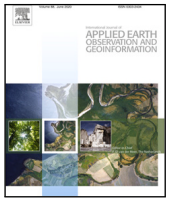
International Journal of Applied Earth Observation and Geoinformation 132 (2024) 104014

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



DEM super-resolution guided by shaded relief using attention-based fusion

Wenjun Huang, Qun Sun *, Wenyue Guo, Qing Xu, Jingzhen Ma, Tian Gao, Anzhu Yu

PLA Strategic Support Force Information Engineering University, Zhengzhou, 450001, China

ARTICLE INFO

Dataset link: https://doi.org/10.6084/m9.figshare.25590945

Keywords:
DEM super-resolution
Terrain features
Shaded relief
Deep learning
Attention-based feature fusion

ABSTRACT

Deep-learning based approaches have been proven effective for Digital Elevation Model (DEM) super-resolution (SR) tasks. Previous networks typically treat DEM elevation values as single-channel image for input. However, DEM images alone cannot fully capture spatial and terrain features. Shaded relief images (SRIs), derived from DEMs, serve as crucial visual cues that intuitively convey terrain characteristics, addressing the limitations of DEM images and providing synergistic benefits for training DL models. The primary challenge in utilizing SRIs for guiding DEM SR lies in accurately selecting a consistent structure to extract and effectively integrate features from SRIs and DEMs. In this study, we propose an Attention-based Hierarchical Terrain Fusion (AHTF) framework for guided DEM SR. Specifically, an Attention-based Feature Fusion Module (AFFM) is designed to efficiently fuse relevant information from LR DEM and SRI, which includes a feature enhancement block to select valuable features and a feature recalibration block to fuse diverse terrain features. Additionally, we optimize the loss function from the perspectives of terrain analysis and visual effects. We validate AHTF on our newly constructed real-world Shade-DEM SR dataset and two open-source DEM SR datasets. Compared to the current state-of-the-art methods, our AHTF achieves the best results in terms of root mean square error (RMSE) for elevation, slope, and aspect. Furthermore, the extracted stream networks are closer to real-world conditions. This study offers new insights and methods for further research and application in the field of DEM super-resolution. Our dataset can be obtained at https://doi.org/10.6084/m9.figshare.25590945.

1. Introduction

Digital Elevation Model (DEM) typically employs discrete elevation data from the earth's surface for terrain representation (Liang et al., 2012). DEM plays pivotal roles across various domains such as hydrology (Borzi et al., 2021), ecology (Pan et al., 2013), and geomorphology (Xiong et al., 2021). Acquiring high-resolution (HR) DEM data poses significant challenges. While traditional methods like aerial photogrammetry (Ouédraogo et al., 2014), InSAR interferometry (Chunxia et al., 2005), and LiDAR measurements (Liu, 2008) can yield high-quality DEM data. However, due to measurement costs and technological limitations, acquiring HR DEMs for global or specific areas still poses certain challenges. With advancements in technology and increasing demand for HR DEMs in various geo-spatial analysis tasks, there is a growing need for methods capable of directly reconstructing HR DEMs from low-resolution (LR) counterparts, as known as super-resolution (SR) methods (Wang et al., 2024).

For DEM SR, traditional approaches relied on interpolation-based methods such as Bicubic (Dunlop, 1980). While these methods are often efficient simple and convenient to use, they may produce HR images with insufficient sharpness and distortion (Wang et al., 2024). Subsequently, research endeavors have explored introducing learning-based methods to address this issue. The first approach involves Convolutional Neural Networks (CNNs), such as the simplistic SRCNN structure, deep residual networks (Jiao et al., 2020) and EDSR (Xu et al., 2019), which have demonstrated superior performance over traditional interpolation methods. As the network architectures for DEM SR become increasingly complex, Zhang et al. (2021) proposed the RSPCN model based on a recursive approach to reduce model complexity. Additionally, in an effort to address the challenge of insufficient capture of underlying feature dependencies in deep learning aggregation processes, Han et al. (2023) introduced a DEM SR reconstruction method (GISR) with global information constraints. Yao et al. (2024) employed implicit neural representation models for DEM SR. The second approach involves leveraging Generative Adversarial Networks (GANs) for DEM SR, such as D-SRGAN (Demiray et al., 2021), ESRGAN model (Wu et al., 2021; Wu and Ma, 2020). The third approach involves the use of Transformers for large-scale DEM SR (Li et al., 2023; Wang et al., 2024), which has been proven to achieve superior reconstruction results.

* Corresponding author.
E-mail address: 13503712102@163.com (Q. Sun).

https://doi.org/10.1016/j.jag.2024.104014
Received 14 April 2024; Received in revised form 4 June 2024; Accepted 1 July 2024
Available online 13 July 2024

In these researches, LR DEMs are typically obtained through a predetermined degradation process from HR DEMs (e.g., Bicubic downsampling) and have not utilized paired DEM datasets from real-world scenarios. The issue of DEM SR reconstruction in real-world scenarios was first addressed (Wu et al., 2021). Additionally, researches have often focused on the elevation features of DEMs. Zhou et al. (2021) proposed the EDEM-SR model, which employs parallel convolutional kernels with different receptive fields to form a dual convolutional residual network, enabling better extraction and fusion of DEM features. However, these approaches overlook the multiscale spatial and terrain characteristics inherent in DEMs. Recent researches have shown that utilizing deeper networks to convey terrain details can improve DEM reconstruction accuracy (Lin et al., 2022). Considering the unique spatial features of DEMs, Zhang et al. (2022) introduced a Terrain Feature-Aware Super-Resolution (TfaSR) method by incorporating deformable convolution network (DCN) (Dai et al., 2017). Jiang et al. (2023) considered terrain parameters by constructing loss functions based on slope and aspect to constrain network learning and convergence. Zhou et al. (2023a) proposed an MTF-SR method that integrates raster terrain features in both input and loss functions and vector terrain features in model output optimization. These researches leverage terrain features and have been proven effective in optimizing DEM SR models. In summary, existing DEM SR reconstruction research has certain limitations.

(1) In terms of DEM dataset preprocessing, the objective is to acquire paired DEM datasets. Current preprocessing methods involve degrading HR DEMs to generate LR DEMs (Zhang and Yu, 2022), resulting in the model fitting the inverse operation of the degradation process. This approach may not be entirely applicable to real-world scenarios, limiting the model's application in real world applications.
(2) In terms of DEM SR models, current approaches typically use the elevation values from LR DEM as single-channel inputs during training to generate HR DEM. These models often consider terrain feature information by using derived slope and aspect values from DEM as loss functions to guide model convergence. However, this method does not fully exploit the inherent terrain features of DEM.

To this end, we reconstructed paired DEM datasets from various sources and resolutions in real-world scenarios, resulting in a new dataset called the Sichuan dataset. Additionally, we introduce shaded relief images (SRIs) to guide the DEM SR task. SRIs simulate the distribution of light and shadows on the earth's surface, providing an intuitive representation of terrain features from large ridges to minor features such as ravines or terraces. Visualizing DEM as shaded relief offers a clearer depiction of terrain characteristics, especially valleys and ridgelines. Therefore, we proposed a terrain-feature-guided approach that supplies the network with richer terrain information. SRIs intuitively express terrain features, addressing the limitations of single-input DEM images and providing synergistic benefits for training deep learning models. The primary challenge lies in accurately selecting a consistent structure to extract and effectively integrate features from both SRIs and DEMs to guide DEM SR. We propose a novel DEM SR method called the Attention-based Hierarchical Terrain Fusion (AHTF) framework, which leverages SRIs to guide DEM SR tasks. SRIs simulate the distribution of light and shadow on the Earth's surface, providing an intuitive representation of terrain elements ranging from large ridges to small features such as valleys or terraces (Jenny et al., 2020). Expressing DEMs as shaded relief can provide a clearer representation of terrain features, particularly valleys and ridgelines (Dahal et al., 2022; Syzdykbayev et al., 2020; Lee et al., 2020). Therefore, this paper proposes a terrain feature-guided approach aimed at providing richer terrain information to the network, thereby ensuring that the generated HR DEMs maintain refined local terrain features.

The proposed AHTF utilizes hierarchical convolutions to extract elevation information from DEM and terrain information from SRIs. However, integrating the elevation information from DEM with the terrain features from SRIs to enhance DEM SR performance is a challenging task. A simple approach would be to add or concatenate the features of both, but this method fails to effectively suppress similar features between DEM and SRIs and promote the fusion of complementary features. Inspired by the Convolutional Block Attention Module (CBAM) (Woo et al., 2018), which utilizes two sequential channel and spatial attention modules to refine intermediate feature maps, we introduce a novel Attention-based Feature Fusion Module (AFFM). It effectively utilizes information from both channels for feature fusion. Specifically, the module consists of a feature enhancement block and a feature recalibration block, which are used to select valuable features and unify the pattern similarity metrics of different appearance features, respectively. Finally, multiple DCNs are connected to better extract these fused deep features. Additionally, we further optimize the collaborative loss function by considering visual effects in addition to existing terrain parameter losses.

The contributions of this paper are summarized as follows:

(1) We constructed a real-world Shade-DEM SR dataset, the Sichuan dataset. Unlike typical datasets where LR DEMs are downsampled from HR DEMs, our dataset includes 90 m LR DEMs and 30 m HR DEMs derived from real-world scenarios. It also contains SRIs generated from DEMs, totaling 2646 image pairs. This dataset covers diverse and complex terrains, including mountainous regions, valleys, depressions, and hills. The dataset is available at https://doi.org/10.6084/m9.figshare.25590945.
(2) To the best of our knowledge, this is the first study to use SRIs derived from DEMs as prior knowledge to guide the DEM SR task. We designed the AHTF framework to effectively integrate elevation information from DEMs and terrain information from SRIs. The fusion module includes a feature enhancement block to select valuable features and a feature recalibration block to fuse diverse terrain features. Additionally, we optimized the loss function from the perspectives of terrain analysis and visual effects to achieve refined terrain feature representation.
(3) We conducted systematic experiments using three datasets with different terrain types and resolutions. The proposed model was compared against traditional Bicubic interpolation and existing deep-learning based methods (EDSR, TfaSR, and SRFormer). Both quantitative (including RMSE-elevation, RMSE-slope, and RMSE-aspect) and qualitative results demonstrate that AHTF outperforms state-of-the-art models. Compared to single DEM image inputs, the SRI-guided method significantly enhances the recovery and reconstruction of HR DEMs.

2. Methodology

We propose an attention-based hierarchical terrain fusion framework (AHTF) for integrating terrain features to guide DEM SR tasks. The overview of this framework is illustrated in Fig. 1. There are four specific modules, elaborated on in detail in this section.

2.1. Hierarchical terrain feature extraction module

The module adopts a hierarchical approach to extract significant features from the upsampled LR DEM and SRI. The proposed encoder architecture consists of four sub-blocks, each comprising 3×3 convolutions followed by non-linear activation functions. We have selected the Parametric Rectified Linear Unit (PReLU) (He et al., 2015) as the activation function, which effectively addresses the overfitting issue without increasing computational complexity. This leads to faster convergence and lower error rates. By fully considering the non-linearity of rectification units, it enables a more robust initialization approach.

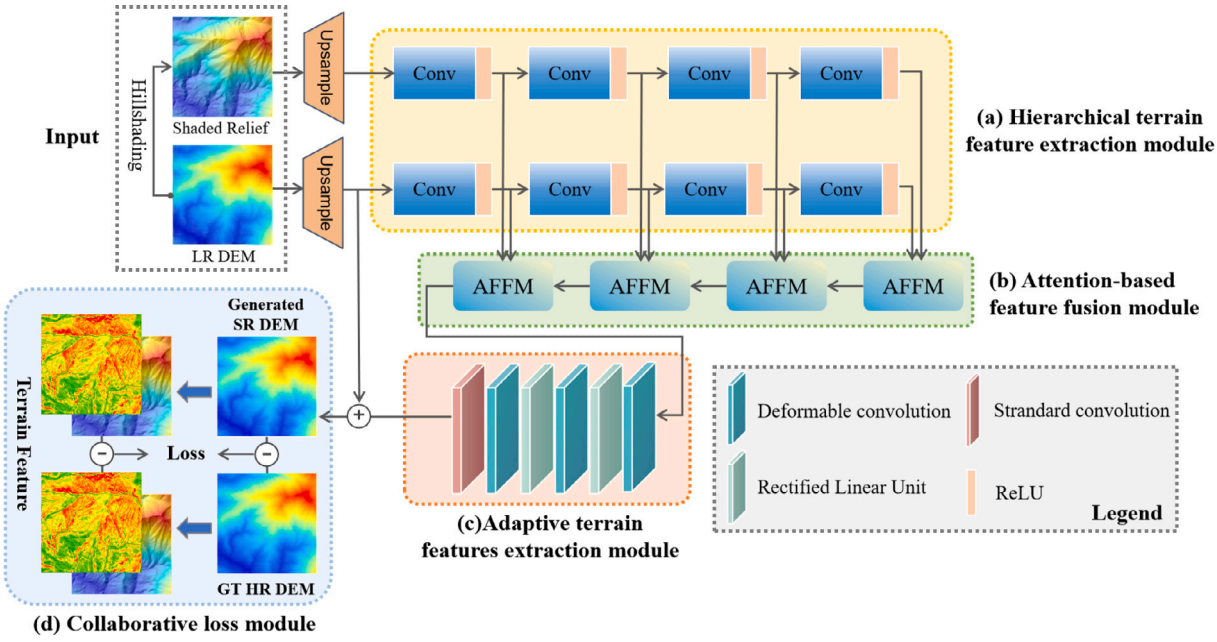


Fig. 1. Overview of the proposed attention-based hierarchical terrain fusion framework (AHTF): (a) the hierarchical terrain feature extraction module, detailed description in Section 2.1; (b) the attention-based feature fusion module, detailed description in Section 2.2; (c) the adaptive terrain features extraction module, detailed description in Zhang et al. (2022); (d) the collaborative loss module, detailed description in Section 2.3. AHTF is a dual-branch input model, with LR DEM and the SRI derived from DEM as its inputs. The generation process is described in Section 3.1.

This methodology allows us to train the AHTF network directly from scratch. The formula for feature extraction is as follows:

$$F_x^i = \begin{cases} \sigma\left(W_x^1 * D_x^{LR} + b_x^1\right), i = 1 \\ \sigma\left(W_x^i * F_x^{i-1} + b_x^i\right), 1 < \mathrm{i} \le 4 \end{cases} \tag{1}$$

where $x \in \{DEM, Shade\}$, W_x^i represents the convolutional kernels for feature extraction in the ith layer, used respectively for DEM and SRI feature extraction. b_x^i denotes the bias terms, and i denotes the number of layers used for feature extraction. σ denotes the PReLU activation function.

2.2. Attention-based feature fusion module

The intermediate feature maps extracted from the DEM and SRI channels of the two encoders possess distinct semantic information, making their fusion challenging. Simple concatenation and addition can lead to insufficient feature fusion. Therefore, we propose an attention-based fusion method, which enhances the representation capacity of DEM and SRI feature maps, capturing their salient features while suppressing unnecessary ones. Our proposed Attention-based Feature Fusion Module (AFFM) comprises a feature enhancement block and a feature recalibration block, as illustrated in Fig. 2. The formula for feature fusion is as follows:

$$F_f^i = AFFM_i\left(F_{DEM}^i, F_{Shade}^i\right), 1 \le \mathrm{i} \le 4 \tag{2}$$

where F_{DEM}^i and F_{Shade}^i represent the hierarchical extracted DEM and SRI features, F_f^i denotes the fused feature, and i indicates the number of feature layers.

- Feature enhancement block: The encoder obtains hierarchical feature maps for DEM and SRI. Inspired by Yu et al. (2019), gating units are employed as the Feature enhancement block to enhance features for each intermediate feature map of the streams. Then, feature maps are fused based on the interdependence between the two feature maps in different channels.
- Feature recalibration block: the intermediate feature maps of each stream represent local descriptors. These statistics of descriptors, including mean and maximum values, serve as effective image representatives. Nonlinear and non-exclusive relationships between the intermediate feature maps of DEM and SRI are established through pooling, nonlinear activation, convolution, and fully connected layers.

2.3. Collaborative loss module

To ensure the preservation of local terrain features during DEM SR, we introduce a collaborative loss module. Grounded in terrain analysis and visual perception, we incorporate RMSE of slope, elevation loss (RMSE and L_1 loss), and visual perception loss to optimize the training.

2.3.1. L_1 loss

In SR tasks, the mean squared error (MSE) is commonly utilized as the loss function. However, compared to MSE, the mean absolute error (MAE) provides uniform penalization for outlier errors, thereby aiding in preserving fine details during DEM SR. Consequently, we opt to utilize the L_1 loss to compute the error between corresponding pixel positions in the SR and the HR. The calculation formula is as follows:

$$L_1^X = \sum_i \left| X_{HR}^{(i)} - X_{SR}^{(i)} \right|, X = Elevation \tag{3}$$

where $X_{HR}^{(i)}$ and $X_{SR}^{(i)}$ respectively represent the values of a terrain feature X for the ith unit in the original HR DEM and the generated SR DEM.

2.3.2. Root mean square error loss

Root Mean Square Error (RMSE) quantifies the magnitude of errors between predicted and actual values. It effectively reflects the accuracy of the SR DEM. RMSE can also be computed for slope generated from terrain analysis, thereby optimizing local DEM terrain features. The calculation formula is as follows:

$$L_{RMSE}^X = \sqrt{\frac{\sum_{i=0}^n \left(X_{HR}^{(i)} - X_{SR}^{(i)} \right)^2}{n}}, X \in \{Elevation, Slope\} \tag{4}$$

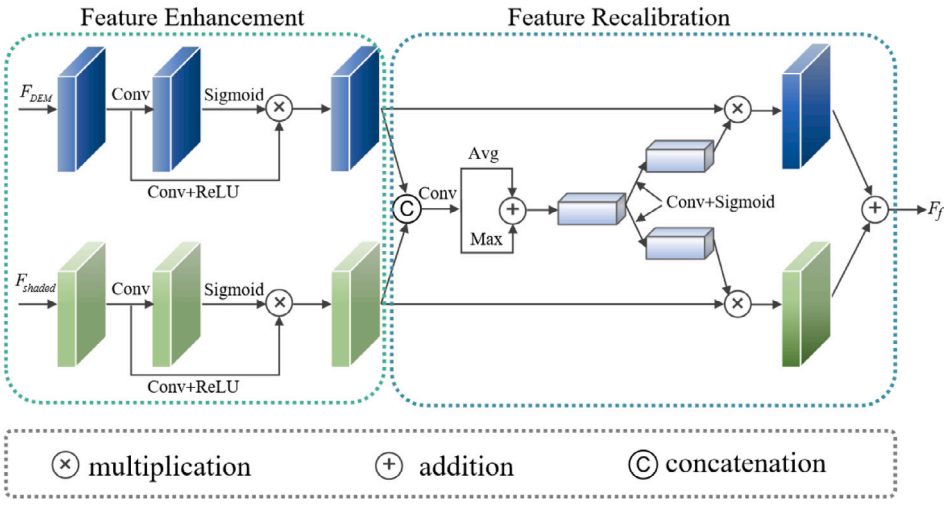


Fig. 2. The Attention-based Feature Fusion Module consists of a feature enhancement block and a feature recalibration block. AFFM takes DEM and SRI features as inputs and outputs the fused feature.

2.3.3. Visual perception loss

In addition to accounting for variations in DEM slope analysis, we also consider the visual representation of the DEM. This is quantified by computing MSE of SRI as part of the loss function.

$$L_{visual} = \frac{1}{n}\sum_{i=0}^{n}\left[X_{HR}^{(i)} - X_{SR}^{(i)}\right]^2, X = Shade \tag{5}$$

2.3.4. Collaborative loss

Considering both the accuracy of the generated DEM and the detail recovery in the DEM super-resolution process, we employ elevation L_1 loss and RMSE loss. Additionally, we integrate terrain feature losses, including the RMSE loss of the slope and terrain feature loss. The trade-offs between these losses are adjusted based on the scale differences of the datasets and the complexity of the tasks. Static weighted learning may constrain the convergence direction of the model, thereby affecting the final performance of the model. Dynamic adjustment of the weights between losses is necessary to better adapt to the DEM SR task. Inspired by Kendall et al. (2018), we adopt uncertainty-based weighting to achieve task weighting. The collaborative loss in our approach is formulated as follows:

$$L_{total} = \sum_{\tau}\left(\frac{1}{2\sigma_\tau^2}L_\tau(W) + \log\sigma_\tau^2\right) \tag{6}$$

where $L_\tau(W) \in \left\{L_1^{Elevation}, L_{RMSE}^{Elevation}, L_{RMSE}^{Slope}, L_{visual}\right\}$. σ_τ is the learnable weight parameter. $log\sigma_\tau^2$ is used as the regularization method to alleviate the influence of parameters with small weights, which optimizes numerical stability during the optimization process.

3. Experiments

3.1. DEM dataset

We selected two publicly available DEM datasets, Austria and TFaSR30. Considering that their LR DEMs are obtained through a certain degradation process from HR DEMs, this may result in the model fitting the inverse operation of the degradation process, thereby limiting its applicability in real-world scenarios. Hence, we constructed a more challenging dataset — the Sichuan Dataset — comprising pairs of DEMs from different sources.

These datasets consist of pairs of LR DEM and HR DEM images. We generate SRIs from the LR DEMs to guide DEM SR. Fig. 3 illustrates the representation of DEMs expressed as shaded relief, demonstrating its ability to articulate terrain features more clearly. For the hillshading process, we employ the equation to generate SRIs (Burrough et al., 2015):

$$\begin{aligned} Hillshade = 255.0 \cdot (&\cos Zenith_{rad} \cdot \cos Slope_{rad} \\ &+ \sin Zenith_{rad} \cdot \sin Slope_{rad} \cdot \cos(Azimuth_{rad} - Aspect_{rad})) \end{aligned} \tag{7}$$

where $Slope_{rad}$, $Aspect_{rad}$, $Zenith_{rad}$, and $Azimuth_{rad}$ represent slope, aspect, solar zenith angle, and solar azimuth angle, respectively. For computational convenience, we set $Zenith_{rad}$ and $Azimuth_{rad}$ to 45 and 315, respectively.

Furthermore, due to the wide range of values in DEM, for better learning optimization by the model, we normalize the input DEM and SRI separately. All data values are scaled between [−1, 1]. The normalization process is as follows:

$$\hat{Y} = 2 \times \frac{Y_{\text{origin}} - Y_{\min}}{Y_{\max} - Y_{\min}} \tag{8}$$

where Y_{origin} represents each original DEM data or SRI data, with $Y_{\max}$ and $Y_{\min}$ denoting the maximum and minimum values of each dataset, respectively. Next, the detailed introduction of the three datasets used in the experiments follows.

3.1.1. Sichuan dataset

To evaluate the robustness of the proposed method, we selected the western region of Sichuan, China, characterized by distinctive ridges and valleys, as the study area (see Fig. 4(a)). We reconstructed a more challenging Shaded-DEM SR dataset, called Sichuan dataset. The HR DEM and LR DEM were sourced from SRTM1 (30 m) and SRTM3 (90 m) data, respectively, both acquired by the Space Shuttle Endeavour. Preprocessing steps, including registration and cropping, were performed to obtain well-paired DEM image pairs. The study area features an elevation range of up to 7000 m, with terrain characteristics illustrated in Fig. 4(b). The western part of the study area consists of highly rugged mountains and valleys, while the eastern part encompasses the Sichuan Basin, including lowlands and hills. This diverse terrain poses significant challenges to the generalization capability of the super-resolution task. The LR DEM and HR DEM were divided into non-overlapping sub-images of 85 × 85 and 255 × 255 pixels, resulting in 2646 image pairs. These pairs were split into training and testing sets with an 8:2 ratio. sub-figures (c) and (d) display the elevation distributions of the training and testing sets, respectively. We have made the data open source and it can be obtained at https://doi.org/10.6084/m9.figshare.25590945.

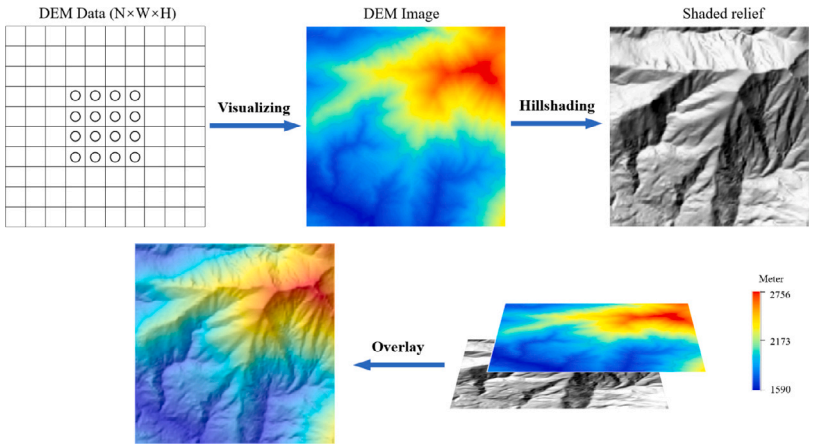


Fig. 3. The process of representing a discrete data-based DEM as SRI involves. Overlaying SRI on the DEM representation allows for a clearer depiction of terrain features, especially ridge lines and valley lines.

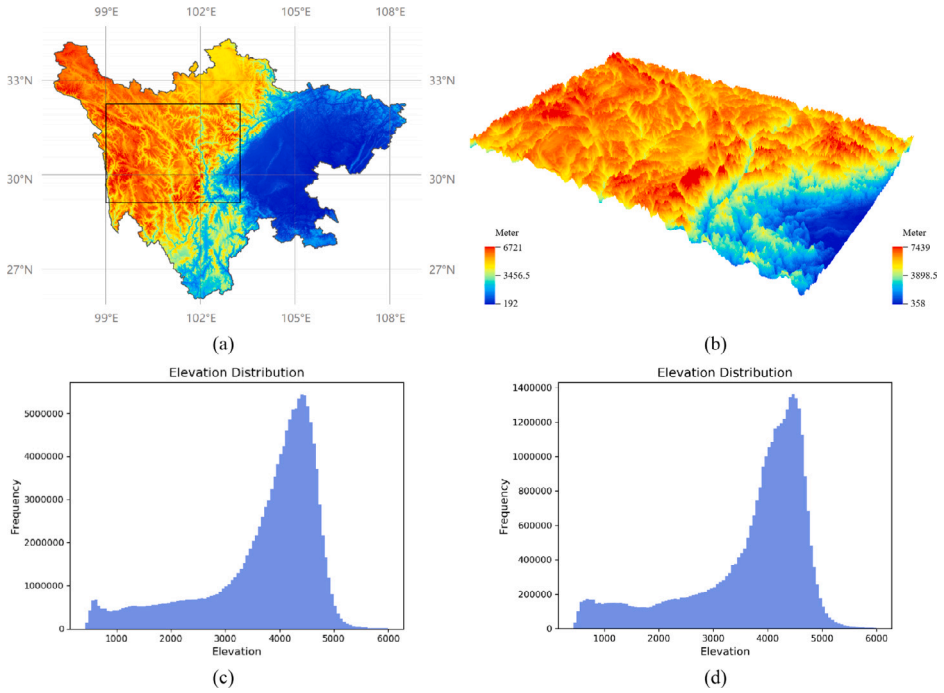


Fig. 4. Visualization of the experimental Sichuan datasets. (a) The study area; (b) Terrain features of the study area; (c) and (d) Elevation distribution maps for the training and testing datasets.

3.1.2. Austria dataset

The Austria dataset, available at https://opendem.info/superResolution.html, is constructed using airborne LiDAR scanners. It consists of 10 m DEM data from the Austria region as the HR DEM, while the LR DEM is obtained by downsampling the 10 m data (300 × 300 pixels) to 30 m (100 × 100 pixels) using nearest-neighbor interpolation, followed by upsampling back to 30 m using trilinear interpolation. The dataset comprises 2003 pairs for the training set and 101 pairs for the test set, with DEM data selected from two distinct regions for both training and testing. The elevation distribution of this region is depicted in Fig. 5, shaped by river and glacier erosion.

3.1.3. TfaSR30 dataset

The TfaSR30 dataset is available from https://doi.org/10.6084/m9.figshare.19225374. It consists of two distinct areas: Train Area and Test Area, which are geographically adjacent but non-overlapping. The dataset is provided in the format of 10 m HR DEM (3528 × 3528 pixels) and 30 m LR DEM (1176 × 1176 pixels). To accommodate the two streams input of AHTF, reprocessing was conducted. SRI was generated from the 30 m LR DEM, and LR, HR, and SRI were grouped and cut into fixed-size sub-DEMs (63 × 63 pixels and 21 × 21 pixels) without overlap. Fig. 6 illustrates their elevation distribution and the data distribution of the train and test sets. The terrain features of this area include mountains, ridges, valleys, and rivers.

3.2. Experimental setup

We selected the interpolation method BiCubic, the EDSR (Xu et al., 2019) based on residual blocks, the TfaSR (Zhang et al., 2022) considering terrain factors, and the state-of-the-art SRFormer (Zhou et al., 2023b) based on transformer as the comparative methods. It should be noted that the methods from related researches in 2024 have not been open-sourced. Therefore, we compared with the state-of-the-art models that have already been open-sourced. The deep learning methods were implemented using the PyTorch framework and trained on an NVIDIA Quadro P6000 24 GB graphics card. All four deep learning models

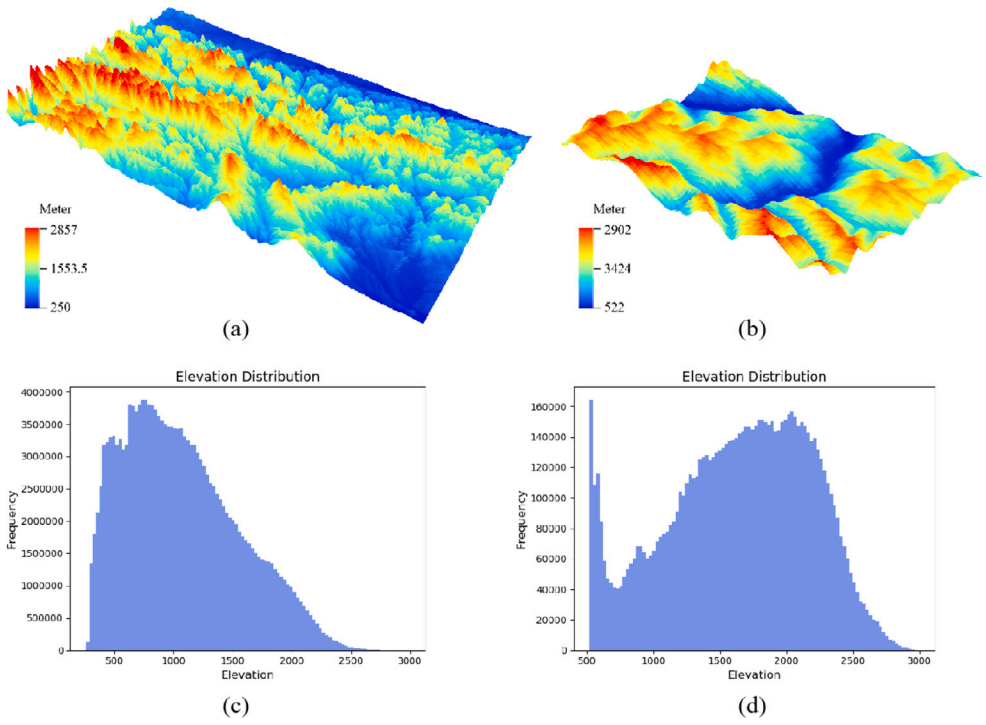


Fig. 5. Visualization of the experimental Austria datasets. (a) and (b) depict the terrain features of the training and testing areas, respectively; (c) and (d) show the elevation distribution maps for the training and testing datasets.

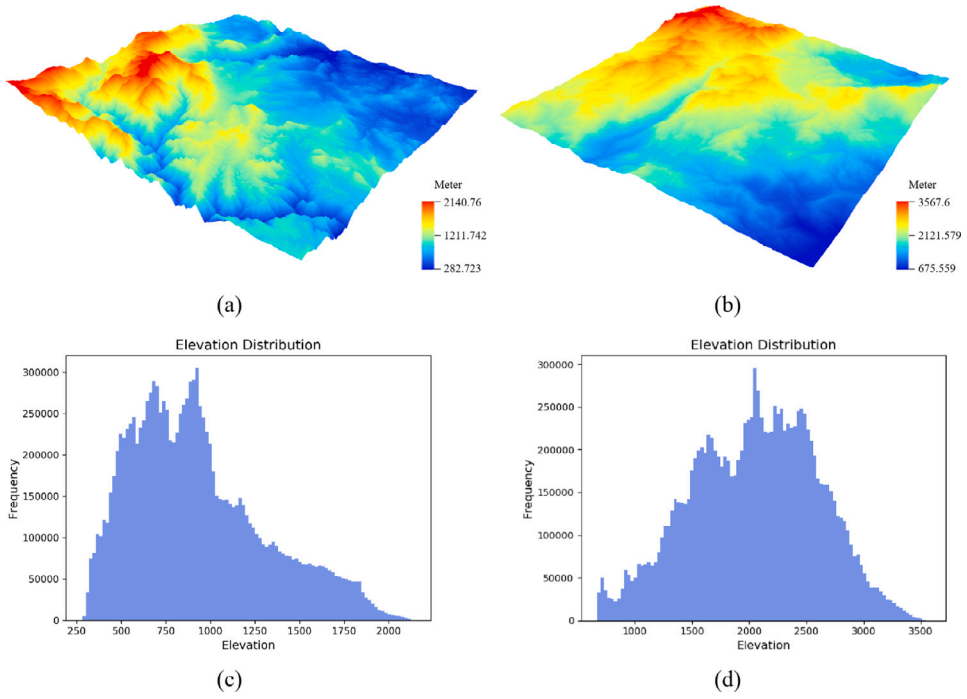


Fig. 6. Visualization of the experimental TfaSR30 dataset. (a) and (b) depict the terrain features of the training and testing areas, respectively; (c) and (d) show the elevation distribution maps for the training and testing datasets.

used the same hyperparameters with the batch size as large as possible. During training, the Adam optimizer was chosen for 200 epochs with a learning rate of 0.0001, and the loss function was the Collaborative loss described in Section 3.4. It is worth noting that, for consistency, TfaSR differed from the original research (Zhang et al., 2022) by omitting the loss of the near-stream area segmentation. For computational efficiency, a lightweight version of SRFormer was selected. According to the research (Zhou et al., 2023b), SRFormer-light can significantly reduce computational costs while maintaining almost the same accuracy. Additionally, batch sizes were set to 10, 128, and 12 for input image sizes of 100×100, 21×21, and 85×85, respectively.

Additionally, RMSE-elevation is employed to measure the changes in the DEM generated during each epoch to assess the training performance of the model. Three terrain indices, namely RMSE-elevation, RMSE-slope, and RMSE-aspect, are selected as evaluation metrics for

terrain reconstruction results. Lower RMSE values indicate better performance.

3.3. Training procedure

In the subsection, the training details of different methods are described, including EDSR, TfaSR, SRFormer, and the proposed AHTF. The RMSE of the recovered elevation during the training process on train sets of the Austria, TfaSR30, and Sichuan dataset are illustrated in Fig. 7. It is noteworthy that the average RMSE elevation values produced by the Bicubic method for the three training datasets are 3.142 m, 2.919 m, and 32.947 m, respectively. Overall, the RMSE elevation gradually decreases over 200 epochs for all four models. However, EDSR, TfaSR, and SRFormer exhibit higher RMSE in the initial 1–10 epochs due to random initialization of network weights, resulting in significant deviations between initial predictions and ground truth. In contrast, AHTF utilizes upsampling on LR DEM and features extracted by the Adaptive Terrain Features Extraction Module, followed by skip connections to generate the final SR results. This effectively avoids the aforementioned issue, giving AHTF an advantage in the early stages of training.

From Fig. 7(a)(d)(g), it can be observed that after 50 epochs of training, there is not much decrease in RMSE for all three datasets. However, when the training curves for epochs 50–200 are shown in subplots (b)(e)(h), it can be seen that the RMSE for all three models continues to decrease, indicating no overfitting. AHTF significantly outperforms the other three models, with TfaSR exhibiting better convergence compared to EDSR, especially on the Sichuan dataset, where TfaSR approaches the performance of AHTF. This demonstrates the effectiveness of the Adaptive Terrain Features Extraction Module utilizing deformable convolutions. It is worth noting that the RMSE curves of AHTF, TfaSR, and EDSR models exhibit smaller oscillations compared to the larger oscillations seen in the TfaSR model. This instability in the TfaSR model is attributed to the smaller batch size during training. SRFormer shows rapid convergence in the initial 1–100 epochs but stabilizes between 150–200 epochs, likely due to the limited size of the dataset. This indicates that Transformer requires a larger dataset compared to CNN for DEM SR tasks.

Furthermore, subplots (c)(f)(i) illustrates the RMSE curve of AHTF throughout the entire training period. It can be observed that AHTF demonstrates consistently stable performance with a steady decrease in RMSE. This stability is attributed to: (1) the feature fusion performed by the Attention-based feature fusion module, which integrates DEM and SRI features while preserving local terrain characteristics, and (2) the collaborative loss proposed to optimize training.

3.4. Inference results

3.4.1. Overall results

In this subsection, all trained models are applied to three test datasets, and the comparison results are listed in Table 1. From the statistical results, it can be seen that our proposed AHTF outperforms all other methods in terms of all three terrain evaluation metrics, not only for the Austria and TfaSR30 datasets for the task of recovering from 30 m to 10 m resolution but also for the Sichuan dataset for the task of recovering from 90 m to 30 m resolution. However, it should be noted that all five models perform less effectively on the Sichuan dataset compared to the other two datasets. This is because the LR and HR DEM of the Sichuan dataset are derived from two different real measurement data sources, and the experimental area has significant variations in elevation, posing challenges for the DEM SR task.

Additionally, it can be observed that deep learning methods generally exhibit more satisfactory performance and greater stability compared to traditional interpolation methods (BiCubic), highlighting the superiority of deep learning methods for DEM SR tasks. As the latest transformer-based SR model, SRFormer achieves better results than EDSR for DEM SR tasks. Interestingly, while SRFormer exhibits slightly lower training curve values than TfaSR on the Austria dataset, seen from Fig. 7(b), its performance on the test set is not as robust as TfaSR. This could be attributed to overfitting of SRFormer, as it relies on a transformer-based architecture and may require a larger dataset for DEM SR tasks. EDSR and SRFormer do not consider the influence of terrain features. In contrast, TfaSR, which incorporates terrain feature-awareness, demonstrates competitive performance, second only to our AHTF model. TfaSR effectively integrates deep residual modules and DCN modules to extract depth and adaptive terrain features, thereby enhancing terrain feature perception through its loss function. Building upon these insights, AHTF guides the DEM SR network with SRI, facilitating the fusion of multiple terrain and elevation features. Furthermore, AHTF optimizes the TfaSR loss function by incorporating terrain visual loss. Consequently, AHTF offers a more accurate and reliable solution for DEM SR tasks.

3.4.2. Error analysis

To gain a deeper understanding of the results, we conducted a quantitative analysis by presenting boxplots of overall errors for the five models. Fig. 8 illustrates the distribution of overall errors for each model across the three datasets. By comparing the upper and lower boundaries of each box, we can assess the dispersion of model predictions. Notably, the boxplots for AHTF, SRFormer, and TfaSR appear narrower, indicating relatively stable predictions and greater model robustness. In contrast, the boxplots for EDSR and BiCubic are wider, suggesting higher data dispersion and increased uncertainty in predictions. Furthermore, we observe a wider span of boxplot widths on the Sichuan test dataset, indicating that the Sichuan dataset presents greater challenges for DEM SR tasks due to its unique data sources and terrain characteristics. Finally, by computing the average RMSE for each model, we derived the performance curves for prediction accuracy. Considering the comprehensive analysis results, our AHTF demonstrates the capability to recover most of the DEM with high precision.

3.4.3. Qualitative assessment

(1) Elevation assessment

We selected examples of DEMs from three datasets to visually evaluate the reconstructed terrain. Fig. 9(a), (c), and (e) display the SR DEMs generated by different methods. At first glance, all methods are capable of restoring the LR DEM to a higher resolution version. However, there are still significant differences in detail among each generated DEM. To visually display the differences, we subtracted the original HR DEM from the SR DEMs to obtain elevation error maps, as shown in subfigures (b), (d), and (f), and represented the distribution of the difference results using line charts. It can be observed that compared to other models, our AHTF model exhibits a more concentrated distribution of errors around zero, with fewer errors distributed in larger error intervals, consistent with the evaluation indicator results of RMSE-elevation in Table 1. In contrast, the elevation error map of BiCubic shows large areas of blue and red, indicating a distribution of errors in larger numerical intervals. Additionally, the SRFormer model presents a predominantly red elevation error map on the Austria dataset, while the TfaSR model presents a predominantly blue elevation error map on the Sichuan dataset, as evidenced by the right-skewed or left-skewed curves in the error distribution plots, indicating a bias in the learned elevation features. Furthermore, from the error maps, it can be observed that most errors occur in areas with significant terrain fluctuations such as rivers and valleys, where the error maps appear more red or blue. However, our AHTF model exhibits lighter colors in these areas, indicating that our designed modules are able to more accurately preserve the terrain trends.

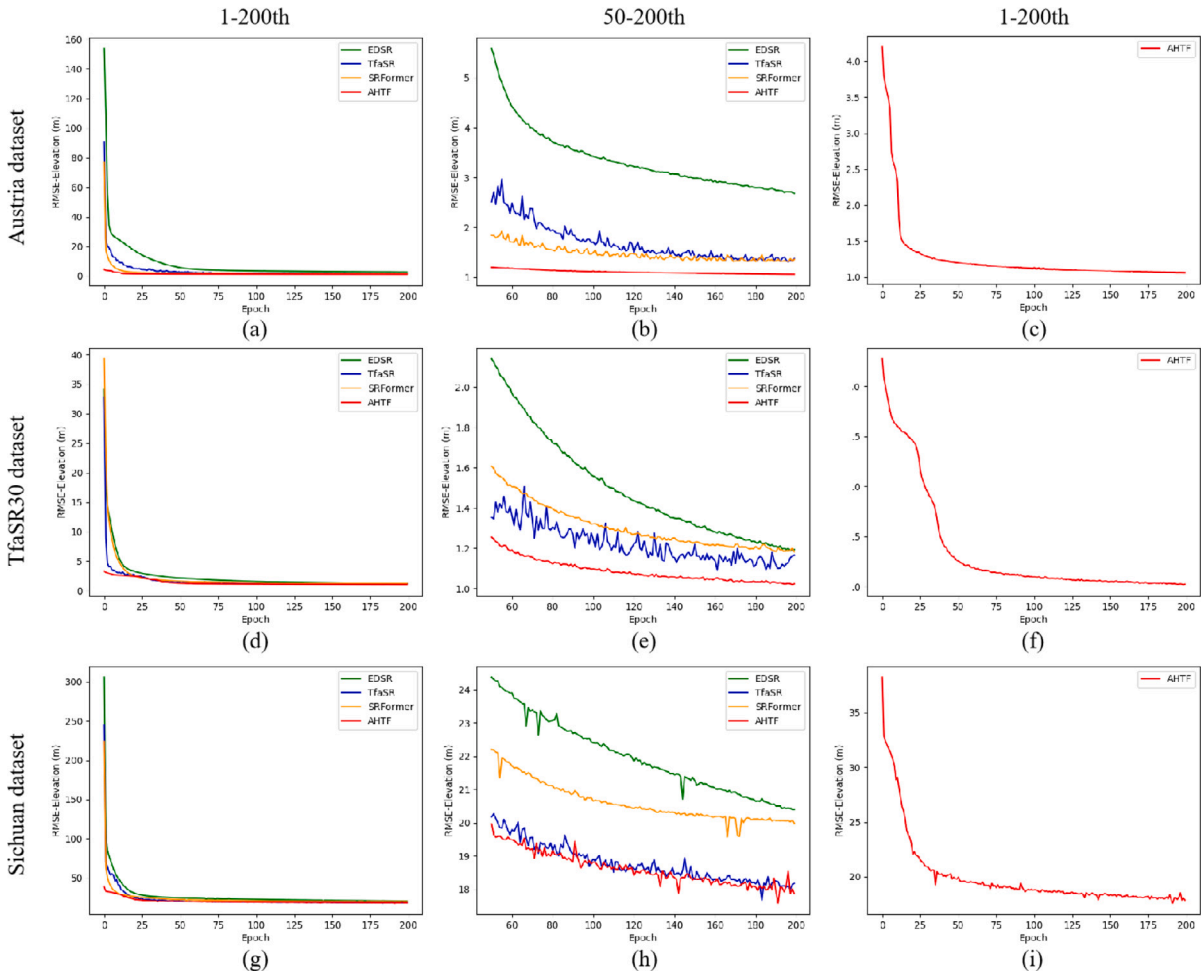


Fig. 7. Training details of the AHTF method and other SR methods on three train datasets.

Table 1
RMSE evaluation results tested on the three datasets.

Dataset	Model	RMSE-Elevation (m)	RMSE-Slope (°)	RMSE-Aspect (°)
Austria (30-10 m)	BiCubic	5.371	6.293	105.665
	EDSR	4.027	2.188	63.527
	TfaSR	1.814	1.433	52.035
	SRFormer	1.859	1.516	53.425
	AHTF (ours)	**1.471**	**1.329**	**50.398**
TfaSR30 (30-10 m)	BiCubic	3.290	1.202	49.542
	EDSR	1.904	0.805	39.752
	TfaSR	1.776	0.795	40.914
	SRFormer	1.791	0.868	39.985
	AHTF (ours)	**1.679**	**0.763**	**38.579**
Sichuan (90-30 m)	Bicubic	32.106	13.870	106.479
	EDSR	16.770	4.375	55.287
	TfaSR	15.808	4.143	53.688
	SRFormer	16.526	4.331	55.172
	AHTF (ours)	**15.470**	**4.111**	**52.723**

(2) Slope assessment

Additionally, we compared the derived slope results, as shown in Fig. 10. It can be observed that a majority of errors manifest in regions with significant slope distributions, as highlighted by the black circles in the figure. These regions typically represent intricate terrain features such as peaks and valleys, rendering the visual interpretation more intuitive compared to Fig. 9. Conversely, in areas characterized by smaller slopes—indicative of less complex terrain—the reconstruction tends to be more accurate. Hence, the reconstruction of complex terrain areas poses a notable challenge for DEM SR. Within these complex regions, our AHTF model tends to exhibit colors tending towards yellow on the error map. This indicates a precise preservation of terrain trends, facilitated by the guidance provided by SRI for DEM SR, which offers more intricate terrain features. Interestingly, the error distribution for BiCubic shows an increasing concentration of errors in the negative value range. This can be attributed to the loss of elevation information inherent in interpolation methods, leading to generated slopes that are smaller than the original ones. In contrast, our method yields smaller error values compared to other methods, with a higher concentration of errors in the (−1, 1) range and fewer errors in other ranges. This

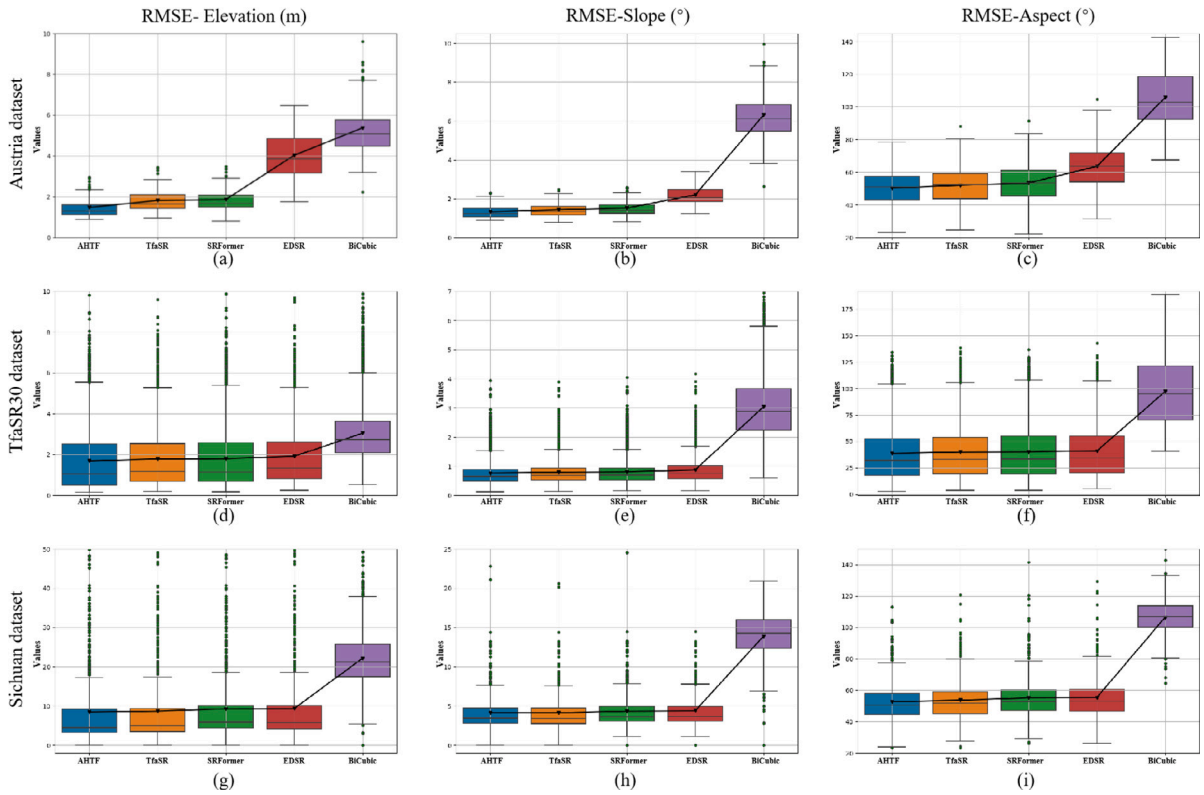


Fig. 8. Comparison of error distribution of different DEM SR models tested on three datasets.

Table 2
The IOU_{stream} results for the extracted stream networks using different methods in study areas (a)-(e).

IOU_{stream} (%)	(a)	(b)	(c)	(d)	(e)
Bicubic	54.54	56.89	58.56	54.35	52.33
EDSR	75.67	78.54	76.33	76.52	77.15
TfaSR	80.60	81.46	82.92	82.88	84.70
SRFormer	78.46	80.63	79.64	83.07	81.68
AHMF(ours)	83.24	85.56	82.57	84.45	86.01

serves to validate that AHTF is more adept at restoring complex terrain areas.

(3) Aspect assessment

Fig. 11 presents the aspect results derived from the SR DEMs generated by different models. Aspect denotes the direction of the maximum change in elevation values. Upon visual inspection of the aspect results, it becomes apparent that our method excels in restoring results most akin to the original DEM. This indicates that AHTF can preserve global trends, whereas other methods may disrupt these trends, particularly the results based on BiCubic, which lose a significant amount of terrain detail. TfaSR and SRFormer relatively maintain terrain trends, suggesting that deep learning methods can learn some relevant knowledge. Notably, terrain-aware TfaSR surpasses other methods. This assertion is further substantiated by the statistical findings presented in Table 1.

3.4.4. Stream network extraction

The position and morphology of stream networks derived from DEM are directly related to the terrain. Correctly representing river networks requires not only adjusting the elevation values of individual pixels but also correcting the relative spatial positions of pixels with different elevation values. Therefore, to assess the extent to which DEM SR preserves detailed terrain information, we extracted stream networks from DEMs reconstructed using different methods. The reference streamlines used for comparison were obtained vector hydrology layers from the open-source platform OpenStreetMap (OSM). These layers were aligned with the DEM through registration correction, ensuring a closer representation of real-world terrain reconstruction applications. The finer the DEM reconstructed by the SR methods, the better it captures terrain features, resulting in a stream network that more closely matches the reference streamlines. Taking a portion of the Austria dataset as an example, the results extracted by the five models are depicted in Fig. 12. The stream network extracted by AHTF closely matches the reference stream network at both main and tributary confluences. In contrast, the results from EDSR and BiCubic show larger deviations from the reference stream network, as shown in Fig. 12(e)(d). To more intuitively reflect the alignment between the extracted stream networks and the actual river streams, we introduce the Intersection over Union (IoU) of the stream buffer area. The calculation formula is presented as follows:

$$IOU_{stream} = \frac{Area_{SR} \cap Area_{GT}}{Area_{SR} \cup Area_{GT}} \times 100\% \quad (9)$$

where $Area_{SR}$ denotes the buffer area of the extracted stream network from different DEM SR methods, and $Area_{GT}$ represents the buffer area of the reference streamlines. We set the buffer width for the stream lines to 30 m. The IOU_{stream} results for the extracted stream networks using different methods in study areas (a)–(e) are shown in Table 2. Our AHTF achieves higher IOU scores, indicating a better alignment between the extracted and actual river lines. This superior performance of AHTF in generating more accurate stream networks is primarily due to its effective integration of terrain features from shaded reliefs and its enhanced control over these features.

4. Discussion

Observing the results from Section 3.4, it is evident that the performance of the AHTF model is optimal across all three datasets, with particularly stable performance on the Austria test dataset. Therefore, the experiments in this subsection are conducted on the Austria dataset for testing purposes.

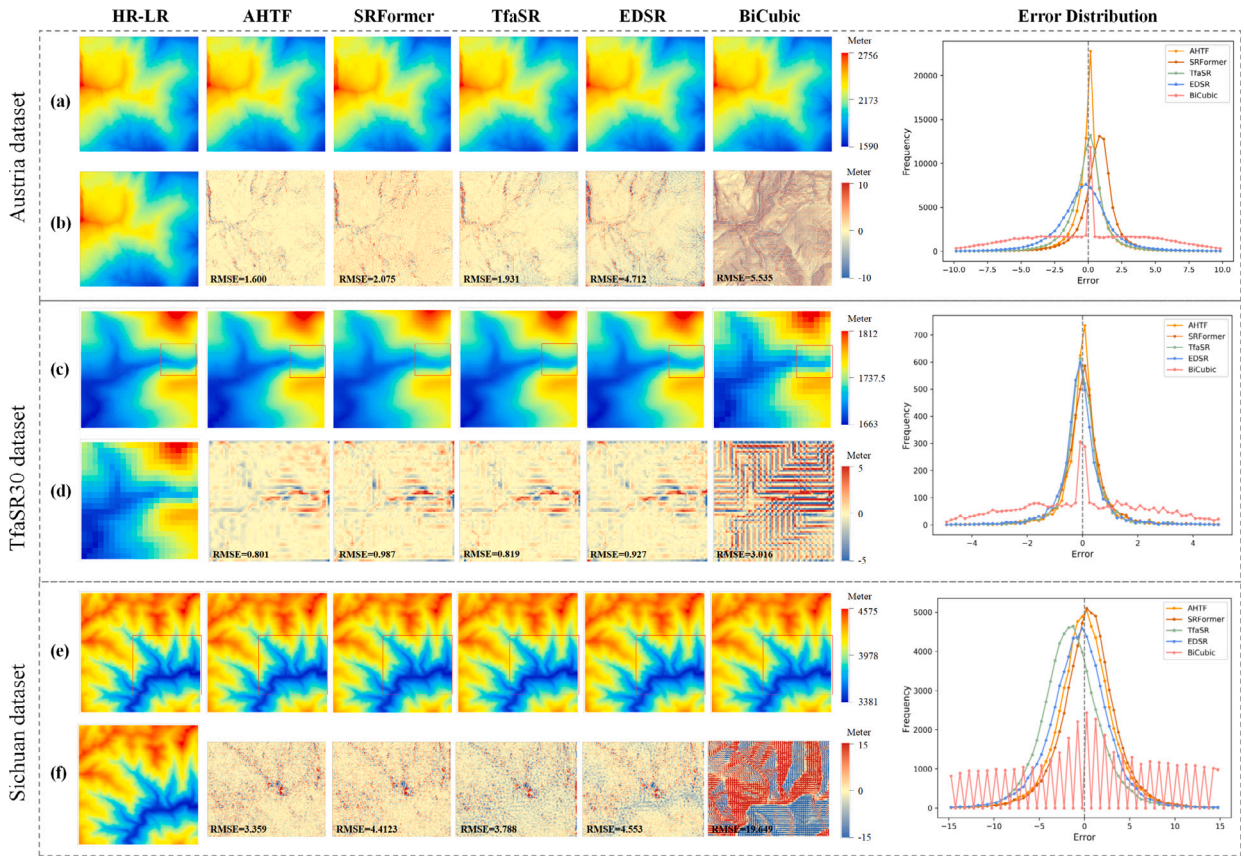


Fig. 9. The comparison of SR DEM results generated by different models, based on selecting one DEM example from three datasets. Rows (a), (c), and (e) present the visual representations of SR DEM generated, while rows (b), (d), and (f) display the error maps derived from the difference between the SR DEM within the red-bordered area and the original HR DEM elevation values. In the error maps, colors closer to red or blue indicate larger errors, while colors closer to yellow indicate smaller errors. The error distribution results are shown on the right side. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
RMSE evaluation results with and without inputting SRI.

Inputs	Epoch	RMSE-Elevation (m)	RMSE-Slope (°)	RMSE-Aspect (°)
LR DEM + SRI	100	1.559	1.400	51.317
	200	1.471	1.329	50.398
LR DEM + LR DEM	100	1.534	1.389	50.905
	200	1.528	1.386	50.838

4.1. The impact of terrain guidance

DEM contains single-channel elevation information, while SRI can better highlight the terrain features of DEM. AHTF guides the DEM SR task with SRIs, enabling the network to learn richer terrain features and thereby restore more detailed DEMs. To demonstrate this conclusion, we conducted the experiment with a set of dual inputs, both LR DEM, and the results are shown in Table 3. The metrics RMSE-elevation, RMSE-slope, and RMSE-aspect are all inferior to the results obtained when using LR DEM and SRI as inputs. This indicates that SRI can complement DEM features, allowing the model to recover more terrain details.

4.2. The impact of the AFFM module

To validate the effectiveness of the fusion method employed by AHTF in integrating SRI and DEM features, we replaced the AFFM module with simple multi-feature addition and concatenation, denoted as $AHTF_{add}$ and $AHTF_{cat}$, respectively. To eliminate other influences, all parameters remained the same as in Section 3.2. The experimental results are listed in Table 4. It can be observed that the results of using the AFFM module outperform those of $AHTF_{add}$ and $AHTF_{cat}$ across all three evaluation metrics, demonstrating the effectiveness of the attention-based feature fusion module.

Table 4
RMSE evaluation results of the fusion method in integrating SRI and DEM features.

	Epoch	RMSE-Elevation (m)	RMSE-Slope (°)	RMSE-Aspect (°)
$AHTF_{add}$	100	1.659	1.437	51.818
	200	1.520	1.381	51.114
$AHTF_{cat}$	100	1.552	1.403	51.364
	200	1.486	1.354	50.652
AHTF	100	1.559	1.400	51.317
	200	1.471	1.329	50.398

To further assess the efficacy of the AFFM module in DEM SR, we visualized the input and output feature maps of addition, concatenation fusion, and AFFM in Fig. 13. Taking the first row as an example, the feature maps of DEM and SRI extracted by the Hierarchical terrain feature extraction module (subplot (a) and subplot (f)) were fused using Eq. (2) to obtain subplot (c). From Fig. 13, it is observed that both addition and concatenation fusion methods tend to emphasize texture features represented by SRI while overlooking elevation information. Particularly, the addition fusion method more significant bias towards

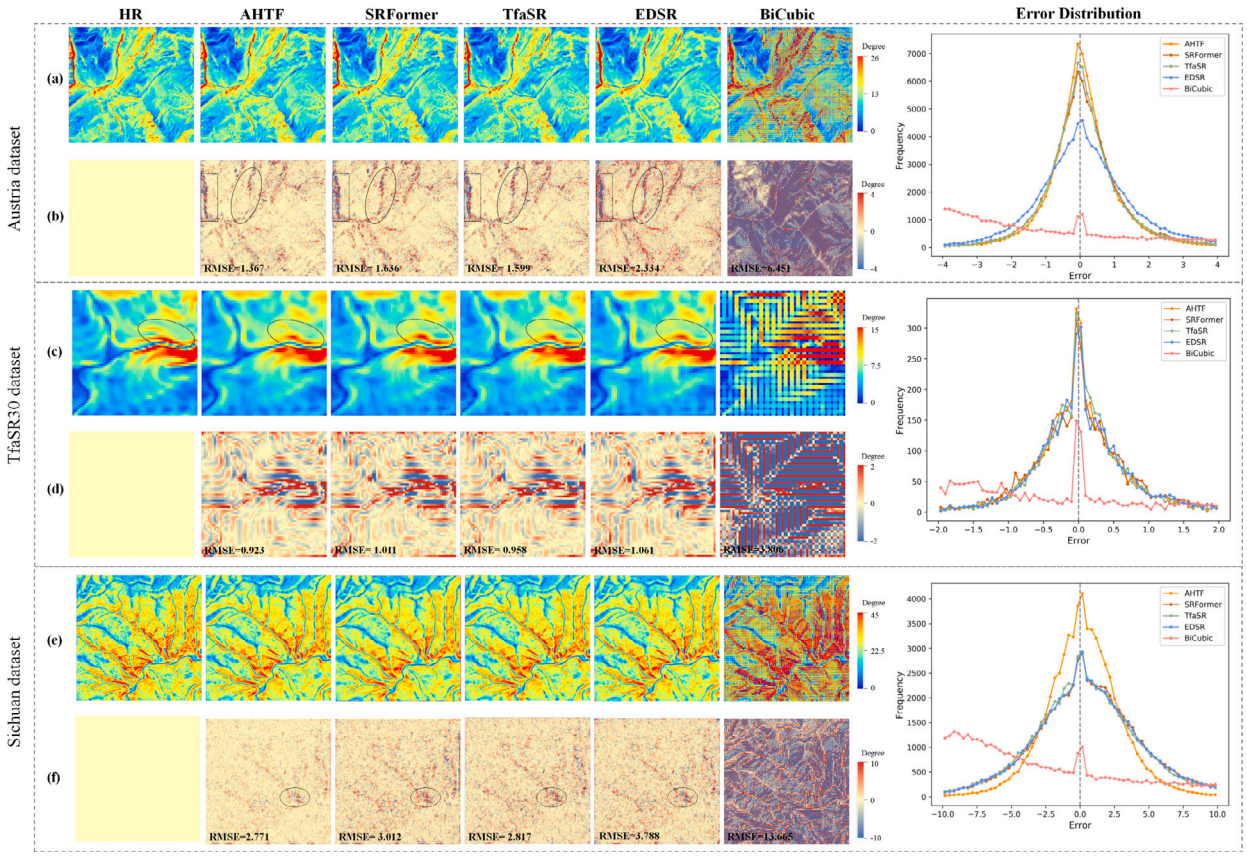


Fig. 10. The comparison of slope derived from SR DEM generated by different models, based on the DEM examples shown in Fig. 9. Rows (a), (c), and (e) present the corresponding visual representations of slope, while rows (b), (d), and (f) display the error maps obtained from the difference between the SR slope values and the original HR slope values. In the error maps, colors closer to red or blue indicate larger errors, while colors closer to yellow indicate smaller errors. The error distribution results are shown on the right side. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SRI, evident from the maximum value of the feature map being close to that of the SRI map and significantly different from the DEM feature map. However, AFFM effectively balances the elevation information extracted from the DEM images and the terrain information extracted from the SRI. This is primarily due to the attention mechanism employed in AFFM, which enhances feature representation capability while leveraging the nonlinear and non-exclusive relationships between the two feature maps. The analysis conclusions are consistent with the quantitative results in Table 4. Therefore, the Attention-based Feature Fusion Module facilitates better terrain recovery by the model.

4.3. The impact of collaborative loss

In this subsection, the influence of different components of the collaborative loss function is investigated, as shown Table 5. Three scenarios are considered: Loss1, which accounts for RMSE loss for elevation values only; Loss2, including RMSE loss for elevation values and L1 loss; and Loss3, where terrain feature loss is excluded. The experimental outcomes, depicted in Eq. (6), reveal that incorporating L1 loss yields superior performance, potentially due to its lower sensitivity to errors. Moreover, considering the RMSE of slope allows the model to optimize for reduced errors. The inclusion of slope loss, derived from DEM data, notably reduces RMSE-slope for the test set. Importantly, by integrating both terrain analysis and visual effect considerations, selecting RMSE of slope, elevation loss (RMSE and L1 loss), and terrain visual loss collectively minimizes errors, thereby preserving local terrain features in the DEM SR task.

Table 5
RMSE evaluation results of collaborative loss.

Loss	RMSE-Elevation (m)	RMSE-Slope (°)	RMSE-Aspect (°)
Loss1	1.484	1.348	50.753
Loss2	1.473	1.347	50.609
Loss3	1.476	1.330	50.419
Collaborative Loss	1.471	1.329	50.398

Table 6
The comparison of model complexity on different models.

	Params(M)	FLOPs(G)
EDSR	1.552	124.324
TfaSR	1.275	**102.144**
SRFormer	**1.181**	121.141
AHTF (Ours)	1.320	104.228

4.4. The assessment of model complexity

We selected a LR DEM input size of 100 × 100 and an upsampling factor of 3. We used the number of multiply-accumulate operations (MACs) and the number of parameters (Params) as metrics for quantitative comparison between AHTF and three other deep learning models, shown in Table 6. It is important to note that AHMF requires two input streams (DEM and SRI), yet its computational complexity does not significantly increase compared to single-input models. Moreover, considering that reconstruction accuracy is more critical than efficiency for DEM SR (Zhang and Yu, 2022), and most downstream geo-spatial tasks do not require real-time updates, our AHMF achieves more satisfactory results.

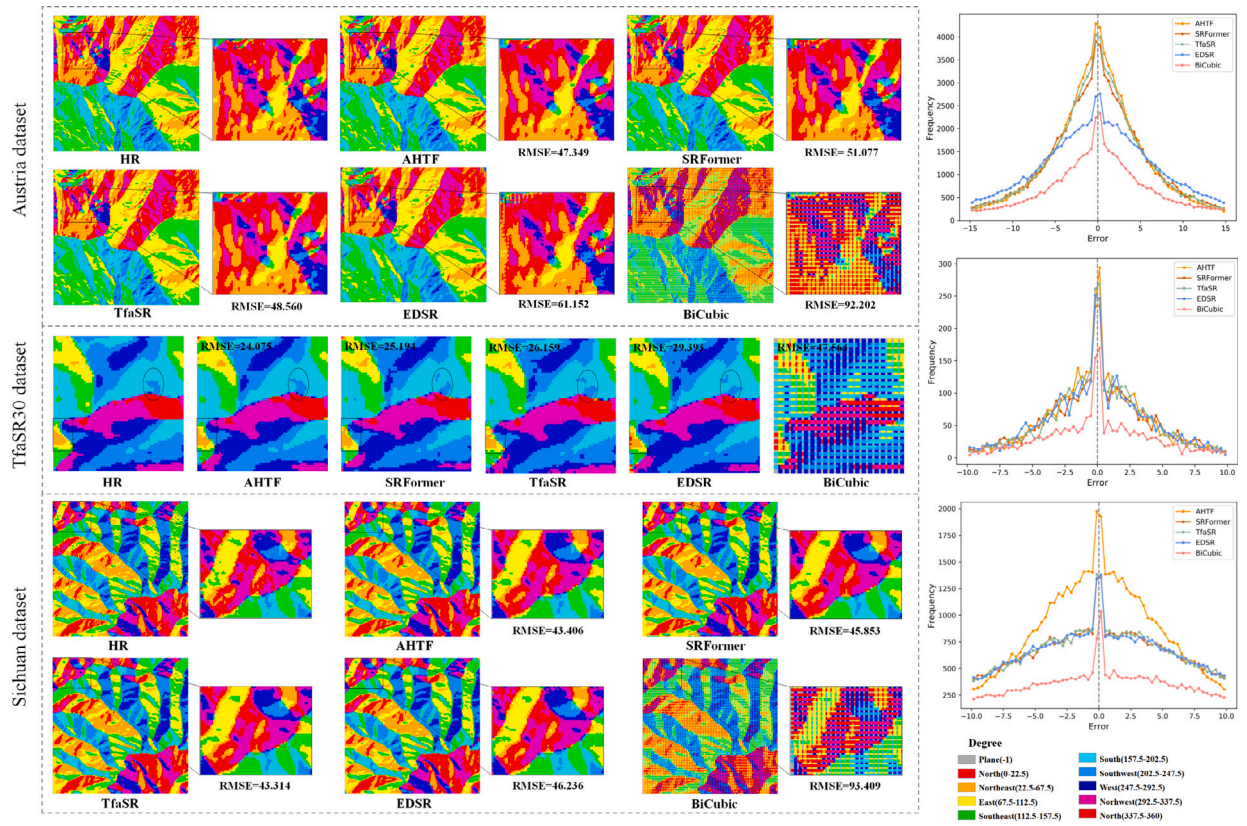


Fig. 11. The comparison of aspect derived from SR DEM generated by different models, based on the DEM examples shown in Fig. 9. The right side presents the distribution of error maps obtained from the difference between the SR aspect values and the original HR aspect values.

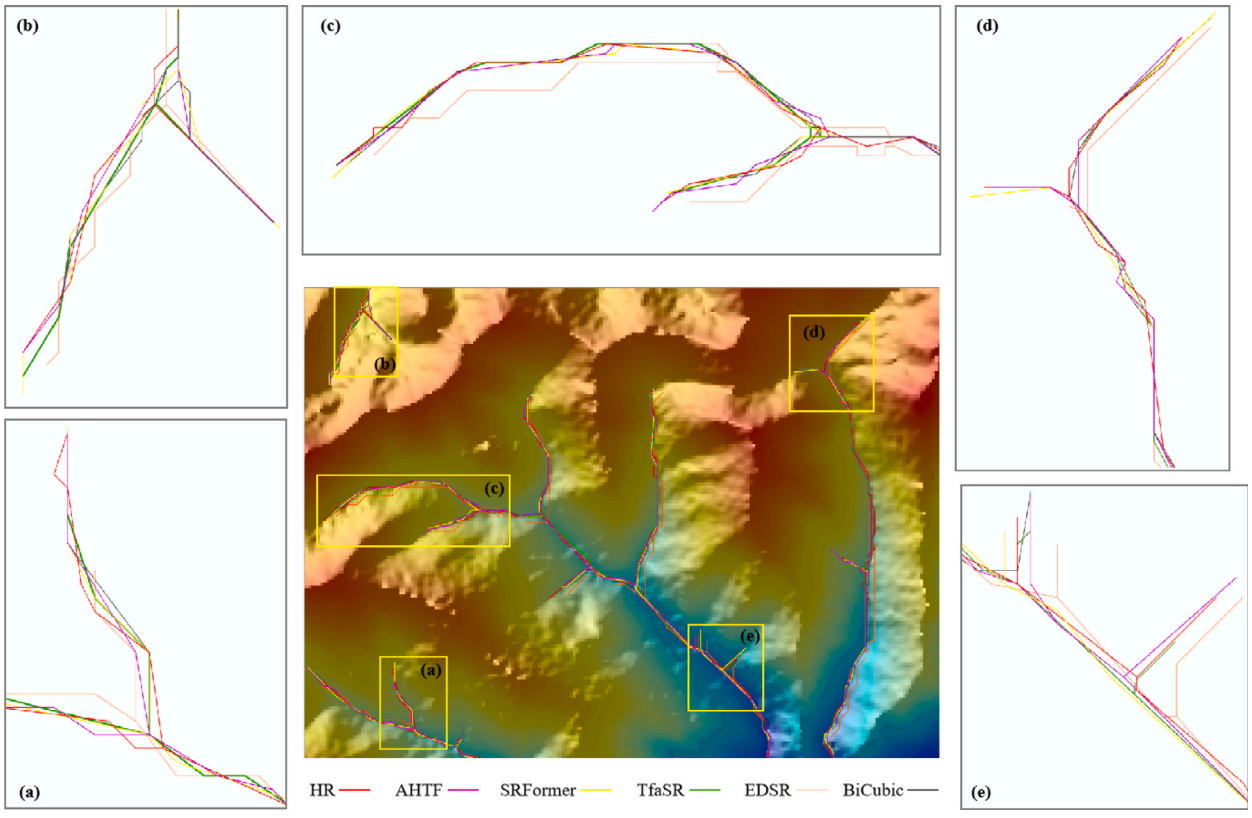


Fig. 12. The comparison of river extraction from SR DEMs produced by different methods.

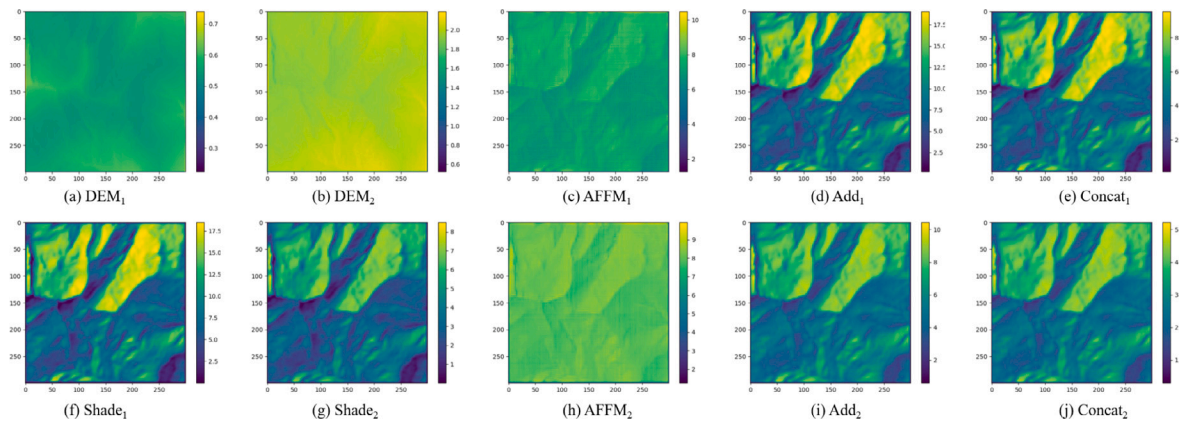


Fig. 13. Visualization of feature maps, exemplified by a region of the Austria dataset. (a) and (f) depict feature maps extracted from the DEM images; (b) and (g) represent feature maps extracted from the SRIs; (c), (d), and (e) show the feature maps obtained by passing the first layer (first column) through the proposed AFFM, addition, and concatenation, respectively; (h), (i), and (j) correspond to the feature maps of the second layer (second column).

4.5. Limitations

Through the integration of SRI and DEM for multi hierarchical feature fusion, as well as the refinement of loss functions, our AHTF model has shown promising results in DEM SR. However, several limitations persist.

Firstly, the model exhibits insufficient robustness in the rugged terrain of the real-world Sichuan dataset. Therefore, acquiring more DEMs of different resolutions through open-source means is one method to enhance the model's robustness.

Secondly, the model guides DEM SR tasks with SRI to supplement terrain features. With the diverse development of measurement methods, consideration can also be given to integrating multi-source remote sensing data. For example, using Sentinel images to guide DEM SR. Beyond raster data, vector information such as open-source river network data should also be considered. This will pose greater challenges for multi-source data fusion methods.

Thirdly, the upsampling factor of the method is fixed rather than arbitrary, which may limit the potential applications of our approach. This aspect will be addressed in our forthcoming research endeavors.

5. Conclusion

In this paper, we introduce SRIs derived from DEMs to guide the DEM SR task for the first time. We propose a novel Attention-based Hierarchical Terrain Fusion framework for this purpose. AHTF integrates four specially designed modules. Initially, elevation and terrain features are extracted from LR DEM and SRI using multi-hierarchical convolutions. Subsequently, an Attention-based feature fusion module thoroughly fuses these features. Then, an adaptive terrain feature extraction module is employed to identify and extract multi-scale terrain features. Finally, the loss function is optimized from the perspectives of terrain analysis and visual effects. The effectiveness of these modules is systematically validated. Testing results on multiple datasets with different resolutions demonstrate the superiority of our approach over methods solely relying on DEM images. In addition, the superiority of AHTF is demonstrated through evaluations and visualizations of SR results in diverse terrains. As remote sensing data becomes more accessible, our research will provide a reference for the fusion of multiple data sources. In the future, we will further explore the feasibility of guiding DEM SR tasks with multi-source data to generate more accurate global HR terrain products.

CRediT authorship contribution statement

Wenjun Huang: Writing – original draft, Validation, Formal analysis. **Qun Sun:** Supervision, Funding acquisition. **Wenyue Guo:** Resources, Investigation. **Qing Xu:** Resources, Data curation. **Jingzhen Ma:** Supervision, Investigation. **Tian Gao:** Methodology, Conceptualization. **Anzhu Yu:** Writing – review & editing, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Our dataset can be obtained at https://doi.org/10.6084/m9.figshare.25590945.

References

Borzi, G., Roig, A., Tanjal, C., Santucci, L., Tejada Tejada, M., Carol, E., 2021. Flood hazard assessment in large plain basins with a scarce slope in the Pampean Plain, Argentina. Environ. Monit. Assess. 193, 1–14.

Burrough, P.A., McDonnell, R.A., Lloyd, C.D., 2015. Principles of Geographical Information Systems. Oxford University Press, USA.

Chunxia, Z., Linlin, G., Dongchen, E., Hsingchung, C., 2005. A case study of using external DEM in InSAR DEM generation. Geo-Spatial Inform. Sci. 8, 14–18.

Dahal, A., van den Bout, B., Westen, C., Nolde, M., 2022. Deep learning-based super-resolution of digital elevation models in Data Poor Regions. http://dx.doi.org/10.31223/X5DD21, EarthArXiv.

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y., 2017. Deformable convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 764–773.

Demiray, B.Z., Sit, M., Demir, I., 2021. D-SRGAN: DEM super-resolution with generative adversarial networks. SN Comput. Sci. 2, 1–11.

Dunlop, G., 1980. A rapid computational method for improvements to nearest neighbour interpolation. Comput. Math. Appl. 6 (3), 349–353.

Han, X., Ma, X., Li, H., Chen, Z., 2023. A global-information-constrained deep learning network for digital elevation model super-resolution. Remote Sens. 15 (2), 305.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. ICCV, In: ICCV '15, IEEE Computer Society, USA, pp. 1026–1034. http://dx.doi.org/10.1109/ICCV.2015.123.

Jenny, B., Heitzler, M., Singh, D., Farmakis-Serebryakova, M., Liu, J.C., Hurni, L., 2020. Cartographic relief shading with neural networks. IEEE Trans. Vis. Comput. Graphics 27 (2), 1225–1235.

Jiang, Y., Xiong, L., Huang, X., Li, S., Shen, W., 2023. Super-resolution for terrain modeling using deep learning in high mountain Asia. Int. J. Appl. Earth Obs. Geoinf. 118, 103296.

Jiao, D., Wang, D., Lv, H., Peng, Y., 2020. Super-resolution reconstruction of a digital elevation model based on a deep residual network. Open Geosci. 12 (1), 1369–1382.

Kendall, A., Gal, Y., Cipolla, R., 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7482–7491.

Lee, D.G., Shin, Y.H., Lee, D.-C., 2020. Land cover classification using SegNet with slope, aspect, and multidirectional shaded relief images derived from digital surface model. J. Sensors 2020, 1–21.

Li, Z., Zhu, X., Yao, S., Yue, Y., García-Fernández, Á.F., Lim, E.G., Levers, A., 2023. A large scale digital elevation model super-resolution transformer. Int. J. Appl. Earth Obs. Geoinf. 124, 103496.

Liang, S., Li, X., Wang, J., 2012. Geometric processing and positioning techniques. In: Advanced Remote Sensing, Elsevier, Oxford, UK, pp. 33–74.

Lin, X., Zhang, Q., Wang, H., Yao, C., Chen, C., Cheng, L., Li, Z., 2022. A DEM super-resolution reconstruction network combining internal and external learning. Remote Sens. 14 (9), 2181.

Liu, X., 2008. Airborne LiDAR for DEM generation: Some critical issues. Prog Phys Geogr. 32 (1), 31–49.

Ouédraogo, M.M., Degré, A., Debouche, C., Lisein, J., 2014. The evaluation of unmanned aerial system-based photogrammetry and terrestrial laser scanning to generate DEMs of agricultural watersheds. Geomorphology 214, 339–355.

Pan, F., Liao, J., Li, X., Guo, H., 2013. Application of the inundation area—lake level rating curves constructed from the SRTM DEM to retrieving lake levels from satellite measured inundation areas. Comput. Geosci. 52, 168–176.

Syzdykbayev, M., Karimi, B., Karimi, H.A., 2020. A method for extracting some key terrain features from shaded relief of digital terrain models. Remote Sens. 12 (17), 2809.

Wang, Y., Jin, S., Yang, Z., Guan, H., Ren, Y., Cheng, K., Zhao, X., Liu, X., Chen, M., Liu, Y., et al., 2024. TTSR: A transformer-based topography neural network for digital elevation model super-resolution. IEEE Trans. Geosci. Remote Sens..

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 3–19.

Wu, Z., Ma, P., 2020. ESRGAN-based DEM super-resolution for enhanced slope deformation monitoring in lantau island of Hong Kong. Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. 43, 351–356.

Wu, Z., Zhao, Z., Ma, P., Huang, B., 2021. Real-world DEM super-resolution based on generative adversarial networks for improving InSAR topographic phase simulation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 8373–8385.

Xiong, L., Tang, G., Yang, X., Li, F., 2021. Geomorphology-oriented digital terrain analysis: Progress and perspectives. J. Geogr. Sci. 31, 456–476.

Xu, Z., Chen, Z., Yi, W., Gui, Q., Hou, W., Ding, M., 2019. Deep gradient prior network for DEM super-resolution: Transfer learning from image to DEM. ISPRS J. Photogramm. Remote Sens. 150, 80–90.

Yao, S., Cheng, Y., Yang, F., Mozerov, M.G., 2024. A continuous digital elevation representation model for DEM super-resolution. ISPRS J. Photogramm. Remote Sens. 208, 1–13.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S., 2019. Free-form image inpainting with gated convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4471–4480.

Zhang, R., Bian, S., Li, H., 2021. RSPCN: Super-resolution of digital elevation model based on recursive sub-pixel convolutional neural networks. ISPRS Int. J. Geo-Inf. 10 (8), 501.

Zhang, Y., Yu, W., 2022. Comparison of DEM super-resolution methods based on interpolation and neural networks. Sensors (Basel) 22 (3), 745.

Zhang, Y., Yu, W., Zhu, D., 2022. Terrain feature-aware deep learning network for digital elevation model superresolution. ISPRS J. Photogramm. Remote Sens. 189, 143–162.

Zhou, A., Chen, Y., Wilson, J.P., Chen, G., Min, W., Xu, R., 2023a. A multi-terrain feature-based deep convolutional neural network for constructing super-resolution DEMs. Int. J. Appl. Earth Obs Geoinf. 120, 103338.

Zhou, A., Chen, Y., Wilson, J.P., Su, H., Xiong, Z., Cheng, Q., 2021. An enhanced double-filter deep residual neural network for generating super resolution DEMs. Remote Sens. 13 (16), 3089.

Zhou, Y., Li, Z., Guo, C.-L., Bai, S., Cheng, M.-M., Hou, Q., 2023b. Srformer: Permuted self-attention for single image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12780–12791.