








A Discontinuity-Guided Two-Dimensional Phase Unwrapping Method for SAR Interferograms

Wenjie Zhong , Jia Li , Xin Li , Senior Member, IEEE, Juanjuan Feng , Lei Guo , Zhiqiang Li, Junhui Wu , and Lingshuai Kong 

Abstract—Phase unwrapping (PU) is one of the core procedures of interferometric synthetic aperture radar. The traditional PU methods estimate the absolute phase gradient based on the assumption that phase is spatially continuous. However, this assumption is often not true due to the large-phase gradient. In this article, a local enhanced U-net (LEU-Net) was developed to determine the phase discontinuity of synthetic aperture radar (SAR) differential interferograms in areas with substantial elevation change. The phase discontinuity in interferograms predicted by this network was used as a priori information for the phase unwrapping maximum flow/minimum cut algorithm (PUMA). Within the LEU-Net, an interactive compression module was used to reduce the loss of detail information and ensure the ability of the model to segment discrete phase discontinuities; and a cross-fusion module was used to fuse multilevel features and highlight the location information while suppressing noise. The phase discontinuity information predicted by the LEU-Net guided the PUMA to obtain the unwrapped phase through iterative optimization; and in the meantime, the information can suppress the propagation of PU errors from areas of low signal-to-noise areas to areas of high quality. Experiments on simulated and real SAR interferograms with large-phase gradients showed that our method has a more robust PU capability.

Index Terms—Deep learning, graph cuts, large gradients, phase unwrapping (PU).

I. INTRODUCTION

INTERFEROMETRIC synthetic aperture radar (InSAR) can work in all weather and is widely used in the production of

Manuscript received 10 January 2024; revised 13 March 2024; accepted 7 April 2024. Date of publication 23 April 2024; date of current version 6 May 2024. This work was supported in part by the National Key R&D plans of China under Grant 2022YFB3903602, in part by the National Natural Science Foundation of China under Grant 42374053, in part by Hunan Provincial Natural Science Foundation under Grant 2023JJ30656, and in part by the Open Fund of the State Key Laboratory of Geodesy and Earth's Dynamics under Grant SKLGED2023-2-3. (Corresponding author: Jia Li.)

Wenjie Zhong, Juanjuan Feng, Lei Guo, Zhiqiang Li, Junhui Wu, and Lingshuai Kong are with the School of Geoscience and Info-Physics, Central South University, Changsha 410083, China, also with the Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, Changsha 410083, China, and also with the Laboratory of Geo-Hazards Perception, Cognition and Predication, Central South University, Changsha 410083, China.

Jia Li is with the School of Geoscience and Info-Physics, Central South University, Changsha 410083, China, also with the Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, Changsha 410083, China, also with the Laboratory of Geo-Hazards Perception, Cognition and Predication, Central South University, Changsha 410083, China, and also with the State Key Laboratory of Geodesy and Earth's Dynamics, Wuhan 430077, China (e-mail: lijia20050710@csu.edu.cn).

Xin Li is with the Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China.

Digital Object Identifier 10.1109/JSTARS.2024.3392637

digital elevation models (DEMs) [1]. Phase unwrapping (PU) is one of the core steps of InSAR technology. The outputs of PU directly affect the accuracy of DEM products. However, PU is an ill-posed problem. In order to obtain the most realistic solution, the traditional PU methods introduce the phase continuity assumption as a constraint, i.e., the absolute phase gradient between adjacent pixels is less than π [2]. If the assumption holds, the wrapped phase gradient between adjacent pixels is equivalent to the absolute phase gradient. In the regions with rough topography, the fringes of InSAR interferograms are dense and the PU is difficult. To address this problem, the original phase is usually differentiated from the topographic phase simulated from an external DEM. It is much easier to unwarp a differential interferogram. The elevation difference is then converted from the unwrapped differential phase, and the new DEM can be obtained by adding the elevation difference to the external DEM [3]. However, when the ground surface elevation changes significantly, there are a large number of phase discontinuities in the differential interferograms, which makes the phase continuity assumption invalid.

In order to weaken the influence of phase discontinuity, many PU methods have been proposed, which can be divided into three categories [4]: the path-following method, the minimum norm method, and the Bayesian method. The path-following method sets reasonable integration paths and obtains the unwrapped phase by simple integration. Goldstein branch cut [5] is a classical path-following method. This method first identifies the phase residues and sets branch cuts to balance the positive and negative residue pairs, and then sets an integration path that does not pass through the branch cuts. Besides, Flynn's minimum discontinuity PU [6] and quality-guided phase unwrapping (QGPU) [7] are also commonly used path-following methods. The path-following method is a local optimization method. Hence, the results are susceptible to noise, and isolated islands are prone to emerge in regions with low signal-to-noise ratios.

The minimum norm method establishes the L^p norm objective functions to find the global optimal solution that minimizes the difference between unwrapped and wrapped phase gradients. The minimum cost flow (MCF) is an L^1 norm-based method [8]. It estimates the unwrapped phase by minimizing the weighted deviation between the real phase gradient and the estimated phase gradient, and has become one of the most commonly used PU methods. However, the unwrapping errors in the regions with large-phase gradients are easy to propagate to the whole image.

The least square (LS) method is an L^2 norm-based method [9]. It realizes the PU by solving Poisson's equation. The results of the LS method are excessively smoothed. Moreover, the phase after rewrapping is inconsistent with the original wrapped phase. The L^0 norm-based method (e.g., statistical-cost, network-flow algorithm for phase unwrapping (SNAPHU) [10]) has better performance; however, the minimization of the L^0 norm is proved to be a nondeterministic polynomial-time hardness problem. At present, only an approximate solution can be obtained [10].

The Bayesian method estimates the unwrapped phase based on the data observation mechanism and the prior knowledge of the phase [11]. The Markov random fields (MRFs) based method, a typical Bayesian method, transforms the PU problem into an energy minimization problem. The iterated conditional modes (ICM) [12], phase unwrapping max-flow/min-cut algorithm (PUMA) [11], loopy belief propagation [13], and sequential tree-reweighted message passing [14] also belong to this kind. Among these four methods, the PUMA has the highest efficiency and the best performance in the regions with large-phase gradients [4].

In the scenes of substantial surface elevation changes (such as glacier surge or glacier collapse), the phase continuity assumption is severely breached. The results of the above traditional methods are prone to containing obvious jumps. In recent years, researchers have put a lot of efforts to overcome the limitations of the phase continuity assumption through deep learning. Three types of deep-learning-based PU methods have been proposed: the direct mapping method, the wrap-counts prediction method, and the phase-gradient information prediction method. The direct mapping method treats PU as a regression problem. Based on the backpropagation of the neural network, the internal parameters are continuously adjusted. The input-output mapping model is then established to predict the unwrapped phase [15]. The DLPU proposed by Wang et al. [16] can achieve fine results in optical interferometric PU experiments. Wu et al. [17] combined DDNet and PUNet. First, the regions with large-scale subsidence were extracted by the DDNet. And then PUNet was used to regress the solution for the interferograms in the regions with large-scale subsidence. In order to ascertain that the phase after rewrapping is close to the original wrapped phase, Zhou et al. [18] proposed a model based on the conditional generative adversarial networks, i.e., PU-GAN. In general, since the direct mapping method only considers the simple mapping relationship between the wrapped phase and the absolute phase, the universality of the models is poor, and the training dataset can significantly influence on the accuracy of the model.

The wrap-counts prediction method treats PU as a semantic segmentation problem. Spoorthi et al. [19] used PhaseNet to predict the wrap counts, and then smoothed the PU results by clustering. The segmentation results were susceptible to noise at the edges of fringes. Afterward, they developed PhaseNet 2.0 [20], replacing the convolutional layers in the original network with dense blocks. A loss function was designed to improve network performance. However, due to the extreme imbalance of classes in the real InSAR interferograms, the model training of this method is very difficult, and therefore, the regions with large-phase gradients still show obvious PU errors.

The phase-gradient information prediction method first predicts the phase-gradient information based on deep-learning models and then combines the predicted information with traditional methods to reconstruct the absolute phase. Zhou et al. [21] first used PGNet to predict the horizontal and vertical wrap-count gradients, and then used L^1 norm to obtain the unwrapped phase. Subsequently, a number of similar studies have been developed [22], [23], [24], [25]. Wu et al. [26] designed a sample simulator that considered multiple phase features and designed DENet to predict the phase discontinuity. After that, the predicted information was converted into weights for the MCF solver of the SNAPHU to obtain the unwrapped phase. This kind of method combines the advantages of deep-learning models and traditional methods, and has great potential for application in complex scenes. However, both PGNet and DENet perform full-resolution prediction of interferograms. In this context, the segmentation capability and computational efficiency of the model cannot be balanced.

In this article, an InSAR phase discontinuity estimation network (LEU-Net) was developed for areas with substantial elevation change based on the classical medical segmentation model—U-Net [27]. The LEU-Net predicted the phase discontinuity in the interferograms, and then the PUMA used the phase discontinuity confidence score as a priori information to obtain the unwrapped phase. Unlike other models, LEU-Net balances computational efficiency and segmentation accuracy. An interactive compression module (ICM) was used in the encoder of the LEU-Net to reduce local information loss caused by downsampling, and a cross-fusion module (CFM) was used in the decoder to extract channel information from shallow features and guide feature decoding. The PUMA has relatively good robustness and computational efficiency in regions with large-phase gradients [4], [28]. Meanwhile, the phase discontinuity information predicted by LEU-Net can effectively suppress the PU error of PUMA. Therefore, the method proposed in this article can achieve robust PU results in the regions with large-phase gradients. The capability of the new method was validated with the simulated data and the real InSAR data.

The rest of this article is organized as follows. Section II introduces the rationale for combining deep learning and traditional algorithms. Section III introduces the details of the LEU-Net architecture and the way of sample simulation. Section IV presents the evaluation of the new method based on the simulated and real InSAR data. Finally, Section V concludes this article.

II. PROBLEM FORMULATION

In the InSAR interferogram $G_0^{M \times N}$, the wrapped phase ψ and the absolute phase ϕ have the following relationship:

$$\phi(i, j) = \psi(i, j) + 2\pi k(i, j) \quad (1)$$

where $(i, j) \in G_0^{M \times N}$ represents the pixel position; and $k(i, j)$ denotes an integer, called the wrap count of pixels (i, j) . The phase-gradient relationship of adjacent pixels can be obtained as follows:

$$\Delta\phi = \text{wrap}(\Delta\psi) + 2\pi\Delta k \quad (2)$$

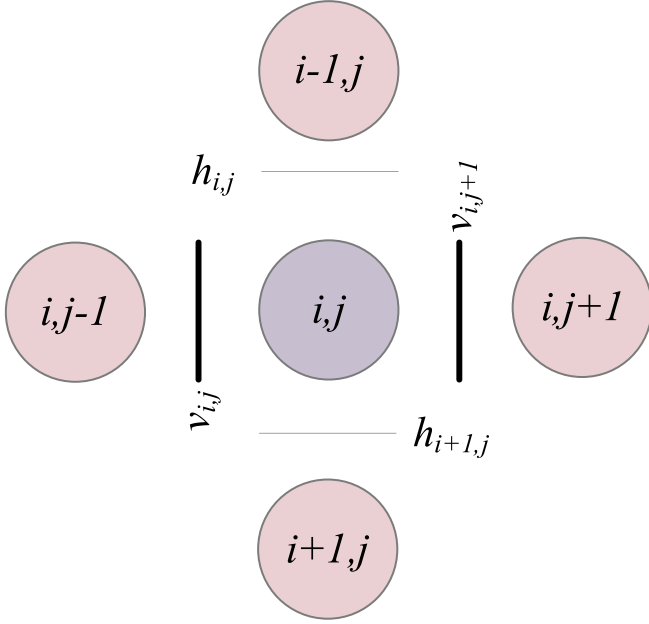


Fig. 1. Display of discontinuities among pixel (i, j) and its first-order neighbors.

where $\text{wrap}(\cdot)$ is a wrapping operator. The PUMA first defines the energy function in the MRF

$$E(k|\psi) \equiv \sum_{ij \in G_0} V(\Delta\phi_{ij}^h) v_{ij} + \sum_{ij \in G_0} V(\Delta\phi_{ij}^v) h_{ij} \quad (3)$$

where $V(\cdot)$ denotes the clique potential that depends only on the site variables indexed by the respective clique elements; $h_{ij}, v_{ij} \in \{0, 1\}$ represent the horizontal and vertical discontinuity on the adjacent pixels (see Fig. 1); $h_{ij}, v_{ij} = 0$ indicates the occurrence of phase discontinuity; and $\Delta(\cdot)^h$ and $\Delta(\cdot)^v$ represent the horizontal and vertical first-order partial derivatives, which can be calculated by

$$\begin{aligned} \Delta\phi_{ij}^h &\equiv [2\pi(k_{ij} - k_{ij-1}) + \Delta\psi_{ij}^h] \\ \Delta\phi_{ij}^v &\equiv [2\pi(k_{ij} - k_{i-1j}) + \Delta\psi_{ij}^v] \\ \Delta\psi_{ij}^h &\equiv \psi_{ij} - \psi_{ij-1} \\ \Delta\psi_{ij}^v &\equiv \psi_{ij} - \psi_{i-1j}. \end{aligned} \quad (4)$$

At time $t + 1$, the wrap count of pixel (i, j) can be expressed as $k_{ij}^{t+1} = k_{ij}^t + \delta_{ij}$. Substituting k_{ij}^{t+1} into (4) gets

$$\begin{aligned} \Delta\phi_{ij}^h &= [2\pi(\delta_{ij} - \delta_{ij-1}) + a^h] \\ \Delta\phi_{ij}^v &= [2\pi(\delta_{ij} - \delta_{i-1j}) + a^v] \\ a^h &\equiv 2\pi(k_{ij}^t - k_{ij-1}^t) + \Delta\psi_{ij}^h \\ a^v &\equiv 2\pi(k_{ij}^t - k_{i-1j}^t) + \Delta\psi_{ij}^v \end{aligned} \quad (5)$$

where $\delta_i \in \{0, 1\}$ is a binary variable. The energy function with respect to k can be transformed into a function with respect to

the binary variable $\delta \in G_1^{M \times N}$

$$\begin{aligned} E(k^t + \delta|\psi) &= \sum_{ij \in G_0} V[2\pi(\delta_{ij} - \delta_{ij-1}) + a^h] v_{ij} \\ &\quad + \sum_{ij \in G_0} V[2\pi(\delta_{ij} - \delta_{i-1j}) + a^v] h_{ij} \end{aligned} \quad (6)$$

$$\begin{aligned} E(k^t + \delta|\psi) &= \sum_{ij \in G_1} E_h^{ij}(\delta_{ij-1}, \delta_{ij}) \\ &\quad + \sum_{ij \in G_1} E_v^{ij}(\delta_{i-1j}, \delta_{ij}) = \sum_{ij \in G_1} E^{ij}(\delta_i, \delta_j). \end{aligned} \quad (7)$$

When a convex clique potential is used, local and global optimality are equivalent and the iterative solution of the binary optimization is convergent. The minimization of (7) with respect to δ can be mapped onto a maximum flow problem. A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is constructed based on the interferogram $G_0^{M \times N}$ and the energy function. \mathcal{V} and \mathcal{E} represent the vertices and edges of the graph. The number of vertices is $2 + M \times N$ because of the two additional terminals, i.e., the source and the sink in the graph. Energy terms E_h^{ij} and E_v^{ij} in (7) can be represented by a basic graph with four vertices of $\{s, t, v, v'\}$. In particular, $\{s, t\}$ represent the source point and the sink point of the graph, and $\{v, v'\}$ represent two adjacent pixels in the interferogram. The clique potential is used to calculate the nonnegative weights of each side. Finally, energy minimization for convex potentials is obtained by iteratively calculating the maximum flow on the graph. When the value is no longer small, the optimal solution is achieved.

When calculating the energy function, previous methods often used coherence to depict the phase discontinuity. The limited number of samples can lead to an overestimation of coherence, while interferometric fringes can lead to an underestimation of coherence [29]. More importantly, it cannot directly reflect the phase discontinuity of adjacent pixels. Inaccurate phase discontinuity information renders low reliability of directed edge weights, which may lead to wrong iterative optimization of maximum flow/minimum cut and, therefore, errors in PU results. To accurately estimate the phase discontinuity, this article presented LEU-Net.

III. PROPOSED METHOD

In the new method, the PU task is divided into two parts. First, the coherence map, residues map, and interferogram are used as the three-channel input of LEU-Net. The output of the model is the confidence of the horizontal and vertical phase discontinuity [$s_{i,j}^h, s_{i,j}^v \in (0, 1)$]. The higher the confidence, the greater the likelihood of a discontinuity. Second, the predicted confidence ($s_{i,j}^h, s_{i,j}^v$) is deemed as h_{ij}, v_{ij} , and the wrap counts are estimated by PUMA.

A. Dataset Generation

Deep-learning model learns by discovering intricate structures in sample data. High-quality training samples reduce the effects of useless information, and therefore are crucial for the

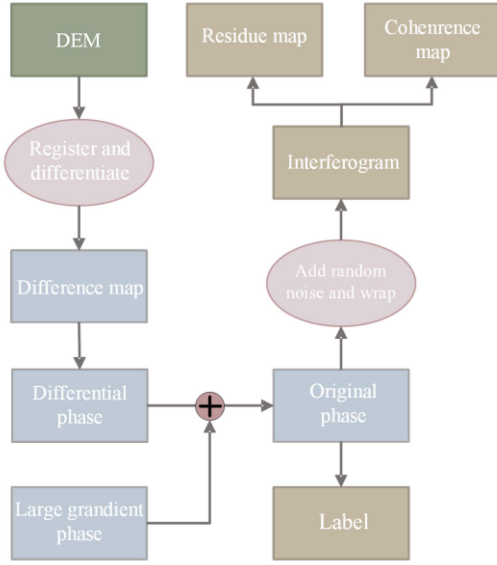


Fig. 2. Flowchart of sample simulation.

effectiveness of model training. As we know, both glacier surge activities and strong ablation can cause significant changes in glacier surface elevation. If the topographic phase simulated from a presurge DEM is subtracted from the postsurge synthetic aperture radar (SAR) interferometric phase, or the topographic phase simulated from a preablation DEM is subtracted from the postablation SAR interferometric phase, it is likely that the phase continuity assumption cannot hold for the differential SAR interferometric phases over the glaciers. In this context, we simulated training samples based on the differences of the SRTM DEM and COP30 DEM in the glacierized areas of high Asia. Random noises and phase of large gradients were considered during the simulation (see Fig. 2). The SRTM DEM was generated from SRTM interferometric SAR pairs acquired in February 2000, and the COP30 DEM was generated from TanDEM-X interferometric SAR pairs acquired during 2010–2015. Glacier surges are common in high Asia, and most of the glaciers in high Asia experienced strong ablation during the early 21st century. As a result, the differences between the SRTM DEM and the COP30 DEM can be significant in some glaciers' surface.

The height datum of the SRTM DEM and the COP30 DEM was unified, and the two DEMs were matched before differencing. The orbital parameters of the TanDEM-X satellite were used to transform the elevation difference map into a differential phase φ

$$\varphi(m, n) = \frac{4\pi}{\lambda} \frac{B_{\perp}}{R} h(m, n) \quad (8)$$

where B_{\perp} is the vertical baseline length (5 ~ 320 m); λ is the wavelength (0.028 ~ 0.052 m); R is the slant range between the SAR antenna and the ground target point (710 000.0 m); and θ is the satellite incidence angle (20° ~ 50°). $h(m, n)$ is generated by geocoding the difference map. To ensure the performance of model, the sizes of differential phase maps were randomly cropped into 256×256 , and the phase values were randomly scaled to $[-6\pi, 6\pi]$.

Since the glacier area occupies a small proportion of the whole area, and not every piece of the glacier area is associated with surge activity or strong ablation, the phase samples of large gradients may be insufficient. In this context, a 2-D Gaussian distribution function was adopted to simulate the phase of large gradients. The simulated phase was added to the sample data with a probability of 30%. The original phase ϕ was obtained by combining the differential phase with the simulated phase of the large gradient.

Decorrelation of the SAR interferometric phase is common in glacierized regions. In order to improve the generalizability of the model, complex Gaussian random noise was added to the original phase. The interferometric phase ψ was calculated as follows:

$$\psi = \text{angle} \left((\cos \phi + D \cdot n_R) + (\sin \phi + D \cdot n_I) \times \sqrt{-1} \right) \quad (9)$$

where $\text{angle}(\cdot)$ calculates the principal value of the complex phase; n_R, n_I denote the real and imaginary parts of the randomly generated complex Gaussian noise; and D controls the amplitude of noises. D consists of a combination of D_{phase} and $D_{\text{decoherence}}$. D_{phase} is related to the original phase value but is randomly scaled. D_{phase} controls the generation of randomly shaped decorrelation phenomena.

Considering that the horizontal and vertical phase discontinuity locations are correlated with the positive and negative phase residues, we calculated the phase residue maps. Coherence maps, residue maps, and interferograms were integrated as the three-channel input data of LEU-Net. Meanwhile, the horizontal and vertical phase discontinuity point maps were integrated as two-channel labels to guide the model training. In total, 10 000 samples were obtained. We divided them into three parts: the training set (8000 samples), the validation set (1000 samples), and the testing set (1000 samples). Three samples are shown in Fig. 3. All three samples contain phase discontinuities caused by glacier changes. In sample 1, the glacier adjacent region is characterized by phase of large gradient. However, the number of noises is small and the interference fringes are distinctive. In sample 2, one glacier tongue and the narrow area ahead of this glacier tongue are covered by the phase of large gradients. In sample 3, the trunk of one glacier is partly covered by phase of large gradient and high noise. It is a challenge to unwrap these three samples correctly.

B. Network Architecture

In the new method, PU is considered a semantic segmentation problem. The interferogram phase discontinuities are predicted by binary classification. Since the distribution of phase discontinuities is uneven, precise target segmentation is challenging. Local and detailed information is vital for predicting phase discontinuity. Inspired by the medical semantic segmentation model, LEU-Net was developed based on the convolutional operation (CNN-based) U-Net [30]. An encoder, a decoder, and skip connections compose LEU-Net, as shown in Fig. 4(a). The encoder and decoder are constructed based on residual blocks [31]. In the encoder, an ICM was designed to preserve the local information as much as possible and to fully extract the features

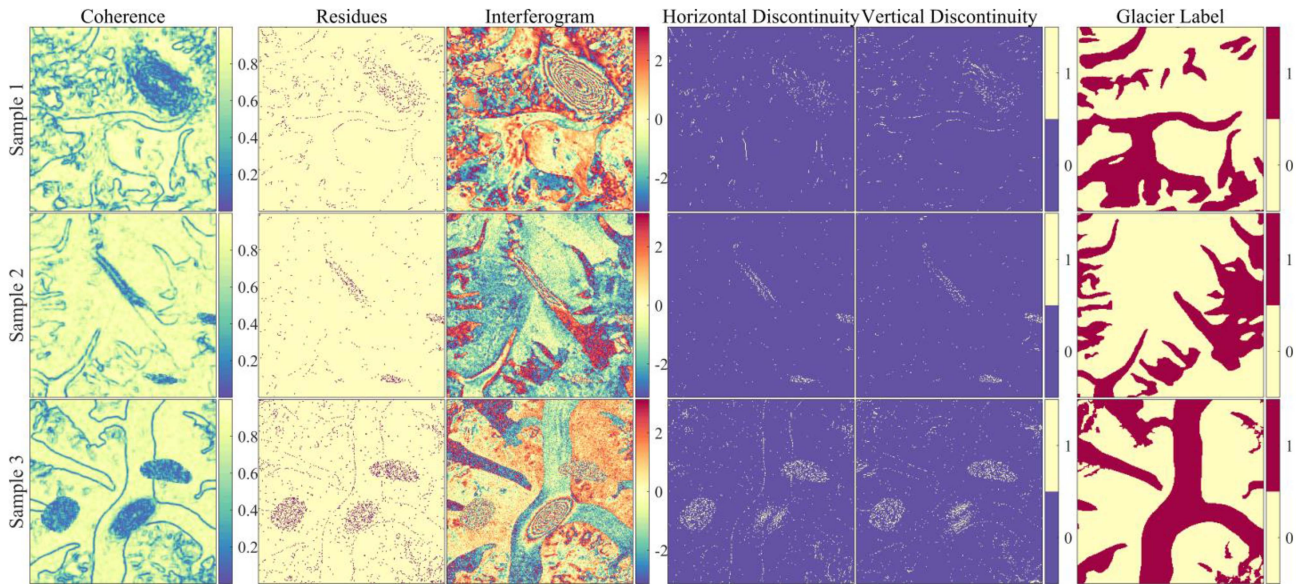


Fig. 3. Example of three simulated samples. The first to the third rows correspond to samples for three regions. The first to the sixth columns correspond to the coherence, phase residues, interferograms, horizontal phase discontinuity, vertical phase discontinuity (0: continuous; 1: discontinuous), and glacier label (0: ice-free area, 1: glacierized area), respectively.

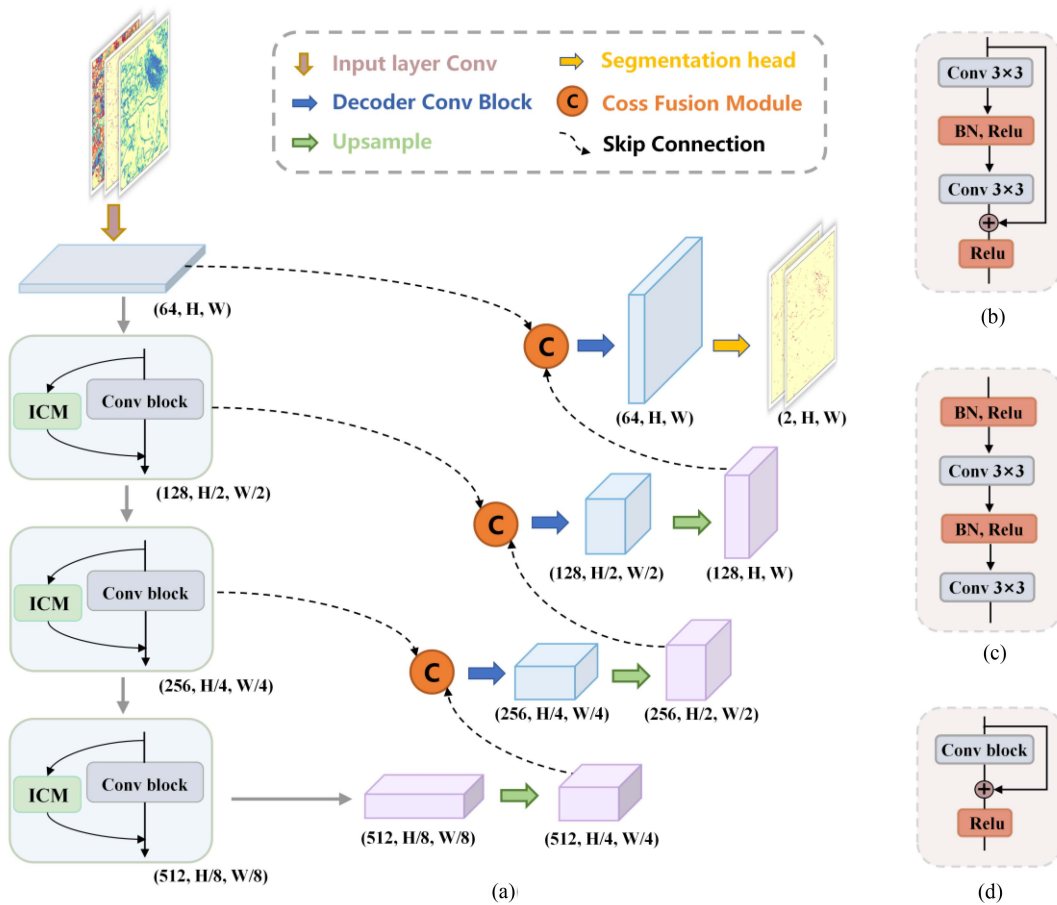


Fig. 4. (a) Overview of LEU-Net architecture. (b) Specific structure of the input layer convolution block. (c) Specific structure of double convolution block. (d) Specific structure of decoder convolution block.

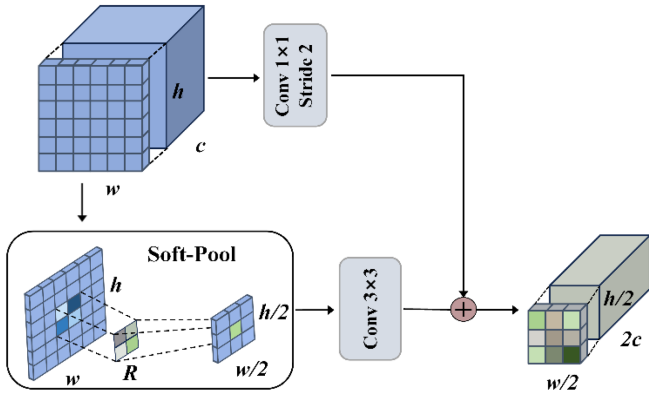


Fig. 5. Structure of the interactive compression module (ICM).

in the image. In the decoder, a CFM was designed based on the squeeze-and-excitation block [32], which can suppress the noise in the shallow features and complement the detailed information in the deep features.

The original image is first put into a convolutional input layer without downsampling, as shown in Fig. 4(b). The obtained shallow features are put into three consecutive modified residual blocks. Each residual block halves the feature resolution and doubles the number of channels to capture deep features. The output feature of residual block no. n can be expressed as $A_n \in \mathbb{R}^{(H/2^n) \times (W/2^n) \times 2^n C}$, $C = 64$. After the encoding process, features $F \in \mathbb{R}^{(H/8) \times (W/8) \times 512}$ can be obtained. In decoder, F is put into a bilinear upsampling layer to increase the resolution, and then fed into CFM to be fused with the low-level feature. Subsequently, the decoder convolution block is used to reduce the dimension of the fused feature, as shown in Fig. 4(d). The above steps are repeated three times to obtain features $F' \in \mathbb{R}^{H \times W \times 64}$ with fully recovered resolution. Finally, the predicted phase discontinuity map is output through a segmentation head that consists of a 1×1 convolution layer.

C. Module Details

1) *Interactive Compression Module (ICM)*: In previous U-shaped networks, downsampling was often achieved by maximum pooling or by adjusting the strided convolutions, which may lead to the loss of local detail. This is disadvantageous for the segmentation of dense, small-scale, and discrete phase discontinuity points. Since local information is vital for segmenting discontinuities, we designed an ICM based on the residual connection (see Fig. 5). Feature $I \in \mathbb{R}^{h \times w \times c}$ is put into the ICM. Here, $h = H/2^{n-1}$, $w = W/2^{n-1}$, $c = 2^{n-1}C$; n denotes the number of residual blocks where the ICM is located. The ICM has two branches. In the residual connection part, the features are put into a 1×1 convolution with a step size of 2 for dimension upscaling and resolution reduction. A batch normalization (BN) is performed to obtain $I_1 \in \mathbb{R}^{h/2 \times w/2 \times 2c}$. The residual connection effectively mitigates gradient vanishing and gradient explosion caused by the increase of the layer number and helps the model to learn the phase-gradient information. The other branch takes sufficient discrete features of phase

discontinuities and introduces the soft pooling [33], which can preserve the local feature information as much as possible. Soft pooling activates the pixels s in the pooling kernel R through exponential weighting

$$\tilde{s} = \sum_{i \in \mathbf{R}} \frac{e^{s_i} * s_i}{\sum_{j \in \mathbf{R}} e^{s_j}}. \quad (10)$$

After pooling, feature \tilde{s} is batch normalized and fed into a convolution layer. Features $I_2 \in \mathbb{R}^{h/2 \times w/2 \times 2c}$ with half of the resolution and twice of channel numbers are obtained. I_2 can be expressed as follows:

$$I_2 = \varphi_1(\text{SoftPool}(s)) \quad (11)$$

where $\varphi_1(\cdot)$ denotes a 3×3 convolution layer with BN and Gaussian error linear unit function [34]. Overall, the residual connection branch is set up to improve the training efficiency of the model, and the soft pooling branch fully uses the data features and preserves local information as much as possible. The two branches are complementary. Finally, the ICM outputs feature F_{ICM}

$$F_{\text{ICM}} = I_1 \oplus I_2. \quad (12)$$

2) *CFM*: Shallow features contain more positional information and details, but are easily affected by noises. Deep features contain stronger semantic information, but the resolution is lower, and the local information is compressed, which makes it difficult to obtain discontinuity information. To effectively fuse shallow and deep features, we proposed CFM (see Fig. 6). Shallow features $L_n \in \mathbb{R}^{h \times w \times c}$ are dimensionally expanded to obtain $L'_n \in \mathbb{R}^{h \times w \times 2c}$. The global maximum pooling and global average pooling strategies are used to compress L'_n along the spatial dimension, which aims to extract the maximum and average numbers with global receptive fields in each channel. The two features $U_{\text{avg}}, U_{\text{max}}$ can be obtained by feeding the above two numbers into a common convolution layer with shared weights. The process is shown as follows:

$$\begin{aligned} U_{\text{avg}} &= \varphi_2(\text{AvgPool}(L'_n)) \\ U_{\text{max}} &= \varphi_2(\text{MaxPool}(L'_n)) \end{aligned} \quad (13)$$

where $\varphi_2(\cdot)$ is a 3×3 convolutional layer with ReLU and BN, which can reduce dimension; $r = 16$ is the dimension reduction times of the convolution layer. Two features $U_{\text{avg}}, U_{\text{max}} \in \mathbb{R}^{1 \times 1 \times 2c/r}$ are then summed and fed into a convolution layer, which allows the network to learn the feature weights of each channel and to enhance the effective features. Weights $U_n \in \mathbb{R}^{1 \times 1 \times 2c}$ are obtained as follows:

$$U_n = \delta(\varphi_3(U_{\text{avg}} + U_{\text{max}})) \quad (14)$$

where δ represents a sigmoid function, and $\varphi_3(\cdot)$ represents a 3×3 convolution layer.

Upsampled deep features $H_n \in \mathbb{R}^{h \times w \times 2c}$ are multiplied by weights U_n to obtain features H'_n . Finally, the refined deep features are concatenated with the original shallow features to generate the output features of the CFM $C_n \in \mathbb{R}^{h \times w \times 3c}$

$$C_n = \text{Concat}(H'_n, L_n). \quad (15)$$

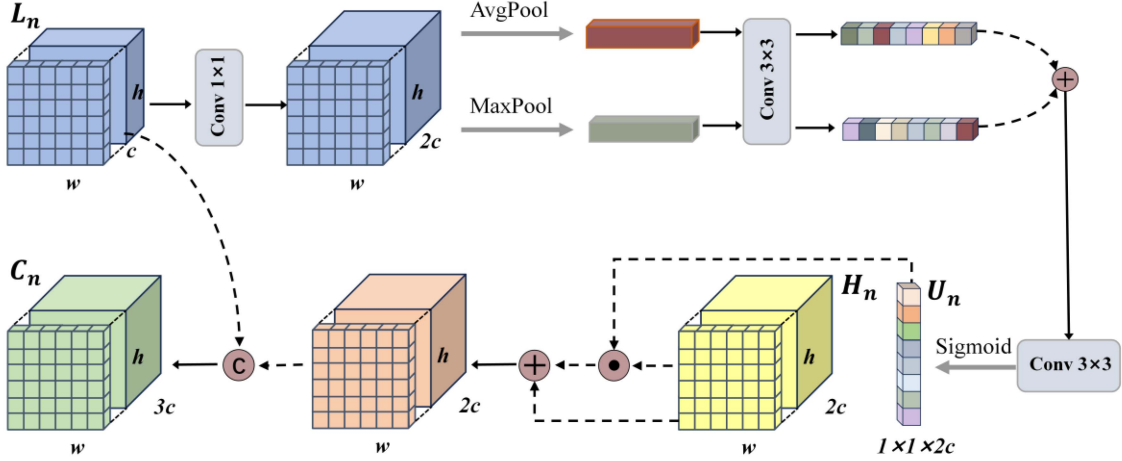


Fig. 6. Structure of the CFM.

D. Training Implementation

1) *Training Settings*: LEU-Net was built using the PyTorch [35] framework based on Python 3.9. The model was deployed on a PC configured with an AMD Ryzen 9 5950X 16-core processor (4.40 GHz), 256 GB RAM, and an NVIDIA GeForce 3090Ti. The batch size was set to 21. The initial learning rate was set to $5e^{-5}$ and the cosine annealing strategy was used to delay the learning rate. An Adam optimizer [36] was used to compute the gradient and to update the network parameters. In total, 150 epochs were trained on the training set. Finally, the model parameters of epoch 132 that showed the best performance on the validation set were selected.

2) *Loss Functions*: As mentioned above, the phase discontinuities are discrete and are unevenly distributed in the interferograms. Moreover, there is an imbalance between the number discontinuous category and the smooth category. The smooth phase can hardly provide effective information for the model learning. Appropriate setting of the loss function is an effective way to deal with the category imbalance of the samples [37]. The dice coefficient is an ensemble similarity metric, which is usually used to calculate the similarity between two samples. It is defined as follows:

$$\text{Dice}(A, B) = 2 \frac{|A \cap B|}{|A| + |B|} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \quad (16)$$

where A and B represent the ground truth of the two channels and the phase discontinuity maps predicted by the model; TP, FN, and FP denote the number of true-positive, false-negative, and false-positive pixels, respectively. Dice loss is a region-based loss function. It can promote the model to mine the foreground region and to focus on the segmentation accuracy of the discontinuous phase. Dice loss [38] is defined as follows:

$$L_{\text{Dice}} = 1 - \text{Dice}(A, B). \quad (17)$$

However, when the phase discontinuities are incorrectly predicted, the dice loss changes significantly. Therefore, in this article, the dice loss is used in combination with the binary cross-entropy (BCE) [39] loss function.

BCE is a commonly used binary classification loss function. It is defined as follows:

$$L_{\text{BCE}} = \text{BCE}(p, y) = \text{BCE}(p_t) = -\log(p_t),$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0. \end{cases} \quad (18)$$

where y is the sample label; 1 and 0 mean the discontinuous and smooth; and p_t denotes the probability that the prediction is consistent with the ground truth.

Finally, the joint loss function used in this article can be expressed as follows:

$$L = L_{\text{Dice}} + L_{\text{BCE}}. \quad (19)$$

Dice loss focuses on pixel-level similarity and can improve the capability of the model to segment phase discontinuities in regions with large-phase gradient. BCE loss focuses on the accuracy of segmentation of phase discontinuity and has fine stability at the beginning of training, which helps to speed up the convergence of the model.

3) *Evaluation Index*: In this article, IoU and Recall were used to evaluate the capability of the model to predict phase discontinuity

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (20)$$

IoU is a metric that measures the performance and accuracy of the model in segmenting phase discontinuities. Recall indicates the sensitivity of the model to phase discontinuity.

IV. EXPERIMENTAL RESULTS

A. Ablation Experiments

In order to evaluate the performance of the model and the two core modules, this article performed ablation experiments using U-Net as a baseline. The results are presented in Table I. Using residual structure instead of VGG in U-Net (i.e., ResUNet)

TABLE I
LEU-NET ABLATION EXPERIMENTS ON THE TEST SET

Model	IoU(%)		Recall(%)	
	Horizontal	Vertical	Horizontal	Vertical
U-Net	56.66	58.61	69.50	72.10
ResUNet	61.85	64.04	74.26	76.24
ResUNet+ICM	64.40	66.69	76.32	78.38
ResUNet+CFM	63.04	65.11	75.96	77.80
ResUNet+ICM+CFM	66.34	68.44	77.41	79.41

The bold font is to highlight the maximum value of the evaluation indicator.

improved mean IoU by 6.29%. Both ICM and CFM proposed in this article improved the segmentation accuracy. The best model performance was achieved when both ICM and CFM were used. In particular, the mean IoU and mean recall reached 67.39% and 78.41%.

B. Results of Phase Discontinuity Prediction

In order to demonstrate the superiority of LEU-Net in phase discontinuity segmentation, this article compared LEU-Net with some advanced U-shaped deep-learning models under the same experimental conditions, including the CNN-based models (U-Net [30], Attention U-Net [40], and UNet++ [41]) and the transformer-based models (TransUnet [42], TransFuse [43], and Swin-Unet [44]). Both TransUnet and TransFuse are network structures based on the hybridization of CNN and transformer. TransUnet embeds a standard ViT in a CNN-based U-shaped network. TransFuse has a CNN branch parallel to the DeiT branch. Swin-Unet is a network composed of pure Swin transformer blocks. In addition to the aforementioned U-type network, another superior performance DENet was also used for comparison. However, due to limited computational resources, the conditions for DENet training are slightly different from the other models, and the size of the training samples is randomly cropped to be 104×104 before inputting into the model. The batch size was set to 9. We used three strategies to evaluate the segmentation accuracy of DENet in the testing phase. The first one is to randomly crop the size of the test samples to 104×104 as well (DENet-104). The second one is to complete the test directly on the 256×256 test set to complete the test (DENet-256). The third one is to use a sliding window to complete the prediction on the test samples. The window size was set to 104×104 , and the repetition rate between each window was set to 0.5.

Table II presents the segmentation results of the above models (for the simulated dataset). LEU-Net has the best ability to segment phase discontinuities in both horizontal and vertical directions. Mean IoU and mean recall reached 67.39% and 78.41%, which proves the effectiveness of LEU-Net. DENet achieves fine performance when using the sliding window prediction strategy.

The segmentation accuracy of the pure CNN-based models is generally better than that of the transformer-based models. It may be due to the reason that the transformer lacks the ability to focus on local information, while the local features and

TABLE II
COMPARISON OF THE ACCURACY OF PHASE DISCONTINUITY PREDICTION BETWEEN THE PROPOSED MODEL AND OTHER MODELS

Model	IoU(%)		Recall(%)	
	Horizontal	Vertical	Horizontal	Vertical
U-Net	56.66	58.61	69.50	72.10
Attention U-Net	56.45	57.61	68.97	70.40
Unet++	55.81	57.90	67.85	71.25
TransUnet	48.42	50.60	62.74	65.61
TransFuse	18.29	18.96	31.72	32.87
Swin-Unet	47.70	48.71	64.27	65.29
DENet-104	63.64	65.66	75.62	77.11
DENet-256	58.22	60.09	73.45	74.71
DENet-slide	59.40	61.22	72.91	74.84
LEU-Net	66.34	68.44	77.41	79.41

The bold font is to highlight the maximum value of the evaluation indicator.

positional information extracted by convolutional operations are beneficial to discontinuity segmentation. The performance of the three CNN-based models is close. Among the three transformer-based models, the accuracy of TransFuse with a parallel structure is significantly lower, probably due to the fusion module of the model, which destroys the detailed positional information of the features. The accuracy of Swin-Unet is comparable with TransUnet. It may be attributed to the use of window attention in the Swin transformer that is able to capture more local image features.

Fig. 7 presents the results of phase discontinuity segmentation in the first sample (see the first row of Fig. 3). In general, phase discontinuity is clustered at glacier boundaries and regions of large-phase gradient. Due to the full-resolution prediction, DENet achieves good results. Benefiting from more adequate local information, LEU-Net can effectively segment discrete phase discontinuity. The confidence level for the segmentation results of these two models is higher than other models. In terms of the local details of the segmentation results (line 4 of Fig. 7), the segmentation results of the CNN-based models and the transformer-based models only cover a part of the phase discontinuities. The ability of DENet in estimating phase discontinuity information near glacier boundaries is clearly inferior to LEU-Net. In contrast, the results of LEU-Net are close to the ground truth. The CNN-based model is sensitive to phase discontinuity. However, there are a large number of false positives in the prediction results, which can lead to serious error propagation in the estimation of winding count estimation. Since the network module is insensitive to detailed information, there are a large number of errors in the prediction results of the transformer-based models, indicating that the transformer-based models cannot be directly used for phase discontinuity segmentation. LEU-Net accurately predicts the phase discontinuity and, therefore, can provide effective prior information for PUMA.

C. Evaluation of PU Performance on Simulated Datasets

In the PU experiments on simulated datasets, our method was compared with five classical traditional methods [including two path-following methods (Branch cut [5] and QGPU [7]), two

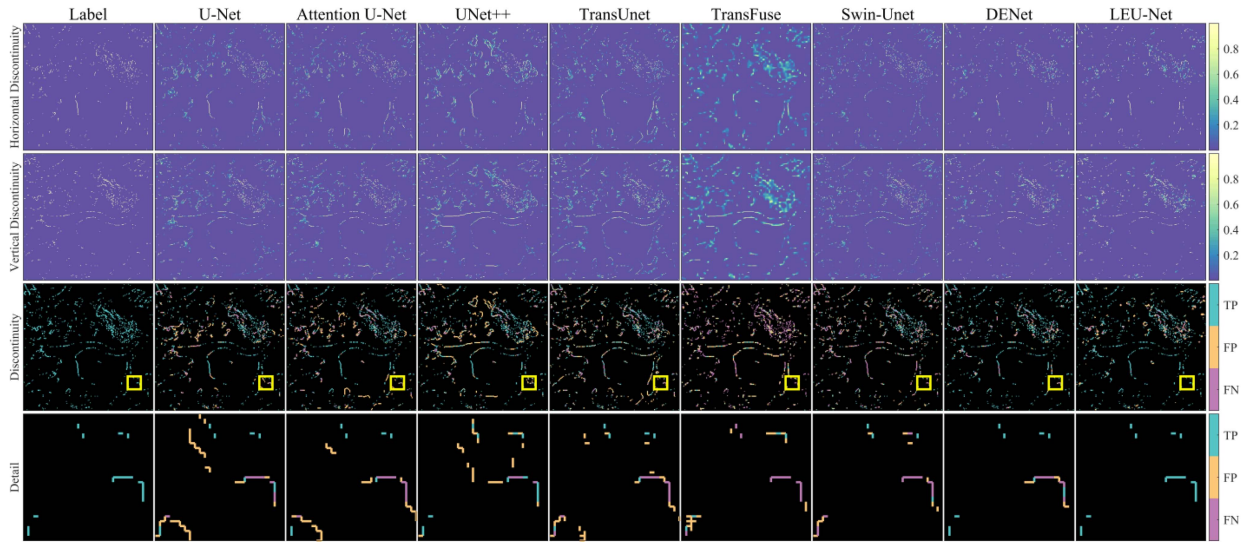


Fig. 7. Phase discontinuity information predicted by different deep-learning models in the first sample (shown in the first row of Fig. 3). The first column is the ground truth. The second to the eighth columns are the prediction results of different models. The first and the second rows show the predicted horizontal and vertical phase discontinuities, respectively. The third row is the location of the phase discontinuity. The fourth row is the zoomed view within the yellow box in the third row, which reflects the segmentation performance of the models at the glacier boundaries. Blue lines: true positives (TP), yellow lines: false positives (FP), and purple lines: false negatives (FN).

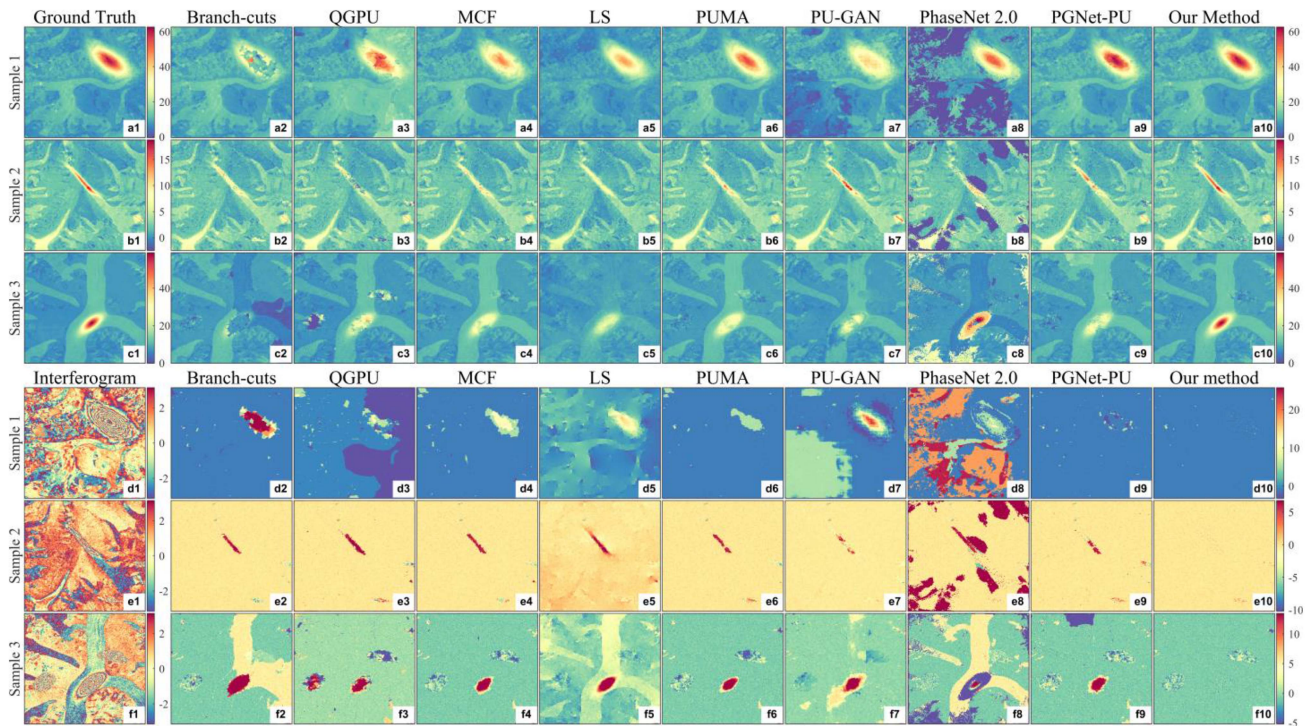


Fig. 8. Comparison of the results of different PU methods for three samples. a1, b1, and c1 represent the absolute phases of the three interferogram samples, i.e., the true phases. a2–a11, b2–b11, and c2–c11 represent the PU results for samples 1, 2, and 3, respectively. d1, e1, and f1 represent the three interferograms. d2–d11, e2–e11, and f2–f11 represent the errors of PU results of different methods (positive values denote underestimation).

minimum norm methods (MCF [45] and LS [9]), and a Bayesian method (PUMA [11]), and three typical deep-learning-based methods [including a direct mapping method (PU-GAN [18]), a wrap-count prediction method (PhaseNet 2.0 [20]), and a phase-gradient information prediction method (PGNet-PU [21]). The clique potential chosen by PUMA and our method was x^2 . The

PU results for three simulated samples (shown in Fig. 3) are shown in Fig. 8. All the PU methods could obtain desirable results in the region of sparse fringes. However, in the regions of large-phase gradients, the performance of these methods showed significant differences. The result of branch cut contained isolated islands. QGPU achieved a fine result in sample

1, but its results in samples 2 and 3 contained obvious jumps, indicating that the path-following method is unreliable in regions with large-phase gradients and high-level noise. The results of norm-based methods (MCF and LS) did not show significant jumps but had severe underestimation in regions with large-phase gradients. The accuracy of LS was lower due to excessive smoothing. Among the traditional methods, PUMA had the best performance. In general, all the traditional methods achieved underestimated unwrapped phase in the regions of large-phase gradients. The wrap-counts prediction method (PhaseNet 2.0) achieved results with remarkable jumps in each sample. The result of direct mapping method (PU-GAN) in sample 1 also had significant jumps. The phase-gradient information prediction method (PGNet-PU) had better performance than the above two methods, but phase jumps and underestimations were still notable in their results. In contrast, the results of our method were close to the true phase.

D. Assessment of PU Performance on Real Interferograms

In this experiment, we selected methods from each of the three types of traditional methods that perform relatively well in the simulation experiments for comparison, i.e., QGPU [7], MCF [45], and PUMA [11]. Deep-learning-based methods PU-GAN [18] and PGNet-PU [21] were also adopted for the comparison. Since the range of wrap counts in the real interferograms cannot be determined, and the range of wrap counts in the simulated interferogram is much different from that in the real interferogram, it is very difficult for PhaseNet 2.0 [20] that is trained by the simulated data to predict the wrap counts of the real interferograms. Therefore, it was not included in this experiment. The experiments covered two scenarios. One contained the phase discontinuities caused by strong ablation in glacier tongues, and the other contained the phase discontinuities caused by glacier surge.

1) *Unwrapping the TanDEM-X Differential Interferogram in the Queer Mountain:* In this experiment, a TanDEM-X bistatic image pair acquired on 14 January 2020 in the Queer Mountain was used to generate an interferogram. The differential interferogram was obtained by differencing the original interferometric phase with the topographic phase simulated from SRTM DEM, as shown in Fig. 9(a). Glaciers are well developed in the Queer Mountain. During 2000–2020, some glaciers' terminus experienced strong ablation. The significant change in glacier surface elevation caused by the strong ablation resulted in phase discontinuities in the differential interferogram. In addition, due to the rough topography, there was spatial decorrelation surrounding the glaciers, as shown in Fig. 9(c). These two factors make it difficult to unwrap the differential interferogram.

The differential interferogram was unwrapped by each PU method (see Fig. 10). In the experimental area, only the surface elevation of the glacierized area changes. The surface elevation difference in the ice-free regions should be close to zero. Hence, the unwrapped phase in the ice-free regions should be also close to zero. In this perspective, MCF, PUMA, and our method had comparable good performance. The unwrapped phase was

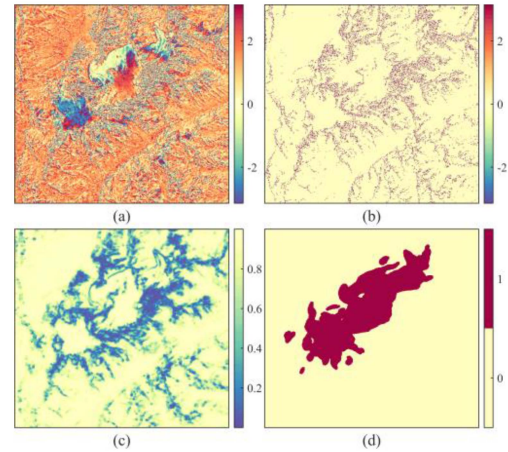


Fig. 9. Experimental data in the Queer Mountain. (a) TanDEM-X differential interferogram. (b) Residue map. (c) Coherence map. (d) Glacier label (0: ice-free area, 1: glacierized area).

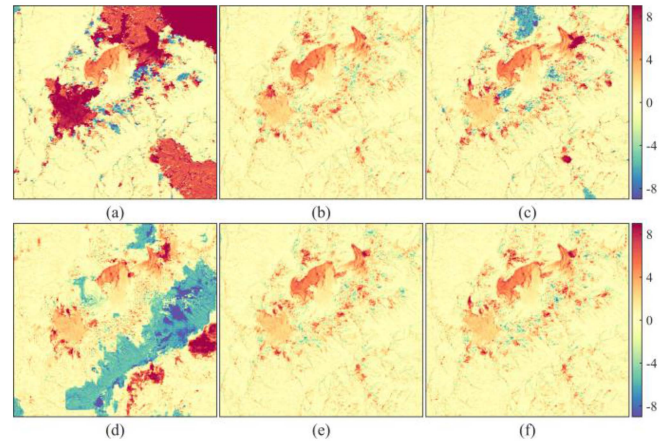


Fig. 10. PU results of different methods for the TanDEM-X interferogram in the Queer Mountain. (a) QGPU. (b) MCF. (c) PGNet-PU. (d) PU-GAN. (e) PUMA. (f) Our method.

converted into the surface elevation difference for further comparison, as shown in Fig. 11(a)–(f). The absolute phase of the real interferogram is unknown, and therefore, the true surface elevation change over glacierized areas is also unknown. In theory, if an optical stereo image pair acquired at a similar time as the TanDEM-X image pair was used to generate a DEM, the difference between this DEM and the SRTM DEM should be close to that converted from the unwrapped differential phase. The archived optical stereo image pair in this experiment area with the closest acquisition time to the TanDEM-X image and the least clouds and shadows in the glacier area is the ASTER stereo pair acquired on 8 April 2020. We used it to generate a DEM, and then differenced this DEM with SRTM DEM. The resulting surface elevation difference was used as a reference, as shown in Fig. 11(g). The accuracy of the unwrapping methods was evaluated by comparing the corresponding surface elevation difference with the reference.

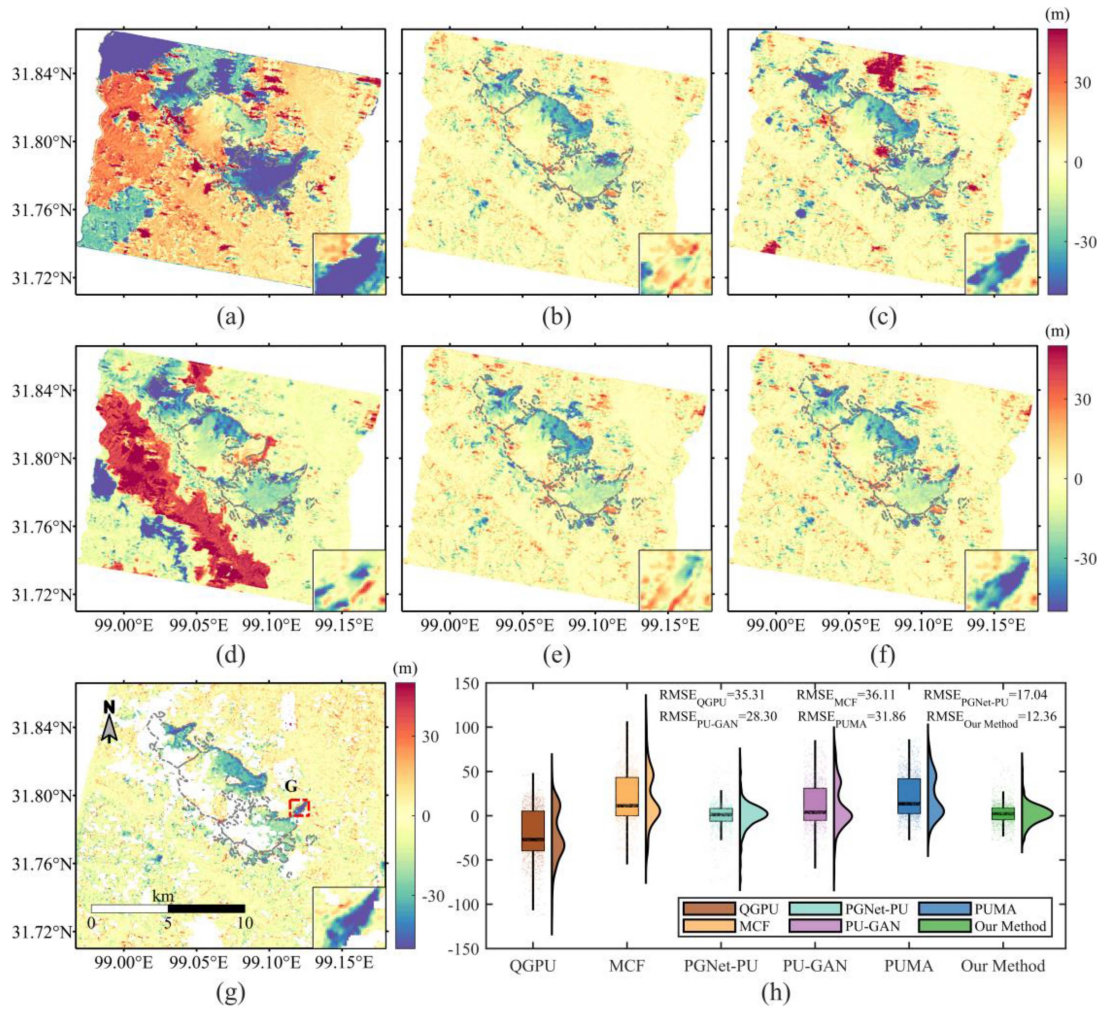


Fig. 11. (a)–(f) Surface elevation difference maps in the Queer Mountain based on different PU methods. (a)–(f) correspond to QGPU, MCF, PGNet-PU, PU-GAN, PUMA, and our method, respectively. The bottom-right subplot in each panel is a zoomed view of the G-region. (g) Surface elevation change obtained by differencing the generated ASTER DEM and the SRTM DEM. The dashed line indicates glacier boundaries. (h) Error statistics of the elevation difference maps in the G-region by using (g) as the reference.

As shown in Fig. 11, the distinction between the elevation difference map based on our method and the reference was the smallest, indicating our method is the most effective. The results based on MCF and PUMA were generally good. However, there were phase jumps in a glacier tongue with strong ablation (G-region). PGNet-PU obtained the correct unwrapped phase in the G-region but obtained the wrong one in the low-coherence region. The results based on QGPU and PU-GAN were poor, both characterized by large errors. For QGPU, noise was the main cause of the unwrapping error. For PU-GAN, it may be due to the weak learning ability of the generator, which makes it difficult to adapt to complex scenarios.

2) *Unwrapping the TanDEM-X Differential Interferogram in the Eastern Pamir Plateau:* Glacier surges are common in the Eastern Pamir Plateau. In this experiment, a TanDEM-X bistatic image pair acquired on 6 June 2011 in the eastern Pamir Plateau was used to generate an interferogram. Likewise, the differential interferogram was obtained by differencing the original interferometric phase with the topographic phase simulated from SRTM DEM, as shown in Fig. 12(a). The topography is also rough and

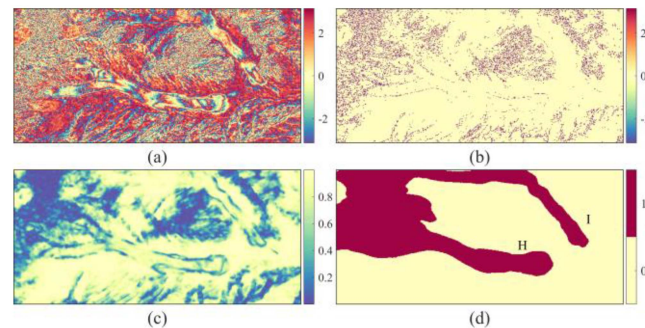


Fig. 12. Experimental data in the Eastern Pamir Plateau. (a) TanDEM-X differential interferogram. (b) Residue map. (c) Coherence map. (d) Glacier label (0: ice-free area, and 1: glacierized area).

the interferometric phase surrounding the glaciers is affected by spatial decorrelation, as shown in Fig. 12(c). Two glaciers (H and I) in this experimental area surged during 2000–2011, resulting in dramatic surface elevation changes in the glacier tongues. The significant change in glacier surface elevation

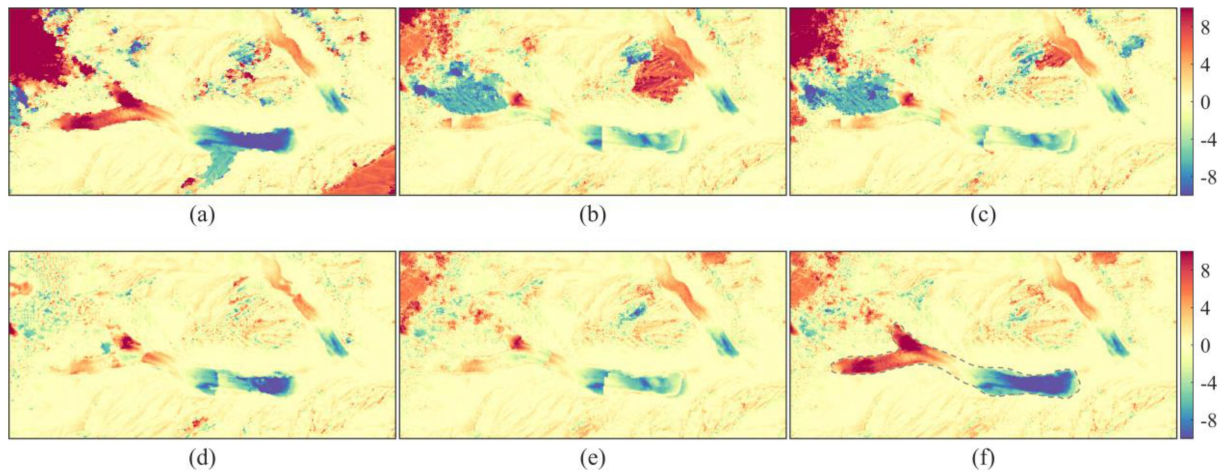


Fig. 13. PU results of different methods for the TanDEM-X interferogram in the Eastern Pamir Plateau. (a) QGPU. (b) MCF. (c) PGNet-PU. (d) PU-GAN. (e) PUMA. (f) Our method.

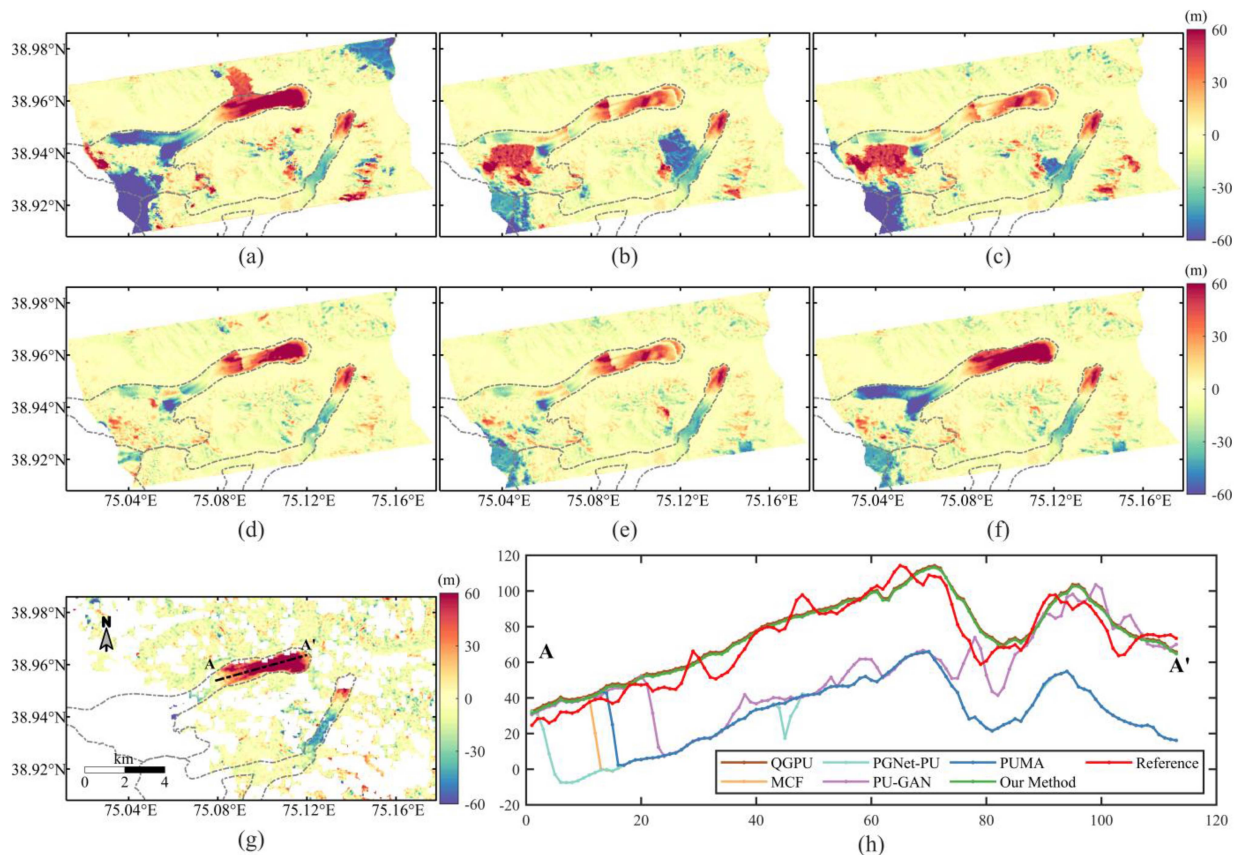


Fig. 14. (a)–(f) Surface elevation difference maps in the Eastern Pamir Plateau. (a)–(f) correspond to QGPU, MCF, PGNet-PU, PU-GAN, PUMA, and our method, respectively. (g) Surface elevation change obtained by differencing the generated ASTER DEM and the SRTM DEM. The dashed line indicates glacier boundaries. (h) Plot of the elevation difference values based on different PU methods and the reference elevation difference values along profile A-A'.

caused by the surge activities resulted in phase discontinuities in the differential interferogram. Hence, unwrap this differential interferogram is also very challenging.

Fig. 13 shows the PU results of different methods. In the upper left of the interferogram, sharp slope caused serious phase aliasing. All methods cannot obtain the correct unwrapped phase

here. In the midstream and downstream parts of glacier H, the area shown by the gray dashed line in Fig. 13(f), the PU results of MCF, PGNet-PU, PU-GAN, and PUMA have obvious phase jumps. The PU result of QGPU was smooth in this area but had obvious jumps below the tongue of glacier H, as well as the lower right of the interferogram. The results of MCF and

PGNet-PU also showed significant jumps in the low-coherence regions. PGNet-PU lost reliability in this experiment, probably due to its shallow network structure.

The accuracy of the PU results of each method was evaluated using the same approach as the previous experiment. The ASTER optical stereo image pair acquired on 10 September 2011 was selected to generate a DEM. This DEM was then differenced with the SRTM DEM. The obtained surface elevation difference was used as the reference, as shown in Fig. 14(g).

From Fig. 14, we can see that the elevation difference corresponding to our method is closest to the reference. For further comparison, a profile (A-A') was extracted in the tongue of glacier H, as shown in Fig. 14(h). Along this profile, the elevation difference obtained from our method and QGPU was almost identical, and tallied with the reference values, indicating that the PU results of these two methods are reliable in the ice tongue. By contrast, the values corresponding to other methods had obvious deviations from the reference values.

The direct mapping method (e.g., PU-GAN) is simple and easy to implement. However, our experiments showed that establishing the mapping relationship between input and output is unstable. No matter how the model is designed, it is heavily dependent on the training data, and its application scene is very limited. The prediction based on this method is completely unreliable if the training data do not match the actual scenario. Meanwhile, the predicted phase is too smooth and inconsistent with the fringes of the interferogram after rewrapping.

The phase-gradient information prediction method is a two-step method. If the predicted phase-gradient information is the wrap-count gradient (e.g., PGNet), it becomes a segmentation of the fringes in the interferograms. This type of method [21], [22], [23], [24], [25] assumes that the wrap-count gradient classes are confined within a range [26]. However, the real wrap-count gradient classes may exceed the range. In addition, such methods generally estimate the wrap counts through the norm-based solution, which tends to propagate the incorrectly predicted information throughout the map. If the predicted phase-gradient information is phase discontinuity, there are only two classes in the segmentation results: smooth phase and discontinuous phase. The phase discontinuity information is used as an a priori condition for statistical model-based PU methods. This type of method does not require any additional artificial conditions, and the high robustness of the wrap-count estimation algorithm ensures the accuracy of PU.

V. CONCLUSION

In this article, we proposed a new PU method for the SAR interferometric phase of large gradient. The deep-learning model LEU-Net was used to predict the phase discontinuity, and the predicted phase discontinuity information was used as a priori condition for the PUMA algorithm to obtain the unwrapped phase. Our method and the method in the literature [26] belong to the phase discontinuity prediction method. In semantic segmentation tasks, downsampling is generally used to increase the receptive field and reduce the computational complexity.

In order to preserve local information, DENet has not introduced downsampling and embedded the ASPP module in the network to increase the receptive field. Given the sparse and discrete phase discontinuities, this idea is feasible, but DENet's requirement of computational resources is too high. It is difficult for regular GPUs to afford the huge amounts of calculations. We, therefore, designed the ICM and CFM accordingly. The ICM aimed to minimize the loss of detailed information during the downsampling process. The CFM was designed to extract the channel features in the low-level features to enhance the detail information to suppress the noise and to fuse the multilevel features. These two modules guaranteed the performance of the model and reduced the computational effort for training. In addition, the datasets simulated based on real DEM differences were closer to the real differential SAR interferograms. These high-quality samples prevented the model from falling into the trap of artificially defined data characteristics and improved the performance of the model. If the proportion of phase discontinuity in the samples is sufficient, the class imbalance is alleviated and the loss function does not require special design. Experiments based on simulated and real data showed that the phase discontinuity information predicted by LEU-Net played an important role in guiding PUMA PU. LEU-Net can predict the phase discontinuity among nine adjacent pixels and in the four directions of the intermediate pixel. PUMA can adjust the clique settings and select the appropriate clique potential to improve the PU results.

The performance of the deep-learning model depends on the quality of the data. However, it is difficult to obtain the ground truth from real InSAR data. To some extent, the simulated datasets contain artificial data characteristics. This will greatly affect the performance of deep-learning models. As mentioned in [46], reinforcement learning may be a key method to improve PU performance. At the same time, self-supervised learning has been continuously developed in recent years. In the field of remote sensing, methods based on self-supervised segmentation have been applied [47], [48]. Self-supervised learning may effectively solve the problem of difficulty in obtaining ground truth. Therefore, the self-supervised PU method is worthy of studying too.

ACKNOWLEDGMENT

The authors would like to thank ESA for providing COP30 DEMs (<https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model>), DLR for providing TanDEM-X CoSSC images (<https://eoweb.dlr.de/egp/>), and LP DAAC for providing AST_L1A (https://lpdaac.usgs.gov/products/ast_11av003/) and 1 arc-second C-band SRTM DEMs (<https://e4ftl01.cr.usgs.gov/MEASURES/SRTMSWBD.003/>).

REFERENCES

- [1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.
- [2] K. Itoh, "Analysis of the phase unwrapping algorithm," *Appl. Opt.*, vol. 21, no. 14, pp. 2470–2470, 1982.

- [3] J. Li et al., "Early 21st century glacier thickness changes in the Central Tien Shan," *Remote Sens. Environ.*, vol. 192, pp. 12–29, Apr. 2017.
- [4] L. Zhou, D. Chai, Y. Xia, P. Ma, and H. Lin, "Interferometric synthetic aperture radar phase unwrapping based on sparse Markov random fields by graph cuts," *J. Appl. Remote Sens.*, vol. 12, no. 1, Jan. 2018, Art. no. 015006.
- [5] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Sci.*, vol. 23, no. 4, pp. 713–720, 1988.
- [6] T. J. Flynn, "Two-dimensional phase unwrapping with minimum weighted discontinuity," *J. Opt. Soc. Amer. A*, vol. 14, no. 10, pp. 2692–2701, 1997.
- [7] T. J. Flynn, "Consistent 2-D phase unwrapping guided by a quality map," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 1996, vol. 4, pp. 2057–2059.
- [8] M. Costantini, "A novel phase unwrapping method based on network programming," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 813–821, May 1998.
- [9] D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," *J. Opt. Soc. Amer.*, vol. 67, no. 3, pp. 370–375, 1977.
- [10] C. W. Chen and H. A. Zebker, "Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization," *J. Opt. Soc. Amer. A*, vol. 18, no. 2, pp. 338–351, 2001.
- [11] J. M. Bioucas-Dias and G. Valadao, "Phase unwrapping via graph cuts," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, 2005, pp. 360–367.
- [12] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc., B*, vol. 48, no. 3, pp. 259–302, 1986.
- [13] J. S. Yedidia, W. Freeman, and Y. Weiss, "Generalized belief propagation," in *Proc. Adv. Neur. Inf. Process. Syst.*, 2000, vol. 13, pp. 1–7.
- [14] M. Wainwright, T. Jaakkola, and A. Willsky, "Tree consistency and bounds on the performance of the max-product algorithm and its generalizations," *Statist. Comput.*, vol. 14, pp. 143–166, 2004.
- [15] Z. Wu, H. Zhang, Y. Wang, T. Wang, and R. Wang, "A deep learning based method for local subsidence detection and InSAR phase unwrapping: Application to mining deformation monitoring," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 20–23.
- [16] K. Wang, Y. Li, Q. Kema, J. Di, and J. Zhao, "One-step robust deep learning phase unwrapping," *Opt. Exp.*, vol. 27, no. 10, pp. 15100–15115, 2019.
- [17] Z. Wu, T. Wang, Y. Wang, R. Wang, and D. Ge, "Deep learning for the detection and phase unwrapping of mining-induced deformation in large-scale interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5216318, doi: [10.1109/tgrs.2021.3121907](https://doi.org/10.1109/tgrs.2021.3121907).
- [18] L. Zhou, H. Yu, V. Pascasio, and M. Xing, "PU-GAN: A one-step 2-D InSAR phase unwrapping based on conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5221510.
- [19] G. Spoorthi, S. Gorthi, and R. K. S. S. Gorthi, "PhaseNet: A deep convolutional neural network for two-dimensional phase unwrapping," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 54–58, Jan. 2018.
- [20] G. E. Spoorthi, R. K. Sai Subrahmanyam Gorthi, and S. Gorthi, "PhaseNet 2.0: Phase unwrapping of noisy data based on deep learning approach," *IEEE Trans. Image Process.*, vol. 29, pp. 4862–4872, Mar. 2020, doi: [10.1109/tip.2020.2977213](https://doi.org/10.1109/tip.2020.2977213).
- [21] L. Zhou, H. Yu, and Y. Lan, "Deep convolutional neural network-based robust phase gradient estimation for two-dimensional phase unwrapping using SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4653–4665, Jul. 2020, doi: [10.1109/tgrs.2020.2965918](https://doi.org/10.1109/tgrs.2020.2965918).
- [22] Y. Gao et al., "Two-dimensional phase unwrapping method using a refined D-LinkNet-based unscented Kalman filter," *Opt. Lasers Eng.*, vol. 152, May 2022, Art. no. 106948, doi: [10.1016/j.optlaseng.2022.106948](https://doi.org/10.1016/j.optlaseng.2022.106948).
- [23] L. Li, H. Zhang, Y. Tang, C. Wang, and F. Gu, "InSAR phase unwrapping by deep learning based on gradient information fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4502305.
- [24] F. Sica, F. Calvanese, G. Scarpa, and P. Rizzoli, "A CNN-based coherence-driven approach for InSAR phase unwrapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4003705, doi: [10.1109/lgrs.2020.3029565](https://doi.org/10.1109/lgrs.2020.3029565).
- [25] H. Wang, J. Hu, H. Fu, C. Wang, and Z. Wang, "A novel quality-guided two-dimensional InSAR phase unwrapping method via GAUNet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7840–7856, Jul. 2021, doi: [10.1109/jstars.2021.3099485](https://doi.org/10.1109/jstars.2021.3099485).
- [26] Z. Wu, T. Wang, Y. Wang, R. Wang, and D. Ge, "Deep-learning-based phase discontinuity prediction for 2-D phase unwrapping of SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5216516, doi: [10.1109/tgrs.2021.3121906](https://doi.org/10.1109/tgrs.2021.3121906).
- [27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
- [28] N. Syakrani, T. L. Mengko, A. B. Suksmono, and E. T. Baskoro, "Comparison of PUMA and CUNWRAP to 2-D phase unwrapping," in *Proc. Int. Conf. Elect. Eng. Inform.*, 2011, pp. 1–6.
- [29] H. A. Zebker and K. Chen, "Accurate estimation of correlation in InSAR observations," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 124–127, Apr. 2005.
- [30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [33] A. Stergiou, R. Poppe, and G. Kalliatakis, "Refining activation downs with SoftPool," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10357–10366.
- [34] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," 2016, *arXiv:1606.08415*.
- [35] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, vol. 32, pp. 8026–8037.
- [36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [37] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, 2019, Art. no. 27.
- [38] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. IEEE 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.
- [39] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, pp. 19–67, 2005.
- [40] O. Oktay et al., "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [41] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, 2018, pp. 3–11.
- [42] J. Chen et al., "Transunet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.
- [43] Y. Zhang, H. Liu, and Q. Hu, "Transfuse: Fusing transformers and CNNs for medical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2021, pp. 14–24.
- [44] H. Cao et al., "Swin-Unet: Unet-like pure transformer for medical image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2023, pp. 205–218.
- [45] U. Wegmuller, "Gamma SAR processor and interferometry software," in *Proc. 3rd ERS Sci. Symp.*, 1997.
- [46] L. Zhou, H. Yu, Y. Lan, and M. Xing, "Artificial intelligence in interferometric synthetic aperture radar phase unwrapping: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 10–28, Jun. 2021, doi: [10.1109/mgrs.2021.3065811](https://doi.org/10.1109/mgrs.2021.3065811).
- [47] H. Li et al., "Global and local contrastive self-supervised learning for semantic segmentation of HR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5618014.
- [48] A. Yu, B. Liu, X. Cao, C. Qiu, W. Guo, and Y. Quan, "Pixel-level self-supervised learning for semi-supervised building extraction from remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Sep. 2022, Art. no. 2507105.



Wenjie Zhong was born in Shaoyang, Hunan, China, in 2000. He received the B.S. degree in geomatics engineering in 2021 from Central South University, Changsha, China, where he is currently working toward the postgraduate degree in geomatics engineering.

His research interests include interferometric synthetic aperture radar data processing and glacier hazard monitoring.



Jia Li received the bachelor's degree in surveying and mapping engineering, the academic master's degree in geodesy and surveying engineering, and the academic doctorate degree in geodesy and surveying engineering from Central South University, Changsha, China, in 2009, 2012, and 2016, respectively.

He was an exchange visitor with Hong Kong Polytechnic University from 2013 to 2015. He is an Associate Professor with the School of Geosciences and Info-Physics, Central South University. His research interests include glacier remote sensing monitoring, geological disaster remote sensing monitoring, InSAR large-scale DOM and DEM production, and InSAR basic theoretical research.



Xin Li (Senior Member, IEEE) graduated from the Department of Geosciences, Nanjing University, Nanjing, China, in 1992, and the Ph.D. degree from the Lanzhou Institute of Glaciation and Permafrost, Chinese Academy of Sciences, Beijing, China, in 1998.

He has made innovative scientific research achievements in river basin integration research, land surface data assimilation, cryosphere remote sensing, and information system research, including leading the implementation of the "Heihe Remote Sensing Experiment" from 2007 to 2017, which was evaluated by international peers as "world-class observation," "China's richest hydrological remote sensing experiment," "developed my country's large-scale land surface data assimilation system and high-resolution basin-scale land surface hydrological data assimilation system, in terms of nonlinear filtering algorithm, multisource remote sensing data assimilation, etc. Achieved leading results. He has authored or coauthored more than 500 academic papers (330+ SCI citations), with a total of 26 000+ citations (14 000+ SCI citations).



Juanjuan Feng was born in Xiangyang, Hubei, China, in 1998. She received the bachelor's degree in geomatics engineering from the East China University of Technology, Nanchang, China, in 2021. She is currently working toward the postgraduate degree in geomatics engineering with Central South University, Changsha, China.

Her research interest is to use deep learning models to predict Arctic sea ice.



Lei Guo received the B.S. degree in geomatics engineering from Chongqing Jiaotong University, Chongqing, China, in 2017, and the M.S. degree in surveying and mapping in 2020 from Central South University, Changsha, China, where he is currently working toward the Ph.D. degree in surveying and mapping.

His main research interests include analyzing the dynamics of glaciers with multiple remote sensing techniques, with a special focus on glacier hazard monitoring.



Zhiqiang Li was born in Yushu, Jilin, China, in 1997. He received the B.S. degree in geology from Lanzhou University, Lanzhou, China, in 2019. He is currently working toward the postgraduate degree in geomatics engineering with Central South University, Changsha, China.

His research interest is glacier thickness estimation.



Junhui Wu was born in Huanggang, Hubei, China, in 2001. He received the B.S. degree in geomatics engineering from the China University of Mining and Technology, Xuzhou, China, in 2022. He is currently working toward the postgraduate degree in geomatics engineering with Central South University, Changsha, China.

His research interests include remote sensing monitoring of glacier instability and simulation of disaster scenarios.



Lingshuai Kong was born in Tianjin, China, in 2000. He received the B.S. degree in geomatics engineering from Hohai University, Nanjing, China, in 2023. He is currently working toward the postgraduate degree in geomatics engineering with Central South University, Changsha, China.

His research interests include developing algorithms for mountain glacier detection and the construction of platforms.