A Wrap-count-based Phase Unwrapping Method for Large-scale, Low Coherence Interferograms Using Deep Learning

Kun Jiang, Wenbin Xu, Andrew J Hooper, *Senior Member, IEEE*, and Lei Xie

***Abstract*—Unwrapping synthetic aperture radar interferograms with extensive low-coherence regions remains challenging, even when the deformation signal has a low gradient, such as that due to interseismic displacement. Here, we present a novel algorithm to unwrap low-gradient interferograms more efficiently and reliably using a semantic segmentation neural network. We first partition large-scale interferograms into overlapping patches and employ a trained Segformer network, making full use of spatial features, to identify decorrelated pixels while predicting the wrap count for coherent pixels. In a further step, we correct wrap counts or re-unwrap certain patches, based on a reliability metric. Finally, the patches are mosaicked to reconstruct the fully unwrapped interferogram by leveraging overlapping areas. The Segformer model is trained on over 20,000 simulated samples with varying decorrelation noise and more than 10,000 real-world samples from the COMET-LiCSAR portal. Synthetic experiments show that our approach significantly reduces the mean absolute error compared to the classical minimum cost flow method. Further validation on thousands of interferograms from the western Altyn Tagh Fault and the western Haiyuan Fault confirms its superior performance in isolated regions separated by decorrelated noise and its ability to suppress the wide-ranging propagation of unwrapping errors, with the proportion of unwrapping errors reduced by 28-83% for real interferograms. These results highlight the potential of our method to be used in large-scale automated processing of straining regions, such as the Alpine-Himalayan Belt.**

***Index Terms*—Deep learning, interferometric synthetic aperture radar (InSAR), large-scale interseismic deformation, phase unwrapping, wrap count.**

I. Introduction

INTERFEROMETRIC synthetic aperture radar (InSAR) has been widely used for observing large-scale surface deformation associated with tectonics, such as interseismic deformation [1, 2, 3, 4, 5]. Phase unwrapping is a crucial step in InSAR data processing to retrieve relative deformation from the interferometric phase, which is inherently wrapped within the range of $-\pi$ to π rad. However, large-scale interferograms related to tectonic studies often contain significant decorrelation due to factors such as surface cover changes (e.g. vegetation growth, snow accumulation and melt, dune migration) and long temporal baselines, posing substantial challenges for accurately extracting slow tectonic deformation [6, 7, 8].

Conventional phase unwrapping is based on the phase continuity assumption , which states that the absolute difference between unwrapped phases of adjacent pixels should generally not exceed π [9, 10]. In other words, the unwrapped gradients are generally expected to be equal to the wrapped gradients, which is reasonable under ideal conditions without severe decorrelation, geometric distortion, or steep deformation gradients. Typically, when this assumption is violated, phase residues—defined as inconsistencies in a square loop of four pixels—arise. Most traditional phase unwrapping methods can be formulated within a generalized L^p -norm optimization framework, where different objective functions are employed to minimize the discrepancy between the wrapped gradients and the unwrapped gradients. This is equivalent to connecting residues with branch cuts, across which the unwrapped phase gradients differ from the wrapped phase gradients by an integer multiple of 2π rad. The branch-cut algorithm is a representative L^0-norm approach [11], which minimizes the total length of branch cuts. The minimum cost flow (MCF) method follows an L^1 -norm formulation, which minimizes the total absolute deviation between wrapped and unwrapped gradients along the branch cuts [12]. The SNAPHU method, however, differing from the L^p -norm methods, treats phase unwrapping as a maximum a posterior probability estimation problem. It constructs statistical, nonlinear cost functions for all gradients using predetermined, shape-adaptive probability distributions [13, 14]. These classic optimization-based methods possess strong resistance to noise, as they aim to find the globally

This work was supported by Natural Science Foundation of Hunan Province (2024JJ3031), National Natural Science Foundation of China (42174023, 42304037), China Scholarship Council (CSC) scholarship (202406370151). This work was also supported by the NERC Centre for the Observation and Modeling of Earthquakes, Volcanoes and Tectonics (COMET) and the EU Horizon 2020 project DEEPVOLC (Grant 866085). (*Corresponding author: Wenbin Xu.*)

Kun Jiang is with the School of Geosciences and Info-Physics, Central South University, Changsha 410083, China, and also with the COMET Institute, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, U.K. (e-mail: kunjiang@csu.edu.cn).

Wenbin Xu is with the School of Geosciences and Info-Physics, Central South University, Changsha 410083, China (e-mail: wenbin.xu@csu.edu.cn).

Andrew J Hooper is with the COMET Institute, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, U.K. (e-mail: a.hooper@leeds.ac.uk).

Lei Xie is with the School of Geosciences and Info-Physics, Central South University, Changsha 410083, China (e-mail: leixie_geo@csu.edu.cn).

optimal solutions of the objective functions. However, the objective functions are built upon arcs which rely on the connectivity between pixels [15]. When substantial regions of decorrelated pixels are present, such as in large-scale interseismic interferograms, these methods struggle to mitigate their impact. The sparse MCF method attempts to address this by applying a coherence threshold to mask low-coherence pixels; however, selecting an appropriate threshold remains a difficult task, and the triangulated arcs used for phase integration may still violate the phase continuity assumption, limiting its effectiveness. For phase unwrapping of multi-temporal interferograms, incorporating temporal constraints offers an enhanced approach [16]. For instance, Hussain *et al.* [17] proposed an iterative weighted unwrapping method based on traditional SNAPHU to progressively reduce large-area unwrapping errors by assigning spatial unwrapping weight according to temporal loop closure errors from previously unwrapped results.

Apart from improving phase unwrapping algorithms, detecting and correcting phase unwrapping errors is another useful strategy for obtaining high-quality tectonic velocity maps. Well-known open-source InSAR time-series processing softwares, such as LiCSBAS and Mintpy, integrate such steps [18, 19]. These error correction techniques can be categorized into temporal and spatial approaches. The temporal approach leverages phase closure and additional regularization constraints to estimate unwrapping errors at the pixel or sub-block level [19, 20, 21], while the spatial approach primarily employs flux analysis to extract error boundaries and segment erroneous regions for correction [22]. However, temporal correction requires plenty of redundant connections across SAR acquisitions, and spatial correction demands connectivity of coherent regions. Both criteria are hard to satisfy in large-scale multi-temporal interferograms with extensive decorrelation, resulting in limited correction performance.

In recent years, many deep learning approaches have been developed for 2-D phase unwrapping, which exploit deep learning's ability in image feature extraction. Unlike traditional methods that strictly adhere to the fixed phase continuity assumption, deep learning-based phase unwrapping is data-driven, learning the mapping between interferograms and key unwrapping features from large datasets to establish a network model based on empirical knowledge [23]. Consequently, it generally outperforms conventional methods in specific scenarios.

Existing deep learning-based approaches can be summarized into one-step and two-step methods. One-step methods utilize deep neural networks to output unwrapped phases all at once, enabling end-to-end processing with high efficiency [24, 25]. However, ensuring complete congruence of fringes between the input and output is theoretically unfeasible, despite efforts to improve it through advanced techniques such as generative adversarial networks (GANs) [26] and by incorporating phase unwrapping knowledge derived from classic methods [27]. The incongruence will introduce unwrapping errors and prevent the phase from completing closed loops in time, strongly affecting the subsequent time series analysis, as well as the final accuracy and uncertainty of slight tectonic deformation.

Two-step approaches, on the other hand, employ neural networks to first predict critical information required for traditional phase unwrapping, such as branch cuts, phase discontinuities, wrap count gradients, and wrap counts, followed by conventional unwrapping algorithms to ensure congruence. The branch cut prediction method uses a convolutional neural network (CNN) to identify branch cuts from residues, hugely reducing the computational time for setting cuts [28]. The phase discontinuity prediction method assigns weights to integer-valued difference parameters to be solved using MCF, based on network-predicted discontinuity probabilities, where a higher probability corresponds to a larger difference solution [29]. The wrap-count gradient method typically substitutes the wrapped gradients obtained by the phase continuity assumption in the MCF objective function with integer-valued wrap-count gradients predicted by neural networks. The L^1 norm is then used to obtain the final unwrapped results [30, 31]. Wrap count methods have been primarily applied to optical interferograms, where fringe patterns are simpler, and noise levels are lower. These methods straightforwardly segment the wrap count for each pixel and add it to the wrapped phase to obtain the unwrapped phase [32, 33]. This kind of approach is conceptually simple and does not rely on traditional path integration or global optimization methods. Compared to other two-step methods, it is more promising for unwrapping low-coherence interferograms as it does not depend on pixel connectivity. However, it often requires post-processing, as the accuracy of phase unwrapping is directly influenced by the reliability of neural network predictions [34], which prevents its application in interferograms produced by SAR satellites.

In this paper, we describe a novel deep learning-based 2-D phase unwrapping method using wrap-count segmentation, which targets large-scale interseismic interferograms, typically containing slow deformation and extensive decorrelation. Unlike existing deep learning-based approaches, our proposed method has four advantages as follows: 1) the deep neural network directly segments wrap counts as a semantic segmentation task, leveraging both global and local context information without being constrained by the connectivity between pixels. This is particularly beneficial for isolated regions commonly found in large-scale, low coherence interferograms; 2) the neural network is designed to automatically identify and mask low-coherence regions, reducing the impact of decorrelated noise while improving processing efficiency; 3) a partitioning-mosaicking strategy is introduced to enable large-scale interferograms unwrapping with limited graphics processing unit (GPU) memory and also effectively suppress the wide-ranging propagation of unwrapping errors; 4) a reliability metric based on predicted probabilities is defined to determine the appropriate post-processing approach for patches, enhancing the robustness of the patch-wise unwrapped results. Our new method, which we

refer to as SWIPHU, (Segmented Wrap-count Interferogram PHase Unwrapping) includes three main steps. First, a Segformer-based semantic segmentation network predicts the wrap count of each pixel in multiple patches of the large-scale interferogram. The segmented wrap count is classified into seven categories (1–6 for wrap counts and 0 for decorrelation). Second, based on a customized reliability metric derived from the network output, post-processing is applied to patches. When the reliability metric is low, the traditional MCF method is employed for re-unwrapping, whereas for moderate reliability, an additional deep convolutional neural network (DCNN) corrects wrap-count errors. Finally, the post-processed wrap-count maps are mosaicked to reconstruct the fully unwrapped map. The Segformer model is trained using a combination of simulated and real samples to increase dataset diversity. A large number of simulated and real large-scale interferograms has verified the effectiveness of our deep learning approach. Results show that it is capable of unwrapping isolated regions with higher accuracy and effectively mitigates the propagation of unwrapping errors in large-scale interferograms.

The remainder of this article is organized as follows. Section II elaborates on the principles and general workflow of the SWIPHU method, including the architecture of Segformer, the dataset generation for wrap-count segmentation, patch post-processing, and the mosaicking strategy. The performance of the SWIPHU method on patch-wise wrap-count prediction and simulated interferograms is shown in Section III. Section IV assesses our deep learning method on real interferograms from the western Altyn Tagh Fault (ATF) and the western Haiyuan Fault (HF). Section V discusses the effectiveness of the customized reliability metric and the mosaicking approach, as well as SWIPHU's potential application in volcanic scenarios. Finally, Section VI provides the conclusions.

II. Methodology

The essence of phase unwrapping can be expressed as:

$$\Psi(s) = \varphi(s) + 2k(s)\pi \tag{1}$$

where s denotes the spatial index of a pixel in the interferogram, $\Psi(s)$ and $\varphi(s)$ represent the unwrapped and wrapped phase, respectively, and $k(s)$ is an integer denoting the unknown wrap count. Traditional methods try to solve this ill-posed problem using the wrapped gradients $\hat{\Delta}_k(s,\ s-1)$ calculated by the phase continuity assumption. In contrast, deep learning-based wrap-count methods directly predict $k(s)$. Given that the size of large-scale interferograms often reaches thousands by thousands of pixels, training and predicting the entire interferogram at once poses a significant burden on limited GPU memory. To address this, we introduce a partitioning-mosaicking strategy, dividing the whole workflow into three parts: wrap-count segmentation, patch post-processing, and patch mosaicking, as shown in Fig. 1.

A. Wrap-count Segmentation

Semantic segmentation, a popular task in computer vision, classifies each pixel for pixel-level understanding and is widely used in remote sensing applications, such as land cover classification [35, 36]. Classic CNN-based architectures, including FCN, U-Net, PSPNet, and DeepLabV3+, have been extensively used for segmentation. With the advancement of deep learning, recent transformer-based architectures outperform CNNs in capturing long-distance dependencies and global context [37, 38, 39], which is particularly beneficial for wrap-count estimation in low-coherence interferograms. Compared with other advanced backbones such as Swin Transformer and ConvNeXt, we adopt SegFormer because, as an end-to-end model designed for semantic segmentation, it exhibits an excellent trade-off between the use of self-attention mechanisms and high computational efficiency. In contrast, ConvNeXt is a purely convolutional network that is less effective at capturing global features, whereas Swin Transformer imposes a heavy computational burden for our large-scale phase-unwrapping task. Segformer is available in seven versions (B0 to B6), with increasing model complexity [40]. We select Segformer-B2 to balance model capacity and computational efficiency for patch-wise wrap-count estimation.

The overall architecture of Segformer is shown in Fig.2(a), consisting of two main components: an encoder and a decoder. The encoder is constructed from four hierarchical Transformer Blocks, each preceded by an Overlap Patching Embedding (OPE) module. These OPE modules progressively reduce spatial dimensions while increasing feature channels, enabling effective multi-scale feature extraction similar to CNNs. The decoder is primarily composed of a pure Multi-Layer Perceptrons (MLPs) architecture, which avoids convolutional operations in the decoding stage compared with commonly used hybrid decoders that integrate CNNs with MLP layers, leading to a lightweight design with reduced computational cost [40]. . The input size for Segformer is expressed as $H \times W \times 3$, where 3 represents RGB three color channels. H and W are set to 512 to facilitate batch training on our GPU. We define seven segmentation classes (N_{cls} = 7): 0 for decorrelated pixels and 1–6 representing different wrap-count values within the patch. Given that slow tectonic deformation typically manifests as sparse fringe patterns, we set the maximum wrap count within a patch to 6. Table I provides the layer configurations and key parameters of Segformer-B2, along with the corresponding feature map sizes. The following section details the structure of its primary modules.

1): Overlap Patching Embeddings: Unlike commonly used transformer models in computer vision (e.g., ViT), Segformer avoids positional encoding when converting an image into sequence. It employs convolution operations to divide the input into multiple overlapping patches. This design enhances robustness to varying input resolutions while preserving spatial

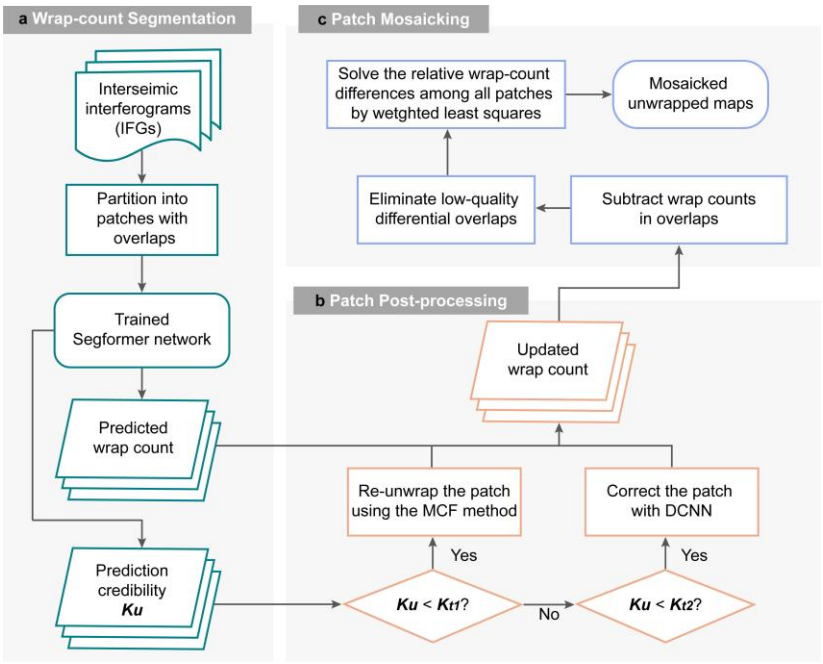

Fig. 1. Flowchart of the SWIPHU proposed method for phase unwrapping.

continuity. As shown in Table I, the OPE module preceding Transformer Block 1 utilizes a larger kernel size and stride, effectively embedding initial patches.

2): Transformer Block: Each transformer block consists of N_i Efficient Self-Attention (Efficient Self-Attn) and Mix-Feedforward Network (Mix-FFN) modules, where i represents the index of the transformer block. To reduce the computational complexity, Segformer modifies the self-attention mechanism by applying a scaling factor R_i. Assume that the Query (Q), Key (K), and Value (V) matrices have the same initial size of $T_i \times C_i/H_i$, where H_i is the number of attention heads, T_i is the sequence length (equal to the product of the height and width of the input feature map), and C_i is the number of channels. As shown in Fig. 2(b), a convolution layer with kernel size of R_i is applied to downsample the input, reducing the size of Q and V matrices to $T_i/R_i \times C_i/H_i$. The self-attention sequence is then computed as follows:

$$\text{Attention}\,(Q,K,V) = \text{Softmax}\,(\frac{QK^T}{\sqrt{C_i/H_i}})V. \qquad (2)$$

The output x_{ESA} from Efficient Self-Attn is then passed through the Mix-FFN module, which combines MLPs with an expansion ratio E_i and a 3 × 3 depth-wise convolution. This process is briefly illustrated in Fig. 2(b) and can be mathematically expressed as:

$$x_{MF} = Linear(C_i \cdot E_i, C_i)$$
$$\left(Act\left(Conv\left(Linear(C_i, C_i \cdot E_i)(x_{ESA})\right)\right)\right) + x_{ESA} \qquad (3)$$

where x_{MF} is the output of Mix-FFN, Linear(C_{in}, C_{out})(x) represents a linear layer that projects an input sequence x of dimension C_{in} into an output sequence of dimension C_{out}, and Act denotes the activation layer.

3): Decoder: The feature maps from the four transformer blocks are first processed through MLP layers, where the channel dimension is linearly projected from C_i to C, and the spatial size is upsampled to $H/4 \times W/4$. These processed feature maps are then concatenated to form a fused representation of size $H/4 \times W/4 \times 4C$. Finally, an MLP followed by an upsample layer transforms this fused representation into a segmentation output of size $H \times W \times N_{cls}$, which is then passed through a softmax function for wrap-count classification.

In phase unwrapping applications for filtered multi-look interferograms, an input size of 512 × 512 typically covers about 30 km × 30 km. Given the extensive spatial coverage and relatively low phase gradient of interseismic interferograms, this input size necessitates dividing the interferogram into numerous patches. However, more patches complicate mosaicking and may limit fringe variation, affecting model training and prediction. To address the issue, our method partitions large-scale interseismic interferograms which are padded with zeros, into patches of either 1024 × 1024 or 2048 × 2048 pixels, depending on the overall interferogram size, with strides of 800 pixels or 1600 pixels, creating overlaps of 224 × 1024 or 448 × 2048 pixels, respectively. These patches are downsampled to 512 × 512 using a nearest-neighbor algorithm for training and prediction.

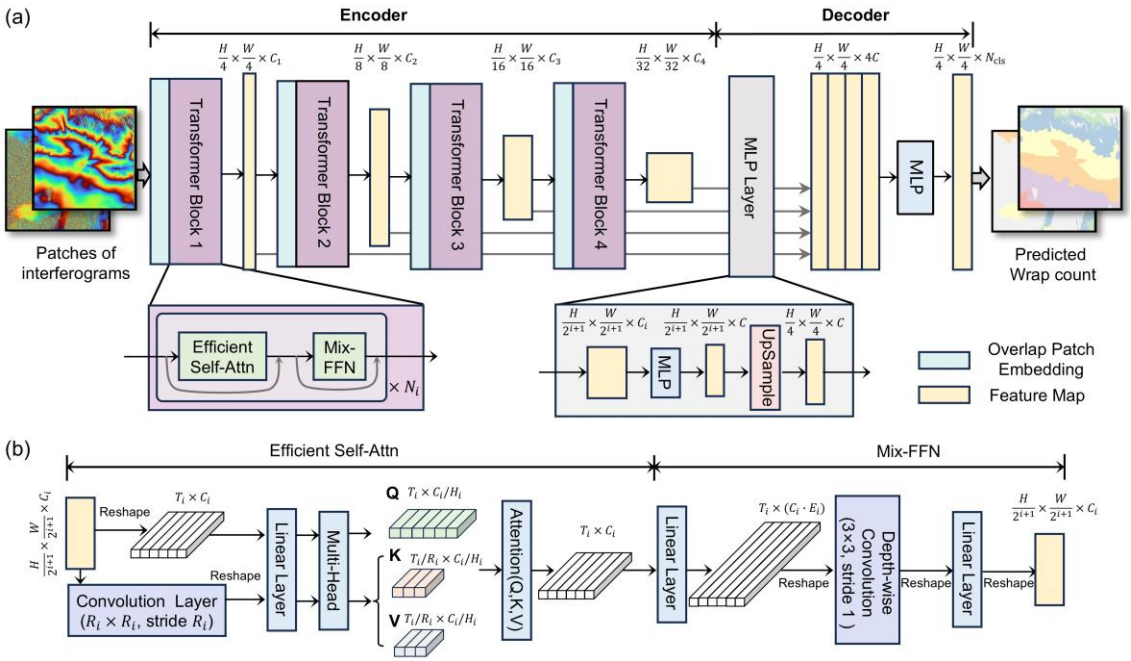

Fig. 2. The schematic diagram of Segformer for wrap-count segmentation. (a) Overall structure. (b) The detailed structure of the Transformer Block.

TABLE I
LAYER AND PARAMETER CONFIGURATION OF MAJOR MODULES IN SEGFORMER-B2

Module		Layers	Parameters	Output Size
Encoder	OPE 1	Convolution	*Kernel_size* = 7, *Stride* = 4, *Padding* = 1, *Channels* = 64	$H/4 \times W/4 \times 64$
	Transfomer Block 1	(Efficient Self-Attn + Mix-FFN) × N_1	$N_1 = 3$, $C_1 = 64$, $H_1 = 1$, $R_1 = 8$, $E_1 = 4$	$H/4 \times W/4 \times 64$
	OPE 2	Convolution	*Kernel_size* = 3, *Stride* = 2, *Padding* = 1, *Channels* = 128	$H/8 \times W/8 \times 128$
	Transfomer Block 2	(Efficient Self-Attn + Mix-FFN) × N_2	$N_2 = 4$, $C_2 = 128$, $H_2 = 2$, $R_2 = 4$, $E_2 = 4$	$H/8 \times W/8 \times 128$
	OPE 3	Convolution	*Kernel_size* = 3, *Stride* = 2, *Padding* = 1, *Channels* = 320	$H/16 \times W/16 \times 320$
	Transfomer Block 3	(Efficient Self-Attn + Mix-FFN) × N_3	$N_3 = 6$, $C_3 = 320$, $H_3 = 5$, $R_3 = 2$, $E_3 = 4$	$H/16 \times W/16 \times 320$
	OPE 4	Convolution	*Kernel_size* = 3, *Stride* = 2, *Padding* = 1, *Channels* = 512	$H/32 \times W/32 \times 512$
	Transfomer Block 4	(Efficient Self-Attn + Mix-FFN) × N_4	$N_4 = 3$, $C_4 = 512$, $H_4 = 8$, $R_4 = 1$, $E_4 = 4$	$H/32 \times W/32 \times 512$
Decoder		MLP + Upsample + Concatenate	$C = 768$	$H/4 \times W/4 \times 3072$
		MLP + Upsample	$N_{cls} = 7$	$H \times W \times 7$

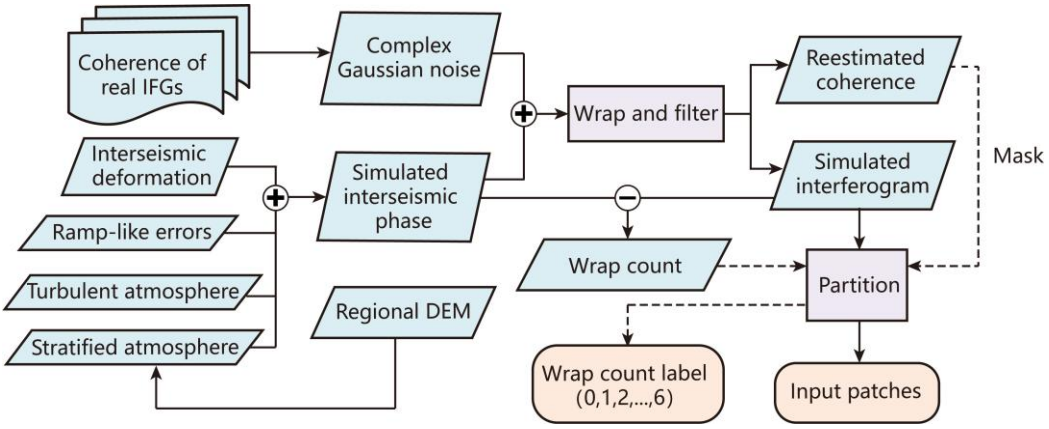

Fig. 3. The flowchart of generating simulated dataset based on regional DEM and real coherence maps.

B. Dataset Construction Strategy

The quality and diversity of the dataset are crucial to the performance of deep learning-based methods. Determining the ground truth wrap count in real interferograms is difficult due to decorrelation noise and unwrapping errors. Here, we propose a dataset simulation workflow based on real digital elevation models (DEMs) and real coherence maps, as shown in Fig. 3. The simulated interseismic phase consists of four components: deformation, topographically-correlated atmosphere, turbulent atmosphere, and ramp-like signals [29, 41]. Deformation is modeled using the Okada model [42]. The modeled fault is constructed based on a segment of the western Altyn Tagh Fault. In the Okada model, the fault length, width, top depth, strike, dip, and rake angle are set to 1000 km, 970 km, 15 km, 55°, 90°, and 0°, respectively. This model assumes that the shallow part of the fault is locked above a depth of 15 km, while the deeper part slips at a long-term rate of 6 mm/yr. The large fault dimensions are chosen to generate an interseismic-like long-wavelength deformation pattern within the simulated domain and to reduce boundary effects. The topographically-correlated atmosphere is generated following Wang *et al.* [43] using real fault zone DEMs multiplied by a power-law function for spatial representation and seasonal trigonometric functions with random perturbations for temporal variation. Turbulent atmospheric effects, featuring inherent randomness with spatial correlation decaying over distance, are simulated using various fractal Perlin noise functions [44].

Ionospheric and plate motion-related signals are introduced as simple first-order ramps. The plate motion-related term describes the large-scale rigid rotation of tectonic plates about a fixed Euler pole in the International Terrestrial Reference Frame (ITRF14), and manifests as a long-wavelength, ramp-like signal in the LOS deformation along the radar range direction [45]. Coherence maps matching the DEM coverage are used to incorporate realistic decorrelation noise. Complex Gaussian noise is added to the raw phase φ_t and wrapped as:

$$\varphi_n = arg\{(cos\varphi_t + A_n n_R) + (sin\varphi_t + A_n n_I)i\},$$
$$A_n = \begin{cases} (1-\gamma) * 8\pi, & \gamma < 0.1 \\ (1-\gamma) * \frac{\pi}{4}, & 0.1 < \gamma < 0.3 \\ (1-\gamma) * \frac{\pi}{8}, & \gamma > 0.3 \end{cases} \quad (4)$$

where γ is the real raw coherence, A_n represents the amplitude of decorrelated noise, n_R and n_I denote random Gaussian noise, arg (·) is the operation of extracting the argument of a complex number, i denotes the imaginary unit (i^2 = -1), and φ_n is the wrapped phase. This equation is adapted from Wu *et al.* [29] based on our experiments to simulate noise as realistically as possible. For coherence values below 0.1—which typically correspond to fully decorrelated areas such as deserts and water bodies—the noise needs to be amplified several times to match real data characteristics. When coherence exceeds 0.1, adding excessive noise would cause the final interferograms to deviate significantly from realistic patterns after filtering. To better replicate noise characteristics in multi-look filtered interferograms, we apply the *adf* function in GAMMA software [46] to filter φ_n and re-estimate its coherence $\hat{\gamma}$, yielding the final simulated interferogram $\hat{\varphi}_n$. Then the wrap count k is calculated by

$$k = round\left\{\frac{(\varphi_t - \hat{\varphi}_n)}{2\pi}\right\} \quad (5)$$

where round (·) denotes the rounding operation. Pixels with re-estimated coherence $\hat{\gamma}$ < 0.7 after filtering in our simulated scenarios, or those in connected regions smaller than 100 pixels, are labeled as decorrelated pixels (i.e., k=0). The simulated interferograms and their corresponding wrap count maps are partitioned into patches, downsampled to 512 × 512, and selected to retain samples with $k_{max} \leq 6$. The simulation generated over 26,000 samples using DEMs from the shuttle radar topography mission (SRTM) and coherence maps from real SAR images over the western Altyn Tagh Fault, and some samples are shown in Fig. 4(a).

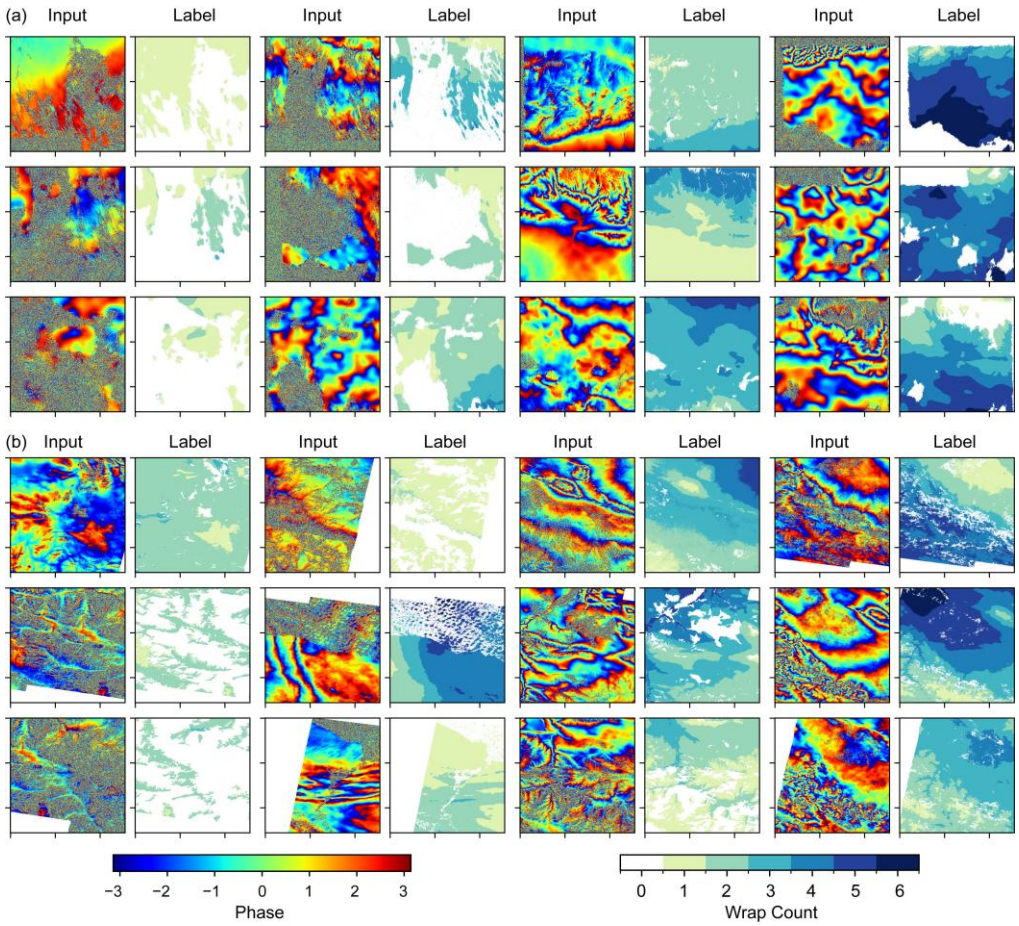

Fig. 4. Sample examples of our dataset for wrap-count segmentation. (a) Simulated samples. (b) Real samples.

To further enrich the dataset, we incorporated real samples. The real interferograms and their corresponding unwrapped results, coherence maps are downloaded from the COMET-LiCSAR platform [47] , derived from three frames covering the western part of the Haiyuan Fault. Since the unwrapped results obtained by SNAPHU may involve errors, we manually exclude interferograms with severe unwrapping artifacts. After identifying decorrelated pixels, computing wrap counts, partitioning, and downsampling as described earlier, we select nearly 13,000 high-quality samples with insignificant unwrapping errors, complicated fringe patterns, and moderate decorrelated noise via visual check. The sample information in the source frames is shown in Table II, and several real samples are presented in Fig 4(b).

TABLE II
THE DATA SOURCE INFORMATION ABOUT REAL SAMPLES

LiCSAR Frame ID	Downloaded Interferograms	Time Span	Selected Samples
033D_05106_131313	765	2014.10-2024.02	5042
033D_05304_131313	877		2679
033D_05503_131313	3048		5788

C. Patch Post-processing

Affected by factors like the limited training data and the inherent deep-learning drawbacks, errors in estimated wrap counts sometimes still occur. As shown in (1), errors in the wrap count k directly determine the unwrapped results. To mitigate these errors, we apply a post-processing step. Automatically judging whether a patch requires post-processing is of great importance. Here, we define a prediction credibility metric K_u based on Segformer's output. Let p_1 and p_2 be the highest and second-highest predicted probabilities for each pixel, respectively. The prediction confidence k_w for each pixel is then computed as the difference between these two values:

$$k_w = p_1 - p_2. \tag{6}$$

A larger k_w indicates that Segformer has more confidence in its prediction, while a smaller k_w suggests that Segformer may have difficulty distinguishing between the two classes, increasing the likelihood of errors [48]. We define the credibility metric for a patch as $K_u = [K_u^1, K_u^2]$, where K_u^1 is the average k_w of all pixels, and K_u^2 is the average k_w of non-decorrelated pixels ($k \neq 0$).

For patches with high K_u (both values $> K_{t1}$), where K_{t1} is set by the user, no post-processing is necessary. For patches with moderate K_u (either value $< K_{t1}$ and both values $> K_{t2}$),

small areas with prediction errors are likely to occur. They are corrected using a DCNN network trained on synthetic samples. During training of the DCNN, random small-area $+/-2\pi$ errors are added to the samples simulated before, and pixels are labeled accordingly. For patches with low K_u (either value $< K_{t2}$), complex and wide-range errors can be present due to the instability of our segmentation model. We re-unwrap these patches using the sparse MCF method in GAMMA software, aiming to get more reliable wrap counts. Extensive experiments suggest optimal thresholds K_{t1} = [0.95, 0.95] and K_{t2} = [0.65, 0.8] for practical applications.

D. Patch Mosaicking

The third step of SWIPHU is to seamlessly mosaic all the patch-wise updated wrap counts. We formulate mosaicking as a linear equation-solving problem, and estimate the relative wrap-count differences between each non-first patch and the first patch based on the observed wrap-count difference in overlapping regions. Theoretically, the wrap count within each patch is relative, meaning the overlap differences should be a constant integer. However, segmentation or post-processing may cause variations across overlapping pixels. We use the mode value $L_{j,k}$ as the relative wrap-count difference between patches j and k, discarding low-quality overlaps where decorrelated pixels exceed 97% or the mode value constitutes less than 80% of pixels.

For an $m \times n$ partitioned interferogram, the unknown parameters X form an $(mn\text{-}1)\times 1$ matrix, while the observation matrix L contains $m(n\text{-}1)+n(m\text{-}1)$ elements if all overlaps are preserved. The linear equation can be expressed as $L = BX$:

$$L = \begin{bmatrix} L_{1,2} \\ L_{2,3} \\ \vdots \\ L_{mn-1,mn} \\ L_{1,n+1} \\ L_{2,n+2} \\ \vdots \\ L_{n(m-1),mn} \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 \\ -1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}, X = \begin{bmatrix} X_{1,2} \\ X_{1,3} \\ X_{1,4} \\ \vdots \\ \vdots \\ X_{1,mn} \end{bmatrix} \quad (7)$$

where B represents the coefficient matrix. Each preserved observation $L_{j,k}$ is weighted by $P_{j,k}$ based on the proportion of mode-value pixels in the overlap. X is then solved using:

$$X = (B^T P B)^{-1} B^T P L. \quad (8)$$

Since all observations and unknowns are integers, the solution should be error-free. If the residual is non-negligible, the estimated parameters may be biased, leading to patch-wise mosaicking errors across the entire interferogram. Thus if the residual exceeds 0.00001 or the coefficient matrix B is rank-deficient due to excessive discarded observations, the interferogram is considered unreliable. In such cases, it either produces mosaicking errors across several patches or cannot be fully mosaicked from all patches, and is therefore discarded from our workflow.. For successfully solved X, we reconstruct the full wrap-count map by aligning patches to the first patch's reference after rounding X to the nearest integer. A median filter is applied to reduce errors along fringe edges, yielding the final unwrapped result.

III. RESULTS

A. Assessment of Patch-wise Wrap-count Prediction

The Segformer model was trained using Pytorch 2.2 on a Quadro RTX 4000 GUP, with a batch size of 2 and an initial learning rate of 0.0001. Considering that transformer-based models face difficulty in training parameters with a limited dataset, we initialized the network using a pre-trained weight file based on the PASCAL VOC 2012 (VOC12) and Semantic Boundaries Dataset (SBD). The model was built after training for more than 10 epochs until the loss function fully converged.

To assess the performance of Segformer-B2, other well-known semantic segmentation networks including U-Net [49], DeepLabV3+ [50], and Segformer-B0 were also trained using the same dataset and optimal training strategy. The wrap counts of nearly 2000 test samples were predicted by different models, and the Intersection over Union (IoU) between predicted results and ground truth for each class was calculated, as shown in Table III.

Segformer-B2 achieved the highest Mean IoU (mIoU) of 91.4%, outperforming U-Net and Segformer-B0, which had mIoUs below 50.0%. For classes with wrap counts greater than 2, U-Net's IoU drops sharply, indicating its poor ability to segment complex fringe patterns, despite having 31,030,983 trainable parameters. Segformer-B0, with 3,715,943 parameters, performs better than U-Net for larger wrap counts but is constrained by its smaller model size. DeeplabV3+ and Segformer-B2 have 2,754,423 and 27,352,007 parameters, respectively. Benefiting from the Transformer-based self-attention mechanism and larger model size, Segformer-B2 outperforms DeeplabV3+, achieving over 2.0% improvement in classes with wrap counts of 3, 4, and 6.

Further comparisons of predicted wrap counts and prediction confidence from Segformer-B2 and DeeplabV3+ on representative test samples are presented in Fig. 5. As shown in Fig. 5(a) and Fig. 5(c), DeeplabV3+ produces wrap count biases between isolated regions. While its predictions within coherent regions are relatively accurate with high confidence, it fails to reveal errors between isolated coherent regions, as seen in Fig. 5(a). The overestimated prediction confidence prevents necessary post-processing, allowing unwrapping errors to persist. For more complex fringe patterns affected by decorrelation noise, as shown in Fig. 5(b) and Fig. 5(e), DeeplabV3+ struggles with segmentation, whereas Segformer demonstrates greater robustness. Moreover, DeeplabV3+ tends to misclassify slightly noisy regions as class 0, whereas Segformer-B2 preserves more coherent pixels, as observed in Fig. 5(d). These results all underscore the superior robustness of Segformer-B2 in SAR interferogram wrap-count segmentation, consistent with its performance in natural scene images [40]. This suggests that the trained model has reached

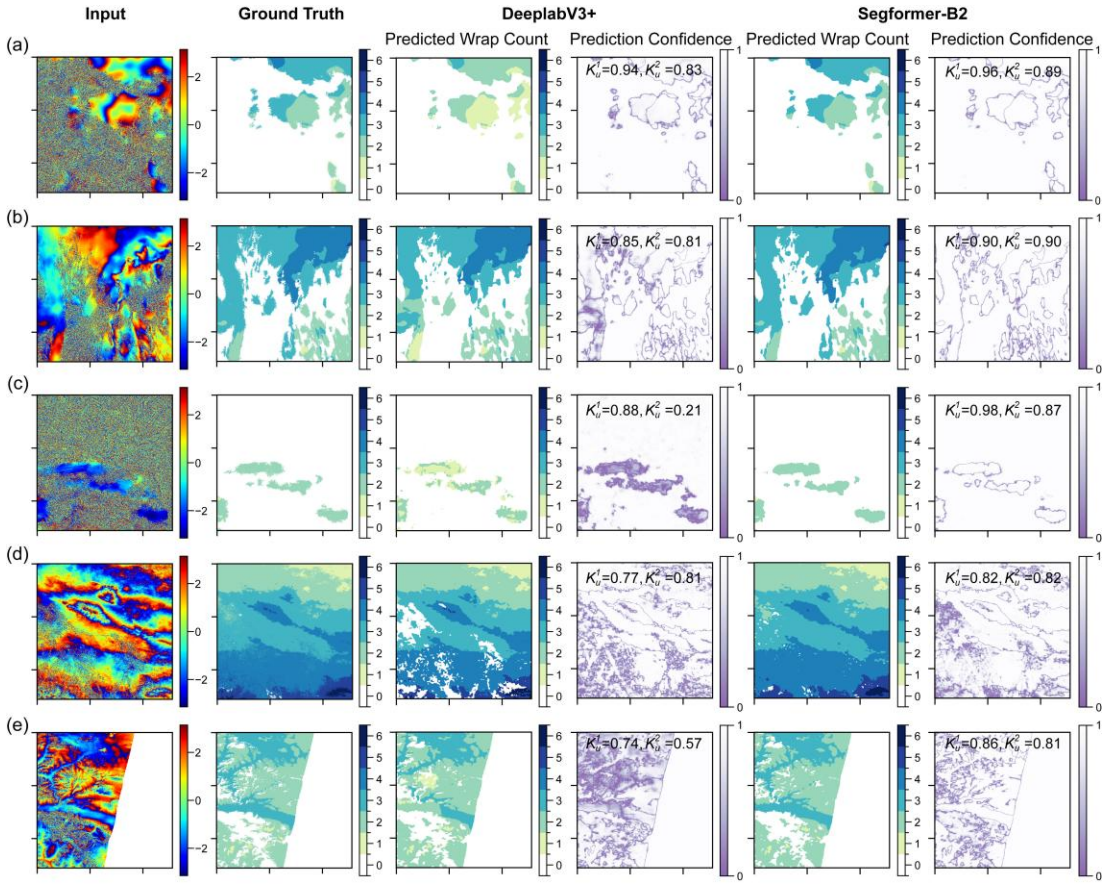

Fig. 5. Comparison of wrap count and prediction confidence from DeeplabV3+ and Segformer on five representative patch-wise test samples.

optimal stability with the existing training dataset and can be integrated into the proposed workflow.

TABLE III
WRAP-COUNT PREDICTION PERFORMANCE OF DIFFERENT NETWORKS ON THE TEST DATSET

Network	IoU for different classes (%)							Mean IoU
	0	1	2	3	4	5	6	
U-Net	92.2	62.1	50.6	16.0	4.7	3.3	3.8	33.2
DeepLab V3+	91.6	94.1	92.4	88.4	87.4	89.2	85.2	89.7
Segformer-B0	87.5	80.3	69.5	36.4	33.0	31.4	0.0	48.3
Segformer-B2	91.8	94.8	93.4	90.9	91.2	90.2	87.5	91.4

B. Algorithm Performance on Synthetic Interferograms

To verify the effectiveness of the whole SWIPHU algorithm, we conducted simulated experiments in this section. A total of 134 large-scale interseismic interferograms were generated using the strategy described in Section II B. This simulation is based on the western end of the Haiyuan Fault, covered by Sentinel-1's descending orbit path 33. Real projected DEM in the radar coordinate system and hundreds of coherence maps derived from Sentinel-1 interferometry were utilized to generate synthetic interferograms, which were then processed using SWIPHU. Fig. 6(a) shows a typical synthetic interseismic interferogram, characterized by severe decorrelated noise and considerable tropospheric delay. To keep a constant patch step size and maintain consistent mosaicking across interferograms of different sizes, the interferogram was first padded with zeros and then divided into 10 patches, each measuring 2048 × 2048 pixels, before being input into the Segformer network. Following wrap-count segmentation by Segformer, post-processing was applied based on the prediction credibility metric K_u. The post-processing details for Patches No. 4 and No. 3 in Fig. 6(a) are illustrated in Fig. 6(g) and Fig. 6(h), respectively. As shown, a moderate K_u value in patch No. 4 corresponds to minor prediction errors, and the trained DCNN plays an essential role in correcting such small-region errors. In contrast, the lower K_u value in patch No. 3 indicates larger and more complex prediction errors, which in some cases require re-unwrapping using the traditional MCF method. Fig. 6(b) displays the final whole wrap count map after mosaicking, which aligns well with the fringe patterns in Fig. 6(a), demonstrating the validity of SWIPHU. Notably, only three out of 134 simulated interferograms were discarded during mosaicking, indicating strong agreement in wrap count across overlapping regions after segmentation and post-processing.

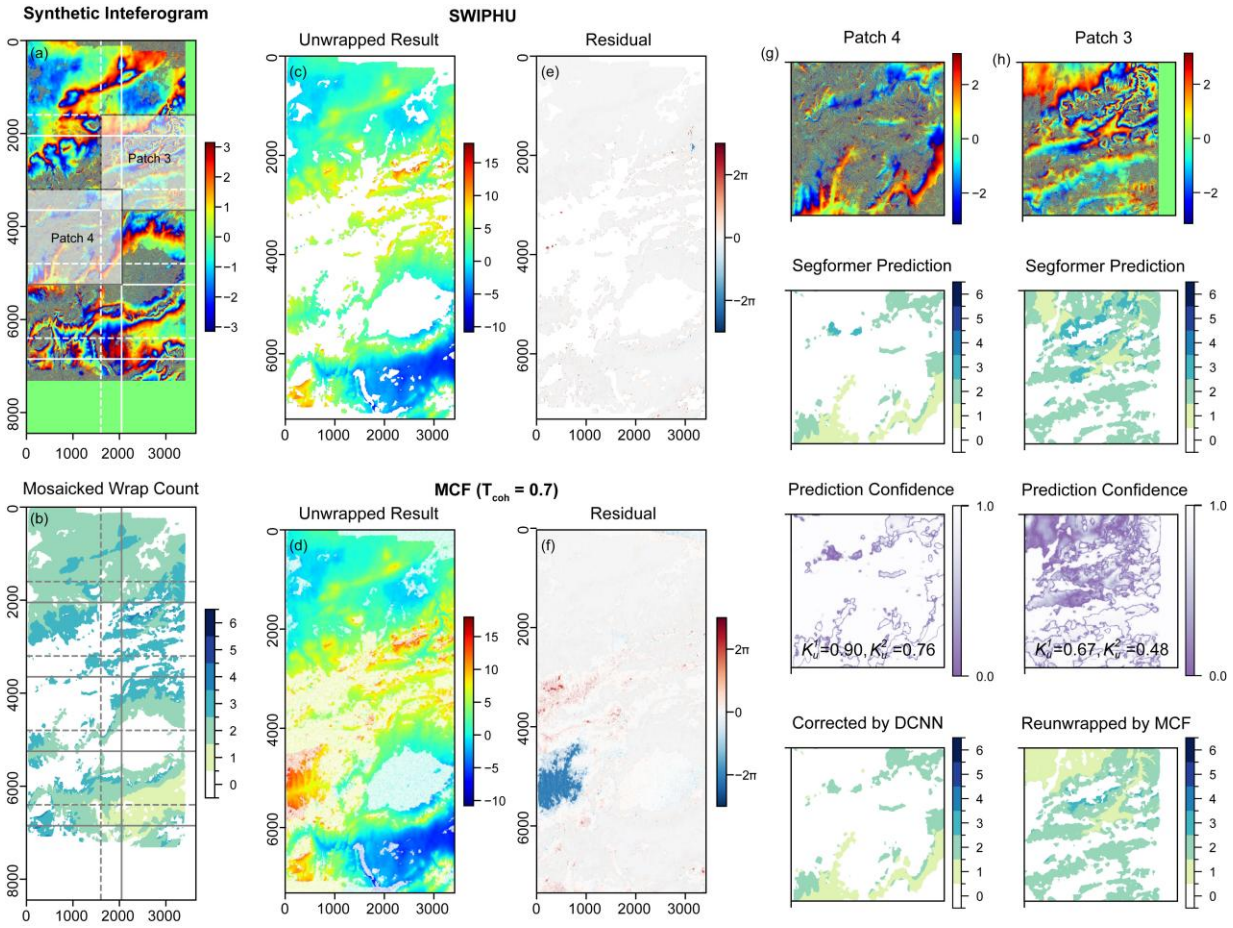

Fig. 6. Performance of the SWIPHU method on a typical simulated large-scale interferogram compared with MCF. (a) Simulated interferogram. (b) Mosaicked wrap count by SWIPHU. (c)-(d) Unwrapped results of (a). (e)-(f) Residual maps of (c) and (d), respectively. (g)-(h) Comparison of two patches in (a) before and after different post-processing in SWIPHU. In Patch 4, DCNN correction is applied due to a moderate credibility metric K_u, while in Patch 3, MCF re-unwrapping is used due to a low K_u.

Given that the MCF method is widely recognized as one of the most robust and effective approaches for large-scale low-coherence interferograms in 2-D phase unwrapping [29], with extensive validation in numerous applications, we compare our algorithm with the MCF implementation in GAMMA [46] using different masking strategies. We employed two types of mask files: (1) a mask derived from reestimated coherence maps using a coherence threshold (T_{coh}) of 0.7, and (2) a mask generated based on the wrap-count maps predicted by the Segformer network. The 131 interferograms retained in our workflow were then unwrapped using MCF with these two masking configurations.

The mean absolute error (MAE) between the unwrapped results and the ground truth is computed to assess the accuracy of each unwrapping approach. The MAEs are 0.69, 0.58, and 0.31 rad for MCF with T_{coh}=0.7, MCF with Segformer-derived masking and SWIPHU, respectively. These results show that SWIPHU significantly improves unwrapping accuracy of synthetic interferograms compared to MCF. Fig. 6(c)-(f) highlights the unwrapping results of Fig. 6(a) and the corresponding residual maps using MCF with T_{coh}=0.7 and SWIPHU. The unwrapping errors by the traditional MCF method are concentrated in the isolated patch surrounded by decorrelated noise in the lower left, as well as in some fragmented decorrelated pixels that are not masked. In contrast, our deep-learning-based method automatically masks low-coherence pixels and reduces errors in isolated patches by leveraging global feature extraction and a long-distance receptive field. Note that there are some tiny regions with unwrapping errors in the SWIPHU method's results compared to MCF, particularly in highly complex fringe patterns and severely decorrelated mountainous areas with dense vegetation. Additionally, the residual map reveals minor scattered errors across the interferogram, which can be attributed to interpolation artifacts introduced during Segformer's multiple downsampling and upsampling operations, particularly at wrap-count discontinuities along fringes edges.

To ensure a fairer comparison, we refine the MAE calculation for MCF with T_{coh}=0.7 by excluding fragmented decorrelated pixels identified by the Segformer mask. This adjustment reduces its MAE to 0.40 rad. Nevertheless, the residual error remains higher than that of SWIPHU, illustrating that SWIPHU achieves at least a 23% improvement in unwrapping accuracy over MCF.

V. Application to Real Interferograms

A. Large-scale Interseismic Interferograms in the Western Altyn Tagh Fault

The Altyn Tagh Fault (ATF) is a 1600-km sinistral strike-slip fault along the northern margin of the Qinghai-Tibet Plateau, with a deep slip rate ranging from ~10 mm/yr to ~3 mm/yr [51, 52, 53]. The western segment of the Altyn Tagh Fault is characterized by dramatic topographic relief of up to 3 km on both sides of the fault and complex landforms, including deserts, plateau permafrost, and seasonal snow cover, as shown in Fig. 7(a). These factors contribute to severe atmospheric delays and decorrelation. The presence of distinct fringe patterns, decorrelated noise and isolated regions in desert areas makes this region an ideal case to test the performance of the SWIPHU algorithm.

We processed two frames from Sentinel-1 descending orbit track 85, which crosses the western Altyn Tagh Fault, spanning the period from January 1, 2019 to June 30, 2022. A total of 763 large-scale interseismic interferograms with spatial perpendicular baseline < 20 m and temporal baseline > 300 days were formed after interferometry. While using short temporal baselines only can simplify unwrapping, it can introduce a phase bias (also known as fading signal [54]). Including longer temporal baselines is an effective way to mitigate phase bias and better constrain long-term interseismic velocities. The baseline network is presented in Fig. 7(b). All interferograms were unwrapped using SWIPHU and six variations of the MCF algorithm implemented in GAMMA, incorporating different coherence threshold masks (0.5, 0.7, and 0.9) and two distinct triangulation modes: filled triangulation (FT) and Delaunay triangulation (DT). In the workflow of our method, the full-size interferograms were partitioned to eight patches of 2048 × 2048 pixels in the first segmentation step, and 729 interferograms were preserved after mosaicking, while 34 interferograms were discarded due to unresolved wrap-count contradiction in overlaps of adjacent patches.

Triangular loop closure analysis is utilized to evaluate unwrapping errors in real scenarios [6]. However, nonzero loop closure is not solely attributed to unwrapping errors; some deviations arise from fading signals (also known as a phase bias) [54], while others result from decorrelation noise. To better quantify unwrapping errors within triangular loops, we define an index named unwrapping error proportion (UEP) rather than using the RMS. UEP is calculated as the ratio of pixels with unwrapping errors to the total number of unmasked pixels within a loop. A pixel is considered to have an unwrapping error if its absolute closure error exceeds π rad and belongs to a connected region of at least 100 pixels. By applying a minimum size threshold, UEP minimizes the influence of fragmented incoherent pixels introduced by different coherence masking and of tiny scattered errors, providing a more fair metric for performance comparison. The mean UEPs of all 1295 triangular loops formed by 726 unwrapped interferograms using different MCF configurations and SWIPHU are compared in Table IV. For the FT mode of MCF, the lowest mean UEP is observed at a coherence threshold of 0.7, whereas in the DT mode, the mean UEP decreases as the coherence threshold increases, reaching a minimum of 0.9. The SWIPHU method significantly outperforms both MCF modes, achieving a mean UEP of 1.0%, which represents a 28% and 65% reduction compared to the optimal FT (1.4%) and DT (2.9%) configurations, respectively.

TABLE IV
MEAN UEP OF DIFFERENT UNWRAPPING RESULTS ON THE WESTERN ALTYN TAGH FAULT INTERFEROGRAMS (%)

MCF	Coherence threshold (T_{coh})			SWIPHU
	0.5	0.7	0.9	
FT	1.5	1.4	1.7	1.0
DT	4.8	4.3	2.9	

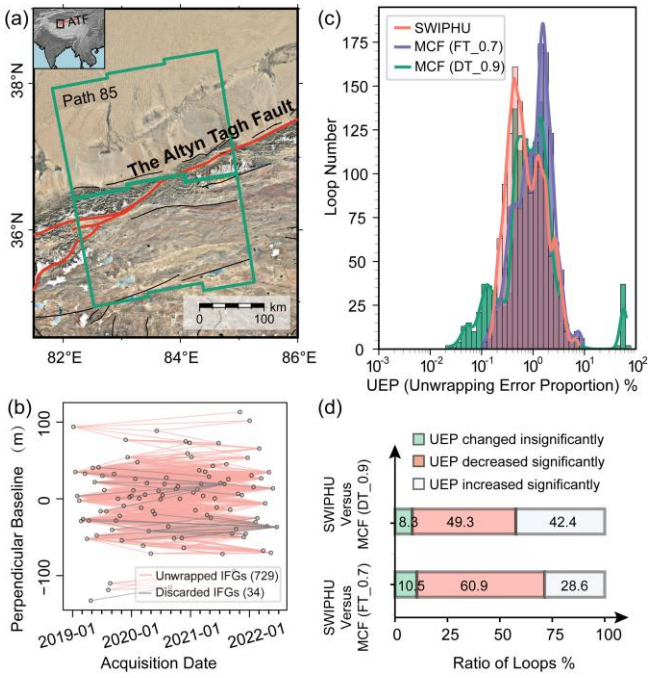

Fig. 7. Verification on real data over the western Altyn Tagh Fault (ATF). (a) Study area of the western ATF. The green rectangles indicate the coverage of the ascending Sentinel-1 large-scale interferograms. The red lines outline the trace of ATF. (b) Baseline network of the interferograms. (c) Histograms of Unwrapping Error Proportion (UEP) for loops unwrapped by SWIPHU and two the optimal MCF configurations. (d) The UEP variations after applying SWIPHU compared with two optimal MCF configurations.

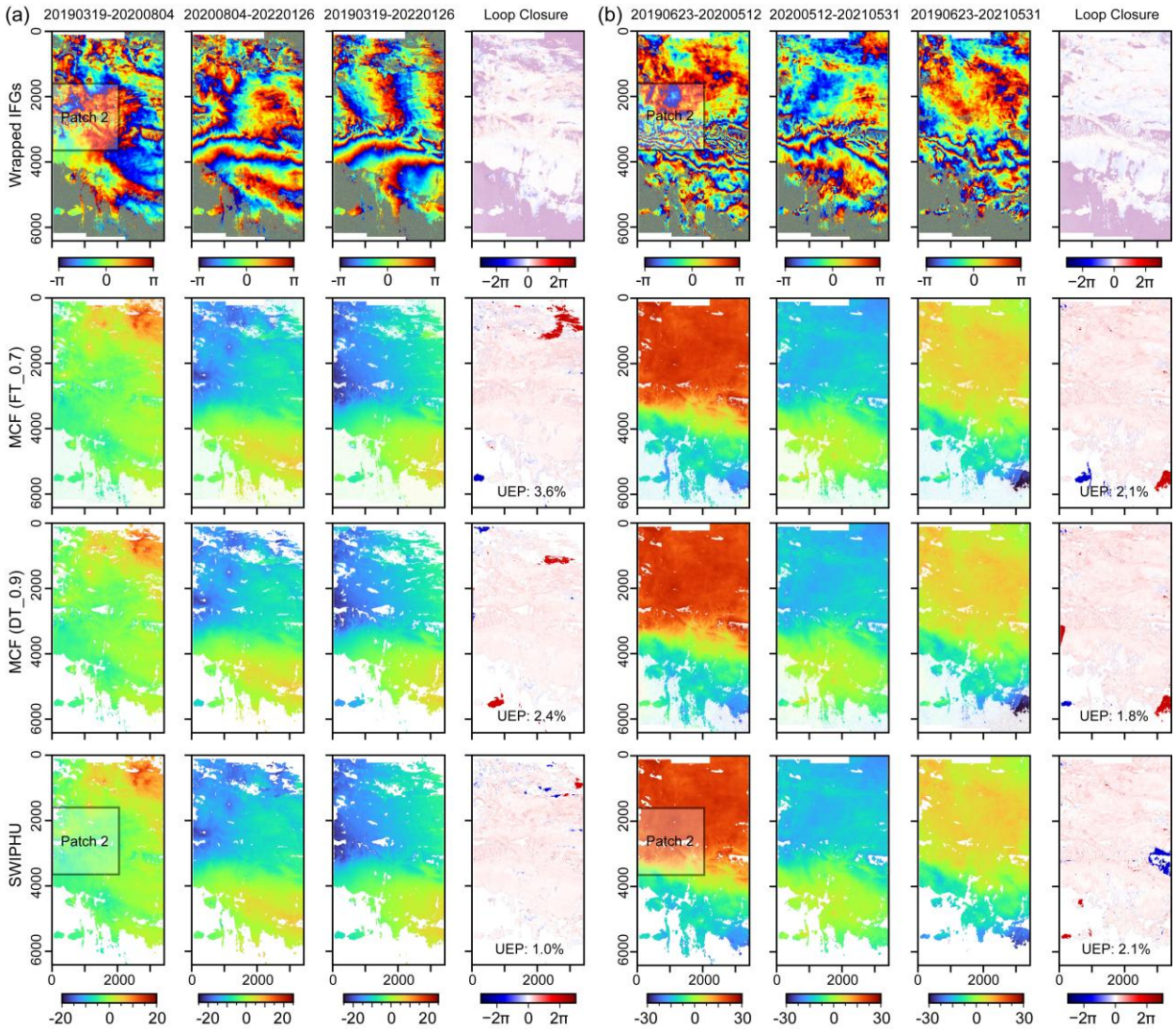

Fig. 8. Unwrapping results of different methods for two loop examples from the western Altyn Tagh Fault (ATF). (a) Loop 20190319-20200804-20220126. (b) Loop 20190623-20200512-20210531. The gray squares indicate the patches requiring MCF re-unwrapping in the SWIPHU method.

Fig. 7(c) highlights the UEP distribution across all loops for two optimal MCF methods (FT_0.7 and DT_0.9) and the SWIPHU algorithm. Though the UEPs of SWIPHU and MCF (FT_0.7) are both primarily concentrated between 0.1% and 10.0%, a higher proportion of UEP values obtained by SWIPHU fall below 1.0%, whereas most loops from MCF (FT_0.7) lie within the 1.0%–8.0% range. In contrast, MCF (DT_0.9) demonstrates wider variance, with some interferograms exhibiting UEP values exceeding 50.0%, indicating higher instability in this MCF configuration. To analyze the differences more clearly, Fig. 7(d) quantifies UEP variations between the two MCF methods and SWIPHU for each loop. A UEP relative change threshold of 10% is set, considering UEP variations below this value insignificant. That is, if the UEP obtained using our method increases or decreases by more than 10% of that obtained with the original MCF method, we consider this change to be significant. The results show that, compared to MCF (DT_0.9), 49.3% of loops processed using SWIPHU exhibit a significant reduction in UEP, while 8.3% show negligible changes. Similarly, relative to MCF (FT_0.7), only 10.5% of loops show insignificant changes, while 60.9% experience a remarkable UEP decline. These results further demonstrate the effectiveness of SWIPHU, notably enhancing the phase unwrapping accuracy for over 45% of the large-scale interferograms in the western Altyn Tagh Fault. However, it should be noted that although half of the loops show better performance, approximately one-third of the loops present inferior performance compared to the two MCF methods, indicating the limitations of SWIPHU, which are analyzed in the following text.

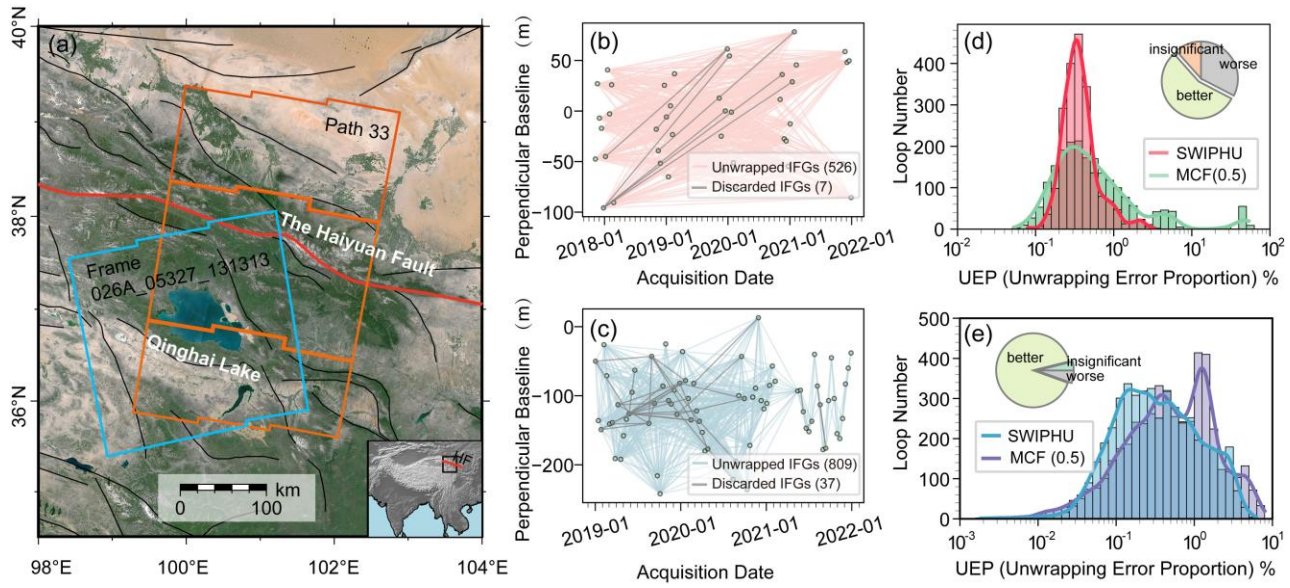

Fig. 9. Verification on real data over the western Haiyuan Fault (HF). (a) Study area of the western HF. The orange and blue rectangles indicate the coverages of the first and the second sets of interferograms, respectively. The red lines outline the trace of HF. (b)-(c) Baseline networks of the first and second interferogram sets, respectively. (d)-(e) Histograms of Unwrapping Error Proportion (UEP) distributions and pie charts of UEP variations for loops unwrapped using SWIPHU and MCF, corresponding to the first and second interferogram sets, respectively.

Fig. 8 depicts a detailed comparison of two loop cases with different performances. From Fig. 8(a), it is evident that in the MCF methods, 2π closure errors in the upper right corner of the loop closure diagram originate from the second interferogram in the triangular loop, where fringe aliasing occurs due to snow and ice coverage. -2π and 2π closure errors appear in the bottom left of the two MCF loops, stemming from the third and the second interferogram, respectively, where desert-induced decorrelation leads to localized isolated unwrapping errors. In contrast, our approach effectively mitigates these two types of unwrapping errors in both the upper right and lower left regions, resulting in a UEP decline of 2.6% and 1.4% compared to the two MCF methods. Additionally, Fig. 8(a) shows that only one patch required re-unwrapping via MCF in the post-processing step of SWIPHU among these three interferograms, suggesting that the observed UEP reduction in this case is mainly attributed to wrap-count segmentation by Segformer rather than post-processing corrections. Conversely, another loop case where UEP increases is shown in Fig. 8(b). It can be observed that the MCF methods introduce unwrapping errors in the desert regions and along interferogram edges. While SWIPHU partially reduces the desert-related errors and completely removes edge errors, it introduces a new -2π closure error in the middle-right section of the loop. By comparative analysis of Fig. 8(b), it can be seen that this new error stems from the first interferogram in the loop, where atmospheric-related dense fringes trouble the output of our trained Segformer. Additionally, the interferometric patterns of Patch No. 2 in the first interferogram which requires MCF-based re-unwrapping, reveal the method's limitations in handling complex and densely fringed regions.

B. Two Sets of Interferograms in the Western Haiyuan Fault

Situated along the northeast margin of the Qinghai-Tibet Plateau, the Haiyuan Fault (HF) is a left-lateral strike-slip fault with a slip rate of ~2-7 mm/yr, and includes creeping segments. It is considered a high seismic hazard due to the presence of a noticeable seismic gap [55, 56, 57]. The western HF features wholesale vegetation cover and runs close to Qinghai Lake, the largest inland saline lake in China, as shown in Fig. 9(a). To further verify the generalization and robustness of our deep-learning-based algorithm across diverse environments and data sources, two sets of multitemporal interferograms in the western Haiyuan Fault are tested in this section.

The first set of interferograms covers a larger area than the interferograms of the western Altyn Tagh Fault, consisting of three frames from the Sentinel-1 descending orbit named Path 33. It was generated using SAR images acquired from November to February of the following year between 2017 and 2021. To maximize the number of interferograms, only the temporal baseline was constrained to be greater than one year, while the spatial baseline threshold was not restricted, yielding 533 large-scale interferograms. These were then unwrapped using SWIPHU with a patch size of 2048 × 2048 pixels, and 527 interferograms were successfully retained, while six were discarded during mosaicking, as shown in Fig. 9(b). For comparison, the same interferograms were also unwrapped using the MCF method in FT mode, applying coherence mask thresholds of 0, 0.1, 0.3, 0.5, 0.7, and 0.9. Since the DT mode of MCF failed to unwrap certain interferograms at higher coherence thresholds and generally exhibited lower accuracy than the FT mode (as demonstrated in Section IV.A), it is excluded from further analyses in this paper.

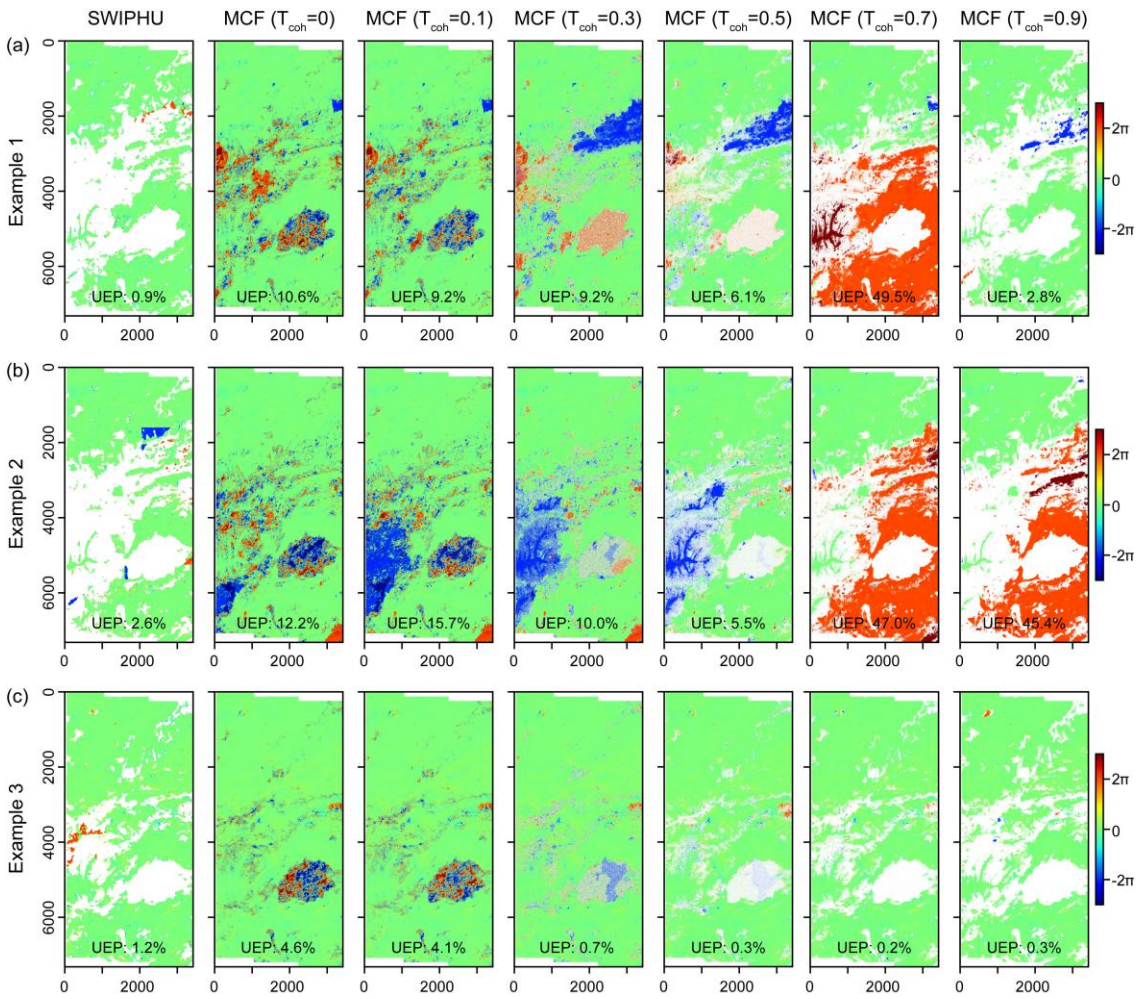

Fig. 10. Three examples comparing loop closure performance of SWIPHU and MCF under varying coherence mask threshold from the first set of interferograms in the western Haiyuan Fault (HF). (a)-(b) Examples 1 and 2, in which SWIPHU outperforms MCF. (c) Example 3, in which SWIPHU shows inferior performance.

TABLE V
MEAN UEP OF DIFFERENT UNWRAPPING RESULTS FOR TWO SETS OF INTERFEROGRAMS ON THE WESTERN HAIYUAN FAULT (%)

Set	Coherence threshold (T_{coh}) of MCF						SWIPHU
	0	**0.1**	**0.3**	**0.5**	**0.7**	**0.9**	
1	9.9	8.8	3.6	2.4	6.0	11.2	0.4
2	2.3	2.3	1.2	1.0	1.1	1.6	0.7

The accuracy of these results is assessed using 2042 formed triangular loops, with the mean UEPs summarized in Table V. Among the MCF results, the coherence threshold of 0.5 yields the lowest average UEP, confirming that either lower or higher thresholds lead to more unwrapping errors—lower thresholds incorporate more decorrelated pixels, while higher thresholds result in more separated regions. The SWIPHU method significantly outperforms MCF, achieving a UEP of 0.4%, representing an 83% reduction compared to the minimum UEP of 2.4% obtained with MCF. The UEP distribution histograms of these two results are plotted in Fig. 9(d). SWIPHU predominantly yields UEP values between 0.1% and 1.0%, whereas the MCF results exhibit a broader distribution, with hundreds of loops reaching up to 30.0% UEP. A specific loop-wise UEP variation comparison between the two methods is also shown in Fig. 9(d). As mentioned in the previous section, a relative UEP change exceeding 10% is regarded as significant. Among the 2,042 loops, 1139 loops demonstrate improved performance with SWIPHU, accounting for 55.8%, and 666 loops exhibit worse accuracy, representing 32.6%.

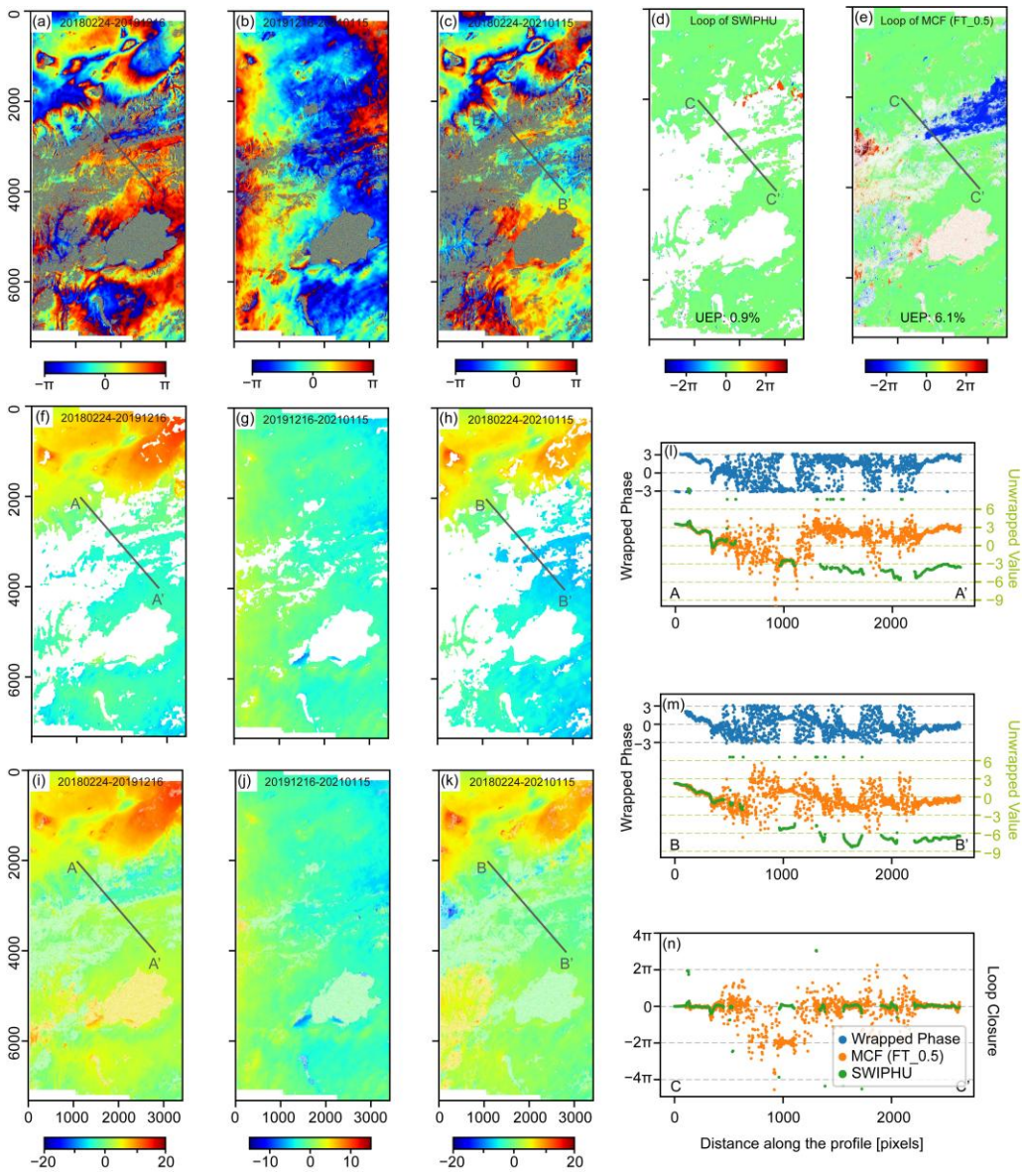

Fig. 11. Detailed comparison of example 1 (Loop 20180224-20191216-20210115) from Fig. 10. (a)-(c) Wrapped interferograms involved in the loop. (d)-(e) Loop closure results using SWIPHU and MCF (T_{coh}=0.5). (f)-(h) Unwrapped results of (a)-(c) using SWIPHU. (i)-(k) Unwrapped results of (a)-(c) using MCF (T_{coh}=0.5). (l)-(m) Profile comparisons of the first interferogram, the third interferogram, and the loop closure, respectively, unwrapped by different methods.

Fig. 10 compares different unwrapping approaches for three representative triangular loop cases. Example 1 depicts a loop with a low UEP of 0.9% processed using SWIPHU. In contrast, the loop closure maps from different MCF methods, with coherence mask threshold ranging from 0 to 0.9, contain significantly more unwrapping errors. Especially, when T_{coh} = 0.7, the UEP increases to 49.5%, exhibiting widespread errors across the large-scale image. Example 2 demonstrates a case where SWIPHU introduces localized stitching-related unwrapping errors, resulting in a UEP of 2.6%. However, the MCF results produce even higher UEPs (5.5%–47.0%), reinforcing that while our deep learning approach may not completely eliminate unwrapping errors in some cases, it still provides a meaningful improvement over traditional methods, which tend to introduce large-scale errors. Example 3 reveals a case where results are worse using SWIPHU, with the MCF results being nearly error-free, except over Qinghai Lake where there is decorrelation. The SWIPHU method introduces a 2π loop closure for a number of small areas concentrated along decorrelated regions, manifesting as localized patchy and stitching-related errors.

Since the loop closure map reflects the combined influence of all involved interferograms rather than a direct unwrapping error map, we conducted a detailed analysis of the three original interferograms in Example 1 to compare the SWIPHU method with the optimal MCF method. As shown in Fig. 11(a) and Fig. 11(c), a prominent vegetation-induced decorrelation region extends from the central left to the upper right of the

interferogram. This region makes MCF prone to large-scale unwrapping errors. Fig. 11(f)–(k) reveals a clear discrepancy in the lower half of the first and third unwrapped maps produced by the SWIPHU method and the MCF method. A profile across the prominent decorrelation region is extracted from the unwrapped results, along with the corresponding wrapped interferograms and loop closure maps, to facilitate a detailed comparison. In Fig. 11(l), the first interferogram's unwrapped results begin to diverge around the 1200th pixel along the profile, where the MCF method introduces a clear 2π jump, while SWIPHU maintains the gradually decreasing trend of variation. Similarly, in Fig. 11(m), discrepancies emerge around the 1000th pixel in the third interferogram, deviating by approximately π and -π relative to the left adjacent high-coherence points, respectively. It is difficult to directly identify which result is erroneous. However, by examining the loop closure values in Fig. 11(n) and cross-referencing with Fig. 11(l), we can infer that the MCF method introduces significant unwrapping errors in the third interferogram, whereas the SWIPHU method remains error-free.

This analysis shows that although the loop closure values of the MCF method may be zero, substantial unwrapping errors can actually occur in the loop's interferograms, offsetting each other within the loop. It further highlights that traditional methods tend to propagate errors when encountering decorrelated noise that extends across the interferogram, ultimately leading to large-scale unwrapping failures. In contrast, our computer vision-based approach operates on a patch-level segmentation rather than relying on pixel-wise connectivity, enabling it to automatically detect decorrelated regions, substantially enhancing the unwrapping accuracy in interferograms with widespread low-coherence pixels.

The second set of interferograms in the western Haiyuan Fault originates from LiCSAR ascending frame, 026A_05327_131313, covering the Qinghai Lake region, as shown in Fig. 9(a). All available 846 interferograms from January 2019 to December 2021 were downloaded from the LiCSAR platform, with varying spatial and temporal baselines. Unlike the first set of large-scale interferograms in the radar coordinate system, these are in the geographic coordinate system and have a smaller spatial coverage. Given their size, they were unwrapped using SWIPHU by partitioning them into 1024 × 1024 pixel patches. As shown in Fig. 9(c), 809 interferograms were preserved. The set of interferograms was also unwrapped using the MCF method with different coherence mask thresholds, and the statistical results of quantitative accuracy on all 4730 loops are summarized in Table V. The mean UEP of SWIPHU is 0.7%, marking a 30% reduction compared to the lowest mean UEP of 1.0% among the MCF unwrapping results. Fig. 9(e) depicts the UEP distribution histograms of these two methods and the UEP changes after applying SWIPHU. Most loops processed with our method exhibit lower UEP values, with the majority falling below 1.0%, whereas the MCF method results in UEP values ranging from 1.0% to 10.0% for numerous loops. Compared to MCF with a coherence threshold of 0.5, SWIPHU maintains accuracy for 3.9% of the loops and significantly improves in 90.6% of them. However, as in previous cases, some of the loops (5.5%) experience a decrease in accuracy.

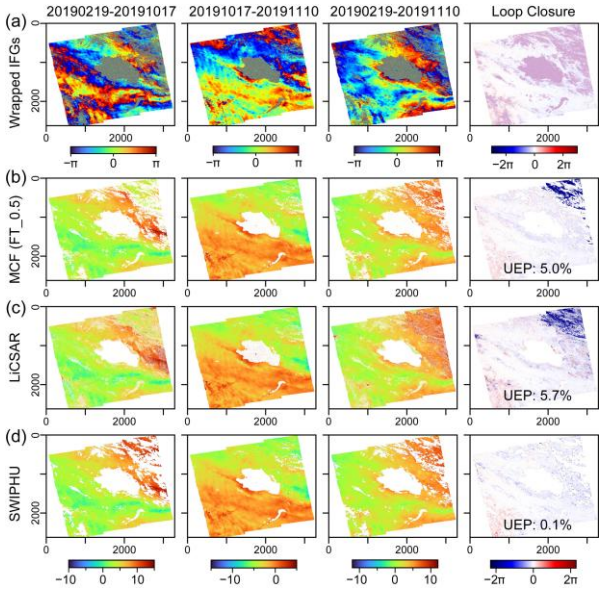

Fig. 12. Unwrapped results and loop closure comparison for Loop 20190219-20191017-20191110 from the second interferograms set in the western Haiyuan Fault (HF). (a) Wrapped interferograms. (b) Unwrapped results using MCF. (c) Unwrapped results from LiCSAR. (d) Unwrapped results using SWIPHU.

Fig. 12 displays the unwrapping results of three interferograms from a specific loop, including those obtained from MCF, LiCSAR, and SWIPHU. Note that the results provided by LiCSAR are unwrapped using the SNAPHU method [47]. Differences among the three methods primarily emerge in the upper-right corner of the first interferogram, which are also reflected in the loop closure maps. The errors in this interferogram indicate that both the traditional MCF and SNAPHU methods introduce unwrapping errors when crossing decorrelated regions, whereas SWIPHU method effectively avoids such issues. These results further demonstrate the strong generalizability and robustness of our deep learning model for unwrapping different interseismic interferograms, underscoring its potential for practical applications.

V. Discussion

A. Assessing the Reliability of the Prediction Credibility K_u

The effectiveness of the second post-processing step in the SWIPHU method primarily relies on the prediction credibility index K_u derived from the prediction confidence k_w. Fig. 5 illustrates that the pixel-wise k_w, output by Segformer, represents the uncertainty of the predicted wrap count. To quantitatively assess whether the patch-wise K_u is a reliable indicator for determining post-processing methods in real situations, the relationship between K_u and unwrapping errors is analyzed using loop closures as external validation.

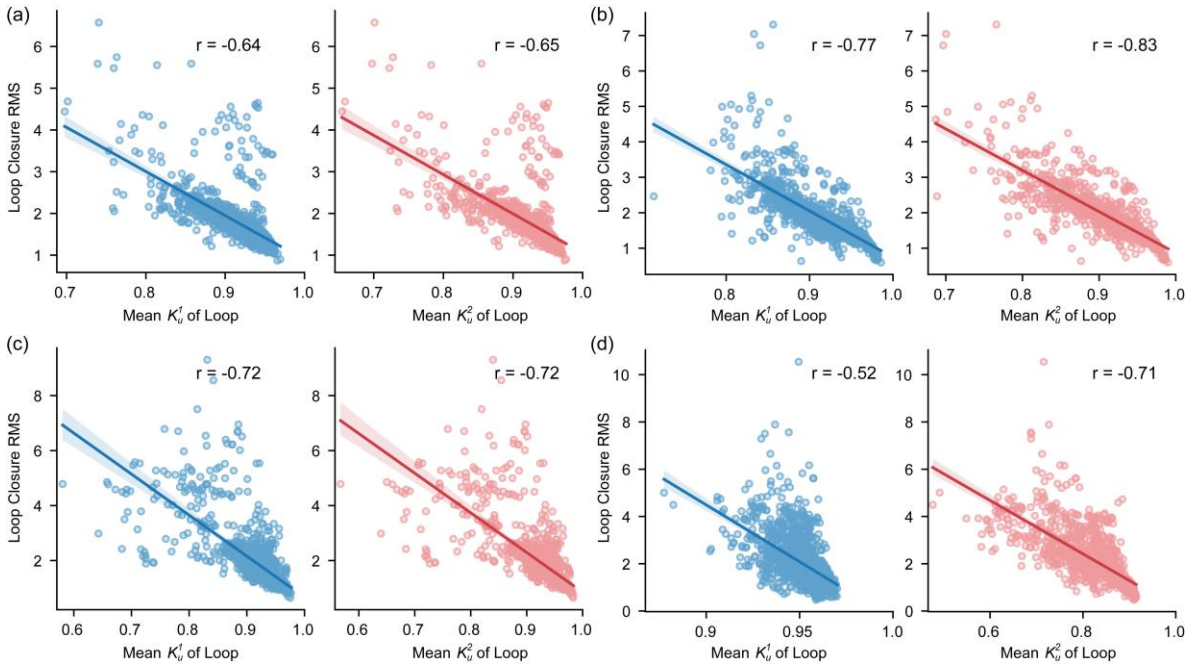

Fig. 13. Scatter plots showing the correlation between the prediction credibility scores (K_u^1 and K_u^2) and loop closure RMS on four representative patches from the interferograms over the western Altyn Tagh Fault (ATF). (a) Patch 0. (b) Patch 1. (c) Patch 2. (d) Patch 6.

We selected four patches that have undergone frequent post-processing during the proposed workflow, namely patches 0, 1, 2, and 6, from the real interferograms at the western end of Altyn Tagh Fault, and conducted the tests separately on each. For each patch, the initial unwrapped results of 763 interferograms were obtained using the Segformer model, and loops were constructed. Since the UEP index ignores the impact of fragmented unmasked pixels, and our objective here is not to compare different unwrapping methods with varying masking strategies, we directly compute the RMS of closure within each patch-wise loop as a proxy for unwrapping error. The K_u^1 and K_u^2 of the three interferograms involved in each loop were averaged to represent the overall prediction credibility. The scatter of the loop closure RMS against the mean K_u^1 and K_u^2 for each patch is plotted in Fig. 13. A clear negative linear correlation is observed across all patches, and the strongest correlation is found in patches 1 and 2, with the Pearson correlation coefficient reaches approximately -0.7 to -0.8. Notably, K_u^2 seems to exhibit a stronger correlation with loop closure RMS than K_u^1, suggesting that the mean k_w of non-decorrelated pixels may serve as a more accurate indicator for characterizing unwrapping errors than the mean k_w of all pixels. Despite minor differences, combining K_u^1 and K_u^2 effectively guides post-processing decisions, ensuring a more stable and accurate wrap-count correction in the second stage of the SWIPHU method.

B. Impact of Partitioning Strategy in Different Methods

In two-dimensional phase unwrapping, larger interferograms are more prone to unwrapping errors, resulting in partitioning strategies being incorporated into some traditional methods [46, 58]. To figure out whether partitioning improves the conventional unwrapping for interseismic interferograms and to compare them with our proposed partitioning-based method, the second set of interferograms in the western Haiyuan Fault was unwrapped again using partitioning modes of MCF and SNAPHU. The patch size used in these methods was matched to that of SWIPHU for fair comparison. The coherence threshold for masking was set to 0.7, as it yielded a mask range most similar to that automatically identified by SWIPHU. The mean UEP for all loops unwrapped via SNAPHU and MCF partitioning methods is 3.6% and 5.4%, respectively, both of which are significantly higher than the 0.7% achieved by our partitioning-based method, as well as higher than the global SNAPHU and MCF methods. This suggests that the partitioning modes integrated in SNAPHU and MCF fail to reduce unwrapping errors and actually introduce additional stitching errors.

To further investigate the impact of our mosaicking algorithm, the patch-wise unwrapped results from MCF were mosaicked using the algorithm described in Section II.D. This process discarded 137 of 846 interferograms due to contradictions in overlaps, highlighting a higher inconsistency rate than with Segformer. The remaining 709 interferograms formed 3074 loops, reducing the mean UEP from 3.5% (using MCF partitioning) to 1.6%. This improvement demonstrates that our mosaicking algorithm enhances accuracy more effectively than MCF's approach and imposes stricter quality control on overlap discrepancies. For those 3074 loops obtained by our whole workflow, we observe a significantly lower mean UEP of 0.6%. This confirms that though the accuracy gains of

SWIPHU can be primarily attributed to the deep learning-based wrap-count segmentation, the third mosaicking step also plays a role in enhancing its accuracy.

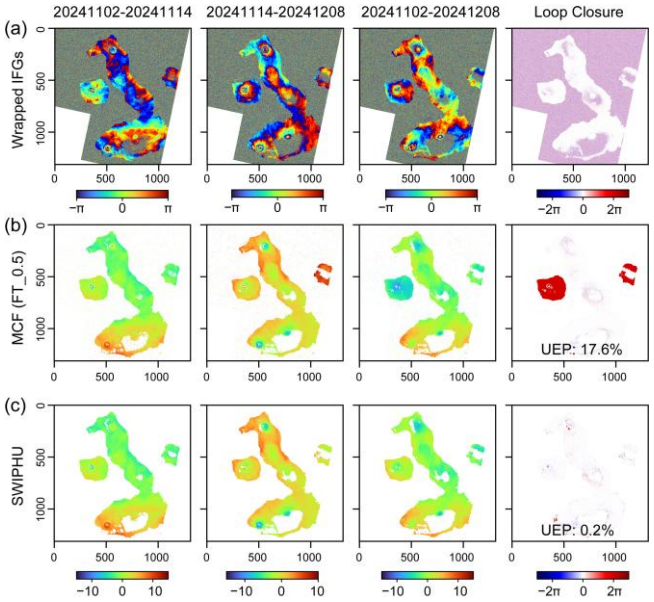

Fig. 14. Unwrapped results and loop closure comparison for Loop 20241102-20241114-20241208 over the western Galápagos Islands. (a) Wrapped interferograms. (b) Unwrapped results using MCF. (c) Unwrapped results using SWIPHU.

C. Application in Volcanic Scenarios

The core principles of the SWIPHU method are universally applicable and can be extended to interferograms containing any type of deformation signal. However, due to the limited training dataset designed for large-scale interseismic interferograms, its primary application in this study has been focused on slow-deforming interseismic cases. To explore the broader applicability of the method, we applied it to interferograms from a volcanic region.

A total of 34 unfiltered interferograms, spanning a four-month period from October 21, 2024, to March 1, 2025, were downloaded from a descending frame covering the western Galápagos Islands volcano region in the LiCSAR portal. These interferograms exhibit relatively slow phase variations compared to those acquired during eruptive events, making them more suitable for testing SWIPHU. Given their smaller size, a reduced patch size of 512 × 512 pixels was used during the unwrapping process. Results show that SWIPHU successfully unwrapped 33 interferograms and formed 41 triangular loops, achieving a mean UEP of 0.6%, a significant improvement over the 4.4% mean UEP obtained using the traditional MCF method with an optimal coherence threshold of 0.5. Fig. 14 depicts the unwrapped results from these two methods within a selected loop. It is obvious that MCF leads to unwrapping biases between different islands in the second and third interferograms due to being surrounded by decorrelated noise, whereas SWIPHU effectively addresses the unwrapping problems of isolated regions. Interestingly, in this case, SWIPHU achieves higher accuracy on unfiltered interferograms than on filtered ones, aligning well with its original design goal—to improve phase unwrapping for interferograms with extensive low-coherence pixels. These results demonstrate that the SWIPHU method also holds great potential for phase unwrapping beyond interseismic studies, particularly in volcanic environments, warranting further exploration and development.

D. Impact on interseismic velocities and tectonic implications

To better quantify the geophysical benefits of the proposed SWIPHU method, we utilized LiCSBAS [18] to derive the large-scale interseismic velocity of the western Haiyuan Fault based on the set of unwrapped interferograms obtained by both MCF (T_{coh} = 0.5) and SWIPHU. Fig. 15 (a) and(b) show the number of unclosed loops identified by LiCSBAS for MCF and SWIPHU, respectively, defined as the number of loops with absolute closure greater than π on a pixel-by-pixel basis. The results confirm that the unwrapping errors of SWIPHU are significantly reduced compared with MCF, as the majority of pixels exhibit fewer than 30 unclosed loops for SWIPHU, whereas more than half of the pixels in MCF exhibit more than 50 unclosed loops. This difference in unclosed loops directly leads to a difference in the RMS of time series inversion: the mean RMS of SWIPHU is 1.2 mm, while that for MCF is 3.0 mm (Fig. 15 (c)-(d)). We removed the Eurasian plate motion using the ITRF Plate Motion Model [45, 59] and fitted a constant between the resulting InSAR velocities and available GNSS stations to obtain GNSS-referenced InSAR LOS velocities in the Eurasian reference frame, as shown in Fig. 15 (e)-(f). The RMS values between the GNSS LOS velocities and the unified InSAR velocities are 1.13 mm/yr and 1.16 mm/yr for SWIPHU and MCF, respectively. Fig. 15 (g) depicts the difference between these two InSAR velocity fields, showing that the difference in long-term linear interseismic velocities reaches up to -0.5 mm/yr to 0.5 mm/yr due to the use of different phase-unwrapping algorithms. To further explore the impact of unwrapping errors on the time series, we plotted the MCF- and SWIPHU-derived InSAR time series at two GNSS stations relative to a common reference point in Fig. 15 (h)-(i), and we assume that the GNSS time series is purely linear. It is evident that the two time series obtained with SWIPHU are more tightly clustered around the linear trend, especially in early 2018, whereas the MCF time series is more scattered and deviates from the GNSS observations. The RMS values between the InSAR and GNSS time series are 4.9 mm and 5.4 mm for SWIPHU and MCF at Point 1, and 7.7 mm and 8.4 mm at Point 2. In other words, the misfit of the time series shows a ~9% reduction when using our proposed SWIPHU phase-unwrapping method, demonstrating improved accuracy of interseismic observations.

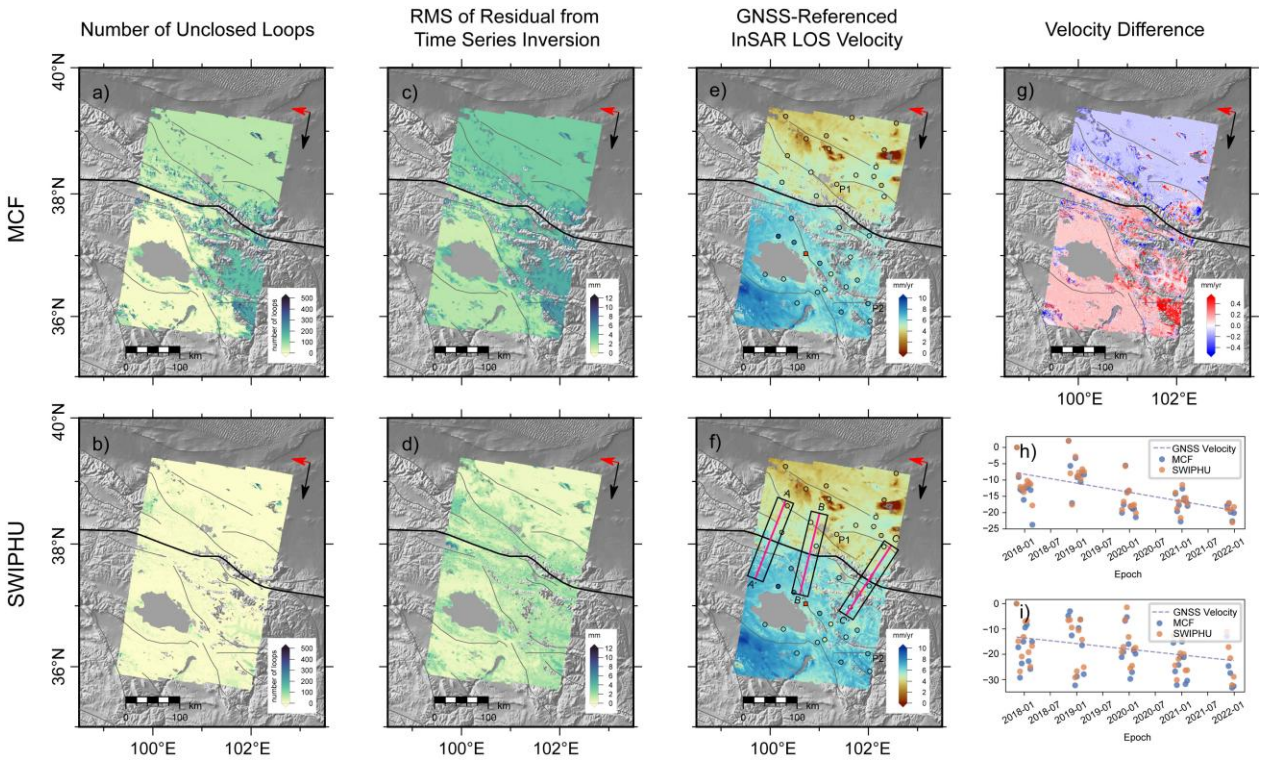

Fig. 15. Comparison of MCF and SWIPHU on time series results. (a)-(b) Number of unclosed loops. (c)-(d) RMS of residual after LiCSBAS time series processing. (e)-(f) InSAR LOS linear velocity after removing plate motion and tied with GNSS in the Eurasia reference. The colored circles display the GNSS velocities projected to the LOS direction. (g) Difference between MCF and SWIPHU-derived InSAR LOS velocity. (h)-(i) Inverted time series between MCF and SWIPHU of P1 and P2 in (f), respectively, where the reference point is the red square in (f).

We extracted three profiles perpendicular to the Haiyuan Fault from the final GNSS-referenced InSAR LOS velocity field derived using SWIPHU to invert interseismic parameters. Each profile is 150 km long and 30 km wide. We then modeled these three profiles using a simple elastic dislocation model [60] with four parameters: s, d, l, and a, denoting the strike slip rate, locking depth, fault location offset, and constant offset, respectively. The inversion is implemented using a Bayesian approach with Markov Chain Monte Carlo sampling [61, 62]. The optimal parameters and their uncertainties for the three profiles are shown in Fig. 16. The strike slip rate near 100°E is 4.4 mm/yr, which is consistent with the 3.5 mm/yr reported by [55] in the same region. For 101°E and 102°E, the strike slip rate remains nearly unchanged, at approximately 2.2 mm/yr. Although the strike slip rates are slightly smaller than existing geodetic and geological slip rate estimates (3-5 mm/yr), the difference is reasonable considering the limited constraint from a single descending orbit and the different strategies for correcting long-wavelength signals. The inferred locking depth decreases from west to east and then increases again, a trend that is in good agreement with previous study [55]. Overall, the inversion results highlight that SWIPHU improves both the stability and tectonic interpretability of InSAR time series, underscoring its potential for broader application to interseismic deformation studies along other active fault systems.

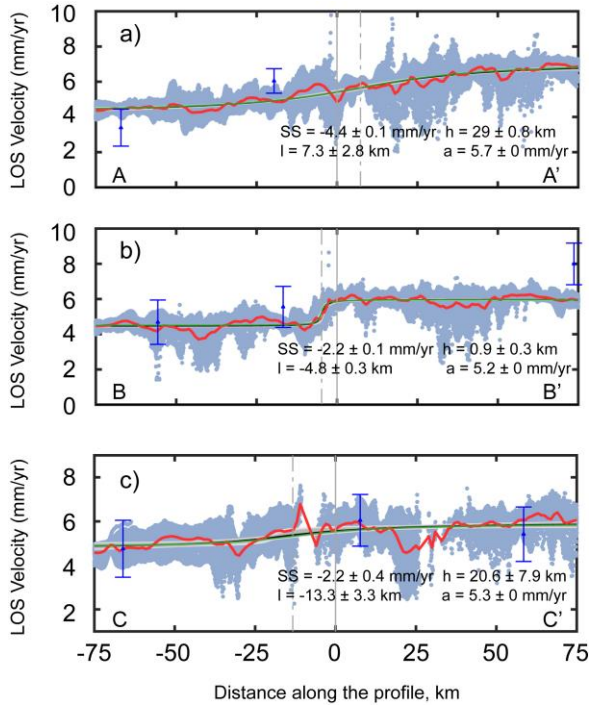

Fig. 16. Fault-perpendicular profiles of descending LOS velocities with inverted results. The locations of the profiles are shown in Fig. 15 (f). Light-blue scatters represent velocities from InSAR measurements. Red dots are the binned average values every 1 km along the profiles. Green lines are the best-

fitting models. Gray dashed lines show the inverted fault locations. Dark-blue points with error bars are GNSS velocities and their uncertainties projected to the LOS direction.

VI. Conclusion

In this work, we describe a novel wrap-count-based phase unwrapping method (SWIPHU) tailored for large-scale, low coherence interferograms. SWIPHU is a machine learning-based approach to phase unwrapping that relies less on traditional approaches than most current neural network-based methods. It eliminates the need for absolute phase gradient estimation and the dependence on the connectivity of pixels in traditional methods, and alleviates the impact of decorrelated pixels on unwrapping accuracy. The SWIPHU method consists of three steps: wrap-count segmentation, post-processing, and mosaicking, enabling deep learning to be effectively leveraged for large-scale interferograms. To facilitate wrap-count segmentation, a comprehensive dataset integrating a diverse set of simulated and real samples was constructed. The transformer-based Segformer model was utilized for the wrap-count segmentation task, demonstrating superior capability in learning from the training dataset compared to other well-established semantic segmentation models. Extensive evaluations on both simulated and real interferograms confirmed that SWIPHU exhibits strong generalization and robustness across thousands of interferograms with varying spatial sizes and different extents of decorrelation. With the advantage of predicting wrap count directly and identifying decorrelated pixels automatically, the phase unwrapping accuracy is significantly improved compared to traditional MCF and SNAPHU methods, especially in isolated regions of interferograms with substantial low-coherence pixels. Furthermore, the reliability analyses of the second post-processing and the third mosaicking steps in the SWIPHU method enhance the overall unwrapping performance. The successful application of this method to volcanic interferograms further underscores its potential for broader geophysical applications. However, SWIPHU has certain limitations. The complexity and variability of real-world data, the limited number of wrap counts that can be identified within a single patch, potentially inaccurate post-processing choices, and stitching errors may all contribute to deviations in the final unwrapped results. Despite some limitations that warrant further refinement, our deep learning-based method advances the applicability of AI for large-scale InSAR phase unwrapping and will lead to improved accuracy in the observation and mapping of diverse earth deformation processes that use geodetic satellite data.

Acknowledgment

The Sentinel-1 SAR images were provided by the European Space Agency (ESA). Some interferograms was downloaded from the COMET-LiCSAR portal (https://comet.nerc.ac.uk/COMET-LiCS-portal/). This work was implemented in the Pytorch version 2.2 and the TensorFlow version 2.0 based on Python 3.7.

References

[1] T. J. Wright, B. Parsons, P. C. England, and E. J. Fielding, "InSAR observations of low slip rates on the major faults of western Tibet," *Science,* vol. 305, no. 5681, pp. 236-239, 2004, doi: 10.1126/science.1096388.

[2] J. Elliott, R. Walters, and T. Wright, "The role of space-based observation in understanding and responding to active tectonics and earthquakes," *Nat. Commun.,* vol. 7, no. 1, p. 13844, 2016, doi: 10.1038/ncomms13844.

[3] W. Xu *et al.*, "Interseismic ground deformation and fault slip rates in the greater San Francisco Bay Area from two decades of space geodetic data," *J. Geophys. Res., Solid Earth,* vol. 123, no. 9, pp. 8095-8109, 2018, doi: 10.1029/2018JB016004.

[4] X. Liu *et al.*, "The Role of Post-Earthquake Fluid Pressure in Driving the 2021 Thessaly (Greece) Aftershock Sequence," *Journal of Earth Science,* vol. 36, no. 6, pp. 2836-2841, 2025, doi: 10.1007/s12583-025-2041-0.

[5] X. Liu *et al.*, "Rheological and tectonic implications of eastern Tibet: Insights from early aftershock sequences driven by afterslip following three 2021–2022 moderate-large events," *International Journal of Applied Earth Observation and Geoinformation,* vol. 143, p. 104770, 2025, doi: 10.1016/j.jag.2025.104770.

[6] J. Biggs, T. Wright, Z. Lu, and B. Parsons, "Multi-interferogram method for measuring interseismic deformation: Denali Fault, Alaska," *Geophys. J. Int.,* vol. 170, no. 3, pp. 1165-1179, 2007, doi: 10.1111/j.1365-246X.2008.03932.x.

[7] J. Fang *et al.*, "Strain partitioning in the Southeastern Tibetan Plateau from kinematic modeling of high-resolution Sentinel-1 InSAR and GNSS," *Geophys. Res. Lett.,* vol. 51, no. 19, p. e2024GL111199, 2024, doi: 10.1029/2024GL111199.

[8] H. Liu, L. Xie, G. Zhao, E. Ali, and W. Xu, "A joint InSAR-GNSS workflow for correction and selection of interferograms to estimate high-resolution interseismic deformations," *Satellite Navigation,* vol. 4, no. 1, p. 14, 2023, doi: 10.1186/s43020-023-00105-6.

[9] K. Itoh, "Analysis of the phase unwrapping algorithm," *Appl. Opt.,* vol. 21, no. 14, pp. 2470-2470, 1982, doi: 10.1364/AO.21.002470.

[10] H. Yu, Y. Lan, Z. Yuan, J. Xu, and H. Lee, "Phase unwrapping in InSAR: A review," *IEEE Geosci. Remote Sens. Mag.,* vol. 7, no. 1, pp. 40-58, 2019, doi: 10.1109/MGRS.2018.2873644.

[11] 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, doi: 10.1029/RS023i004p00713.

[12] M. Costantini, "A novel phase unwrapping method based on network programming," *IEEE Trans. Geosci. Remote Sens.,* vol. 36, no. 3, pp. 813-821, 1998, doi: 10.1109/36.673674.

[13] C. W. Chen and H. A. Zebker, "Network approaches to two-dimensional phase unwrapping: intractability and two new algorithms," *J. Opt. Soc. Am. A, Opt. Image Sci.,* vol. 17, no. 3, pp. 401-414, 2000, doi: 10.1364/JOSAA.17.000401.

[14] 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. Am. A, Opt. Image Sci.,* vol. 18, no. 2, pp. 338-351, 2001, doi: 10.1364/JOSAA.18.000338.

[15] K. Jiang, W. Xu, and L. Xie, "Unwrap intractable C-band coseismic interferograms: An improved SNAPHU method with range offset gradients as prior information," *J. Geophys. Res., Solid Earth,* vol. 129, no. 10, p. e2024JB028826, 2024, doi: 10.1029/2024JB028826.

[16] A. Hooper and H. A. Zebker, "Phase unwrapping in three dimensions with application to InSAR time series," *J. Opt. Soc. Am. A, Opt. Image Sci.,* vol. 24, no. 9, pp. 2737-2747, 2007, doi: 10.1364/JOSAA.24.002737.

[17] E. Hussain, A. Hooper, T. J. Wright, R. J. Walters, and D. P. Bekaert, "Interseismic strain accumulation across the central North Anatolian Fault from iteratively unwrapped InSAR measurements," *J. Geophys. Res., Solid Earth,* vol. 121, no. 12, pp. 9000-9019, 2016, doi: 10.1002/2016JB013108.

[18] Y. Morishita, M. Lazecky, T. J. Wright, J. R. Weiss, J. R. Elliott, and A. Hooper, "LiCSBAS: An open-source InSAR time series analysis package integrated with the LiCSAR automated Sentinel-1 InSAR processor," *Remote Sens.,* vol. 12, no. 3, p. 424, 2020, doi: 10.3390/rs12030424.

[19] Z. Yunjun, H. Fattahi, and F. Amelung, "Small baseline InSAR time series analysis: Unwrapping error correction and noise reduction," *Comput. Geosci.,* vol. 133, p. 104331, 2019, doi: 10.1016/j.cageo.2019.104331.

[20] Z. Ma *et al.*, "Go extra miles: An additional error correction procedure aimed to further improve phase unwrapping accuracy and reduce creep model uncertainty," *J. Geophys. Res., Solid Earth,* vol. 127, no. 1, p. e2021JB022478, 2022, doi: 10.1029/2021JB022478.

[21] X. Yang, H. Wang, C. Pagli, A. H.-M. Ng, and Q. He, "Closure-based correction of InSAR phase unwrapping errors by integrating block-wise Tikhonov regularization and flux analysis," *IEEE Trans. Geosci. Remote Sens.,* vol. 62, Aug. 2024, Art no. 5219215, doi: 10.1109/TGRS.2024.3437757.

[22] A. Benoit, B. Pinel-Puysségur, R. Jolivet, and C. Lasserre, "CorPhU: An algorithm based on phase closure for the correction of unwrapping errors in SAR interferometry," *Geophys. J. Int.,* vol. 221, no. 3, pp. 1959-1970, 2020, doi: 10.1093/gji/ggaa120.

[23] L. Zhou, H. Yu, and Y. Lan, "Artificial intelligence in interferometric synthetic aperture radar phase unwrapping: A review," *IEEE Geosci. Remote Sens. Mag.,* vol. 9, no. 2, pp. 10-28, 2021, doi: 10.1109/MGRS.2021.3065811.

[24] K. Wang, Y. Li, Q. Kemao, J. Di, and J. Zhao, "One-step robust deep learning phase unwrapping," *Opt. Express,* vol. 27, no. 10, pp. 15100-15115, 2019, doi: 10.1364/OE.27.015100.

[25] 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, pp. 1-18, 2021, doi: 10.1109/TGRS.2021.3121907.

[26] L. Zhou, H. Yu, V. Pascazio, 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, pp. 1-10, 2022, doi: 10.1109/TGRS.2022.3145342.

[27] L. Zhou and H. Yu, "MoDL-PU: Model-Based Deep Learning for InSAR Phase Unwrapping," *IEEE Trans. Geosci. Remote Sens.,* 2025, doi: 10.1109/TGRS.2025.3549607.

[28] L. Zhou, H. Yu, Y. Lan, and M. Xing, "Deep learning-based branch-cut method for InSAR two-dimensional phase unwrapping," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1-15, 2021, doi: 10.1109/TGRS.2021.3099997.

[29] 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, pp. 1-16, 2021, doi: 10.1109/TGRS.2021.3121906.

[30] 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, 2020, doi: 10.1109/TGRS.2020.2965918.

[31] 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. Top. Appl. Earth Observ. Remote Sens.,* vol. 14, pp. 7840-7856, 2021, doi: 10.1109/JSTARS.2021.3099485.

[32] T. Zhang *et al.*, "Rapid and robust two-dimensional phase unwrapping via deep learning," *Opt. Express,* vol. 27, no. 16, pp. 23173-23185, 2019, doi: 10.1364/OE.27.023173.

[33] G. Spoorthi, R. K. S. S. 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, 2020, doi: 10.1109/TIP.2020.2977213.

[34] D. Zhou and Z. Zhao, "Fast InSAR phase unwrapping method for complex mountainous areas with high noise and large gradient changes," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.,* 2024, doi: 10.1109/JSTARS.2024.3417417.

[35] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS-J. Photogramm. Remote Sens.,* vol. 145, pp. 60-77, 2018, doi: 10.1016/j.isprsjprs.2018.04.014.

[36] W. Chen, Z. Jiang, Z. Wang, K. Cui, and X. Qian, "Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8924-8933.

[37] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inform. Process. Syst.,* vol. 30, 2017.

[38] K. Han *et al.*, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 45, no. 1, pp. 87-110, 2022, doi: 10.1109/TPAMI.2022.3152247.

[39] X. He, Y. Zhou, J. Zhao, D. Zhang, R. Yao, and Y. Xue, "Swin transformer embedding UNet for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1-15, 2022, doi: 10.1109/TGRS.2022.3144165.

[40] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Adv. Neural Inform. Process. Syst.,* vol. 34, pp. 12077-12090, 2021.

[41] Z.-F. Ma, S.-J. Wei, Y. Aoki, J.-H. Liu, and T. Huang, "A new spatiotemporal InSAR tropospheric noise filtering: An interseismic case study over central san Andreas fault," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1-16, 2022, doi: 10.1109/TGRS.2022.3206540.

[42] Y. Okada, "Surface deformation due to shear and tensile faults in a half-space," *Bull. Seismol. Soc. Amer.,* vol. 75, no. 4, pp. 1135-1154, 1985, doi: 10.1785/BSSA0750041135.

[43] Y. Wang *et al.*, "Refined InSAR tropospheric delay correction for wide-area landslide identification and monitoring," *Remote Sens. Environ.,* vol. 275, p. 113013, 2022, doi: 10.1016/j.rse.2022.113013.

[44] K. Perlin, "An image synthesizer," *ACM Siggraph Computer Graphics,* vol. 19, no. 3, pp. 287-296, 1985, doi: 10.1145/325165.325247.

[45] O. L. Stephenson, Y. K. Liu, Z. Yunjun, M. Simons, P. Rosen, and X. Xu, "The impact of plate motions on long-wavelength InSAR-derived velocity fields," *Geophys. Res. Lett.,* vol. 49, no. 21, p. e2022GL099835, 2022, doi: 10.1029/2022GL099835.

[46] C. Werner, U. Wegmüller, T. Strozzi, and A. Wiesmann, "Gamma SAR and interferometric processing software," in *Proceedings of the ers-envisat symposium, Gothenburg, Sweden*, 2000, vol. 1620: Citeseer, p. 1620.

[47] M. Lazecký *et al.*, "LiCSAR: An automatic InSAR tool for measuring and monitoring tectonic and volcanic activity," *Remote Sens.,* vol. 12, no. 15, p. 2430, 2020, doi: 10.3390/rs12152430.

[48] C. Huynh, A. T. Tran, K. Luu, and M. Hoai, "Progressive semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 16755-16764.

[49] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 2015: Springer, pp. 234-241.

[50] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Eur. Conf. Comput. Vis.*, 2018, pp. 801-818.

[51] C. Liu, L. Ji, L. Zhu, C. Xu, and J. Qiu, "Interseismic strain rate distribution model of the Altyn Tagh Fault constrained by InSAR and GPS," *Earth Planet. Sci. Lett.,* vol. 642, p. 118884, 2024, doi: 10.1016/j.epsl.2024.118884.

[52] L. Shen, A. Hooper, J. R. Elliott, and T. J. Wright, "Variability in interseismic strain accumulation rate and style along the Altyn Tagh Fault," *Nat. Commun.,* vol. 15, no. 1, p. 6876, 2024, doi: 10.1038/s41467-024-51116-z.

[53] D. Wang, J. R. Elliott, G. Zheng, T. J. Wright, A. R. Watson, and J. D. McGrath, "Deciphering interseismic strain accumulation and its termination on the central-eastern Altyn Tagh fault from high-resolution velocity fields," *Earth Planet. Sci. Lett.,* vol. 644, p. 118919, 2024, doi: 10.1016/j.epsl.2024.118919.

[54] Y. Maghsoudi, A. J. Hooper, T. J. Wright, M. Lazecky, and H. Ansari, "Characterizing and correcting phase biases in short-term, multilooked interferograms," *Remote Sens. Environ.,* vol. 275, p. 113022, 2022, doi: 10.1016/j.rse.2022.113022.

[55] Z. Huang, Y. Zhou, X. Qiao, P. Zhang, and X. Cheng, "Kinematics of the∼1000 km Haiyuan fault system in northeastern Tibet from high-resolution Sentinel-1 InSAR velocities: Fault architecture, slip rates, and partitioning," *Earth Planet. Sci. Lett.,* vol. 583, p. 117450, 2022, doi: 10.1016/j.epsl.2022.117450.

[56] R. Jolivet *et al.*, "Shallow creep on the Haiyuan fault (Gansu, China) revealed by SAR interferometry," *J. Geophys. Res., Solid Earth,* vol. 117, no. B6, 2012, doi: 10.1029/2011JB008732.

[57] Q. Ou *et al.*, "Large-scale interseismic strain mapping of the NE Tibetan Plateau from Sentinel-1 interferometry," *J. Geophys. Res., Solid Earth,* vol. 127, no. 6, p. e2022JB024176, 2022, doi: 10.1029/2022JB024176.

[58] C. W. Chen and H. A. Zebker, "Phase unwrapping for large SAR interferograms: Statistical segmentation and generalized network models," *IEEE Trans. Geosci. Remote Sens.,* vol. 40, no. 8, pp. 1709-1719, Aug. 2002, doi: 10.1109/TGRS.2002.802453.

[59] Z. Altamimi, L. Métivier, P. Rebischung, H. Rouby, and X. Collilieux, "ITRF2014 plate motion model," *Geophys. J. Int.,* vol. 209, no. 3, pp. 1906-1912, 2017, doi: 10.1093/gji/ggx136.
[60] J. Savage and R. Burford, "Geodetic determination of relative plate motion in central California," *J. Geophys. Res.,* vol. 78, no. 5, pp. 832-845, 1973, doi: 10.1029/JB078i005p00832.
[61] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The journal of chemical physics,* vol. 21, no. 6, pp. 1087-1092, 1953, doi: 10.1063/1.1699114.
[62] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika,* vol. 57, no. 1, pp. 97-109, 1970, doi: 10.1093/biomet/57.1.97.

Kun Jiang received the bachelor's degree from the School of Geosciences and Info-Physics, Central South University, Changsha, China, in 2021, and he is pursuing his Ph.D. degree in geodesy and surveying engineering at Central South University. Since October 2024, he has also been a visiting Ph.D. student with COMET, School of Earth and Environment, University of Leeds, U.K.

His research interests include InSAR phase unwrapping, the application of deep learning in InSAR data processing, and large-scale interseismic deformation monitoring.

Wenbin Xu received the Ph.D. degree in geophysics from King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, in 2015.

Between 2016 and 2019, he was with the Berkeley Seismological Laboratory, University of California, Berkeley and the Department of Land Surveying and Geo-Informatics, the Hong Kong Polytechnic University, Kowloon, Hong Kong. Since June 2019, he has been a Professor at the School of Geoscience and Info-Physics, Central South University, Changsha, China, where he leads the volcano and earthquake research laboratory. His primary research interests focus on the application of space geodesy to study crustal deformation associated with a variety of geophysical and anthropogenic processes. Dr. Xu was the recipient of the John Wahr Eearly Career Award by the American Geophysical Union, which recoginizes early or mid-career scientists who have made outstanding contributions to geodesy.

Andrew Hooper (Senior Member, IEEE) received the M.S. and Ph.D. degrees in geophysics from Stanford University, Stanford, CA, USA, in 2002 and 2006, respectively, for his development of SAR persistent scatterer interferometry algorithms for geophysical applications.

Between 2002 and 2003, he was with the German Space Center (DLR) on algorithms for the TerraSAR-X satellite. From 2006, he was with University of Iceland, before moving to Delft University of Technology in 2008. Since 2013, he has been a Professor of geophysics and geodesy with the University of Leeds, Leeds, U.K. His research interests include measuring and modeling deformation due to volcanic and tectonic processes, and the application of machine learning techniques to deformation data. Dr. Hooper was the recipient of the James B. Macelwane medal by the American Geophysical Union, given to outstanding early career scientists who have shown depth, breadth, impact, creativity and novelty in their research, in 2016. He is the Co-Founder and the Director of university spin-out SatSense.

Lei Xie received the bachelor's degree in surveying and mapping engineering from China University of Geosciences, Wuhan, China, in 2014, the master's degree in geomatics engineering from the University of Calgary, Calgary, AB, Canada, in 2018, and the Ph.D. degree in geodesy from the Hong Kong Polytechnic University, Hong Kong, in 2022.

He is currently a Lecturer with the School of Geosciences and Info-Physics, Central South University, Changsha, China. His research interests include InSAR, geodynamics, and applications in infrastructure monitoring.