENT METHODS ON DATASET R5

Methods	Pre(%)	Rec(%)	OE	F1(%)	PCC(%)	$\kappa \times 100$	Time(s)
HLT [19]	65.74	80.94	22443	72.55	90.48	66.87	46.80
SE [20]	75.73	79.27	16906	77.46	92.83	73.20	6.71
TCD-Net [32]	74.48	80.79	17186	77.51	92.71	73.17	2370.89
EEMCNN [33]	78.25	77.98	16012	78.12	93.21	74.10	3661.41
DDNet [28]	78.89	76.79	16038	77.83	93.20	73.81	1067.45
PCANet [26]	82.00	71.57	16178	76.43	93.14	72.44	2876.56
RUSACD [30]	77.58	78.34	16237	77.96	93.11	73.88	418.96
NJCDN [34]	77.81	77.90	16242	77.85	93.11	73.78	3141.01
DA-GLN	80.06	77.39	15354	78.70	93.49	74.86	426.80

The bolded values have the highest precision compared to other values.

the effects of the noise, but there are still some isolated pixels in their backgrounds (areas farmed in red). Although RUSACD

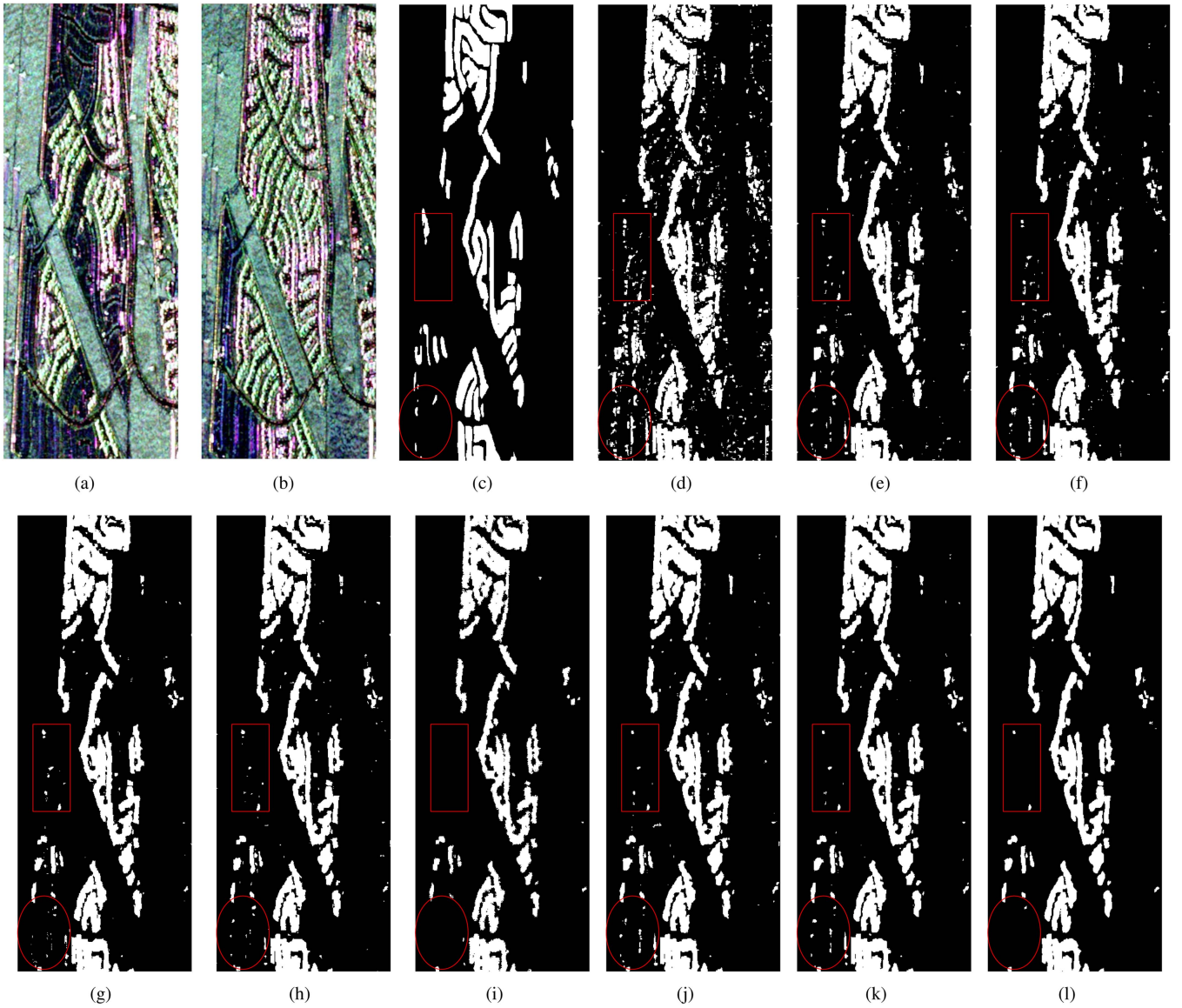


Fig. 18. CD results of different methods on Dataset R5. (a) Prechanged image. (b) Postchanged image. (c) Ground-truth image. (d) HLT. (e) SE. (f) TCD-Net. (g) EEMCNN. (h) DDNet. (i) PCANet. (j) RUSACD. (k) NJCDN. (l) DA-GLN.

and NJCDN perform better than the above methods, their ability to suppress noise as well as describe edge detail information still needs to be improved. PCANet is effective in suppressing noise, whereas its visual results suffer from incomplete detection. DA-GLN is robust to noise (as shown in the red boxes) and has the fewest missed detections, consistent with the results in Table XII.

Taken together, our DA-GLN outperforms other methods in both objective and visual comparisons, which proves its robustness on datasets with different scenarios. Remarkably, the DA-GLN suppresses speckle noise and accurately localizes changed areas, demonstrating the effectiveness of global and local feature fusion using GLCFF. Moreover, the introduction of the focal loss function attenuates the impact of the unbalanced samples and improves the detection accuracy. During the above experiments, we also find some weaknesses in the proposed method. First,

DA-GLN can not effectively detect some nonnegligible changed areas with complex scenarios or similar scattering characteristics. Second, our method has higher computational complexity than other CNN-based methods owing to the integration of PiT. In future research, we will try to utilize model compression methods to accurately detect changes in a shortertime.

V. CONCLUSION

In this article, we develop a novel unsupervised method based on dual attention to tackle the PolSAR image CD tasks, which can suppress speckle noise and improve the discrimination of change-aware features. In the proposed method, the developed DA-GLN extracts local-global features by DRSN and PiT, respectively. The designed GLCFF strategy constructs a feature constraint loss function to remove redundancy and exploits

the complementarity between global and local features, thus obtaining compact and nonredundant fusion features. The focal loss function is introduced as part of the proposed FC-F loss function to solve the sample imbalance problem. Ablation and comparison experiments are tested using five real spaceborne PolSAR datasets. Concretely, the ablation study demonstrates the effectiveness of introducing PiT, GLCFF, and focal loss function. Furthermore, the visual results of comparison experiments indicate that our dual attention-based approach can effectively improve localization accuracy and efficiently suppress noise. Quantitatively, our method can achieve up to 98.51% accuracy on four datasets, which is higher than other related state-of-the-art methods.

In future work, rich polarization information could be further utilized to implement semantic CD (SCD) based on the existing binary CD (BCD) results. Compared with BCD, SCD can provide richer land transition information.

ACKNOWLEDGMENT

The authors would like to thank the Aerospace Information Research Institute, Chinese Academy of Science for providing the Gaofen3 satellite datasets. We also thank all the anonymous reviewers for their valuable comments that helped to improve the quality of this article.

REFERENCES

[1] C. Wu, B. Du, and L. Zhang, "Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9774–9788, Aug. 2023.

[2] M. Hu, C. Wu, B. Du, and L. Zhang, "Binary change guided hyperspectral multiclass change detection," *IEEE Trans. Image Process.*, vol. 32, pp. 791–806, 2023.

[3] R. V. Fonseca, R. G. Negri, A. Pinheiro, and A. Atto, "Wavelet spatio-temporal change detection on multitemporal SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4013–4023, 2023.

[4] Y. Cao and X. Huang, "A full-level fused cross-task transfer learning method for building change detection using noise-robust pretrained networks on crowdsourced labels," *Remote Sens. Environ.*, vol. 284, 2023, Art. no. 113371.

[5] Y. Zhang, R. Miao, Y. Dong, and B. Du, "Multi-order graph convolutional network with channel attention for hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2023.

[6] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "Multiscale diff-changed feature fusion network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023.

[7] X. Long, W. Zhuang, M. Xia, K. Hu, and H. Lin, "SASiamNet: Self-adaptive siamese network for change detection of remote sensing image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1021–1034, 2024.

[8] A. Rosenqvist, M. Shimada, N. Ito, and M. Watanabe, "ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3307–3316, Nov. 2007.

[9] Z. Yang, L. Fang, B. Shen, and T. Liu, "PolSAR ship detection based on azimuth sublook polarimetric covariance matrix," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8506–8518, 2022.

[10] G. Gao, Q. Bai, C. Zhang, L. Zhang, and L. Yao, "Dualistic cascade convolutional neural network dedicated to fully PolSAR image ship detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 202, pp. 663–681, 2023.

[11] Y. Cao, Y. Wu, M. Li, M. Zheng, P. Zhang, and J. Wang, "Multifrequency PolSAR image fusion classification based on semantic interactive information and topological structure," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, 2023.

[12] Y. Jiang et al., "Semisupervised complex network with spatial statistics fusion for PolSAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9749–9761, 2023.

[13] H. Shi et al., "Soil moisture retrieval over agricultural fields from l-band multi-incidence and multitemporal PolSAR observations using polarimetric decomposition techniques," *Remote Sens. Environ.*, vol. 261, 2021, Art. no. 112485.

[14] X.-F. Kuang, J. Guo, H.-Y. Wang, and H. Wang, "Agricultural field boundary delineation using a cascaded deep network model from polarized SAR and multispectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7228–7247, 2023.

[15] S. Mahdavi, B. Salehi, W. Huang, M. Amani, and B. Brisco, "A PolSAR change detection index based on neighborhood information for flood mapping," *Remote Sens.*, vol. 11, no. 16, 2019, Art. no. 1854.

[16] P. Ferrazzoli, S. Paloscia, P. Pampaloni, G. Schiavon, S. Sigismondi, and D. Solimini, "The potential of multifrequency polarimetric SAR in assessing agricultural and arboreous biomass," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 5–17, Jan. 1997.

[17] V. Akbari, A. P. Doulgeris, and T. Eltoft, "Monitoring glacier changes using multitemporal multipolarization SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3729–3741, Jun. 2013.

[18] M. Liu, H. Zhang, C. Wang, and Y. Tang, "Change detection of polarimetric sar images applied to specific land cover type," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 6329–6332.

[19] V. Akbari, S. N. Anfinsen, A. P. Doulgeris, T. Eltoft, G. Moser, and S. B. Serpico, "Polarimetric SAR change detection with the complex hotelling-lawley trace statistic," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3953–3966, Jul. 2016.

[20] A. D. Nascimento, A. C. Frery, and R. J. Cintra, "Detecting changes in fully polarimetric SAR imagery with statistical information theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1380–1392, Mar. 2018.

[21] P. R. Kersten, J.-S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 519–527, Mar. 2005.

[22] J. Geng, X. Ma, X. Zhou, and H. Wang, "Saliency-guided deep neural networks for SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7365–7377, Oct. 2019.

[23] Y. Pei, Y. Huang, Q. Zou, X. Zhang, and S. Wang, "Effects of image degradation and degradation removal to CNN-based image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1239–1253, Apr. 2019.

[24] X. Jia et al., "Semi-supervised multi-view deep discriminant representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2496–2509, Jul. 2020.

[25] Z. Li, F. Wang, L. Cui, and J. Liu, "Dual mixture model based CNN for image denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 3618–3629, 2022.

[26] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, Dec. 2016.

[27] A. B. Campos, M. I. Pettersson, V. T. Vu, and R. Machado, "False alarm reduction in wavelength-resolution SAR change detection schemes by using a convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2020.

[28] X. Qu, F. Gao, J. Dong, Q. Du, and H.-C. Li, "Change detection in synthetic aperture radar images using a dual-domain network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[29] X. Zhang, X. Su, Q. Yuan, and Q. Wang, "Spatial-temporal gray-level co-occurrence aware CNN for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[30] X. Zhang, H. Su, C. Zhang, X. Gu, X. Tan, and P. M. Atkinson, "Robust unsupervised small area change detection from SAR imagery using deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 79–94, 2021.

[31] F. Liu, L. Jiao, X. Tang, S. Yang, W. Ma, and B. Hou, "Local restricted convolutional neural network for change detection in polarimetric SAR images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 818–833, Mar. 2018.

[32] R. Habibollahi, S. T. Seydi, M. Hasanlou, and M. Mahdianpari, "TCD-Net: A novel deep learning framework for fully polarimetric change detection using transfer learning," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 438.

[33] S. T. Seydi, M. Hasanlou, and M. Amani, "A new end-to-end multi-dimensional CNN framework for land cover/land use change detection in multi-source remote sensing datasets," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 2010.

[34] C. Wang, W. Su, and H. Gu, "A joint change detection method on complex-valued polarimetric synthetic aperture radar images based on feature fusion and similarity learning," *Int. J. Remote Sens.*, vol. 42, no. 13, pp. 4864–4881, 2021.

- [35] Y. Gao, F. Gao, J. Dong, and S. Wang, "Change detection from synthetic aperture radar images based on channel weighting-based deep cascade network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4517–4529, Nov. 2019.
- [36] Y. Gao, F. Gao, J. Dong, Q. Du, and H.-C. Li, "Synthetic aperture radar image change detection via siamese adaptive fusion network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10748–10760, 2021.
- [37] C. Zhao, L. Ma, L. Wang, T. Ohtsuki, P. T. Mathiopoulos, and Y. Wang, "SAR image change detection in spatial-frequency domain based on attention mechanism and gated linear unit," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [38] M. Zhang, H. Zheng, M. Gong, Y. Wu, H. Li, and X. Jiang, "Self-structured pyramid network with parallel spatial-channel attention for change detection in VHR remote sensed imagery," *Pattern Recognit.*, vol. 138, 2023, Art. no. 109354.
- [39] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.
- [40] Y. Feng, H. Xu, J. Jiang, H. Liu, and J. Zheng, "ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.
- [41] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998–6008, 2017.
- [42] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. 9th Int. Conf. Learn. Representations*, 2021, pp. 1–22.
- [43] T. Saleh, X. Weng, S. Holail, C. Hao, and G.-S. Xia, "DAM-Net: Global flood detection from sar imagery using differential attention metric-based vision transformers," 2023, *arXiv:2306.00704*.
- [44] Y. Du, R. Zhong, Q. Li, and F. Zhang, "TransUNet++SAR: Change detection with deep learning about architectural ensemble in SAR images," *Remote Sens.*, vol. 15, no. 1, 2022, Art. no. 6.
- [45] L. Ding et al., "Looking outside the window: Wide-context transformer for the semantic segmentation of high-resolution remote sensing images," 2021, *arXiv:2106.15754*.
- [46] Z. Xue, X. Tan, X. Yu, B. Liu, A. Yu, and P. Zhang, "Deep hierarchical vision transformer for hyperspectral and LiDAR data classification," *IEEE Trans. Image Process.*, vol. 31, pp. 3095–3110, 2022.
- [47] P. Lv, W. Wu, Y. Zhong, F. Du, and L. Zhang, "SCViT: A spatial-channel feature preserving vision transformer for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.
- [48] W. Wang, C. Tang, X. Wang, and B. Zheng, "A ViT-based multiscale feature fusion approach for remote sensing image segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [49] Y. Zhou, S. Chen, J. Zhao, R. Yao, Y. Xue, and A. El Saddik, "CLT-Det: Correlation learning based on transformer for detecting dense objects in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.
- [50] Y. Wang et al., "Spectral-spatial-temporal transformers for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [51] Y. Dai, T. Zheng, C. Xue, and L. Zhou, "MViT-PCD: A lightweight ViT-based network for martian surface topographic change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [52] H. Zhou, F. Luo, H. Zhuang, Z. Weng, X. Gong, and Z. Lin, "Attention multi-hop graph and multi-scale convolutional fusion network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023.
- [53] K. Ding, T. Lu, W. Fu, S. Li, and F. Ma, "Global-local transformer network for HSI and LiDAR data joint classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.
- [54] T. Guo, R. Wang, F. Luo, X. Gong, L. Zhang, and X. Gao, "Dual-view spectral and global spatial feature fusion network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023.
- [55] X. Liu, Y. Wu, W. Liang, Y. Cao, and M. Li, "High resolution SAR image classification using global-local network structure based on vision transformer and CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [56] Y. Duan, F. Luo, M. Fu, Y. Niu, and X. Gong, "Classification via structure-preserved hypergraph convolution network for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023.
- [57] Z. Teng, Y. Duan, Y. Liu, B. Zhang, and J. Fan, "Global to local: Clip-LSTM-based object detection from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021.
- [58] L. Bao et al., "Aggregating transformers and CNNs for salient object detection in optical remote sensing images," *Neurocomputing*, vol. 553, 2023, Art. no. 126560.
- [59] P. Song, J. Li, Z. An, H. Fan, and L. Fan, "CTMFNet: CNN and transformer multiscale fusion network of remote sensing urban scene imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2022.
- [60] Q. Li, R. Zhong, X. Du, and Y. Du, "TransUNetCD: A hybrid transformer network for change detection in optical remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022.
- [61] W. Li, L. Xue, X. Wang, and G. Li, "ConvTransNet: A CNN-transformer network for change detection with multi-scale global-local representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, 2023.
- [62] Y. Cui, Y. Yamaguchi, H. Kobayashi, and J. Yang, "Filtering of polarimetric synthetic aperture radar images: A sequential approach," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 3138–3141.
- [63] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A sift-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2014.
- [64] D. Xu, M. Li, Y. Wu, P. Zhang, X. Xin, and Z. Yang, "Difference-guided multiscale graph convolution network for unsupervised change detection in PolSAR images," *Neurocomputing*, vol. 555, 2023, Art. no. 126611.
- [65] B. Heo, S. Yun, D. Han, S. Chun, J. Choe, and S. J. Oh, "Rethinking spatial dimensions of vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 11936–11945.
- [66] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2019.
- [67] K. Isogawa, T. Ida, T. Shiodera, and T. Takeguchi, "Deep shrinkage convolutional neural network for adaptive noise reduction," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 224–228, Feb. 2017.
- [68] N. Lin, G. Chen, Q. Zhou, and C. Liu, "Dilated residual shrinkage network for SAR image despeckling," in *Proc. IEEE 6th Int. Conf. Signal Image Process.*, 2021, pp. 503–507.
- [69] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2018, pp. 7132–7141.
- [70] M. Gong, T. Gao, M. Zhang, W. Li, Z. Wang, and D. Li, "An M-Nary SAR image change detection based on GAN architecture search," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, 2023.
- [71] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [72] Y. Cao, Y. Wu, M. Li, W. Liang, and X. Hu, "DFAF-Net: A dual-frequency PolSAR image classification network based on frequency-aware attention and adaptive feature fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022.
- [73] S. N. Anfinsen, R. Jenssen, and T. Eltoft, "Spectral clustering of polarimetric SAR data with Wishart-derived distance measures," in *Proc. POLinSAR*, vol. 7, 2007, pp. 1–9.
- [74] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 4–19, Jan. 2003.



Dazhi Xu (Graduate Student Member, IEEE) received the B.S. degree in electrical and information engineering from the Lanzhou University of Technology, Lanzhou, China, in 2019. He is currently working toward the Ph.D. degree in signal and information processing with the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an, China.

His main research interests include polarimetric synthetic aperture radar image analysis and interpretation and deep learning.



Ming Li (Member, IEEE) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in signal processing from Xidian University, Xi'an, China, in 1987, 1990, and 2007, respectively.

In 1987, he joined the Department of Electronic Engineering, Xidian University, where he is currently a Professor with the National Key Laboratory of Radar Signal Processing. His research interests include adaptive signal processing, detection theory, ultrawideband, and synthetic aperture radar image processing.



Peng Zhang (Member, IEEE) received the B.S. degree in electronic and information engineering and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 2006, 2009, and 2012, respectively.

He is currently an Associate Professor with the National Key Laboratory of Radar Signal Processing, Xidian University. His main research interests include SAR image interpretation and statistical learning theory.



Yan Wu (Member, IEEE) received the B.S. degree in information processing and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1987, 1998, and 2003, respectively.

From 2003 to 2005, she was a Postdoctoral Fellow with the National Key Laboratory of Radar Signal Processing, Xi'an. Since 2005, she has been a Professor with the Department of Electronic Engineering, Xidian University. Her research interests include remote sensing image analysis and interpretation, data

fusion of multisensor images, synthetic aperture radar autotarget recognition, and statistical learning theory and application.



Xinyue Xin received the B.S. degree in electrical information science and technology from Shaanxi Normal University, Xi'an, China, in 2019. She is currently working toward the Ph.D. degree in signal and information processing with the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an.

Her main research interests include polarimetric synthetic aperture radar image analysis and interpretation and deep learning.