An Intelligent Learning Reconfiguration Model Based on Optimized Transformer and Multisource Features (TMSFs) for High-Precision InSAR DEM Void Filling

Tengfei Zhang, *Graduate Student Member, IEEE*, Yumin Chen, Rui Zhu, John P. Wilson, Jun Song, Ruoxuan Chen, Licheng Liu, and Lanhua Bao

Abstract—The synthetic aperture radar (SAR) systems can provide submeter terrain mapping and accurate point elevation information quickly and efficiently. The interferometric SAR (InSAR) technology has proven to be a powerful method for producing digital elevation models (DEMs). However, DEM generation using InSAR technology is limited by mountain shadow overlap, atmospheric noise, low backscatter coefficient, and spatiotemporal incoherence, leading to the problem of voids. This article proposes an intelligent learning reconfiguration model based on optimized transformer and multisource features (TMSFs). First, the intelligent learning reconfiguration model based on the transformer and convolutional neural network (CNN) was constructed, and the multisource feature connection module was used for feature supervision and loss function optimization. Then, the relationship of nonvoid areas between the low-resolution (LR) DEM and the high-resolution (HR) InSAR DEM was found, and the voids were intelligently filled. The experiments used 19 TerraSAR-X images in San Diego (SD), USA, and 18 PAZ images in Yan'an (YA), China, to generate high-precision InSAR void DEMs and intelligently fill the voids. Compared with traditional interpolation or deep learning models, modeling accuracy improved by 11.31%–45.74% and 2.32%–8.78% in the SD and YA areas, respectively. Using the photogrammetric DEM to evaluate the accuracy of the filled DEM, the new method showed improvements of 15.64%–25.91% and 5.60%–28.26%, respectively. In addition, 122 Ice, Cloud, and land Elevation Satellite (ICESat)/Geosciences Laser Altimeter System (GLAS) points collected in the YA area were further validated, with an improvement of 4.40%–22.28%. The generated DEM has considerable advantages for terrain feature preservation and river network extraction, and the new method can provide technical support for DEM void filling.

Tengfei Zhang, Yumin Chen, Ruoxuan Chen, Licheng Liu, and Lanhua Bao are with the School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China (e-mail: tengfeizhang@whu.edu.cn; ymchen@whu.edu.cn; 2023202050017@whu.edu.cn; 2020302051126@whu.edu.cn; lanhuabao@whu.edu.cn).

Rui Zhu is with the Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore 138632 (e-mail: zhur@ihpc.a-star.edu.sg).

John P. Wilson is with the Spatial Sciences Institute, University of Southern California, Los Angeles, CA 90089 USA (e-mail: jpwilson@usc.edu).

Jun Song is with the Department of Geography, Hong Kong Baptist University, Hong Kong, China (e-mail: junsong@hkbu.edu.hk).

Digital Object Identifier 10.1109/TGRS.2025.3535546

I. INTRODUCTION

HIGH-PRECISION digital elevation models (DEMs) play an essential part in geographic information architecture, widely used in urban operations management, flood simulation and prediction, landslide and mudslide risk assessment, and 3-D real scene construction [1], [2], [3]. Traditional high-precision DEM acquisition methods include topographic mapping, aerial photogrammetry, and light detection and ranging (LiDAR) [4]. However, their susceptibility to weather conditions, poor coverage, and high cost have limited their use in large-scale, high-precision DEM production. The synthetic aperture radar (SAR) technology can operate in all weather conditions, all day, and is unaffected by clouds, rain, and fog. In addition, the high-resolution (HR) satellite-based SAR systems can provide submeter terrain mapping and quickly and effectively acquire the point elevation information needed to construct high-precision DEMs [5].

The main methods that use interferometric SAR (InSAR) technology to produce high-precision DEMs include two-pass, multibaseline, and bistatic InSAR. The accuracy of traditional two-pass InSAR is limited and affected by factors, such as SAR platform geometry, spatiotemporal incoherence, atmospheric delay and noise, and unwrapping errors [6]. Currently, multibaseline and bistatic InSAR are the main methods for producing high-precision DEMs using InSAR technology. In terms of multibaseline InSAR, Eineder and Adam proposed a method to generate DEMs without phase unwrapping, and the results show that this is more applicable in rugged mountainous terrain [7]. Zhao et al. [8] incorporated Shuttle Radar Topography Mission (SRTM) DEM data into the process of DEM reconstruction from TerraSAR-X data, which effectively improved the accuracy of the generated DEM. Du et al. [9] used simulated data to analyze the errors of traditional InSAR DEM generation algorithms, such as linear and polynomial models. Multibaseline InSAR DEM generation methods have achieved many successes, but they are still affected by atmospheric delays and noise, baseline errors, and deformation modeling errors. The bistatic InSAR technology

was introduced with the launch of the TerraSAR-X/TanDEM-X satellites. The main characteristics of the satellite are that it can acquire interferometric pairs with a "zero" time baseline and long spatial baseline, effectively overcoming the effects of spatiotemporal incoherence, atmospheric delay and noise, and deformation modeling errors. Lachaise et al. [10] investigated the TerraSAR-X/TanDEM-X data processing flow, which shows high accuracy in areas with simple topography. Lachaise et al. [11] propose a dual-baseline phase unwrapping (DB-PUC) framework, and experimental results show that the DB-PUC framework processes TanDEM-X data with high efficiency and accuracy. Eldhuset [12] estimate the coherence of different components and extract more topographic information from TerraSAR-X/TanDEM-X interferometric data, which is essential for improving the quality of high-precision DEMs. However, the DEMs generated by the above-mentioned methods are subject to the problem of voids due to the limitations of the terrain conditions, the scattering characteristics of the features, and the temporal phase variations in the SAR images.

The presence of voids in DEMs is an inherent limitation of optical remote sensing, aerial photogrammetry, and SAR techniques, and significantly affects the quality of current DEMs. Fig. 1 shows the void areas in the 3-m DEM generated from TerraSAR-X data. Currently, single SAR data fusion and multisource data fusion are the main methods used to fill the voids in DEMs. In terms of single SAR data fusion, most studies focus on the fusion of ascending and descending data by selecting the optimal angle of incidence according to the geometric relationship of the satellite. Deo et al. [13] determine the optimal fusion incidence angle based on the statistical and topographic information to achieve weighted DEM fusion. Zhang et al. [14] used an improved polynomial model combined with ascending and descending data fusion to achieve DEM generation for Heifangtai, China. Dong et al. [15] used a cascaded multibaseline interferometric approach with the weighted fusion of ascending and descending data to achieve high-precision DEM generation over Songshan, China. The ascending and descending data fusion methods are more effective in dealing with mountain shadow and overlap, but is not well suited for areas affected by atmospheric and noise, low backscatter coefficient, and spatiotemporal incoherence problems. In terms of multisource data fusion, Shen et al. [16] proposed an elevation difference fitting neural network method to fuse advanced spaceborne thermal emission and reflection radiometer (ASTER) global digital elevation model (GDEM) and SRTM source data to generate seamless DEM. Yue et al. [17] used a regularized framework to fuse DEM data from multiple sources and scales to produce seamless DEMs. Liu et al. [18] extracted a high-precision DEM of the Antarctic area from TanDEM-X interferometric data using an iterative sub-InSAR technique. The multisource data fusion method can fill the void areas, but the existing multisource data have different accuracies, inconsistent data formats, and large geographical location deviations. In addition, the fusion of the acquired high-precision DEM has the problems of elevation error, insufficient enhancement of high-frequency information, over-smoothing of complex topographic areas, and loss of topographic features.
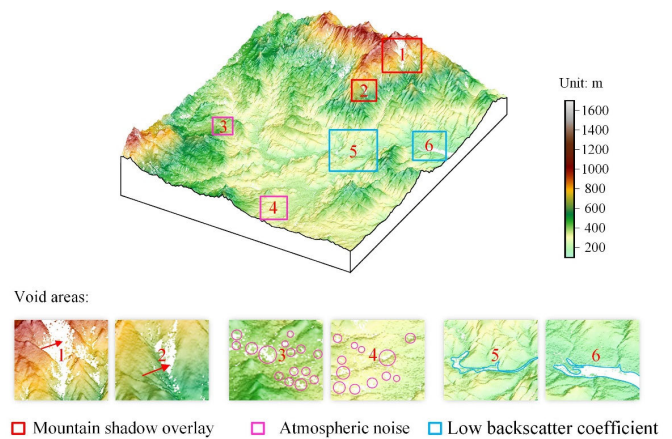

Fig. 1.   Void areas in a 3-m DEM generated with TerraSAR-X data.

With the development of geospatial artificial intelligence (GeoAI), DEM super-resolution (SR) reconstruction plays an important role in high-precision DEM generation. Chen et al. [19] used a three-layer convolutional neural network (CNN) to achieve reconstruction from low-resolution (LR) to HR DEM. Zhou et al. [20] introduced deep CNN and multiterrain features to achieve high-precision DEM reconstruction. Researchers have attempted to demonstrate the feasibility of using these deep learning networks to fill DEM voids [20]. Liu et al. [21] proposed a deep CNN that combines intensity information from SAR data and interferometric phases to generate high-precision DEMs. Chen et al. [22] introduced a residual network with an attention module (AM) for DEM reconstruction, which was used to fill the DEM voids at 5-m resolution. Compared with existing fusion approaches, deep learning-based DEM-SR reconstruction has achieved better performance, but the accuracy of the reconstruction process needs further improvement. In addition, the high-frequency information in the void areas is insufficient, leading to inconsistencies between the results and the actual terrain.

Based on the above analysis, the main challenges for improving the quality of high-precision DEM using AI are as follows.

1) The missing data problem severely affects the DEM quality, and the reasons for this problem in InSAR DEMs are complex and multifaceted, due to problems such as mountain shadow overlap, atmospheric noise, low backscatter coefficient, and spatiotemporal incoherence. Dealing with different types of missing data is extremely challenging.

2) The traditional void-filling method lacks the high-frequency information of the void areas, leading to a large number of elevation errors. The effectiveness of the DEM void filling needs to be further improved.

3) The method based on DEM-SR reconstruction used to fill the void areas needs further refinement and improvement.

To address this limitation, we propose an intelligent learning reconfiguration model based on optimized transformer and multisource features (TMSFs) to fill the void problem in high-precision InSAR DEMs. First, the intelligent learning reconfiguration model based on the transformer and CNN was constructed, and the multisource feature connection module
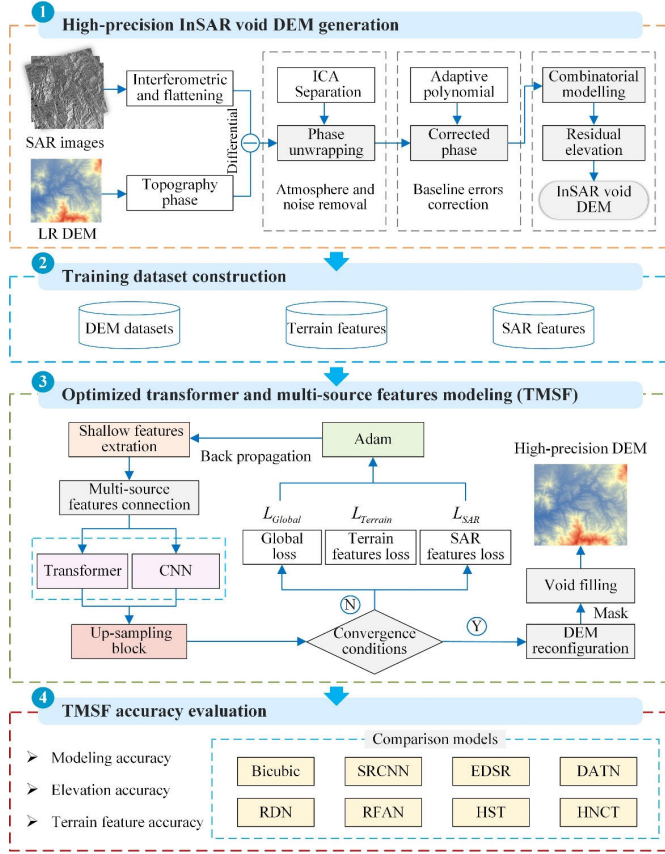
Fig. 2. Workflow of optimized TMSFs.

was used for feature supervision and loss function optimization. Then, the relationship of nonvoid areas between the LR and HR InSAR DEMs was found, and the voids were intelligently filled by TMSF. The multisource feature connection module can effectively increase the high-frequency information in the void areas, and significantly improve the quality of the DEM. The generated high-precision DEM products have significant advantages in elevation accuracy and terrain feature preservation. It provides high-precision DEM 3-D modeling products for geological disasters, environmental protection, soil and water conservation, and other applications.

II. METHODOLOGY

In this article, a TMSF framework is proposed to solve the void problem in high-precision InSAR DEMs. Specifically, the research work includes the following steps: 1) high-precision InSAR void DEM generation; 2) training dataset construction; 3) optimized TMSFs modeling; and 4) TMSF accuracy evaluation. The research workflow is shown in Fig. 2.

A. High-Precision InSAR Void DEM Generation

Multibaseline and bistatic InSAR are the main methods for high-precision InSAR DEM generation. During SAR data preprocessing, the dense interference fringes caused by the topography phase can seriously affect the accuracy of the phase unwrapping. Researchers have proposed to remove the topographic phase using the LR DEM [expressed as $h_{\text{ref}}(x, y)$], and then use the differential interferometric phase to solve the residual elevation between the HR and LR DEMs

[expressed as $\Delta h(x, y)$]. The interferometric phase is written as follows [23], [24], [25], [26]:

$$\delta \varphi_i^m = \delta \varphi_{\text{def}}^i + \delta \varphi_{\text{topo}}^i + \delta \varphi_{\text{orbit}}^i + \delta \varphi_{\text{atm}}^i + \delta \varphi_{\text{noise}}^i + \delta \varphi_{\text{hp}}^i \quad (1)$$

where $\delta \varphi_i^m$ represents the phase at the ith pixel on the mth interferogram ($1 \leq m \leq M$), M represents the number of interferometric pairs. $\delta \varphi_{\text{def}}^i$ represents the deformation phase $\delta \varphi_{\text{def}}^i = 4\pi \cdot \Delta d(x, y)/\lambda$, λ is the radar wavelength, $\Delta d(x, y)$ is the line-of-sight (LOS) deformation. $\delta \varphi_{\text{orbit}}^i$ represents the orbit error. $\delta \varphi_{\text{atm}}^i$ and $\delta \varphi_{\text{noise}}^i$ represents the atmospheric delay and noise phase, respectively. $\delta \varphi_{\text{hp}}^i$ represents the high-pass (HP) deformation phase. $\delta \varphi_{\text{topo}}^i$ represents the topography phase, which can be written as follows:

$$\delta \varphi_{\text{topo}}^i = \frac{4\pi B_\perp^m}{\lambda R \sin \theta} \Delta h(x, y) \quad (2)$$

where R represent the sensor-target distance. θ and $B_\perp^m$ represent the perpendicular baseline and radar incidence angle, respectively. It is worth noting that the bistatic InSAR can calculate the residual elevation directly using (2). The multi-baseline InSAR technique requires the combination of (1) and (2) to build an InSAR model that considers linear deformation, seasonal deformation, and environmental factors [27], and then solves the deformation parameters and residual elevation by least squares (LSs) and singular value decomposition (SVD) [24], [28]. In addition, the independent component analysis (ICA) was used to overcome the effects of atmosphere and noise, and the polynomial fitting was used to remove the baseline error. After obtaining the residual elevations, the elevation on the surface can be written as follows [29]:

$$z(x, y) = h_{\text{ref}}(x, y) + \Delta h(x, y). \quad (3)$$

The elevation information of the high-precision InSAR DEM can be obtained using (3), and then, the high-precision InSAR void DEM corresponding to the actual geography can be generated by geocoding and reprojection.

B. Training Dataset Construction

The training dataset consists of LR and HR training datasets, as shown in Fig. 3. Specifically, it includes the DEM, terrain feature, and SAR feature training datasets.

1) DEM Training Datasets: A wide range of LR DEMs are available for free download worldwide, including 90-, 30-, and 12.5-m DEMs. The suitable data can be selected according to the study area and actual needs. In this study, the high-precision InSAR void DEM obtained in Section II-A was used as the HR DEM, and the advanced land observing satellite (ALOS) 12.5-m DEM was used as the LR DEM. To avoid the influence of forests and buildings in different bands of SAR data (L, X, and P) on the continuity and accuracy of the elevations in the DEMs, filtering or the triangulated irregular network (TIN) method can be used after generating high-precision InSAR void DEMs. To facilitate the calculation of the deep learning network, the resolution of the LR and HR DEMs must satisfy the multiplier relation. Therefore, the interpolation or resampling method is chosen to make the DEM data meet the multiplier requirements [20]. The LR and HR DEMs were clipped to a pixel size of 64 × 64 and 64 × 64 s, respectively, where s is the ratio of the resolution between the LR and HR DEMs.
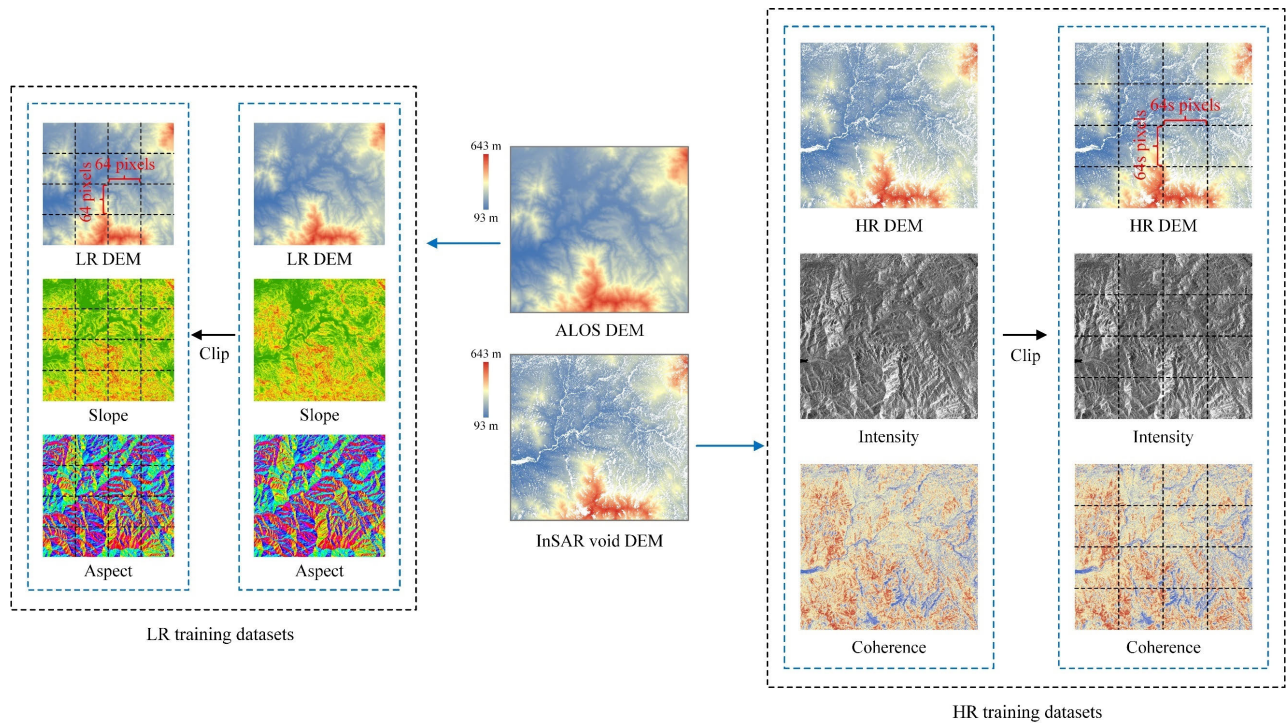
Fig. 3. Training dataset construction.

2) Terrain Feature Training Datasets: Calculate the terrain features corresponding to the LR DEM using the feature extraction operators. Clip the terrain feature using the LR DEM training datasets clip method, and then add the clipped terrain feature to the training datasets. The main terrain features used are slope and aspect, which can improve the accuracy of intelligent learning reconfiguration and preserve the detail and structure of the landform.

3) SAR Feature Training Datasets: The intensity map and coherence map generated in Section II-A were used as SAR features. The intensity feature represents the intensity of the response of the features and landforms in the SAR image to the echo signals, calculates the average of the intensity information at different times, and performs the normalization process. The coherence feature reflects the stability of the InSAR phase, calculates the average of the intensity information at different times, and performs the normalization process. The SAR features were cropped and added to the training datasets according to the cropping method of the HR DEM training datasets.

C. Optimized TMSF Modeling

Combining the DEM datasets and the characteristics of the voids caused by the missing data area, the intelligent learning reconstruction model of TMSF was constructed, as shown in Fig. 4. Specifically, it includes the shallow feature extraction, multisource features connection, transformer and CNN feature fusion block, up-sampling block, and loss function block.

1) Shallow Feature Extraction: The design structure consists of a 3×3 convolutional layer and a shallow feature matrix. Shallow features are extracted from the LR DEM by a 3×3 convolutional layer, which processes the DEM

information from low-to-high dimensions and generates the corresponding shallow feature matrix. Assuming that the input LR DEM is D_{LR}, the extracted shallow feature F_0 can be expressed as follows [22]:

$$F_0 = H_{SFE}(D_{LR}) = W_0 * D_{LR} \tag{4}$$

where H_{SFE} represents the convolutional layer with the weight of W_0, W_0 represents the weight of the 3×3 convolutional for shallow feature extraction, and $*$ represents the convolutional operation.

2) Multisource Features Connection: The multisource features connection module was introduced into the TMSF model to improve the training accuracy. The multisource features accessed include terrain (slope and aspect) and SAR features (intensity and coherence). The feature matrices F_{TS} are obtained after normalization of the terrain and SAR features, respectively. Since the resolution of SAR features differs from that of terrain features, they cannot be fused with terrain features by simple convolution, so the terrain and SAR features were fused using the feature fusion module. The feature fusion module contains two convolutional layers and a rectified linear unit (ReLU), as shown in Fig. 5. The fused features F_{TS} are then concatenated with the shallow feature matrix F_0 to generate a new shallow feature F_{New}, which can be expressed as follows:

$$F_{New} = W_{New}([F_0, F_{TS}]) + b \tag{5}$$

where b represent the weights of the 3×3 convolutional layers for multisource features fusion.

3) Transformer and CNN Feature Fusion Block: To better extract the deep features and enhance the information of the DEM, an attention transformer and CNN feature fusion
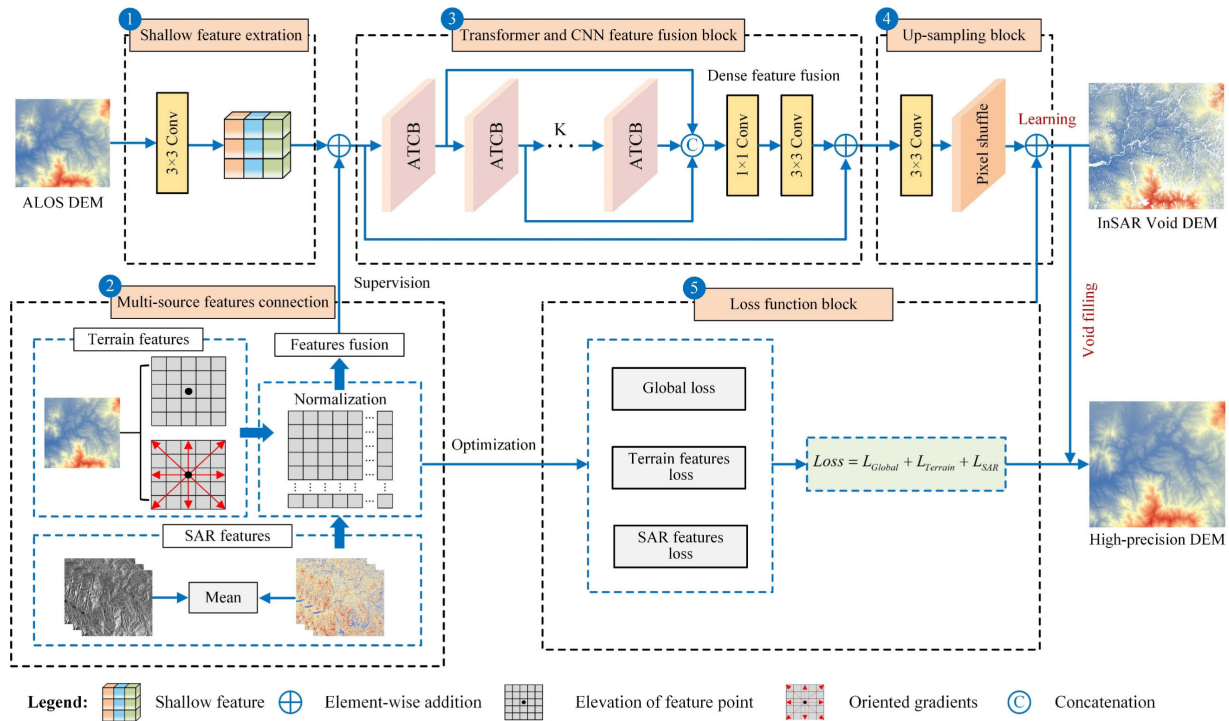
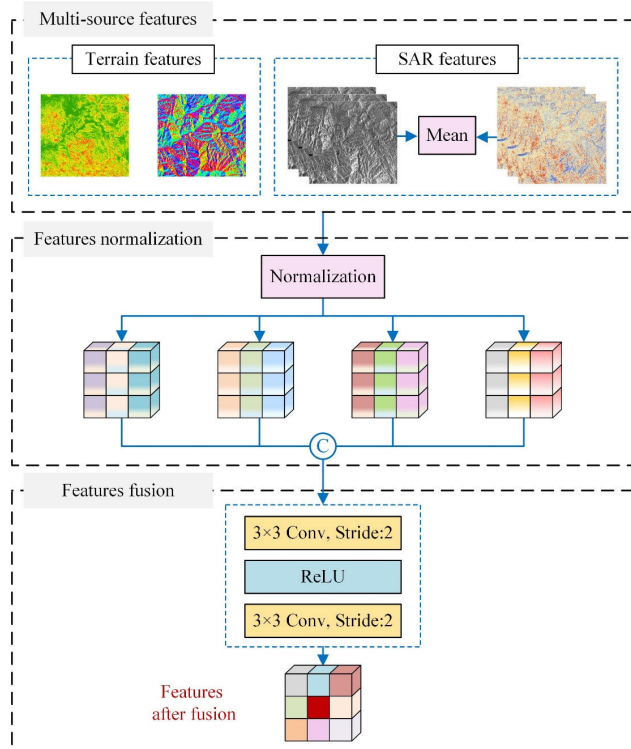Fig. 4. Architecture of the proposed TMSF for DEM void filling.

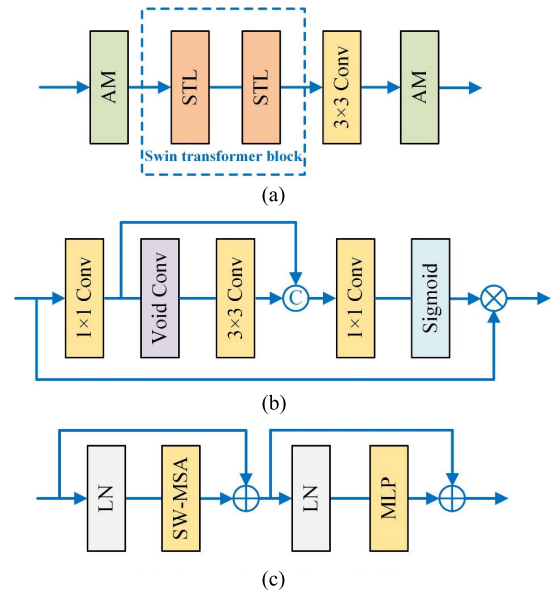Fig. 5. Architecture of the multisource features connection.

Fig. 6. (a) ATCB, (b) AM, and (c) STL.

the deep feature matrix was obtained. ATCB module, AM, and swin transformer layer (STL) module are shown in Fig. 6.

The ATCB module consists of two AMs, a swin transformer block (STB) consisting of two STLs, and a 3×3 convolutional layer. The feature F_k output from the kth ATCB ($1 \leq k \leq K$) module can be expressed as follows:

$$F_k = f_k^{\text{ATCB}}\left(f_{k-1}^{\text{ATCB}}, \ldots, \left(\left(f_1^{\text{ATCB}}(F_{\text{New}})\right)\right)\right) \tag{6}$$

where f_k^{ATCB}, f_{k-1}^{ATCB}, and f_1^{ATCB} represent the nonlinear functions of the features extracted by the kth, $(k-1)$th, and lth ATCB module, respectively. When a feature F_{k-1} is input,

module were designed, the main component of which is the attention transformer and CNN block (ATCB). The new shallow feature F_{New} was added to the transformer and CNN feature fusion block. After passing through K ATCB modules, a 1×1 convolutional layer, and a 3×3 convolutional layer,

the kth ATCB module first extracts the important features from the attention structure AM, then extracts the intermediate features through the STL structure, and then passes through a 3×3 convolutional layer to preserve performance. Finally, it introduces an attention structure AM to weight the locally important areas and improve the smoothness of the terrain. The AM consists of convolutional layers with different sizes and windows, and a sigmoid activation function, which is used to enhance the representational information of the features and increase the capability and efficiency of training. The STL consists of a two-layer nodes (LNs), a shifted windowing-based multihead self-attention (SW-MSA), and a multilayer perceptron (MLP).

Finally, the features output from each ATCB module are combined using deep feature fusion and effectively spliced in the channel dimension. The deep feature fusion consists of a 1×1 convolutional layer and a 3×3 convolutional layer to fuse features. In addition, the residual aggregation structure was used to facilitate learning. The feature after deep feature fusion F_D can be expressed as follows [30]:

$$F_D = W_1(\text{PReLU}(W_2 * [F_1, F_2, \ldots, F_K])) + F_{\text{New}} \quad (7)$$

where W_1 and W_2 represent the weights of the 3×3 convolutional and the 1×1 convolution, respectively, for deep feature fusion. PReLU represents the parametric ReLU activation function and $[F_1, F_2, \ldots, F_K]$ represents the output features of each ATCB module.

4) Upsampling Block: The upsampling module includes a 3×3 convolutional layer and a subpixel shuffle, which are used to map multiple LR DEM feature maps onto a single HR DEM [31]. The DEM after SR reconstruction can be expressed as follows:

$$D_{\text{SR}} = F_{\text{UP}}(W_3 * F_D) \quad (8)$$

where F_{UP} represents the pixel shuffling operation and W_3 represents the weights of the 3×3 convolutional layers for upsampling. The subpixel shuffle first converts an LR DEM feature map into an HR DEM feature map using a convolutional kernel, and then reorganizes the feature map using a pixel mapping relationship to stitch it into the HR DEM feature map. The subpixel shuffle first converts an LR DEM feature map into C^2 LR DEM feature maps through C^2 convolutional kernels, and then reorganizes and merges these LR DEM feature maps into a single HR DEM feature map through the mapping relationship of the pixels. Finally, multiple HR DEM feature maps are merged into a single HR DEM.

5) Loss Function Block: The loss function calculation block was used to calculate the error between the input and the output DEMs, and the model parameters were continuously iterated and updated by the error back-propagation function so that the error was gradually reduced [32]. In the loss function calculation block, the void areas in the DEM were masked out, and the mapping relationship of nonvoid areas was found. Meanwhile, to improve the training ability and enhance the effect of multisource features, the loss function calculation module was designed to include global loss, terrain feature loss, and SAR feature loss. The loss function, Loss, can be

written as follows:

$$\text{Loss} = L_{\text{Global}} + \lambda_1 L_{\text{Terrain}} + \lambda_2 L_{\text{SAR}} \quad (9)$$

where L_{Global} represents the global loss, L_{Terrain} represents the terrain feature loss, L_{SAR} represents the SAR feature loss, and λ_1 and λ_2 represent the weights of terrain and SAR feature loss, respectively.

The global loss L_{Global} can be calculated based on the model-predicted SR DEM and the HR InSAR DEM elevation of the nonvoid area, as follows:

$$L_{\text{Global}} = \frac{1}{J} \sum_{j=1}^{J} |f_{\text{mask}}\left(D_{\text{SR}}^j - D_{\text{SAR}}^j\right)| \quad (10)$$

where J is the number of DEM training datasets. f_{mask} is the void mask function. D_{SR}^j and D_{SAR}^j represent the elevation of the jth SR DEM and HR InSAR DEM training datasets, respectively.

The terrain feature loss L_{Terrain} can be calculated using the error between the terrain features generated by the model-predicted SR DEM and the HR InSAR DEM, as follows:

$$L_{\text{Terrain}} = \frac{1}{J} \sum_{j=1}^{J} \left| T_{\text{SR}}^j - T_{\text{SAR}}^j \right|. \quad (11)$$

The essence of the SAR features was to differentiate the characteristics of different feature landscapes. For model training, the SAR features can be used as prior information to monitor and guide the model. The SAR features were used as weights in the DEM to measure the SAR feature loss, as follows:

$$L_{\text{SAR}} = \frac{1}{J} \sum_{j=1}^{J} \left| f_{\text{mask}}\left(S^j \times D_{\text{SR}}^j - S^j \times D_{\text{SAR}}^j\right) \right| \quad (12)$$

where S^j represents the pixel values of the jth SAR features training datasets, and the SAR features include intensity and coherence features.

The loss function can be calculated by substituting (10)–(12) into (9), considering the global loss, terrain feature loss, and SAR feature loss together. If the loss function does not satisfy the discriminant convergence condition, the model is retrained by error backpropagation. If the loss function satisfies the discriminative convergence condition, the predicted high accuracy DEM is output. The high-precision DEM product can be generated by combining the intelligently filled and predicted missing data area DEM from Section II-C with the high-precision InSAR void DEM generated in Section II-A. The generated high-precision DEM is then subjected to steps, such as coordinate transformation and reprojection to produce a DEM product that corresponds to the actual geographic location.

D. TMSF Accuracy Evaluation

The TMSF accuracy evaluation spanned the modeling, elevation, and terrain feature accuracy. The modeling accuracy was assessed using root mean square error (RMSE) of elevation and mean absolute error (MAE) of elevation and slope between the SR and HR InSAR void DEM. The

evaluation of DEM elevation accuracy after void filling relied on the RMSE, MAE, and line error at 90% probability (LE90) between the generated DEM and photogrammetric DEM. The elevation was further validated using the RMSE and mean error according to Ice, Cloud, and land Elevation Satellite (ICESat)/Geosciences Laser Altimeter System (GLAS). In addition, the MAE of slope and aspect were used to assess the preservation of terrain features and are discussed further. The main classical interpolation or deep learning algorithms for comparison include the following.

1) Bicubic [33]: The bicubic interpolation is a basic image interpolation algorithm. Many images or DEM deep learning algorithms are based on the bicubic algorithm for improvement and development. It offers computational efficiency and is simple to implement.

2) Super-Resolution CNN (SRCNN) [34]: The SRCNN learns the mapping information from LR to HR directly through deep learning. The model structure is simple, and it is the first deep learning model to apply CNN to SR reconstruction. This approach shows better performance in DEM-SR reconstruction.

3) Enhanced Deep SR (EDSR) [35]: The EDSR uses the ResNet residuals concept by removing the regularization layer from the residual structure, which can improve the quality of the results.

4) Deep Asymmetric Transfer Network (DATN) [36]: The DATN is part of the semi-supervised transfer learning framework. In various application scenarios, DATN demonstrates its effectiveness in overcoming the problem of unbalanced domain adaptation.

5) Residual Dense Network (RDN) [37]: The core part of the RDN is the dense residual connectivity structure. Compared with the traditional structures, the dense residual connectivity in RDN allows the extracted feature representation information to propagate quickly through the network structure. In addition, RDN and EDSR serve as the basic methods for model accuracy verification and calibration.

6) Residual Feature Aggregation Network (RFAN) [38]: The RFAN improves the performance of SR image reconstruction through a novel RFA framework and enhanced spatial attention (ESA), the core idea of which is the effective use of residual features. It has been widely applied and has significantly improved image quality.

7) Hierarchical Swin Transformer (HST) [39]: The HST improves representation at any scale by using hierarchical feature representation. It has achieved remarkable results in creating SR images from highly distorted compressed images.

8) Hybrid Network of CNN and Transformer (HNCT) [30]: The HNCT combines the ability of CNN to capture local features with the ability of transformers to model the global context, leading to a more comprehensive understanding of image information.

III. Experiments and Results

A. Study Area and Data Preprocessing

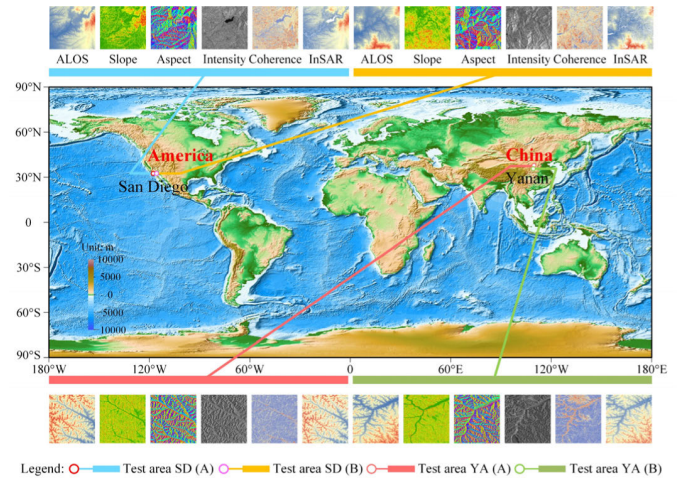The experiments used a total of 19 views of TerraSAR-X images from the San Diego (SD) area of the United States

Fig. 7. Maps of the study areas and datasets [including test areas SD(A), SD(B), YA(B), and YA(B)].

(https://sss.TerraSAR-X.dlr.de/) and 18 views of PAZ images from the city of Yan'an (YA) on the Loess Plateau of China (https://www.inta.es/paz-ciencia/en/AO/ao-006/) to generate high-precision InSAR DEMs. The LR DEM data are based on ALOS 12.5-m DEMs (https://search.asf.alaska.edu). These SAR data and LR DEM are available from the German Aerospace Center and the European Space Agency. The locations of the study areas and corresponding datasets are shown in Fig. 7. The study areas SD(A) and SD(B) are located at (117.078°W∼116.937°W, 32.614°N∼32.734°N) and (116.908°W∼116.767°W, 32.590°N∼32.709°N), respectively. The study areas YA(A) and YA(B) are located at (109.460°E∼109.590°E, 36.662°N∼36.806°N) and (109.430°E∼109.558°E, 36.467°N∼36.656°N), respectively. The four test areas are rich in topographical features, including mountains, rivers, and cities, and contain typical landforms, such as plateaus, mountains, plains, hills, and basins. Each of these study areas contains a different type of void problem (mountain shadow overlap, atmospheric noise, low backscatter coefficient, and spatiotemporal incoherence). The complex topography provides suitable conditions for high-precision DEM research.

The data preprocessing was performed using GAMMA software to format $N + 1$ views of SAR satellite data acquired for the same study area into single-look complex (SLC) images. The preprocessing mainly includes the spatial–temporal baseline connection, image registration, interferogram, and flat-Earth phase removal, using the ALOS 12.5-m DEM to remove the topographic phase [40], [41] and filtering to generate intensity maps, coherence maps, and differential interferograms, the minimum cost flow (MCF) was used for phase unwrapping [42]. High coherence points were selected using thresholding. Atmospheric delays and noise were removed using ICA [43]. Baseline errors were corrected using a polynomial fitting with adaptive discrimination [44], [45]. Deformation errors were modeled using a combined InSAR model. The acquired residual elevations are summed with the LR DEM elevations to generate high-precision InSAR DEM according to (3). Then, the DEM training datasets,
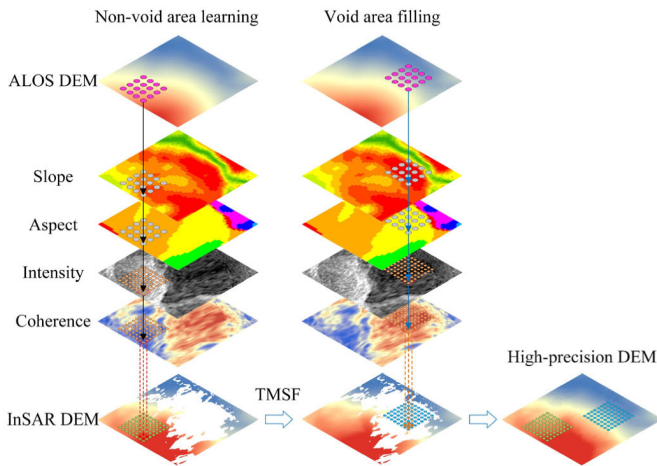
Fig. 8. Dynamic processes of intelligent learning reconfiguration in high-precision InSAR DEM void areas.
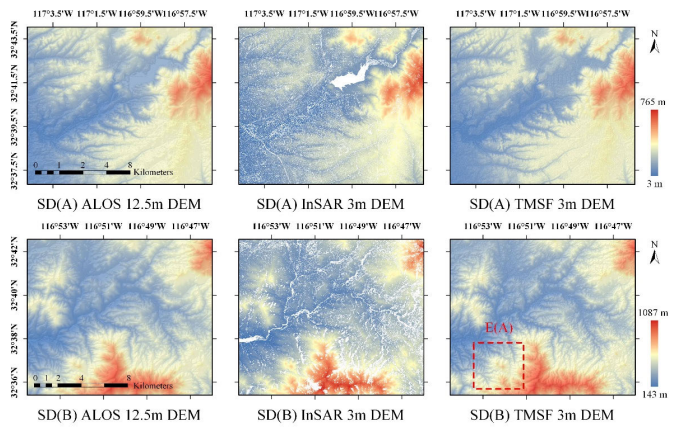

Fig. 9. Intelligent learning reconfiguration using TMSF in the void areas of the SD area DEM.

Fig. 10. Elevation distribution of the ALOS 12.5 m and TMSF 3-m DEMs in the SD area.

terrain feature training datasets, and SAR feature training datasets were constructed. Finally, a total of 4050 samples in the SD area and 5080 samples in the YA area were generated.

B. Nonvoid Area Learning and Void Area Filling by TMSF

The nonvoid area learning and void area filling by TMSF are shown in Fig. 8. The void areas in the training datasets obtained in Section III-A were masked, and the first 80% of these datasets were used as the training datasets, while the last 20% were used as the validation datasets. The TMSF model was trained using the training datasets, and the network parameters were initialized to achieve a training converged model. The TMSF model in this article contains four ATCB modules, where each ATCB module contains two STL structures and two AMs. The number of attention heads is set to 4, and the number of channels in the feature map is set to 64. The window size is set to 8. The experiments are performed by rotating the images 90°, 180°, and 270° to achieve improvement. The Adam optimizer hyperparameters are set to $\beta 1=0.9$, $\beta 2=0.999$, and $\varepsilon=1e-8$. The learning rate is set to 1e-4 and a simulated cosine annealing curve is used to optimize the learning rate. λ_1 and λ_2 are considered as unit weights, respectively [30]. The TMSF model was trained using six Nvidia A6000 graphics cards for model training with a training number of 1000. The void area in the DEM is used as the test dataset. The LR DEM data, terrain feature data (slope and aspect), and SAR feature data (intensity and coherence) from the test datasets are input into the trained model to achieve intelligent filling of the high-precision InSAR DEM.

C. High-Precision DEM Generation

The high-precision DEM product can be generated by combining the DEM of the void area intelligently filled by TMSF with the high-precision InSAR DEM generated in Section II-A. The coordinates of the generated high-precision DEM are then transformed and reprojected to produce DEM products. The results of the high-precision DEM generation for the study areas SD(A) and SD(B) and the elevation
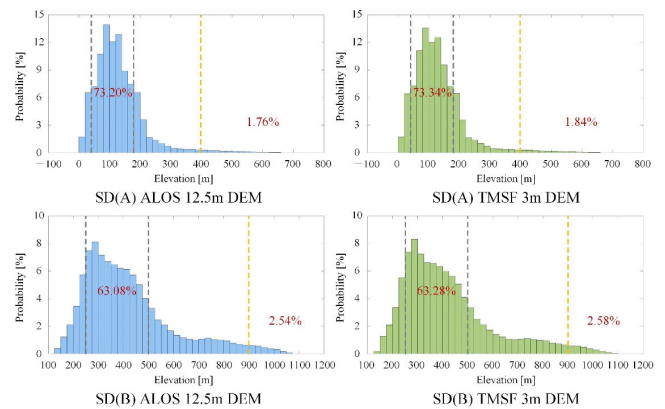
distributions of the ALOS 12.5-m DEM and the TMSF 3-m DEM are shown in Figs. 9 and 10, respectively. Fig. 11 shows the high-precision DEM generation for the study areas YA(A) and YA(B), and Fig. 12 shows the elevation distributions of the ALOS 12.5-m DEM and the TMSF 6-m DEM.

It can be seen from Figs. 9 and 11 that the InSAR 3-m DEMs can better represent the textural details of the feature landforms compared with the ALOS 12.5-m DEMs, but they have large DEM voids. The DEMs filled by TMSF show richer topographic and geomorphological features, with clearer representations of the mountains, rivers, and urban areas. As shown in Fig. 10, for SD(A), the elevation of the pixels in the ALOS 12.5-m DEMs 40- and 180-m span 73.20% of the pixels, and for the TMSF 3-m DEMs 73.34%. The pixels of ALOS 12.5-m DEM with elevations >400 m represent 1.76%, and for the TMSF 3-m DEMs, they represent 1.84%. For the SD(B) area, the pixels in the ALOS 12.5-m DEMs elevation between 250- and 500-m spanned 63.08%, and for the TMSF 3-m DEMs 63.28%; and the pixels in the ALOS 12.5-m DEMs with elevations >900 m represent 2.54%, and for the TMSF 3-m DEMs, they represent 2.58%. This shows that the high-accuracy DEM products produced in the SD area span a smaller range of elevations and have better continuity. In addition, the YA area has similar distribution characteristics, as shown in Fig. 12.
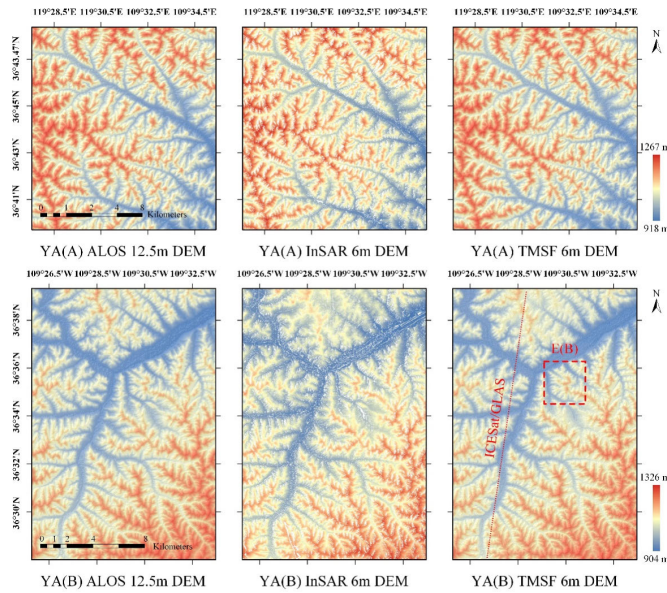
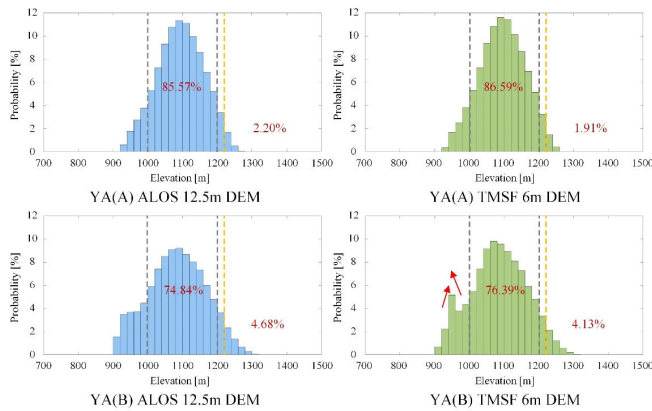Fig. 11. Intelligent learning reconfiguration using TMSF in the void areas of the YA area DEM.

Fig. 12. Elevation distribution of the ALOS 12.5 m and TMSF 6-m DEMs in the YA area.

TABLE I
VOID RATIO OF THE FOUR TEST AREAS

Test area	SD(A)	SD(B)	YA(A)	YA(B)
Void ratio	17.88%	21.24%	19.17%	17.19%

However, an interesting phenomenon can be seen from the red arrows in the YA(B) area, as shown in Fig. 12. The percentage of pixels with elevations between 900 and 940 m and 960 and 1000 m decreases significantly, while the percentage of pixels with elevations between 940 and 960 m increases significantly. We suspect that this phenomenon is related to the fact that the PAZ data used in YA City was collected between 2022 and 2023, while the ALOS 12.5-m DEM was collected in 2008. During this period, the government proposed the mountain excavation and city construction (MECC) project in December 2011, which was reported to have started in April 2012 [46], and urban construction is likely to lead to this outcome [47]. In addition, to better illustrate the effect of TMSF void filling, Table I shows the void ratio of four

test areas. First, the DEM void pixels in complex terrain areas caused by missing data accounted for 21.24% of the total, and second, the TMSF was able to fill the void areas.

D. TMSF Accuracy Evaluation

1) Accuracy Evaluation of TMSF Modeling: The experiment uses the last 20% of the training dataset as the validation dataset. The higher the accuracy of the validation datasets, the better the ability of the TMSF to perform intelligent learning reconfiguration, indicating the higher the potential of the TMSF model to fill the void area. Table II shows the performance of the TMSF model and traditional interpolation or deep learning methods. From Table II, the TMSF model has significantly improved the RMSE and MAE of the elevation, and the MAE of the slope attributes compared with the traditional interpolation or deep learning models. For the SD area, the RMSE of the TMSF elevation is 2.415 m, while the bicubic model is 4.449 m, showing an improvement of 45.71%. For the YA area, the RMSE of the TMSF model elevation is 5.210 m, while the bicubic model is 5.712 m, showing an improvement of 8.7%. The overall effect of the TMSF model is greater in the SD area than in the YA area. We suspect that this phenomenon is mainly due to the following reasons. Frist, the topography of the YA area, located on the Loess Plateau, is dominated by mountainous areas and the overall error of the DEM is large. The topography of the SD area includes mountains and plains, and the overall error is smaller. Through the quantitative analysis, the elevation of the SD area is in the range of 3–1087 m, while the elevation of the YA area is in the range of 904–1326 m. The RMSE of the TMSF model accounts for only 0.22% and 0.39% of the maximum elevation in these areas, respectively, which indicates that the TMSF model has a high modeling accuracy in both study areas. Second, the high-precision DEM generated from TerraSAR-X data in the SD area has a resolution of 3 m and was acquired in 2009 and 2010. The high-precision DEM in the YA area, generated from PAZ data, has a resolution of 6 m and was acquired in 2022 and 2023. The reference DEMs used in the experiments are 12.5-m resolution ALOS data acquired in 2008, and the elevation changes caused by the long time series are greater in the YA area than in the SD area. In addition, the TMSF model requires a ×4 reconstruction in the SD area, while only a ×2 reconstruction is required in the YA area, and the TMSF has a better ability to learn reconstruction modeling of HR DEMs with higher accuracy requirements. Two experiments confirmed the strong generalization of the TMSF model and demonstrated its promising applications in other areas.

2) Accuracy Evaluation Using Photogrammetric DEM: To verify the accuracy of the DEM products after void-filling, photogrammetric DEMs with 3-m resolution (https://datagateway.nrcs.usda.gov/) in the SD area and 5-m resolution in the YA area were acquired for validation at locations E(A) and E(B) shown in Figs. 9(f) and 11(f), respectively. To ensure the consistency of geographical location and spatial reference, the DEMs were geocoded and projected to the WGS-84 coordinate system [48]. The DEM comparison of different models after void filling in the SD area is shown in

TABLE II
PERFORMANCE OF DEM INTELLIGENT LEARNING RECONFIGURATION APPROACH

| Model	SD(A)/ SD(B)			YA(A)/ YA(B)		
	RMSE-elevation (m)	MAE-elevation (m)	MAE-slope (°)	RMSE-elevation (m)	MAE-elevation (m)	MAE-slope (°)
Bicubic	4.449	3.280	1.363	5.712	4.620	2.066
SRCNN	4.435	3.270	1.353	5.665	4.590	2.019
EDSR	3.911	2.857	1.358	5.634	4.567	1.965
DATN	4.181	3.107	1.335	5.440	4.391	1.938
RDN	3.741	2.740	1.294	5.392	4.327	2.040
RFAN	3.570	2.603	1.144	5.286	4.296	1.876
HST	3.663	2.668	1.156	5.342	4.309	1.989
HNCT	2.723	1.873	1.084	5.334	4.301	1.913
TMSF	**2.415**	**1.733**	**0.927**	**5.210**	**4.280**	**1.868**
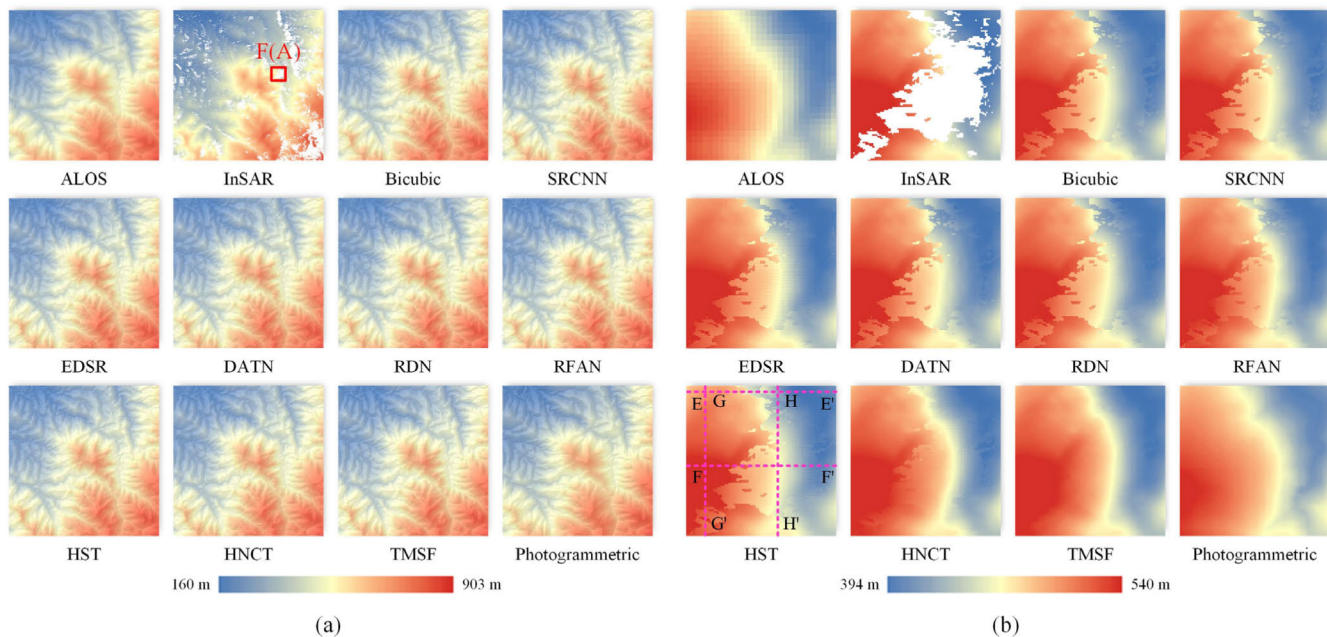
Fig. 13. Comparison of DEM void filling for different models in the SD area. (a) E(A) area DEM. (b) F(A) area DEM.

Fig. 13(a). To clearly show the effect of void filling, the F(A) in the E(A) area was selected for visualization and analysis, as shown in Fig. 13(b). The errors between different models and photogrammetric DEMs and their distributions are shown in Fig. 14. Similarly, the DEM comparison and visualization analysis of the F(B) in the YA area and the errors and their distributions are shown in Figs. 15 and 16, respectively.

It can be seen from Figs. 13 and 15 that the DEM void filling by the TMSF model is significantly improved, and the cracks and edge jagging effects in the void area are gradually eliminated. By introducing multisource features into the TMSF model, the high-frequency information in the void area are effectively enhanced and the texture detail of the DEM becomes richer and closer to the photogrammetric DEM. From Figs. 14 and 16, the error of the TMSF gradually decreases. The quantitative analysis shows that for the SD area, the probability of pixel points with an error distribution between −10 and 10 m is 67.79% for the TMSF model and 39.86% for the bicubic model. For the YA area, the probability of pixel points with an error distribution between −10 and 20 m is 94.53% for the TMSF model, while the bicubic model is 91.75%. This shows that the error distribution of the TMSF model points to smaller values.

To further assess the accuracy of the TMSF model in the SD and YA areas, the RMSE, MAE, and LE90 between different models and the photogrammetric DEM in the E(A)
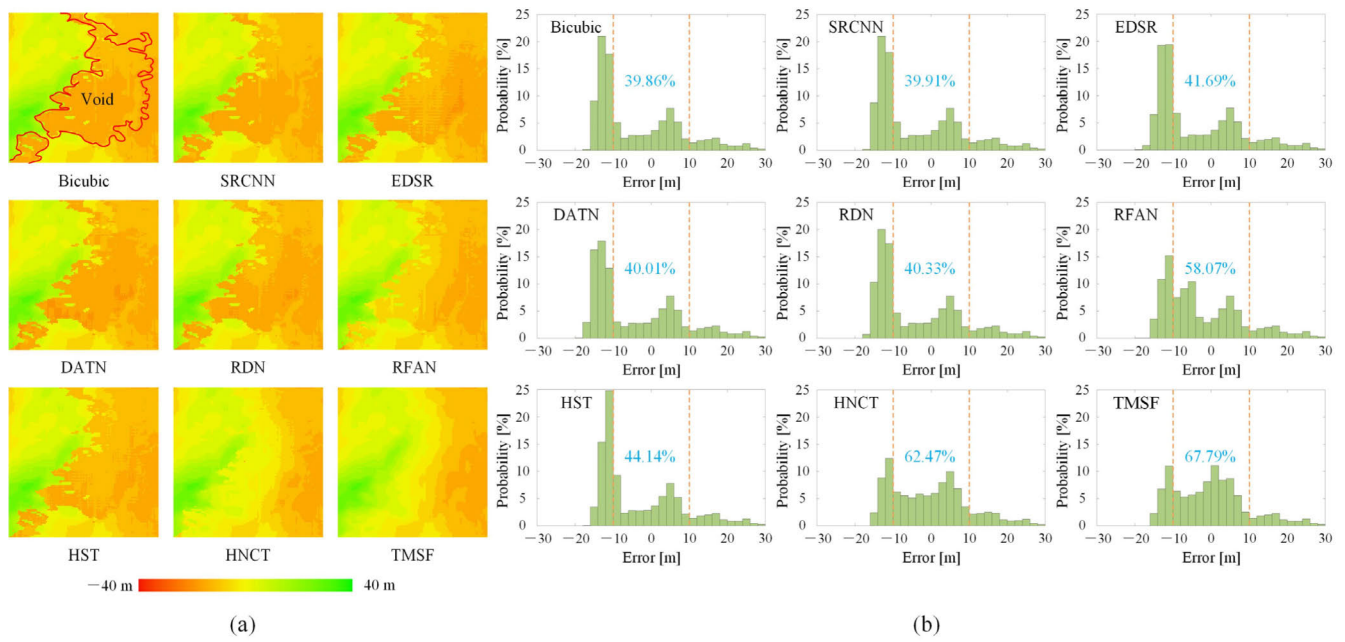
Fig. 14. DEM error and distribution between different models and photogrammetry in the F(A) area. (a) F(A) area error. (b) F(A) area error distribution.
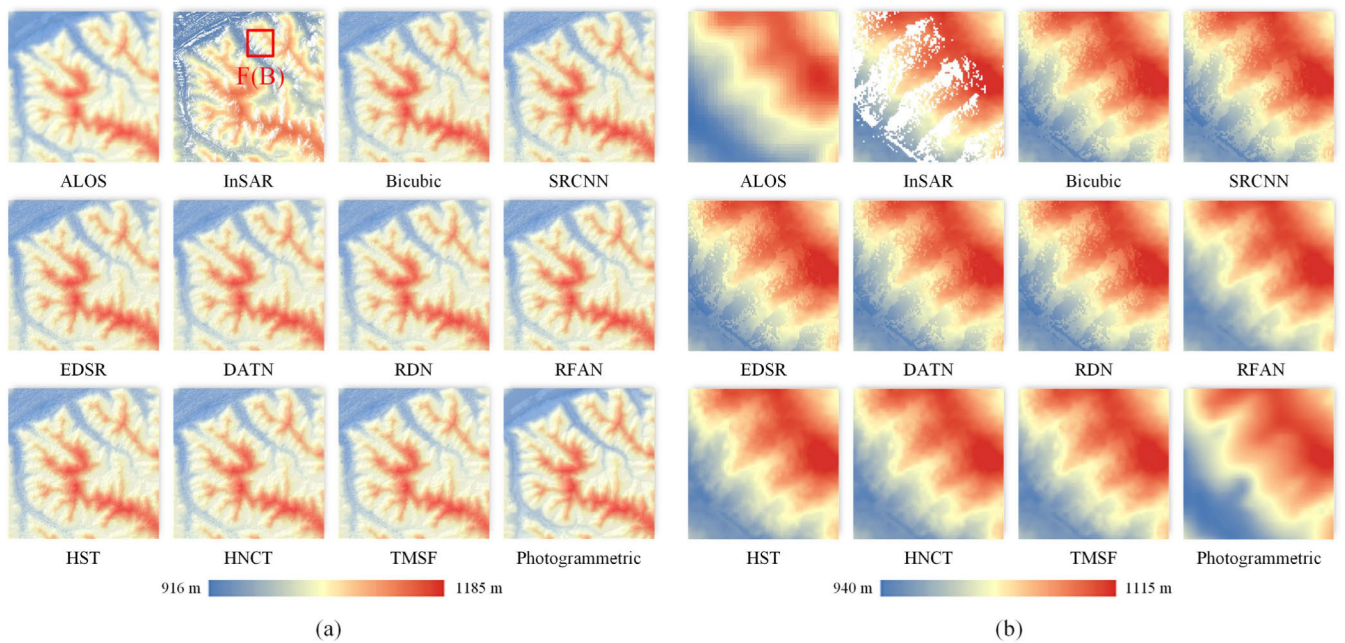

Fig. 15. Comparison of DEM void filling for different models in the YA area. (a) E(B) area DEM. (b) F(B) area DEM.

and E(B) validation areas were calculated. From Table III, the accuracy of the TMSF is significantly improved compared to the traditional interpolation or deep learning model. For the SD area, the RMSE between the TMSF void-filled DEM and the photogrammetric DEM is 2.679 m, while that of the bicubic is 3.616 m, with an improvement of 25.91%. For the YA area, the RMSE between the TMSF void-filled DEM and the photogrammetric DEM is 3.168 m, while that of the bicubic is 4.416 m, with an improvement of 28.26%. This shows that the TMSF model has a significant advantage in high-accuracy DEM void filling.

3) Accuracy Evaluation Using ICESat/GLAS Elevation Data: To further evaluate the accuracy of the TMSF model DEM void filling, 122 ICESat/GLAS points were collected in the YA area for further validation, as shown in Fig. 11. Since the ICESat/GLAS point has a diameter of about 65 m and the high-precision InSAR DEM in the YA area is 6 m, the center position ($H1$), minimum ($H2$), maximum ($H3$), and average ($H4$) of the DEMs in the 9×9 window after void filling are used for comparison to avoid errors due to imprecise matching of point positions [15]. Table IV shows that the variability of the mean error between different model DEMs
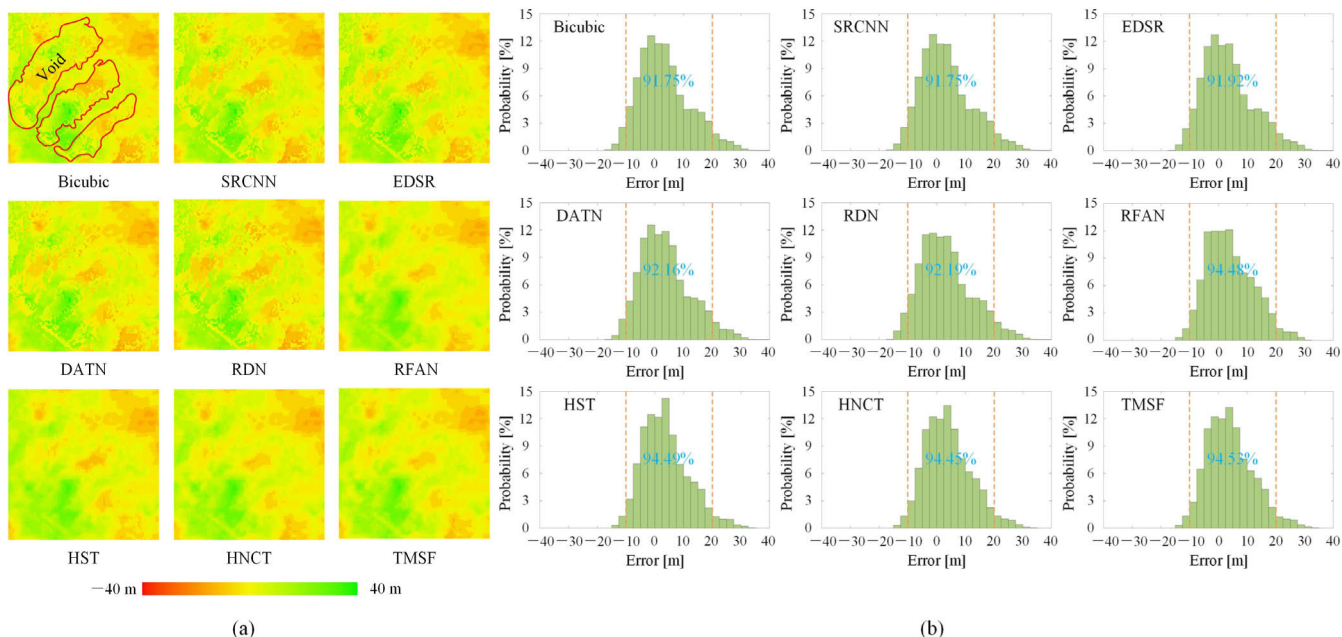
Fig. 16. DEM error and distribution between different models and photogrammetry in the F(B) area. (a) F(B) area error. (b) F(B) area error distribution.

TABLE III
ELEVATION ACCURACY INDICES BETWEEN DIFFERENT MODELS AND THE PHOTOGRAMMETRIC DEM

Model	SD area			YA area		
	RMSE (m)	MAE (m)	LE90 (m)	RMSE (m)	MAE (m)	LE90 (m)
Bicubic	3.616	3.758	6.818	4.416	5.939	8.659
SRCNN	3.465	3.648	6.465	4.371	5.885	8.634
EDSR	3.345	3.558	6.187	4.384	5.701	8.627
DATN	3.455	3.643	6.431	3.966	5.624	8.341
RDN	3.304	3.526	6.098	3.619	4.998	8.110
RFAN	3.197	3.419	5.943	3.468	4.238	7.385
HST	3.240	3.462	6.001	3.577	4.453	7.467
HNCT	3.176	3.418	5.843	3.356	4.193	7.239
TMSF	**2.679**	**2.921**	**4.324**	**3.168**	**4.116**	**7.216**

and ICESat/GLAS is small, demonstrating the accuracy of the geographic position of the generated high-precision DEM products. The RMSE of the TMSF model at H1–H4 can be improved by 4.34%–22.21%, 4.46%–23.37%, 4.31%–22.47%, and 4.40%–22.28%, respectively, compared with the traditional models. This indicates that the void DEM filled by the TMSF model has high accuracy.

IV. DISCUSSION

A. Analysis of the Consistency of Terrain Features

Terrain feature consistency is an important index for terrain analysis, and the higher the terrain feature consistency, the better the quality of the generated high-precision DEM. To fully demonstrate the consistency of the DEM terrain features after void filling by the TMSF model, the DEM in the SD area was analyzed. The slope and aspect are shown in Fig. 17, and the river network is shown in Fig. 18.

From Fig. 17, the slope and aspect generated by TMSF are in good agreement with those generated by photogrammetry. The slope and aspect generated by the LR ALOS DEM are simple, while the slope and aspect generated by the high-precision InSAR DEM can reflect more texture detail. In addition, the slope results show that the traditional model cannot accurately fill the void area, and a distinct red crack appears at the boundary of the void area. As the model improves, the red crack at the boundary gradually disappears, and the TMSF model becomes closer to the photogrammetric

TABLE IV
ELEVATION ACCURACY INDICES BETWEEN DIFFERENT MODELS AND THE ICESAT/GLAS

Model	H1 (m)		H2 (m)		H3 (m)		H4 (m)	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
Bicubic	-0.517	4.587	-1.814	4.724	0.814	4.747	-0.525	4.501
SRCNN	-0.511	4.534	-1.792	4.669	0.803	4.690	-0.519	4.450
EDSR	-0.504	4.487	-1.764	4.610	0.792	4.634	-0.512	4.402
DATN	-0.491	4.426	-1.714	4.530	0.800	4.559	-0.491	4.344
RDN	-0.486	4.349	-1.697	4.436	0.785	4.468	-0.494	4.274
RFAN	-0.427	3.962	-1.514	4.039	0.707	4.074	-0.441	3.898
HST	-0.440	4.081	-1.560	4.160	0.728	4.196	-0.454	4.015
HNCT	-0.399	3.730	-1.408	3.789	0.704	3.846	-0.398	3.659
TMSF	**-0.382**	**3.568**	**-1.345**	**3.620**	**0.671**	**3.680**	**-0.380**	**3.498**
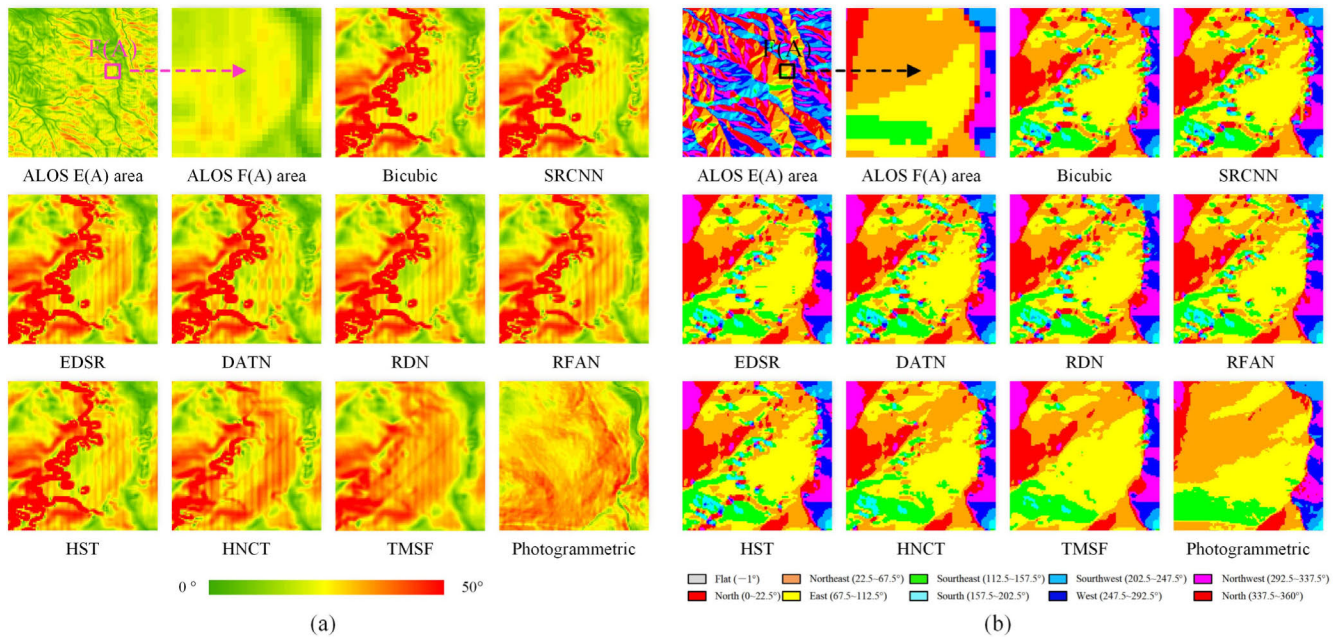
Fig. 17. Comparison of the slope and aspect generated by the DEM after different models of void filling. (a) Slope. (b) Aspect.

DEM. From the aspect results, the traditional model aspect of the void area is more chaotic, while the TMSF model is better at improving the stability of the aspect in the void area. From Fig. 18, the river network generated by the LR ALOS DEM loses a lot of topographic detail, whereas the river network generated by the HR DEM is rich in detail. Due to the presence of voids in the high-precision InSAR DEM, the river network exhibits obvious breaks that are visible, whereas the filled DEM can better connect the river network. The river network in the traditional model creates errors or distortions, and the TMSF model reduces these problems. In addition, to fully demonstrate the results of the consistency of the terrain features, the MAEs of slope and aspect were calculated for the E(a) area, as shown in Table V. Through the

quantitative analysis, the MAEs for slope and aspect generated by TMSF and photogrammetric DEM show improvements of 11.26%–19.14% and 9.34%–20.86%, respectively, compared with the bicubic model. This shows that the TMSF has a better ability to maintain the terrain features.

B. Analyzing the Ability of TMSF to Repair DEMs

To better represent the effect of the TMSF model on the void-filled DEM area, the horizontal and vertical profiles in the DEM [EE', FF', GG', and HH' as shown in Fig. 13(b)] are selected for further discussion. By calculation, the maximum differences between the TMSF model and the bicubic model at EE', FF', GG', and HH' are 28.38, 36.96, 45.17, and 18.02 m, respectively. From Fig. 19, the TMSF void-filled DEM profile

TABLE V
TERRAIN FEATURE CONSISTENCY ACCURACY INDICES (UNIT: °)

Indices	Bicubic	SRCNN	EDSR	DATN	RDN	RFAN	HST	HNCT	**TMSF**
MAE_{Slope}	5.468	5.442	5.396	5.446	5.441	5.010	5.259	4.982	**4.421**
MAE_{Aspect}	16.312	16.295	16.164	16.271	16.473	14.240	15.931	14.439	**12.909**
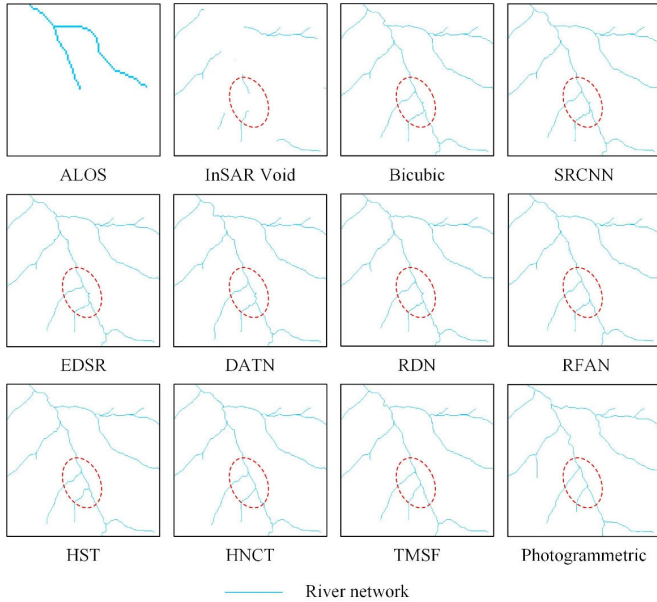
Fig. 18. Comparison of the river network generated by the DEM after different models of void filling.
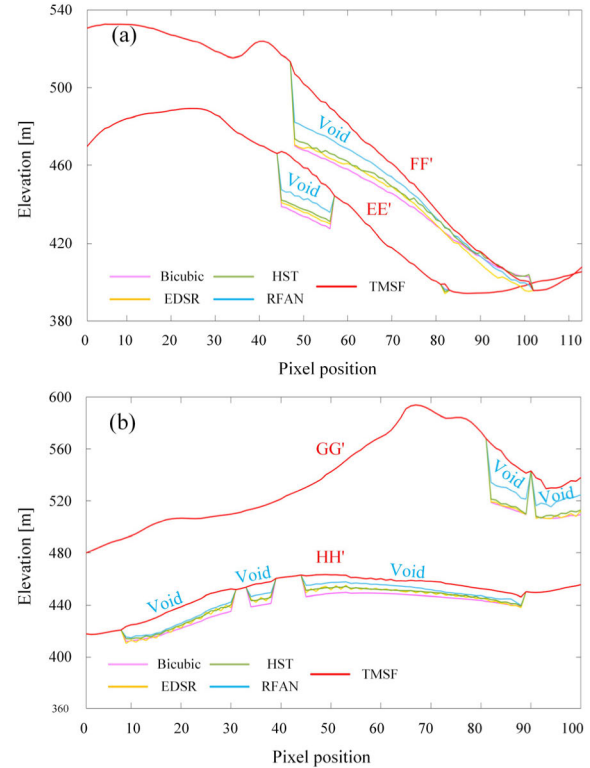

Fig. 19. Comparison between different DEMs using the horizontal and vertical profiles. (a) Horizontal profile EE' and FF'. (b) Vertical profile GG' and HH'.

has better continuity, while the traditional interpolation and deep learning method void areas have an obvious scale cut-off problem.

In addition, to fully demonstrate the ability of TMSF to repair the DEM void problem caused by various factors, different landscape settings are extracted from the SD area for further analysis. These include the mountain shadow overlay, atmospheric noise, and low backscatter coefficient areas, as shown in Fig. 20. As can be seen from the mountain shadow overlay area, the high-precision DEM generated by InSAR has clearer mountain contours than the ALOS DEM, but the problem of voids is serious. Traditional interpolation and deep learning methods will exhibit obvious cracks and a jagged edge phenomenon after void filling. The TMSF model can not only better fill the void areas of the DEM but also add richer high-frequency information. From the atmospheric noise area, it can be seen that the TMSF model can improve the speckle noise in the DEM and fill the DEM better. From the river area, it can be seen that the TMSF can better repair the river area via the contour of the river bank.

C. Analysis of TMSF Interpretability and Ablation Experiments

To investigate the effectiveness of the multisource features and the ATCB module in the TMSF model, two experiments were designed for the ablation of multisource features and

the ablation of the ATCB module, respectively, as shown in Fig. 21. Multisource feature ablation experiments are useful to explore the effectiveness of introducing multisource features and to understand which features have a greater impact on filling DEM voids. The multisource feature ablation experiments from models 1–6 were designed to progressively ablate the loss function and to fully test the advantages of the TMSF model for terrain and SAR features. The ATCB module ablation experiments help to better understand the impact of the transformer and CNN modules on the network performance and the DEM intelligent learning reconfiguration modeling. To test the effectiveness of the transformer and ensure the robustness of the experiments, model 7 replaces the STL module with two 3 × 3 convolutional layers and ReLU activations, while model 8 replaces the STL module with one 3 × 3 convolutional layer and ReLU activation. To test the effectiveness of the AM structure, models 9 and 10 remove one AM and two AM modules, respectively. To test the comparison between the AM structure and the conventional
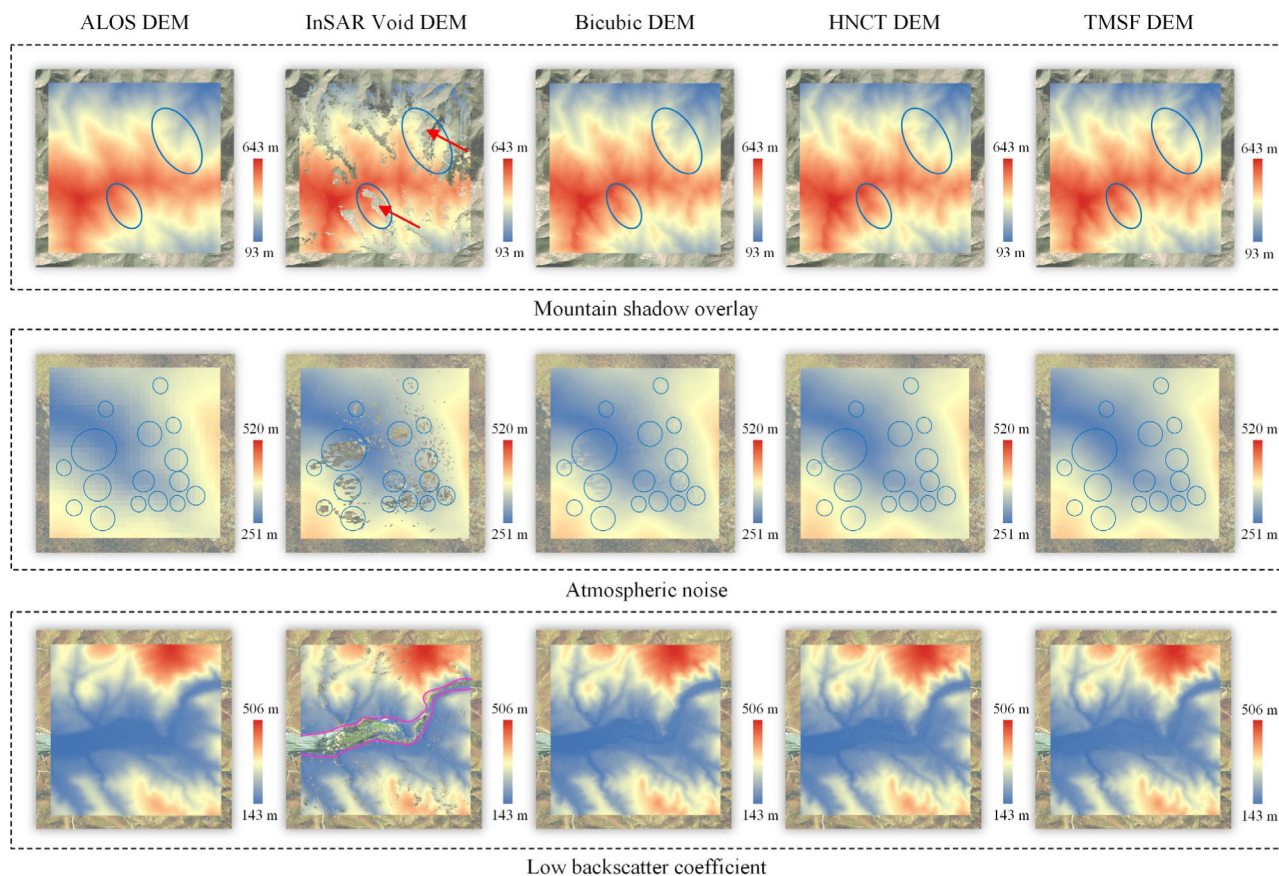
Fig. 20.   Ability of TMSF to repair voids in different landscape settings.

TABLE VI
ACCURACY INDEX FOR MULTISOURCE FEATURES
ABLATION EXPERIMENTS

Model	RMSE-elevation (m)	MAE-elevation (m)	MAE-slope (°)
Model 1	2.474	1.776	0.961
Model 2	2.855	2.061	1.053
Model 3	2.903	2.103	1.073
Model 4	2.474	1.769	0.963
Model 5	2.552	1.798	0.965
Model 6	4.303	3.162	1.356
TMSF	**2.415**	**1.733**	**0.927**

TABLE VII
ACCURACY INDEX FOR ATCB MODULE ABLATION EXPERIMENTS

Model	RMSE-elevation (m)	MAE-elevation (m)	MAE-slope (°)
Model 7	2.433	1.764	0.934
Model 8	2.574	1.816	0.974
Model 9	2.432	1.761	0.942
Model 10	2.567	1.806	0.984
Model 11	2.494	1.795	0.968
Model 12	2.499	1.814	0.943
TMSF	**2.415**	**1.733**	**0.927**

ESA structure, models 11 and 12 replace one and two AM structures, respectively, with one ESA structure.

Tables VI and VII show the accuracy indices for multisource feature and ATCB module ablation, respectively. In the multisource feature ablation experiments, the accuracy indices of models 1–6 are all reduced relative to the TMSF, confirming the effectiveness of the terrain feature, SAR feature, and loss functions. In the ATCB module ablation experiments, the accuracy metrics of models 7–12 are all reduced compared to the TMSF model. In addition, models 7 and 8 confirm the effectiveness of the transformer module insertion. Models 9 and 10 confirm the effectiveness of the AM module. Models 11 and 12 confirm the effectiveness of the AM over the ESA. Therefore, the proposed multi-source feature fusion module and the ATCB module are reasonable and robust and can improve the elevation accuracy. Besides, in the design of the multi-source feature connection module, we have reserved space for additional data to be added or replaced. In this article, we only use four features (slope and aspect in terrain features, intensity, and coherence in SAR features) that are easy to acquire and collect for experiments and
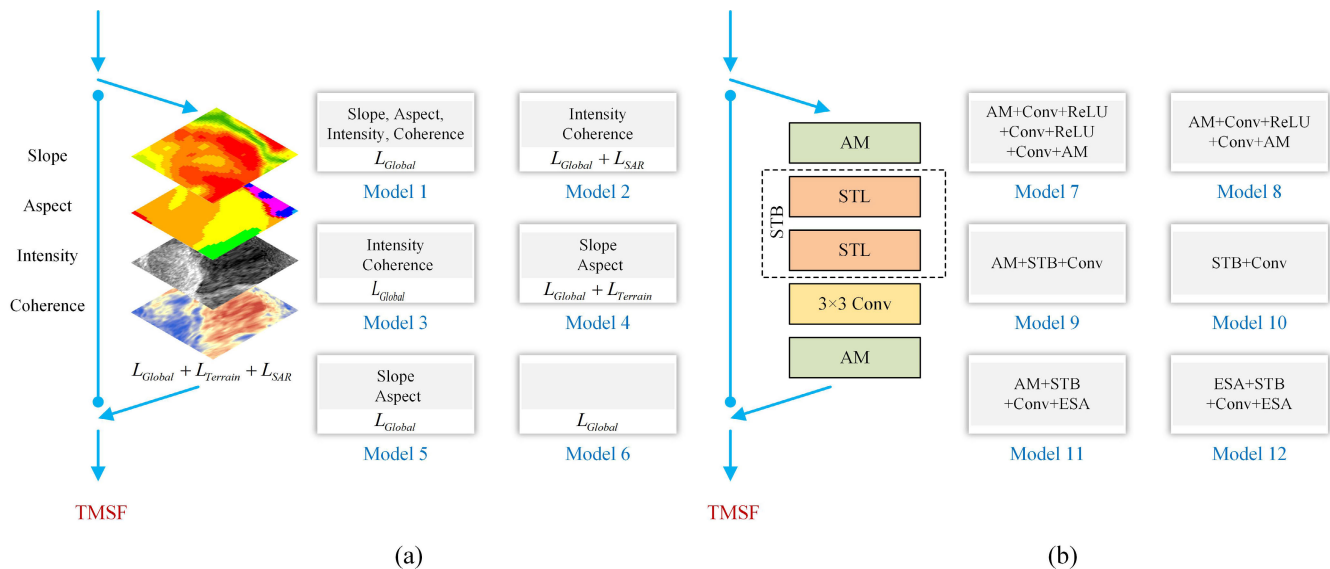
Fig. 21.   (a) Multisource feature module. (b) ATCB module ablation.

applications. In subsequent studies, we will further experiment with the introduction of the other feature data and explore their feasibility.

V. CONCLUSION

This article combines high-precision InSAR DEM generation with intelligent learning reconfiguration modeling to produce high-precision DEM products. The InSAR technology can effectively compensate for the shortcomings of traditional DEM generation. Intelligent learning reconfiguration modeling based on deep learning can significantly increase the information of DEM void areas. The experiments using TerraSAR-X and PAZ data covering the SD and YA areas, respectively, confirmed the effectiveness of the proposed method for intelligent void filling. The main improvements and advances of this work are as follows.

1) An intelligent learning reconfiguration model based on optimized TMSFs was proposed. Through the shallow feature extraction, the multisource feature connection, the transformer and CNN feature fusion, the up-sampling, and the loss function, the intelligent learning reconfiguration model from LR DEM to HR DEM was achieved, which improved the modeling accuracy.

2) The TMSF model shows a significant improvement in accuracy compared with traditional bicubic and deep learning methods. The modeling accuracy of SD and YA areas can be improved by 11.31%–45.74% and 2.32%–8.78%, respectively. Using photogrammetric DEM for validation, the elevation accuracy can be improved by 15.64%–25.91% and 5.60%–28.26%, respectively. Further validation using 122 ICESat/GLAS data, with an improvement of 4.40%–22.28%. The generated high-precision DEM has significant advantages for terrain feature preservation, with improvements of 11.26%–19.14% in slope and 9.34%–20.86% in aspect.

3) The multisource feature connection module was used for feature supervision and loss function optimization. The multisource features (slope and aspect in terrain features,

intensity, and coherence in SAR features) can enhance the high-frequency information of void areas. The generated high-precision DEM can effectively fill DEM voids. It expands the applications of InSAR technology and deep learning, provides important basic data support for geological disasters and environmental protection, and provides a theoretical and methodological system that can be used for the DEM modeling problems in related fields of geoscience.

There are still some limitations in this article that need to be further addressed: 1) this article only uses the multibaseline InSAR to generate the DEM to fill the void area, but the effectiveness of the bistatic InSAR TanDEM-X data needs to be further tested by collecting new data and 2) the photogrammetric DEM accuracy of the validation data used in this article has some limitations, while InSAR has a strong ability to capture the geomorphological features. The texture details beyond the validation data need further clarification.

REFERENCES

[1] T. H. Tarekegn, A. T. Haile, T. Rientjes, P. Reggiani, and D. Alkema, "Assessment of an ASTER-generated DEM for 2D hydrodynamic flood modeling," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 12, no. 6, pp. 457–465, Dec. 2010, doi: 10.1016/j.jag.2010.05.007.

[2] Y. Chen et al., "The simulation of surface flow dynamics using a flow-path network model," *Int. J. Geographical Inf. Sci.*, vol. 28, no. 11, pp. 2242–2260, May 2014, doi: 10.1080/13658816.2014.917312.

[3] Z. Yang, Q. Zhang, X. Ding, and W. Chen, "Analysis of the quality of daily DEM generation with geosynchronous InSAR," *Engineering*, vol. 6, no. 8, pp. 913–918, Aug. 2020, doi: 10.1016/j.eng.2020.07.003.

[4] Z. Li, W. Zhu, C. Yu, Q. Zhang, and Y. Yang, "Development status and trends of Imaging Geodesy," *Acta Geodaetica et Cartographica Sinica*, vol. 52, no. 1, pp. 1805–1834, 2023, doi: 10.11947/j.AGCS.2023.20230003.

[5] T. Toutin and L. Gray, "State-of-the-art of elevation extraction from satellite SAR data," *ISPRS J. Photogramm. Remote Sens.*, vol. 55, no. 1, pp. 13–33, Feb. 2000, doi: 10.1016/s0924-2716(99)00039-8.

[6] W.-J. Lee, H.-S. Jung, and Z. Lu, "Application of ERS and ENVISAT cross-interferometry to generation and accuracy assessment of digital elevation model over northern Alaska," *J. Appl. Remote Sens.*, vol. 9, no. 1, Apr. 2015, Art. no. 096065, doi: 10.1117/1.jrs.9.096065.

[7] N. A. M. Eineder, "Avoiding phase unwrapping in DEM generation by fusing multi frequency ascending and descending interferograms," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 1, Dec. 2004, pp. 477–480, doi: 10.1109/IGARSS.2004.1369067.

[8] C. Zhao, F. Qu, Q. Zhang, and W. Zhu, "A combined multi-interferogram algorithm for high resolution DEM reconstruction over deformed regions with TerraSAR-X data," *J. Geodynamics*, vol. 61, pp. 148–153, Oct. 2012, doi: 10.1016/j.jog.2012.04.004.

[9] Y. Du, L. Zhang, G. Feng, Z. Lu, and Q. Sun, "On the accuracy of topographic residuals retrieved by MTInSAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 1053–1065, Feb. 2017, doi: 10.1109/TGRS.2016.2618942.

[10] M. Lachaise, U. Balss, T. Fritz, and H. Breit, "The dual-baseline interferometric processing chain for the TanDEM-X mission," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 5562–5565, doi: 10.1109/IGARSS.2012.6352345.

[11] M. Lachaise, T. Fritz, and R. Bamler, "The dual-baseline phase unwrapping correction framework for the TanDEM-X mission—Part 1: Theoretical description and algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 780–798, Feb. 2018, doi: 10.1109/TGRS.2017.2754923.

[12] K. Eldhuset, "Using the Pauli scattering mechanisms for analysis of polarimetric SAR images from RADARSAT constellation and TanDEM-X missions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5799–5810, 2022, doi: 10.1109/JSTARS.2022.3190311.

[13] R. Deo, C. Rossi, M. Eineder, T. Fritz, and Y. S. Rao, "Framework for fusion of ascending and descending pass TanDEM-X raw DEMs," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3347–3355, Jul. 2015, doi: 10.1109/JSTARS.2015.2431433.

[14] W. Zhang et al., "Improved DEM reconstruction method based on multi-baseline InSAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Apr. 2022, doi: 10.1109/LGRS.2021.3069239.

[15] Y. Dong, L. Zhang, H. Jiang, T. Balz, and M. Liao, "Cascaded multi-baseline interferometry with bistatic TerraSAR-X/TanDEM-X observations for DEM generation," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 224–237, Jan. 2021, doi: 10.1016/j.isprsjprs.2020.11.012.

[16] H. Shen, L. Liu, L. Yue, X. Li, and L. Zhang, "A multi-source DEM fusion method based on elevation difference fitting neural network," *Acta Geodaetica et Cartographica Sinica*, vol. 47, pp. 854–863, Jun. 2018, doi: 10.11947/j.agcs.2018.20180135.

[17] L. Yue, H. Shen, Q. Yuan, and L. Zhang, "Fusion of multi-scale DEMs using a regularized super-resolution method," *Int. J. Geographical Inf. Sci.*, vol. 29, no. 12, pp. 2095–2120, Jul. 2015, doi: 10.1080/13658816.2015.1063639.

[18] Z. Liu et al., "Obtaining high-precision digital elevation model in Antarctica based on TanDEM-X images and ICESat-2 data," *Bull. Surveying Mapping*, vol. 4, pp. 72–76, Apr. 2022, doi: 10.13474/j.cnki.11-2246.2022.0113.

[19] Z. Chen, X. Wang, Z. Xu, and W. Hou, "Convolutional neural network based DEM super resolution," *ISPRS Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 247–250, Jun. 2016, doi: 10.5194/isprs-archives-XLI-B3-247-2016.

[20] A. Zhou, Y. Chen, J. P. Wilson, G. Chen, W. Min, and R. Xu, "A multi-terrain feature-based deep convolutional neural network for constructing super-resolution DEMs," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 120, Jun. 2023, Art. no. 103338, doi: 10.1016/j.jag.2023.103338.

[21] Y. Liu, J. Qian, L. Wang, and Y. Wang, "Elevation reconstruction combining SAR intensity and interferometric phase data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2022, pp. 1107–1110, doi: 10.1109/IGARSS46834.2022.9883500.

[22] G. Chen, Y. Chen, J. P. Wilson, A. Zhou, Y. Chen, and H. Su, "An enhanced residual feature fusion network integrated with a terrain weight module for digital elevation model super-resolution," *Remote Sens.*, vol. 15, no. 4, p. 1038, Feb. 2023, doi: 10.3390/rs15041038.

[23] A. Ferretti, C. Prati, and F. Rocca, "Multibaseline phase unwrapping for INSAR topography estimation," *II Nuovo Cimento C.*, vol. 24, no. 1, pp. 159–176, Jan. 2001.

[24] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2375–2383, Nov. 2002, doi: 10.1109/TGRS.2002.803792.

[25] D. Li, K. Z. Deng, X. X. Gao, and H. P. Niu, "Monitoring and analysis of surface subsidence in mining areas based on SBAS-InSAR," *Geom. Inf. Sci. Wuhan Univ.*, vol. 43, no. 10, pp. 1531–1537, Oct. 2018, doi: 10.13203/j.whugis20160566.

[26] Y. Zhang, X. Meng, C. Jordan, A. Novellino, T. Dijkstra, and G. Chen, "Investigating slow-moving landslides in the Zhouqu region of China using InSAR time series," *Landslides*, vol. 15, no. 7, pp. 1299–1315, Feb. 2018, doi: 10.1007/s10346-018-0954-8.

[27] R. Zhao, Z.-W. Li, G.-C. Feng, Q.-J. Wang, and J. Hu, "Monitoring surface deformation over permafrost with an improved SBAS-InSAR algorithm: With emphasis on climatic factors modeling," *Remote Sens. Environ.*, vol. 184, pp. 276–287, Oct. 2016, doi: 10.1016/j.rse.2016.07.019.

[28] Q. Tao, L. Ding, L. Hu, Y. Chen, and T. Liu, "The performance of LS and SVD methods for SBAS InSAR deformation model solutions," *Int. J. Remote Sens.*, vol. 41, no. 22, pp. 8547–8572, Sep. 2020, doi: 10.1080/01431161.2020.1782504.

[29] M. Liao, H. Jiang, Y. Wang, T. Wang, and L. Zhang, "Improved topographic mapping through high-resolution SAR interferometry with atmospheric effect removal," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 72–79, Jun. 2013, doi: 10.1016/j.isprsjprs.2013.03.008.

[30] J. Fang, H. Lin, X. Chen, and K. Zeng, "A hybrid network of CNN and transformer for lightweight image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2022, pp. 1103–1112.

[31] Z. Xu, Z. Chen, W. Yi, Q. Gui, W. Hou, and M. Ding, "Deep gradient prior network for DEM super-resolution: Transfer learning from image to DEM," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 80–90, Apr. 2019, doi: 10.1016/j.isprsjprs.2019.02.008.

[32] Y. Wang et al., "TTSR: A transformer-based topography neural network for digital elevation model super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4403719, doi: 10.1109/TGRS.2024.3360489.

[33] D. Han, "Comparison of commonly used image interpolation methods," in *Proc. 2nd Int. Conf. Comput. Sci. Electron. Eng. (ICCSEE)*, Mar. 2013, pp. 1556–1559, doi: 10.2991/iccsee.2013.391.

[34] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015, doi: 10.1109/TPAMI.2015.2439281.

[35] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[36] D. X. Wang, P. Cui, and W. W. Zhu, "Deep asymmetric transfer network for unbalanced domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2018, vol. 32, no. 1, pp. 443–450, doi: 10.1609/aaai.v32i1.11267.

[37] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[38] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2359–2368.

[39] B. Li, X. Li, Y. Lu, S. Liu, R. Feng, and Z. Chen, "HST: Hierarchical Swin transformer for compressed image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Jan. 2022, pp. 651–668.

[40] R. M. Goldstein and C. L. Werner, "Radar interferogram filtering for geophysical applications," *Geophys. Res. Lett.*, vol. 25, no. 21, pp. 4035–4038, Nov. 1998, doi: 10.1029/1998GL900033.

[41] Q. Wang et al., "Investigation of the ground displacement in Saint Petersburg, Russia, using multiple-track differential synthetic aperture radar interferometry," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 87, May 2020, Art. no. 102050, doi: 10.1016/j.jag.2020.102050.

[42] Z.-W. Li, X.-L. Ding, D.-W. Zheng, and C. Huang, "Least squares-based filter for remote SensingImage noise reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2044–2049, Jul. 2008, doi: 10.1109/TGRS.2008.916981.

[43] X. Xing et al., "Independent component analysis (ICA) based method for estimating the deformation of highways in permafrost region (HPICA)—A case study of Maduo section of Gongyu highway," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 970–984, 2024, doi: 10.1109/JSTARS.2023.3336916.

[44] G. Zhang and Y. Z. Zheng, "Optimization estimation for spaceborne InSAR baseline using quadratic polynomial," *J. Appl. Sci.*, vol. 32, no. 4, pp. 409–415, Oct. 2014, doi: 10.3969/j.issn.0255-8297.2014.04.011.

[45] J. H. González et al., "Bistatic system and baseline calibration in TanDEM-X to ensure the global digital elevation model quality," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 3–11, Sep. 2012, doi: 10.1016/j.isprsjprs.2012.05.008.

[46] P. Li, H. Qian, and J. Wu, "Environment: Accelerate research on land creation," *Nature*, vol. 510, no. 7503, pp. 29–31, Jun. 2014, doi: 10.1038/510029a.

[47] X. Hu et al., "Remote sensing characterization of mountain excavation and city construction in loess Plateau," *Geophys. Res. Lett.*, vol. 48, no. 21, pp. 1–11, Oct. 2021, doi: 10.1029/2021gl095230.

[48] L. Zhang, X. He, T. Balz, X. Wei, and M. Liao, "Rational function modeling for spaceborne SAR datasets," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 1, pp. 133–145, Jan. 2011, doi: 10.1016/j.isprsjprs.2010.10.007.

John P. Wilson received the Ph.D. degree from the University of Toronto, Toronto, ON, Canada, in 1986.

He is currently a Professor of spatial sciences and sociology with the Dana and David Dornsife College of Letters, Arts and Sciences, University of Southern California (USC), Los Angeles, CA, USA, where he is the Founding the Director of the Spatial Sciences Institute. He was the Chair of the Department of Geography, USC. He has published numerous books and articles on these topics, including two edited volumes, *Terrain Analysis: Principles and Applications* (John Wiley and Sons, 2000) and the *Handbook of Geographic Information Science* (Blackwell, 2007), and *Environmental Applications of Digital Terrain Modeling* (John Wiley and Sons, 2018). His research focuses on the modeling of environmental systems and makes extensive use of geographic information system (GIS) software tools, fieldwork, spatial analysis techniques, and computer models.

Dr. Wilson currently serves as an Editor-in-Chief for the UCGIS GIST Body of Knowledge Project and is an active participant in the UNIGIS International Network, a worldwide consortium of universities focused on online geographic information science academic programs.

Tengfei Zhang (Graduate Student Member, IEEE) received the B.S. degree in geomatics and surveying from Zhengzhou Technology and Business University, Zhengzhou, China, in 2020, and the M.S. degree in surveying and mapping science and technology from Changsha University of Science and Technology, Changsha, China, in 2023. He is currently pursuing the Ph.D. degree with the School of Resource and Environment Science, Wuhan University, Wuhan, China.

His research focuses on the application of interferometric synthetic aperture radar (InSAR) technology for high-precision digital elevation model (DEM) reconstruction and deformation monitoring in salt solution mining.

Jun Song received the Ph.D. degree from the Imperial College London, London, U.K., in 2021.

He is currently a Research Assistant Professor with Hong Kong Baptist University, Hong Kong, China. He majors in studying the urban complex system including urban environment and transportation. His research interests include AI for geoscience, machine learning, and urban computing.

Dr. Song was awarded "The China Youth Entrepreneurship Award" in 2017.

Yumin Chen is currently a Professor and a Doctoral Supervisor with the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China. Her research focuses on the dynamic simulation of surface processes, spatial information artificial intelligence, remote sensing applications to environment issues, and big data intelligent processing and analysis.

Ruoxuan Chen received the B.S. degree in geographic information science from Yunnan University, Kunming, China, in 2023. She is currently pursuing the M.S. degree with the School of Resource and Environment Science, Wuhan University, Wuhan, China.

Her research focuses on the super-resolution reconstruction of digital elevation model (DEM).

Rui Zhu received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, China, in 2018.

He was a Research Assistant Professor with The Hong Kong Polytechnic University, and a Post-Doctoral Associate with the MIT Senseable City Laboratory, Cambridge, MA, USA. He is currently a Senior a Scientist with the Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore. His study focused on GIScience, urban informatics, and solar energy with a publication of more than 70 SCI articles in journals, such as: *Nature Communications*, *The Innovation*, and *Science Bulletin*. His study has been reported by international media, such as Singapore TV, Lianhe Zaobao, and MIT News.

Dr. Zhu is an Associate Editor of *Computer Science* (Springer Nature), an Editor of *Big Earth Data and Advances in Applied Energy*, and the Young Editor of *The Innovation*. He is also the PI/Co-I for several research grants and the Board of the Director Member of the Chinese Professional in Geographic Information Systems (GISs).

Licheng Liu received the B.S. degree in geographic information science from Wuhan University, Wuhan, China, in 2024, where he is currently pursuing the M.S. degree with the School of Resource and Environmental Sciences.

His research focuses on the fusion of optical remote sensing imagery and synthetic aperture radar (SAR) imagery for building extraction.

Lanhua Bao received the B.S. degree in geographic information science from China University of Geosciences, Wuhan, China, in 2024. She is currently pursuing the M.S. degree with the School of Resource and Environmental Sciences, Wuhan University, Wuhan.

Her research focuses on intelligent processing of spatial data.