Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Highlights

A large scale Digital Elevation Model super-resolution Transformer

International Journal of Applied Earth Observation and Geoinformation xxx (xxxx) xxx

Zhuoxiao Li, Xiaohui Zhu*, Shanliang Yao, Yong Yue, Ángel F. García-Fernández, Eng Gee Lim, Andrew Levers

- A overlapped window attention module enables the model to use more elevation points which considers a larger contextual range when learning terrain features, thereby enhancing accuracy and robustness in identifying and generating complex terrain patterns.
- A module for generating super-resolution coordinates ensures that the super-resolution DEM can be directly used in the GIS systems, emphasizing the real-world value of our research, which offers not only technological innovation but also practical applicability.
- Our proposed method consistently outperforms other super-resolution methods across different resolutions. This is demonstrated both visually and quantitatively, with DSRT producing DEMs that closely resemble the original in appearance and achieving the lowest Root Mean Square Error values across all tests.
- Our study is unique in its focus on the performance across different terrain types such as urban, water, canyon and plateau regions. We find that the performance of these methods is influenced by the specific characteristics of the terrain, with DSRT demonstrating particular strength in complex terrains.

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

A large scale Digital Elevation Model super-resolution Transformer

Zhuoxiao Li [a,b], Xiaohui Zhu [a,*], Shanliang Yao [a,b], Yong Yue [a], Ángel F. García-Fernández [b,c], Eng Gee Lim [a], Andrew Levers [d]

[a] *School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China*
[b] *Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, L69 3GJ, UK*
[c] *ARIES Research Centre, Universidad Antonio de Nebrija, Madrid, 263001, Spain*
[d] *Digital Innovation Facility, University of Liverpool, Liverpool, 215123, UK*

ARTICLE INFO

MSC:
0000
1111

Keywords:
DEM super-resolution
Spatial interpolation
Shifted window
Transformer
Convolutional neural networks
Generative adversarial networks

ABSTRACT

The Digital Elevation Model (DEM) super-resolution approach aims to improve the spatial resolution or detail of an existing DEM by applying techniques such as machine learning or spatial interpolation. Convolutional Neural Networks and Generative Adversarial Networks have exhibited remarkable capabilities in generating high-resolution DEMs from corresponding low-resolution inputs, significantly outperforming conventional spatial interpolation methods. Nevertheless, these current methodologies encounter substantial challenges when tasked with processing exceedingly high-resolution DEMs (256×256,512×512, or higher), specifically pertaining to the accurate restore maximum and minimum elevation values, the terrain features, and the edges of DEMs. Aiming to solve the problems of current super-resolution techniques that struggle to effectively restore topographic details and produce high-resolution DEMs that preserve coordinate information, this paper proposes an improved DEM super-resolution Transformer(DSRT) network for large-scale DEM super-resolution and account for geographic information continuity. We design a window attention module that is used to engage more elevation points in low-resolution DEMs, which can learn more terrain features from the input high-resolution DEMs. A GeoTransform module is designed to generate coordinates and projections for the DSRT network. We conduct an evaluation of the network utilizing DEMs of various types of terrains and elevation differences at resolutions of 64×64,256×256 and 512 × 512. The network demonstrated leading performance across all assessments in terms of root mean square error (RMSE) for elevation, slope, aspect, and curvature, indicating that Transformer-based deep learning networks are superior to CNNs and GANs in learning DEM features.

1. Introduction

The Digital Elevation Model (DEM) symbolizes the Earth's 3D terrain surface, finding varied uses in topographical mapping, geology, hydrology, civil engineering, and remote sensing (Moore et al., 1991). Elevation data, integral to DEMs, can be harnessed from various sources such as aerial photogrammetry (Ouédraogo et al., 2014; Uysal et al., 2015), satellite imagery (Fran and Martin, 2006; Shean et al., 2016), and airborne LiDAR surveys (Liu, 2008). These data inputs are processed into a standard grid of elevation points, typically in raster format, wherein each cell represents a unique elevation value (Moore et al., 1991). DEMs primarily serve as tools for generating topographical maps and rendering terrain visualization (Smith and Clark, 2005).

As technological advancements progress and the requisites of various geospatial analytical tasks intensify, there is a mounting need for high-resolution DEMs (Zhang and Yu, 2022; Demiray et al., 2021; Xu et al., 2019). Granular terrain information can yield more precise outcomes in urban planning, flood simulations, and landslide predictions (Yang et al., 2015). However, many extant DEMs, constrained by data acquisition techniques, fall short of offering the desired level of detail. This underscores the exigency for super-resolution (SR) techniques in DEMs, aiming to augment the resolution of existing data and furnish a more detailed and accurate topographical representation (Hayat, 2018). SR in DEMs encompasses a suite of methodologies designed to amplify the spatial resolution of DEMs, transcending the inherent resolution of the input data (Zhang et al., 2022a; Lin et al., 2022). This enhancement is not merely achieved via rudimentary interpolation but by deploying intricate algorithms and techniques tailored to reconstruct details latent in the original dataset. In essence, the objective of DEM SR is to engender a higher-resolution output from a lower-resolution

[*] Corresponding author.
E-mail address: Xiaohui.Zhu@xjtlu.edu.cn (X. Zhu).

https://doi.org/10.1016/j.jag.2023.103496
Received 12 July 2023; Received in revised form 21 August 2023; Accepted 17 September 2023

input, ensuring the retention of authentic terrain characteristics and precision.

In the field of computer vision, SR can be achieved through a variety of techniques. These include simple interpolation methods like nearest neighbour, bilinear, and bicubic interpolation (Park et al., 2003), as well as more advanced methodologies such as Convolutional Neural Networks (CNNs) (Yang et al., 2019). Generative Adversarial Networks (GANs), in particular, have proven to be highly effective in various image SR tasks, leading to state-of-the-art results (Fu et al., 2020; Singla et al., 2022). Their effectiveness in image super-resolution is attributed to their ability to discern complex relationships between low-resolution and high-resolution images. This unique ability makes GANs a powerful tool for enhancing image resolution and detail.

Research into SR for DEMs often confronts a prevalent challenge in deep learning: many models are difficult to interpret and analyse quantitatively. While we can confirm a model's functionality, understanding its operation and how such insights could drive improvements remains complex. Most techniques based on CNNs prioritize complex architectural designs and incorporate residual blocks (Zagoruyko and Komodakis, 2016) to enhance performance and increase the model's capacity. Despite the significant progress made by CNNs compared to traditional model-based techniques, two fundamental issues often arise from the convolutional layers: firstly, the interaction between images and convolutional kernels, being content-independent, can lead to suboptimal recovery of various image regions with a single kernel (Wan et al., 2021; Schuler et al., 2015). Secondly, CNNs tend to emphasize local features, possibly reducing their effectiveness in modelling long-term dependencies and global features (Hu et al., 2019).

In response to these challenges, Transformers have emerged as potential alternatives to CNNs. Their self-attention mechanism captures global interactions between contexts, demonstrating robust performance across various visual tasks (Han et al., 2022; Khan et al., 2022; Liu et al., 2021; Liang et al., 2021). When using Vision Transformers (ViT), the input image is divided into patches of a predefined size, with each segment processed individually. Despite the superior metric performance of some Transformer-based SR models at this stage, the limited range of information can render their outcomes inferior to those generated by CNNs in certain scenarios. These phenomena indicate that while Transformers excel in modelling local information from DEMs, their capacity to leverage global elevation information for super-resolution remains to be augmented.

In the context of DEM super-resolution, edge recovery becomes even more critical. As the resolution increases, the inaccuracies or distortions at the edges can become more pronounced if not addressed adequately (Da Wang et al., 2019). Successful edge recovery ensures that these boundaries are seamless and maintain the integrity of the elevation data. In the task of DEM SR using machine learning, many SR methods are based on CNNs. Due to the convolutional operations, cells at the edges of the raster may not consider sufficient contextual information during processing, potentially leading to inaccurate edge recovery (Ma et al., 2020). Additionally, the loss functions used in training these methods are not specifically optimized for edge recovery, resulting in models that might perform well in other areas but fall short at the edges (Ma et al., 2022). Moreover, edge recovery is paramount in the domain of DEM mosaicking. As multiple DEMs are combined to form a larger coverage area, any discrepancy in the edge values can lead to visible seams or discontinuities. This affects the visual representation and can also introduce errors in analyses that use the DEM. For instance, an inaccurate boundary can divert the flow path in watershed delineation, leading to incorrect delineation of catchment areas. Similarly, inaccuracies at the edges can lead to false interpretations of slope gradients and potential instability regions in slope stability analysis.

Given the importance of edge recovery in DEM super-resolution, there is a pressing need for more effective methods that can accurately restore edge elevation values. This study presents a DEM super-resolution Transformer network which applies a multi-head attention mechanism using a shifted window approach. We re-design the window attention module, enabling the activation of more elevation information within the DEMs and facilitating more effective learning of geographical features by the network. To alleviate the computational burden associated with the attention mechanism, we turn to the research to optimize and enhance the network architecture specifically tailored for the DEM data structure (Zhang et al., 2022b). As depicted in Fig. 1, the network's architecture consists of four main components: shallow feature extraction, deep feature extraction, GeoTransform, and High-Resolution (HR) generation. To assess the effectiveness of the proposed model, we have chosen bicubic interpolation, SRGAN, ESRGAN, and Tfasr methodologies as experimental benchmarks for comparison.

2. Related work

2.1. Image super-resolution

2.1.1. Interpolation-based image SR

Interpolation-based Image SR also referred to as image scaling, employs each known data point for interpolation, calculating the pixel value to be interpolated (Li et al., 2020). Frequently used interpolation algorithms include nearest neighbour, bilinear, and bicubic interpolation (Han, 2013). These methods apply varying interpolation techniques to fill in pending pixel blocks, exhibiting commendable real-time performance (Su et al., 2011).

However, these methods operate under the assumption of continuity in the image's grey value, leading to inadequate preservation of high-frequency information. Consequently, interpolation-based Image SR demonstrates limited adaptability, particularly when handling edges and textures. This often results in the generation of blurred high-resolution images (Han, 2013).

2.1.2. Deep learning based image SR

SRCNN employs a three-layer neural network to extract feature information from images, achieving nonlinear mapping through activation functions, and ultimately reconstructing images (Dong et al., 2016). Despite SRCNN potentially delivering superior super-resolution results, it is burdened with high computational demands and lacks the capacity to extract global image features (Wang et al., 2020). Addressing the issue of underutilized hierarchical features of Low-Resolution (LR) images during the CNN reconstruction process, Zhang et al. proposed a Residual Dense Network (RDN) (Zhang et al., 2018). This network combines multiple residual dense blocks in series, offering a more nuanced solution. Further advancing the field, Niu et al. introduced a Holistic Attention Network (HAN) to capture the interdependencies across multi-scale layers (Niu et al., 2020). This network adaptively learns the global correlation across different depths and channels, demonstrating the evolution of super-resolution techniques.

Building upon the foundation of CNNs, ResNet inherits the deep feature extraction capability of CNNs. Lim et al. (2017) utilized a residual deep network to learn high-frequency information, achieving superior results compared to SRCNN. To address the challenge of increasing network depth and complexity without improving performance, the Multi-Scale Residual Network (MSRN) was introduced (Li et al., 2018a). This method enhances the residual block by incorporating multi-scale convolution kernels, facilitating adaptive detection of image features across different scales.

In scenarios where a significant scaling factor is involved, the reconstructed Super Resolution (SR) image often lacks texture details. Addressing this issue, GANs have demonstrated remarkable generative capabilities. SRGAN, the first attempt to employ GANs for super-resolution of real-world photos, capitalizes on the mutual antagonism between the generator and discriminator to learn high-frequency texture details of the image (Ledig et al., 2017). Building on the foundation of SRGAN, ESRGAN (Wang et al., 2018) has enhanced capabilities to assess the overall quality of an image accurately, leading to the production of textures with an elevated degree of realism.

2.2. Spatial interpolation

Spatial interpolation, a method used to estimate the values of unknown points based on known sample points, plays a crucial role in Geographic Information System (GIS) systems. Specifically, it is commonly employed to estimate the values in a raster's target area (Mitas and Mitasova, 1999). One such interpolation method is trend surface analysis, which uses a least square fitting approach to replicate the spatial distribution features of known sampling points (Agterberg, 1984). However, to ensure the accuracy of the fit, two issues must be addressed: the setting of the surface function model and the adjustment of the model.

The nearest neighbour (NN) method assigns the value of the target area based on the closest sampling point, employing a more stringent boundary processing technique (Rukundo and Cao, 2012). While this method is effective when there are ample sample points, it may yield less than optimal smoothness at the boundary.

Building on the foundation of variogram theory, the Kriging interpolation method was developed. This method considers both the spatial position relationships among each sample point and the relationships between the interpolation point and the sampling points (Oliver and Webster, 1990). However, Kriging interpolation requires significant computational resources.

Another approach is the Inverse Distance Weighted (IDW) interpolation method, initially proposed by Daly (2006). This method has since been further developed and enhanced (Lu and Wong, 2008; Rahman et al., 2010).

2.3. Deep learning DEM SR

The integration of deep learning into image super-resolution has emerged as a popular research area. However, there are only a few comprehensive studies that examine the feasibility of applying these deep learning techniques specifically to DEM SR. To enhance network efficiency, D-SRGAN, as proposed by Demiray et al. (2021), used SRGAN as the basis of the convolutional neural network, employing a ReLU activation function, and a residual model. This study demonstrated that D-SRGAN outperformed two other CNN-based DEM SR models, namely D-SRCNN (Chen et al., 2016) and DPGN (Xu et al., 2019). However, D-SRGAN exhibited over-smoothing distortion in non-flat terrain due to mean square error (MSE) loss.

In a comparative study by Zhang and Yu (2022), four DEM SR methods, including SRGAN (Ledig et al., 2017), ESRGAN (Wang et al., 2018), and CDEGAN (Zhu et al., 2020), were assessed. The study yielded surprising results: (1) SRGAN outperformed other SR methods across several metrics; (2) although ESRGAN was able to learn a large number of geographical features, many of these features were distorted when tested; (3) while CDEGAN analysed several indicators of geographical variance, the comparison showed that CDEGAN introduced some visual noise signals.

To address these limitations, Lin et al. (2022) proposed a method that combines internal and external learning approaches. This method learns detailed features from the inside out and assigns the learned detailed features to the global feature weight. This approach yielded superior results in mountainous regions with abundant texture and elevation features and improved DEM recovery in flat terrain areas.

To enable the model to understand terrain features, Zhang et al. proposed the Tfasr model (Zhang et al., 2022a). To compensate for terrain feature loss, the model's adaptive terrain feature extraction module uses the Deformable Convnets v2 module (Zhu et al., 2019), the U-Net (Ronneberger et al., 2015) segmentation network, and slope loss as one of the loss functions.

Most of the previous research focused solely on local correlations in DEMs. To address this, Han et al. (2023) proposed a DEM super-resolution model that incorporated global information, demonstrating superior performance across multiple terrains.

3. Methods

The network structure of DSRT is depicted in Fig. 1. The shallow feature extraction module utilizes convolutional layers to preserve low-frequency data. The deep feature extraction module primarily comprises Global Attention Blocks (GABs). Each GAB employs multiple modified lightweight Global Attention Swin Transformer Layers (GA-STLs) to facilitate local attention and cross-window interaction, in alignment with the DEM super-resolution objective. Additionally, we introduce a GeoTransform module capable of natively generating DEM in GeoTiff format. The final block of our network is the high-resolution DEM generation module.

3.1. Shallow feature extraction

The shallow feature extraction performs simple convolution on the input LR DEM to extract lower-level features (edges and textures) (Zhang et al., 2016; Hu et al., 2015). It acts as an initial filtering phase, where simple and easily extractable features such as edges, textures, and basic shapes are gathered from the input data.

In the context of the DSRT model, the shallow feature extraction module uses a convolutional layer to extract these basic features from the input LR DEM. The extracted features are then passed onto the deep feature extraction module for further processing. For the input LR DEM $DEM_{LR} \in \mathbb{R}^{H \times W \times C_{DEM}}$, H and W are the height and width of the input DEM, and C_{DEM} represents the input channel number (normally $C_{DEM} = 1$). A 3 × 3 convolutional layer $H_{SF}(\cdot)$ is used to extract the shallow feature $F_{DEM_{LR}} \in \mathbb{R}^{H \times W \times C}$ of the input LR DEM_{LR}:

$$F_{DEM_{LR}} = H_{SF}(DEM_{LR}) \tag{1}$$

3.2. Deep feature extraction

Deep feature extraction aims to capture high-level semantic information within the DEM. In the context of DEM super-resolution, this could include detecting and representing specific terrain features, patterns, and other topographical information. The output of shallow features $F_{DEM_{LR}}$ can then be used as input for the deeper layers $H_{DF}(\cdot)$ of the network, which performs more complex operations to extract high-level features $F_{DF} \in \mathbb{R}^{H \times W \times C}$:

$$F_{DF} = H_{DF}(F_{DEM_{LR}}) \tag{2}$$

where $H_{DF}(\cdot)$ represents the deeper feature extraction module of the network, and it has n Global Attention Blocks (GABs) (see Fig. 1). The features $F_{GAB_1}, F_{GAB_2}, \cdots, F_{GAB_n}$ can be expressed by the following equation:

$$F_{GAB_i} = H_{GAB_i}(F_{GAB_{i-1}}) \quad i = 1, 2, \ldots, n \tag{3}$$

where H_{GAB_i} denotes the i_{th} Global Attention Block. At the end of the GABs, we use an additional convolutional layer to achieve better performance for aggregating deep features and shallow features (Liang et al., 2021).

3.2.1. Global Attention Block (GAB)

Fig. 2 shows that GAB contains multiple Global Attention Swin Transformer Layers (GA-STLs) and a convolutional layer. The purpose of GAB is to extract the deep feature of the input DEM:

$$F_{DEM_{i,j}} = H_{GA\ \ STL_{i,j}}(F_{DEM_{i,j-1}}), \quad j = 1, 2, \ldots, N \tag{4}$$

where $H_{GA\ \ STL_{i,j}}(\cdot)$ represents the j_{th} GA-STL of the i_{th} GAB.

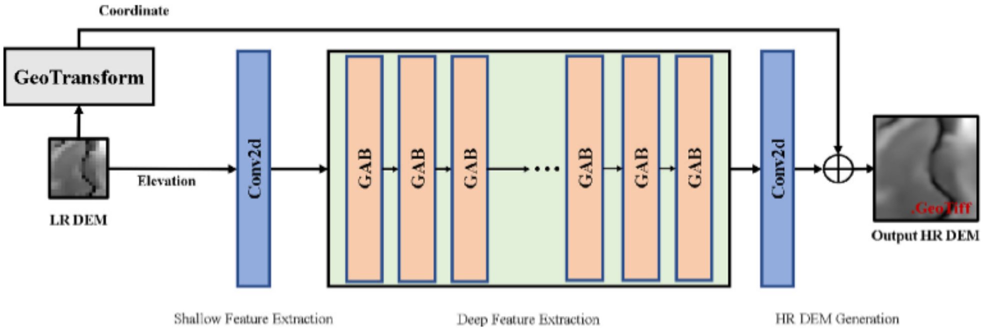


Fig. 1. The overall structure of DSRT comprises four modules.

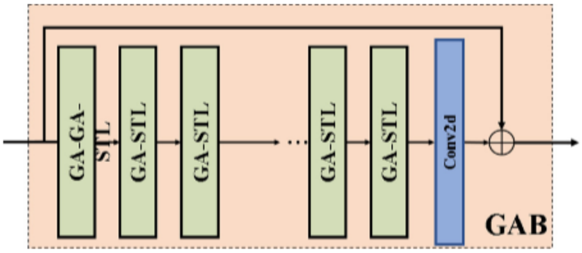


Fig. 2. Overall structure of the Global Attention Block (GAB).

3.2.2. Global Attention Swin Transformer Layer (GA-STL)

Fig. 3(a) shows the overall structure of the modified lightweight Global Attention Swin Transformer Layers. Instead of using layer normalization (LN) (Ba et al., 2016) in the GA-STL, we use batch normalization (BN) (Santurkar et al., 2018) to reduce the tremendous amount of computation brought by the self-attention (SA) (Shaw et al., 2018) and Multi-Head Self-Attention (MSA) (Voita et al., 2019).

In G-MSA (see Fig. 3(b)), the input sequence passes through three distinct linear layers to obtain Query (Q), Key (K), and Value (V) matrices. The weights of these linear layers are learnable parameters. Next, the dot product of Q and K is computed, yielding an attention score matrix. This process describes the degree of association between each element and other elements within the sequence. To maintain computational stability, the attention score matrix is divided by a scaling factor $\sqrt{d}$ and adds a relative position bias $B \in \mathbb{R}^{M^2 \times M^2}$, where M represents the window size, and M^2 represents the number of patches in a window. Subsequently, a $SoftMax$ operation is applied to the attention score matrix, obtaining normalized attention weights. Then, the normalized attention weights are multiplied by the Value (V) matrix. This operation can be considered as a weighted average of the input sequence, emphasizing elements with a stronger association with the current position.

$$Attention(Q, V, K) = SoftMax(\frac{QK^T}{\sqrt{d}} + B)V \tag{5}$$

Lastly, a multi-layer perceptron (MLP) (Taud and Mas, 2018) is added to provide a straightforward nonlinear transformation through the ReLU (Nair and Hinton, 2010) function for activation.

3.3. Geographical coordinate projections and transformations (GeoTranform)

We downsample the resolution of the original DEM to a lower resolution DEM during the network (DSRT) training to ensures the spatial and acquisition time consistency of HR and LR DEMs. Therefore, in the GeoTransform module, two coordinate information processes are involved. As shown in Fig. 4, the first step is to resample the original HR DEM $DEM_{real} \in \mathbb{R}^{H \times W}$ into LR DEM $DEM_{LR} \in \mathbb{R}^{\frac{H}{scale1} \times \frac{W}{scale1}}$, and the second step is to integrate the processed coordinate and projection into the HR DEM $DEM_{HR} \in \mathbb{R}^{H\frac{scale2}{scale1} \times W\frac{scale2}{scale1}}$ generated by the model. Where $scale1$ is the downsampling factor of the real HR DEM and $scale2$ is the upsampling factor of the LR DEM (usually $scale1 = scale2$).

Fig. 5 shows the diagram of transforming and integrating the high-resolution coordinates into the HR DEM. The main idea is to keep the coordinate range of the DEM unchanged and change the cell size of the raster to increase the number of raster cells. Given an LR DEM $DEM_{LR} \in \mathbb{R}^{\frac{H}{scale} \times \frac{W}{scale}}$, the first step is to obtain the elevation values $Elevation_{LR} \in \mathbb{R}^{\frac{H}{scale} \times \frac{W}{scale}}$ and projected coordinate information of each cell. The next step is to calculate the number of rows and columns and the cell size of the high-resolution DEM. Among them, $Columns \times scale$ and $Rows \times scale$ are the height and width of the target HR DEM, $CellSize \times \frac{1}{scale}$ is the cell size and coordinate projection of the target HR DEM. The high-resolution elevation values $Elevation_{HR} \in \mathbb{R}^{H \times W}$ generated by DSRT and HR projected coordinates $GeoTransform_{HR} \in \mathbb{R}^{H \times W}$ are integrated into the final HR DEM $DEM_{HR} \in \mathbb{R}^{H \times W}$ in the last step.

3.4. High-resolution DEM generation

The final process of the network is the generation of the HR DEM with geographical coordinate projections. The high-resolution DEM DEM_{HR} is generated by aggregating shallow and deep features:

$$DEM_{HR} = H_{HR}(F_{DEM_{LR}} + F_{DF}) \tag{6}$$

In Eq. (6), $H_{HR}(\cdot)$ denotes the HR DEM generator of the network.

3.5. Loss function

3.5.1. L_1 loss

Most SR methods used Mean Square Error (MSE) as the loss function (Ledig et al., 2017; Li et al., 2018b; Nagaraj et al., 2020; Daihong et al., 2022). Although direct minimization of MSE loss can get good SR results, avoiding blurring details is difficult (Lim et al., 2017; Anagun et al., 2019; He and Cheng, 2022). Therefore, we use L_1 loss to calculate the error of the corresponding pixel position of SR and HR. The equation is as follows:

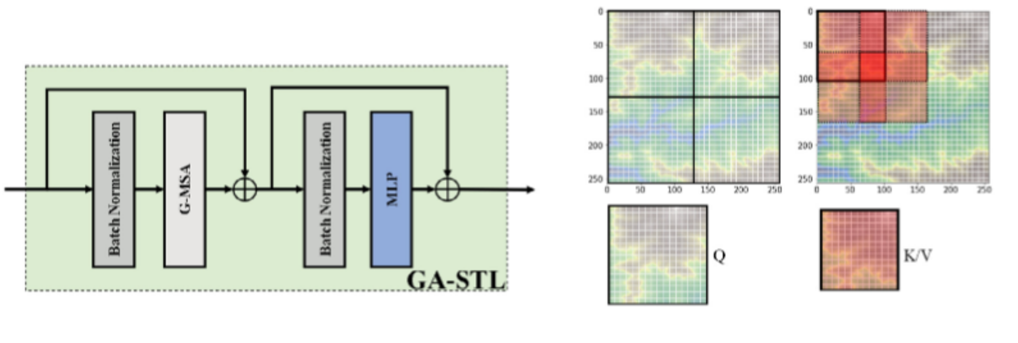
$$L_1(DEM_{real}, DEM_{fake}) = \sum_{i=0}^{m} \left| DEM_{real}^{(i)} - DEM_{fake}^{(i)} \right| \tag{7}$$

3.5.2. Root mean square error

Root mean square error (RMSE) measures the error size of the deviation between the predicted and actual values. RMSE is very sensitive to the error, so it can well reflect the precision of the generated high-resolution DEM. The equation is as follows:

$$RMSE(DEM_{real}^{(i)}, DEM_{fake}^{(i)}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (DEM_{real}^{(i)} - DEM_{fake}^{(i)})^2} \tag{8}$$

where $DEM_{real}^{(i)}$ denotes the elevation value of the real HR DEM and $DEM_{fake}^{(i)}$ means the elevation value generated by the network.



(a) GA-STL

(b) Overlapped Attention

Fig. 3. (a) Overall structure of the Global Attention Swin Transformer Layer (GA-STL), and (b) Overlapped attention.

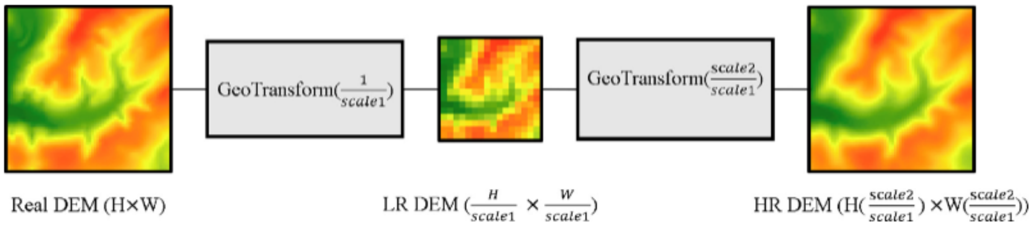

GeoTransform($\frac{1}{scale1}$)
GeoTransform($\frac{scale2}{scale1}$)
Real DEM (H×W)
LR DEM ($\frac{H}{scale1}$ × $\frac{W}{scale1}$)
HR DEM (H($\frac{scale2}{scale1}$) ×W($\frac{scale2}{scale1}$))

Fig. 4. Overall process of the geographical coordinate projections and transformations (GeoTransform) module during the network training.

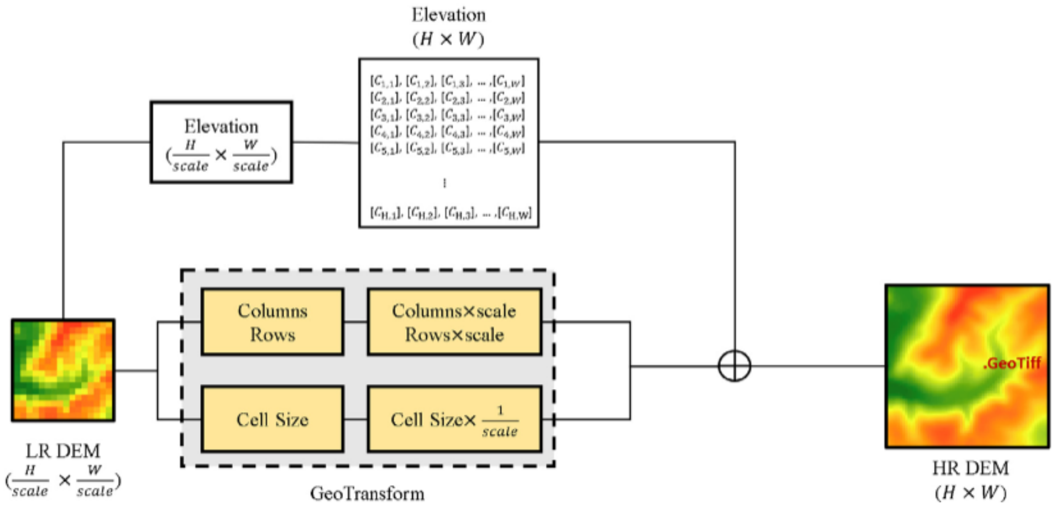

Elevation
($H \times W$)
[$C_{1,1}$], [$C_{1,2}$], [$C_{1,3}$], ... ,[$C_{1,W}$]
[$C_{2,1}$], [$C_{2,2}$], [$C_{2,3}$], ... ,[$C_{2,W}$]
[$C_{3,1}$], [$C_{3,2}$], [$C_{3,3}$], ... ,[$C_{3,W}$]
[$C_{4,1}$], [$C_{4,2}$], [$C_{4,3}$], ... ,[$C_{4,W}$]
[$C_{5,1}$], [$C_{5,2}$], [$C_{5,3}$], ... ,[$C_{5,W}$]
[$C_{H,1}$], [$C_{H,2}$], [$C_{H,3}$], ... ,[$C_{H,W}$]
Elevation
($\frac{H}{scale} \times \frac{W}{scale}$)
Columns
Rows
Columns×scale
Rows×scale
Cell Size
Cell Size× $\frac{1}{scale}$
.GeoTiff
LR DEM
($\frac{H}{scale} \times \frac{W}{scale}$)
GeoTransform
HR DEM
($H \times W$)

Fig. 5. Diagram of using GeoTransform module to integrate coordinate into HR DEM.

3.5.3. DEM feature loss

To better preserve the high-resolution details of the DEM generated by the network, we refer to the idea of perceptual loss (Johnson et al., 2016; Rad et al., 2019; Jo et al., 2020) and VGG-16 feature extraction (Zhang et al., 2019; Ge et al., 2018; Tun et al., 2021) in remote sensing. The equation is as follows:

$$\ell^{\emptyset,j}_{feat}(DEM_{real}, DEM_{fake}) = \frac{1}{C_j H_j W_j} \ || \emptyset_j(DEM_{real}) - \emptyset_j(DEM_{fake}) ||^2 \tag{9}$$

where j represents the j_{th} layer of VGG-16, $\emptyset_j(DEM_{real})$ represents the output of DEM_{real} in VGG-16 intermediate layer j, $\emptyset_j(DEM_{fake})$ represents the output of DEM_{fake} in VGG-16 intermediate layer j, and $C_j H_j W_j$ represent the number of channels, height, and width in VGG-16 middle layer j respectively.

3.5.4. DEM edge loss

This is a specialized loss function that directly targets the edges of the DEMs. It works by first detecting the edges in both the ground truth and the generated DEM, and then computing the difference between the two. E_{pred} and E_{gt} are the edges of the predicted and ground truth images, respectively, the edge loss L_{edge} could be defined as:

$$L_{\text{edge}} = \frac{1}{N} \sum_{i=1}^{N} \left(E_{\text{pred}} - E_{\text{gt}} \right)^2 \tag{10}$$

3.5.5. Global gradient loss

The global gradient loss can play a crucial role in DEM tasks by encouraging the model to generate DEMs that accurately capture the terrain's structure and characteristics, particularly at the edges:

$$L_{\text{grad}} = \frac{1}{N} \sum_{i=1}^{N} \left(\left(\frac{\partial I_{\text{pred}}}{\partial x} - \frac{\partial I_{\text{gt}}}{\partial x} \right)^2 + \left(\frac{\partial I_{\text{pred}}}{\partial y} - \frac{\partial I_{\text{gt}}}{\partial y} \right)^2 \right) \tag{11}$$

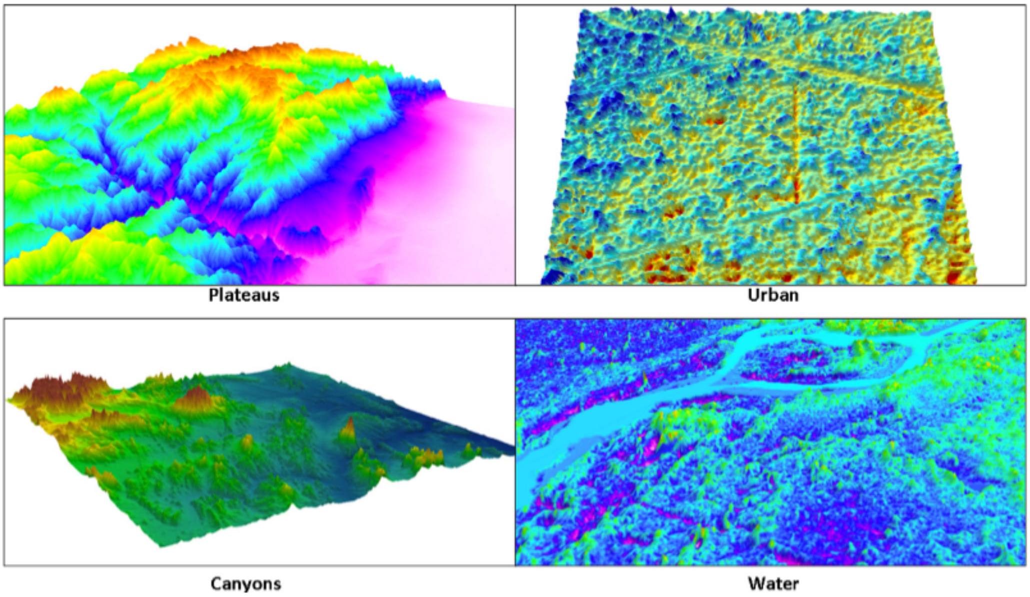


Fig. 6. Four examples of training data from different terrains. (Plateaus: ASTGTMV003_N27E087_dem, Canyons: ASTGTMV003_N27E085_dem, Urban: ASTGTMV003_N26E082_dem, Water: ASTGTMV003_N25E080_dem).

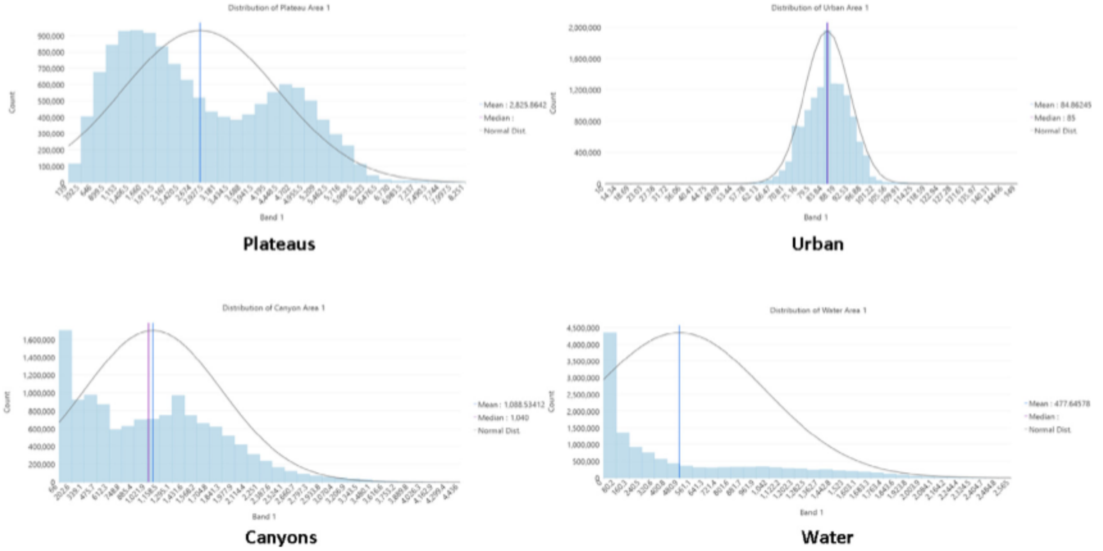


Fig. 7. Four examples of elevation distribution from different terrains.

where I_{pred} and I_{gt} are the predicted and ground truth DEMs, respectively, N is the total number of cells in the DEMs, and the derivatives are computed with respect to the x and y directions.

3.5.6. Collaborative loss

The loss function of the network is a collaboration of the L_1 loss, RMSE, and DEM feature loss, and the equation is as follows:

$$loss = L_1 loss + \alpha \cdot RMSE + \beta \cdot \ell_{feat} + \theta \cdot L_{grad} + \gamma \cdot L_{edge} \tag{12}$$

where α, β, θ and γ are the weight coefficients of the collaborative loss.

4. Experiments

4.1. DEM dataset

The DEM data utilized in this study is sourced from the ASTER GDEM v3 in 16-bit GeoTiff format. The elevation datum is referenced to EGM96, and the reference ellipsoid is WGS84. The DEM boasts a vertical resolution of 7–10 m and a horizontal resolution of 1 arc-second (approximately 30 m). We select 80 experimental regions that encompass various topographical features, with elevations ranging from 0.5 to 8260 m.

For this study, we have curated a diverse set of DEMs representing various terrains, including but not limited to mountains, hills, urban areas, plains, and valleys (see Table 1). Fig. 6 shows an example of each terrain, and Fig. 7 shows the elevation distribution. To conduct a comprehensive analysis, these DEMs were batch cropped into multiple resolutions of 64, 256, and 512 using the ArcPy module in ArcGIS. This approach was adopted with the understanding that DEMs of higher resolution can encapsulate a broader range of elevation differences and retain a greater number of geographical features.

It is noteworthy that generating low-resolution DEMs of the precise site poses a significant challenge, as highlighted in previous research. Consequently, we have also incorporated a downsampling processing technique. This specific approach involves the downsampling of 25 056

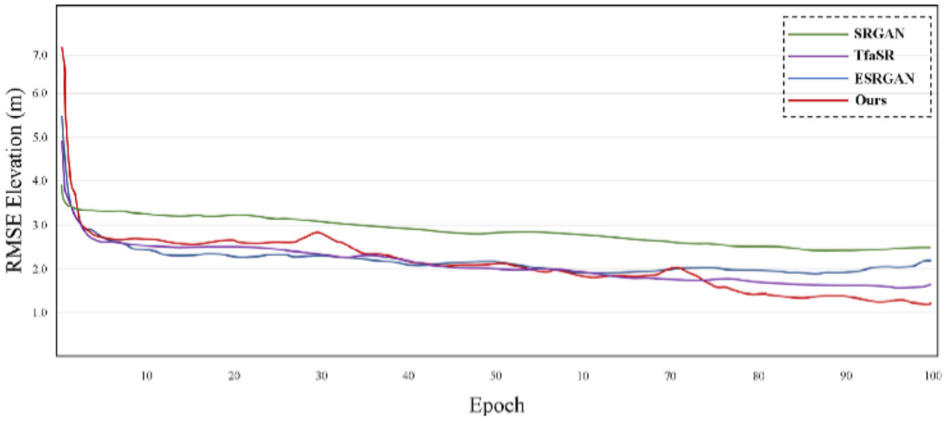


Fig. 8. Elevation loss (RMSE) of four models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Training datasets of different terrains.

Resolution	Elevation difference [m]				
	Mountains	Hills	Valleys	Urban Areas	Plains
64 × 64	100	70	50	30	10
256 × 256	1500	800	400	100	30
512 × 512	3000	1200	500	150	40

high-resolution DEMs four times (down_sampling = 4) using the ArcPy package from ArcGIS.

4.2. Training

Fig. 8 shows the RMSE after recovery of elevation values generated by different methods throughout the training process. First, SRGAN has the lowest RMSE at the beginning of training. However, the feature extraction and perceptual loss used by SRGAN are based on real-world SR tasks, so there is not much downward trend in the RMSE of SRGAN. ESRGAN has a more apparent downward trend, and the final RMSE is better than SRGAN. Tfasr has a very stable performance throughout the training cycle, and the RMSE decreases steadily. This is because: (1) Tfasr used U-Net to pre-train the basin in DEM; and (2) Tfasr used deformable convolution, and this adaptive terrain feature extraction module has certain effectiveness. The model we proposed (DSRT) had the highest RMSE value at the beginning of training (epoch = 0), but as the training progressed, the RMSE gradually dropped to the minimum value among the four models. The RMSE curve has two fluctuations, one when the epoch is 30 and one at 70. These two fluctuations are related to the dynamic learning rate we set. Because (1) we set the learning rate to decrease during $\{35, 70\}th$ epoch, and (2) to reduce the loss as soon as possible in the early stage, we set a larger learning rate within the insurance range.

Fig. 9 show the HR DEMs generated by our model at different epochs. It can be seen that the degree of terrain recovery of the DEM gradually improves throughout the training process. At epoch 5, the accuracy of the DEM generated by our model is not high, and the restoration of terrain features has not been well repaired. At epoch 30, the model we trained has gradually recovered many simple landform features while the elevation error has decreased. Our model has generated terrain features similar to real DEM when trained to epoch 80. Although the RMSE of the model further decreases at epoch100, the effect is minimal.

To demonstrate that our GeoTransform module works correctly, we put the DEM generated by DSRT into ArcGIS (see Fig. 10(a)). The results show that the GeoTransform module can handle coordinates and projections well, and the coordinates and projections of the generated DEMs and the original data coincide perfectly. We can view each cell's elevation value and use ArcGis's 3D Analyst Tool for surface analysis (see Fig. 10(b)).

5. Results

5.1. RMSE assessment

Table 2 provides a comprehensive evaluation of the performance of five different methods – DSRT, Tfasr, SRGAN, ESRGAN, and Bicubic – in generating DEMs of varying resolutions: 64 × 64,256 × 256, and 512 × 512. The performance is assessed based on the Root Mean Square Error (RMSE) of four key parameters: Elevation, Slope, Aspect, and Curvature.

In the 64 × 64 resolution DEMs, the DSRT method exhibits superior performance, achieving the lowest RMSE values across all parameters. This suggests that the DSRT method is most effective at preserving the accuracy of Elevation, Slope, Aspect, and Curvature in the generated DEMs at this resolution. On the other hand, the Bicubic method shows the highest RMSE values for Elevation and Curvature, indicating potential limitations in its ability to accurately represent these parameters.

When the resolution is increased to 256 × 256, the DSRT method maintains its leading performance, again achieving the lowest RMSE values across all parameters. This consistency across different resolutions underscores the robustness of the DSRT method. The ESRGAN method, however, shows the highest RMSE values for Elevation, Slope, and Curvature, suggesting that it may struggle to accurately represent these parameters at higher resolutions. The Bicubic method has the second lowest RMSE for Elevation and Slope, indicating a specific area of strength.

At the highest resolution of 512 × 512, the DSRT method continues to outperform the other methods in terms of Elevation, Slope, and Curvature, reinforcing its effectiveness across different resolutions. Interestingly, the Bicubic method, despite its weaker performance at lower resolutions, achieves the second-lowest RMSE at this resolution. This suggests that the Bicubic method has specific strengths in representing DEMs at higher resolutions. The ESRGAN method, however, exhibits the highest RMSE values across all parameters, indicating a general struggle with accuracy at this resolution.

In conclusion, these results highlight the consistent superiority and robustness of our DSRT method across different resolutions and parameters. In contrast, the other methods demonstrate varying degrees of error across different parameters and resolutions. For instance, the Bicubic method shows the highest RMSE values for Elevation and Curvature in the 64 × 64 resolution DEMs, suggesting potential limitations in accurately representing these parameters. However, it is worth noting that the Bicubic method can achieve good performance when provided with sufficient reference points. This may explain why it achieves the second-lowest results in the 512 × 512 resolution DEMs.

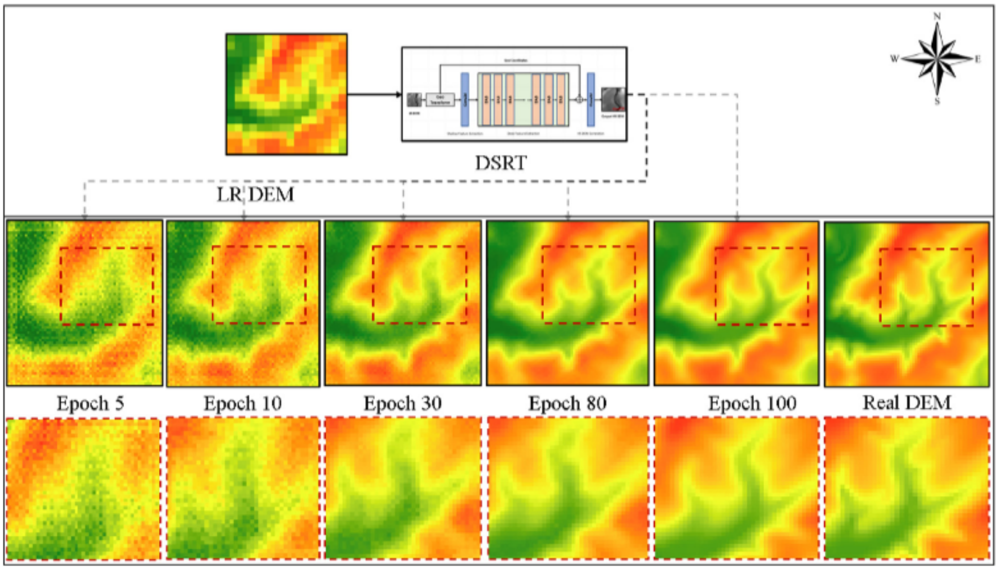


Fig. 9. HR DEMs generated by DSRT at different epochs (red dotted frames are the detail of the DEMs). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

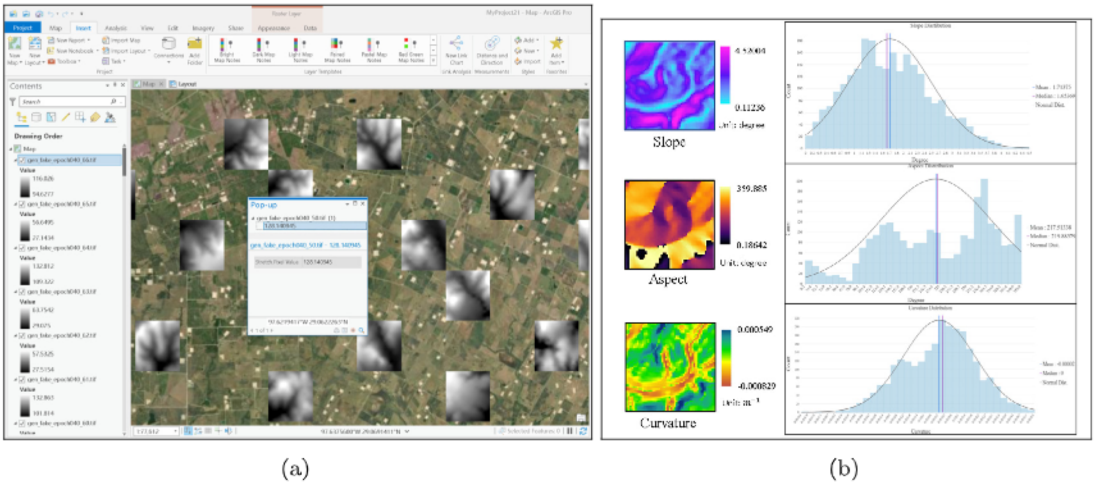


Fig. 10. (a) HR DEMs generated by DSRT in ArcGIS, a pop-up window shows the elevation values of a selected DEM. (b) HR DEM generated by DSRT at epoch 100 (slope, aspect, and curvature rasters processed by ArcGIS, distributions on the right).

Table 2
RMSE assessment results on different resolution DEMs.

$DEM_{64 \times 64}$	RMSE-elevation (m)	RMSE-slope (°)	RMSE-aspect (°)	RMSE-curvature
DSRT	**1.286**	**1.958**	**79.552**	**0.255**
Tfasr	1.493	2.321	82.673	0.356
SRGAN	2.785	3.544	84.987	0.417
ESRGAN	3.146	3.621	84.247	0.474
Bicubic	4.588	3.965	86.433	0.598
$DEM_{256 \times 256}$	RMSE-elevation (m)	RMSE-slope (°)	RMSE-aspect (°)	RMSE-curvature
DSRT	**3.493**	**4.506**	**84.790**	**0.675**
Tfasr	6.254	5.023	86.673	0.729
SRGAN	7.987	6.449	87.770	1.446
ESRGAN	13.223	9.632	89.921	1.547
Bicubic	5.765	4.976	87.425	0.998
$DEM_{512 \times 512}$	RMSE-elevation (m)	RMSE-slope (°)	RMSE-aspect (°)	RMSE-curvature
DSRT	**6.286**	**7.892**	**87.466**	**1.554**
Tfasr	9.909	8.943	90.002	3.566
SRGAN	12.864	11.991	90.945	6.617
ESRGAN	16.677	17.621	97.401	9.474
Bicubic	7.546	7.965	87.839	1.598

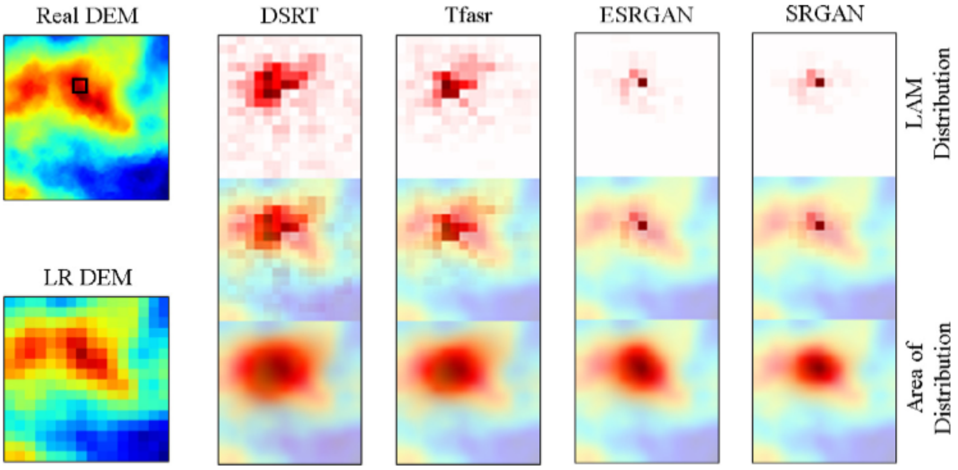


Fig. 11. LAM result of four models on 64 × 64 input DEM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

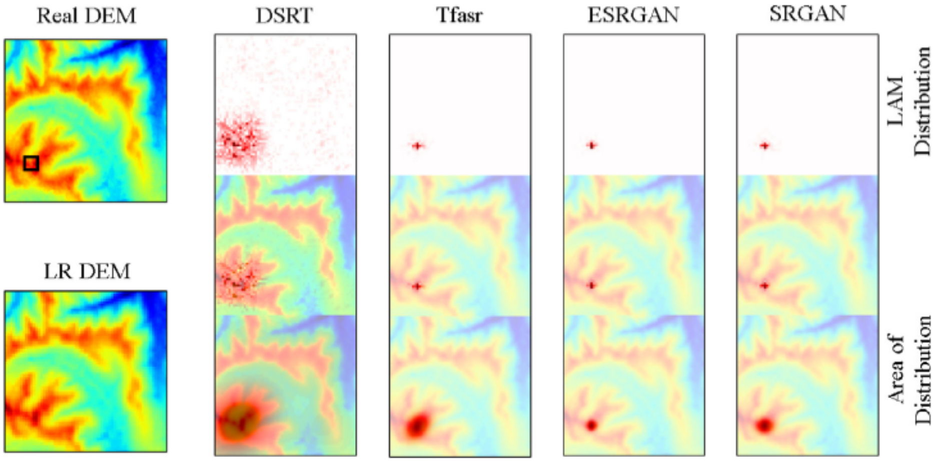


Fig. 12. LAM result of four models on 256 × 256 input DEM.

5.2. LAM results

The Local Attribution Maps (LAM) (Gu and Dong, 2021) method we adopt is based on the Integrated Gradient approach, typically used in attribution analysis within classification problems (Zeiler et al., 2011; Simonyan et al., 2013; Zeiler and Fergus, 2014; Sundararajan et al., 2017). We utilize this method to quantify the extent to which our model has learned the features of surrounding areas when performing super-resolution tasks on specific elevation regions in DEMs. Additionally, as the main challenge in DEM Super Resolution lies in the reconstruction of high-frequency textures, such as ridgelines and river valleys, flat areas can often be restored simply through interpolation. Therefore, we focus on attributing only the complex patches.

Fig. 11 presents the LAM test results for DSRT, Tfasr, ESRGAN, and SRGAN. In the LAM Distribution, red pixels signify the model's capacity to incorporate and compute adjacent elevation values when executing super-resolution tasks at specified elevation points. A greater density of such pixels corresponds to a superior capability of the model to apprehend global characteristics. As evidenced by the LAM Distribution, DSRT learns and utilizes global information from the image when performing SR tasks on the elevation regions highlighted in the black boxes. While activating adjacent elevation values, our model also considers the global feature information of the LR DEM, demonstrating its strong capability to learn global terrain features effectively.

To provide more robust evidence that DSRT can engage a broader set of elevation values in DEM Super Resolution, we utilized DEMs with a resolution of 265 × 256 in our LAM tests (see Fig. 12). Consistently, our model demonstrated a substantial lead in metrics. Upon testing the regions delineated by the black boxes, it was observed that DSRT activated a more extensive array of elevation points and spanned a larger area for SR.

5.3. Visual evaluation

We selected a 64 × 64 DEM to demonstrate the visual evaluation of geographic features (see Fig. 13). Although all models can generate high-resolution DEM similar to real DEM, there are still many differences in the details of each generated DEM. From the perspective of visual evaluation, our method can generate the features most similar to real DEM, and the maximum and minimum elevation values are also the closest to the original DEM. DEM generated by Tfasr is over-smoothed, although some basic features are well preserved. Compared with the real DEM, the DEM generated by Tfasr is too blurry at the lowest elevation value (green part), and a small amount of detail is lost near the highest value (red and orange part). Both ESRGAN and SRGAN can retain as many terrain details as possible, but most details are distorted during the restoration process (excessive smoothing, redundant terrain details, and loss of original terrain details) by ESRGAN. From the training curve (Fig. 8), ESRGAN achieves better results than SRGAN, but from the visual evaluation result, ESRGAN is not competent for DEM super-resolution tasks.

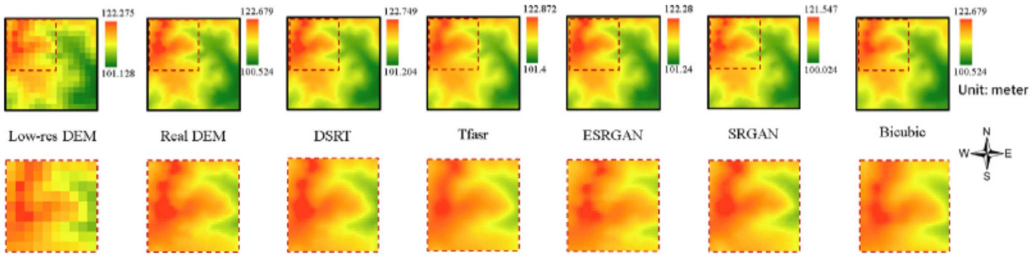


Fig. 13. HR DEM generated by five models (red dotted frames are the zoomed-in detail of the HR DEM generated for each model). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

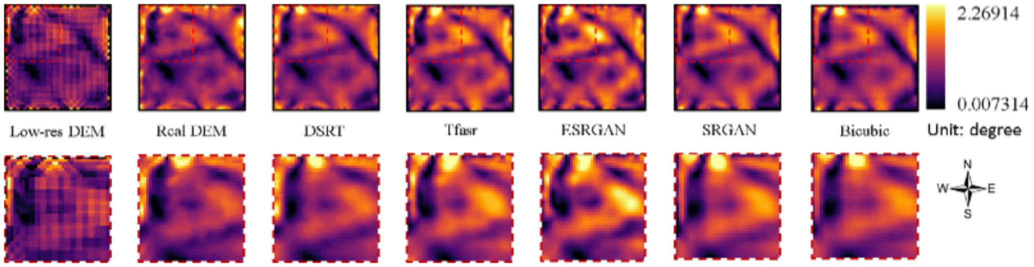


Fig. 14. Slopes generated by five models (red dotted frames are the zoomed-in detail of the slopes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

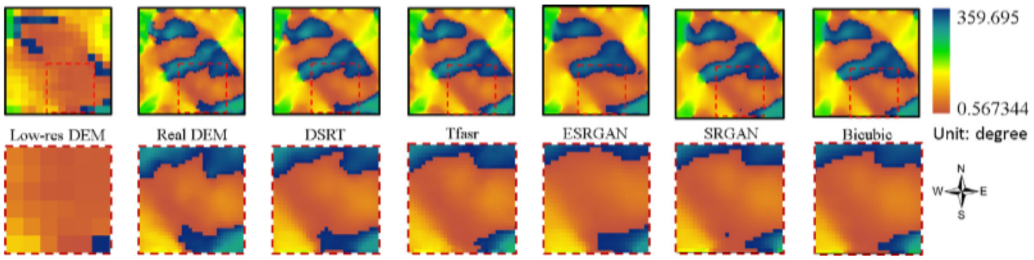


Fig. 15. Aspects generated by five models (red dotted frames are the zoomed-in detail of the aspects). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The slope measures the rate at which the height of a particular area of the ground changes. As shown in Fig. 14, the DEM generated by our model achieves the closest slope map to the original DEM, and the red dotted frames part is the enlarged detail of each generated DEM. Tfasr and ESRGAN achieve the worst results in slope analysis, and all have slope distortion in areas with large slopes (yellow area in Fig. 14). Both have varying slope distortion degrees, corresponding to the elevation analysis results (see Table 2). The HR DEM generated by SRGAN and bicubic cannot fully preserve the terrain features of areas with flatten slopes (such as the purple area in the red box).

Fig. 15 shows the aspects generated by different models. Roughly, each model can generate aspects similar to the original DEM, but the quality of each generated DEM is different in detail. A clear difference can be seen at the bottom right of the DEM (red dotted frames of each DEM). In the 360–300° area (deep blue in Fig. 15), the aspect raster generated by our model is most similar to the original DEM, with almost the same outline. SRGAN performed the worst with one error point. The aspects of the DEM generated by the three models of ESRGAN, Tfasr, and bicubic are very similar, but all have missing aspects.

Fig. 16 presents a three-dimensional visualization of a plateau region, featuring distinct topographical elements such as canyons and mountain peaks. DSRT exhibits exceptional performance in edge reconstruction, accurately delineating the sharp transitions between different topographical features. This is particularly evident in the representation of the abrupt edges of canyons and the pointed peaks of mountains, which are well-preserved in the DSRT-generated DEM. Bicubic interpolation, while not achieving the lowest RMSE, also demonstrates commendable performance. It successfully maintains the continuity of the terrain and does a particularly good job in preserving the shape of mountain peaks. In contrast, the SRGAN and ESRGAN methods show significant shortcomings in edge reconstruction. The DEMs generated by these methods appear blurred and lack the sharpness of the original DEM, especially at the edges of topographical features. This is consistent with the higher RMSE values observed for these methods in our quantitative evaluation. Tfasr also performs well, demonstrating the effectiveness of its design in handling complex terrain features.

6. Discussion

6.1. *The impact of terrain characteristics*

In order to investigate the impact of different terrains on the performance of our model, we categorize the DEMs used in our study into four types: urban areas, rivers, plateaus, and canyons. We then evaluate the RMSE for each of these categories (see Table 3). It allows us to gain a more nuanced understanding of the model's performance across diverse terrains. Each of these terrain types presents unique challenges and features that can influence the model's ability to accurately generate high-resolution DEMs. For instance, urban areas are characterized by man-made structures and flat surfaces, rivers have smooth and gradual changes in elevation but complex shapes, plateaus have high elevation differences, and canyons have steep slopes and sharp changes in elevation. By evaluating the model's performance on each of these terrain types separately, we can better understand its strengths and weaknesses, and identify areas for improvement.

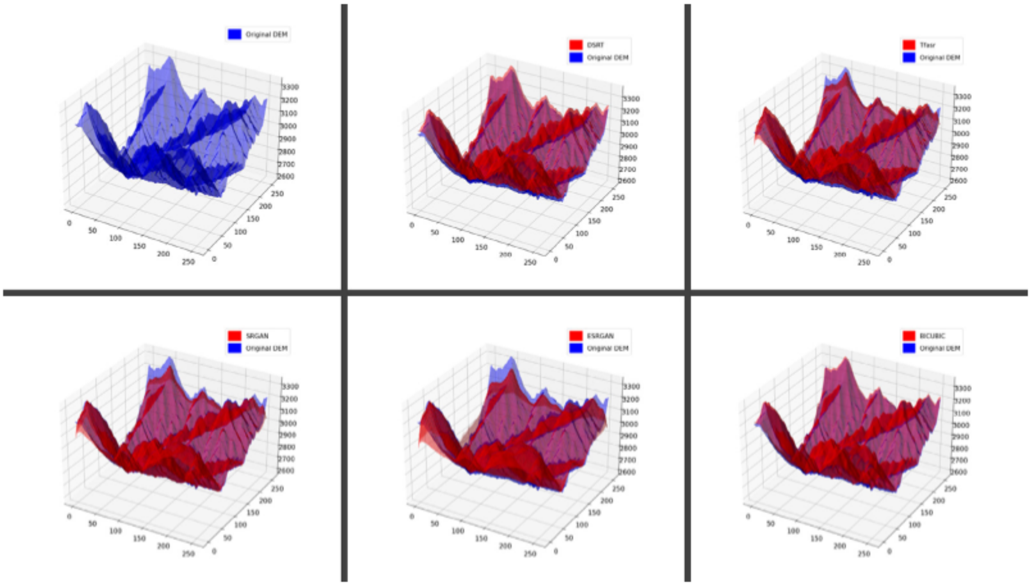


Fig. 16. 3D terrain generated by five models.

Table 3
RMSE-elevation assessment results on different terrain characteristics DEMs.

Urban regions	64 × 64	256 × 256	512 × 512
DSRT	**2.198**	**3.958**	**6.012**
Tfasr	3.569	5.327	7.298
SRGAN	4.299	6.742	8.267
ESRGAN	7.246	9.167	13.597
Bicubic	4.879	5.987	7.329
Water regions	64 × 64	256 × 256	512 × 512
DSRT	**3.497**	**4.205**	**4.969**
Tfasr	3.896	4.411	5.089
SRGAN	4.274	5.132	6.652
ESRGAN	6.009	7.930	11.562
Bicubic	3.765	4.576	5.288
Canyon regions	64 × 64	256 × 256	512 × 512
DSRT	**5.987**	**7.574**	**8.847**
Tfasr	7.058	8.982	9.724
SRGAN	8.572	11.862	14.002
ESRGAN	9.865	13.405	16.274
Bicubic	7.696	8.117	9.025
Plateau regions	64 × 64	256 × 256	512 × 512
DSRT	**7.232**	**8.301**	9.857
Tfasr	8.990	9.211	11.001
SRGAN	9.721	13.000	14.298
ESRGAN	11.521	18.274	25.932
Bicubic	7.900	9.012	**9.530**

6.1.1. Performance evaluation in urban regions

It can be seen from Table 3 that, in urban regions, DSRT consistently achieves the lowest RMSE across all resolutions, indicating that it is the most accurate method for super-resolving DEMs among the methods tested. The Tfasr and SRGAN exhibit moderate RMSE values, suggesting that while they are capable of generating reasonably accurate urban DEMs, their performance is not as robust as the DSRT method. These methods may be more susceptible to errors during the super-resolution process, leading to discrepancies between the generated and real DEMs. The Bicubic method, while not achieving the lowest RMSE, demonstrates a relatively stable performance across different resolutions.

Fig. 17 shows that DEM generated by the DSRT model stands out for its ability to accurately reproduce roads, as evidenced by its minimal RMSE. This suggests that the DSRT model is highly effective in capturing and replicating intricate details from the Real DEM, thus resulting in a high-quality output that closely mirrors the original. On the other end of the spectrum, the ESRGAN model's output is characterized by the highest error and a significant degree of blurriness. This indicates that the ESRGAN model struggles to accurately replicate the features of the Real DEM, leading to a less precise and lower quality output. The performance of the Tfasr and Bicubic models falls somewhere in between. While they do not match the accuracy of the DSRT model, they manage to produce reasonably clear and recognizable features, suggesting a moderate level of effectiveness in DEM super-resolution. However, due to the chaotic and disordered elevation values in the DEM of urban areas with a resolution of 30 m, all methods appear very blurry upon visual inspection.

6.1.2. Performance evaluation in water regions

In this case, the DSRT method consistently achieves the lowest RMSE across all resolutions, indicating superior accuracy in super-resolving DEMs of water regions. The ESRGAN method shows the highest RMSE values, particularly for the 512 × 512 resolution, indicating that it is the least accurate method among those tested.

Fig. 18 shows that the DSRT method, despite having the lowest overall RMSE and effectively restoring the outline of the river, exhibits a higher RMSE in the river area. This suggests that while our method is generally effective at super-resolving DEMs, it may struggle to accurately capture the waterbody's elevation. The Tfasr method, on the other hand, not only restores the river's outline well but also achieves the lowest error in the river area. This superior performance in the river area can be attributed to the specific design of the Tfasr network, which has been tailored to handle watershed areas. However, the Tfasr method's performance appears to falter around the periphery of the water system, indicating a potential limitation in its ability to resolve the surrounding terrain accurately.

6.1.3. Performance evaluation in canyon regions

It is evident in Table 3 that the DSRT method consistently outperforms the other methods across all resolutions, as indicated by its consistently lowest RMSE values. This suggests that the DSRT method is adept at preserving the intricate features of canyon regions during the super-resolution process, thereby producing outputs that closely mirror the original, high-resolution DEMs. The SRGAN and ESRGAN methods, on the other hand, show a trend of increasing RMSE values with increasing resolution. The Bicubic method, despite not achieving the lowest RMSE, demonstrates a level of performance that is relatively

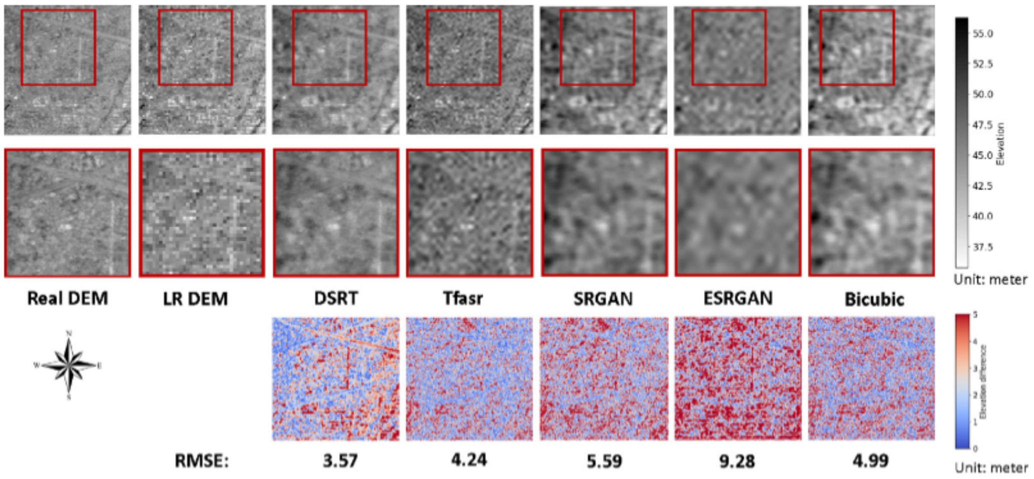


Fig. 17. 256 × 256 resolution urban area DEMs generated by different methods and their elevation differences.

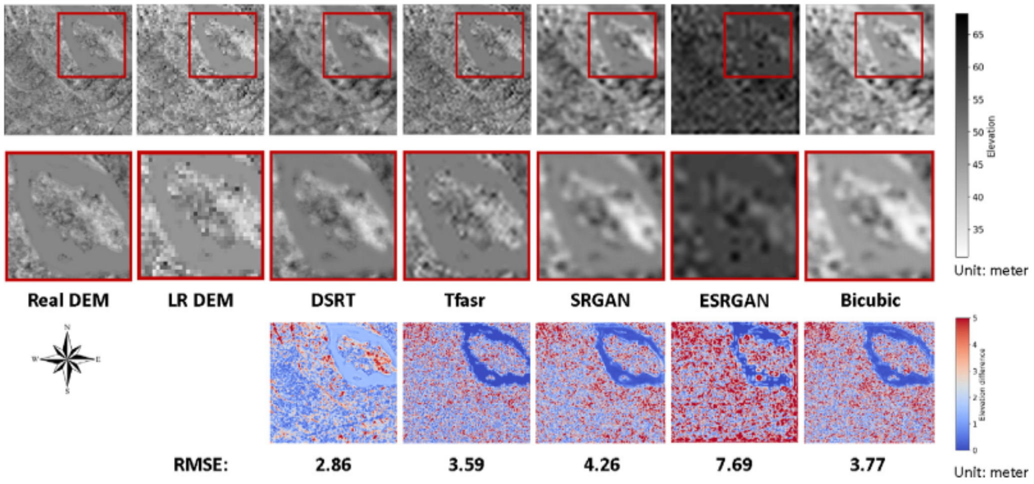


Fig. 18. 256 × 256 resolution water region DEMs generated by different methods and their elevation differences.

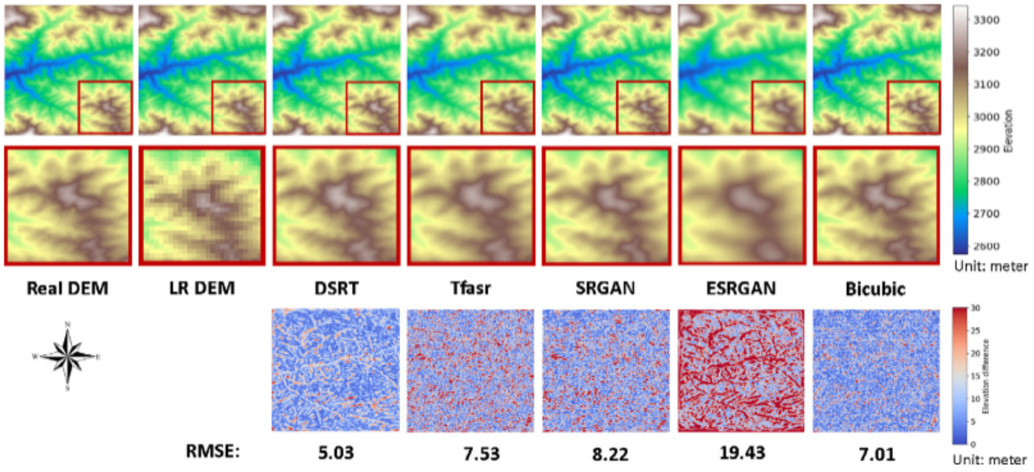


Fig. 19. 256 × 256 resolution canyon region DEMs generated by different methods and their elevation differences.

stable across different resolutions. The Tfasr method, while not the top performer, exhibits a commendable level of stability across different resolutions. This indicates that the Tfasr method is capable of maintaining a consistent level of accuracy irrespective of the resolution of the input DEM, a trait that could be advantageous in certain applications.

In Fig. 19, the DSRT method, despite having the lowest overall RMSE, not only effectively restores the shape of the peaks but also demonstrates a remarkable ability to capture the overall topography of the canyon region. This is a testament to the superior performance of the DSRT method, which, despite minor visual discrepancies, manages

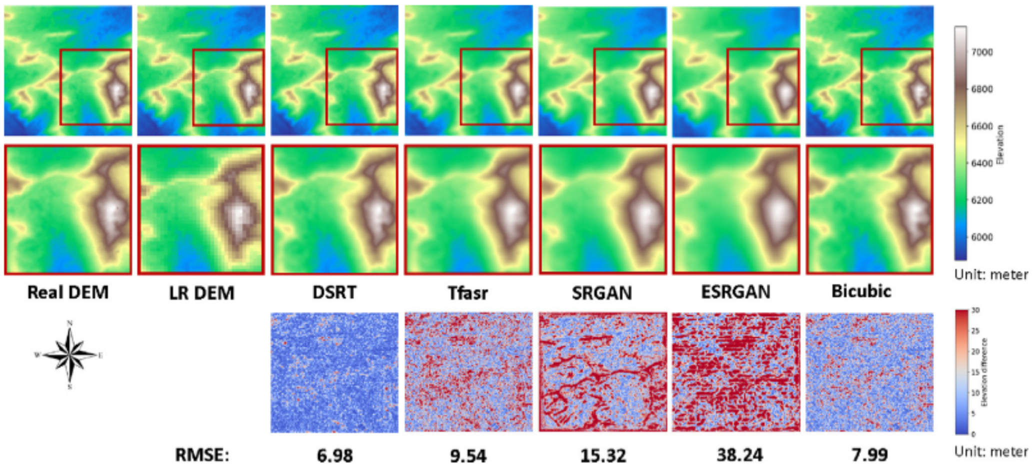


Fig. 20. 256 × 256 resolution plateau region DEMs generated by different methods and their elevation differences.

to generate a DEM that closely mirrors the real one. This is a significant achievement, considering the complexity of the terrain and the challenges associated with accurately capturing the nuances of such landscapes. The Bicubic method, while performing admirably in restoring the outline of the peaks, falls short in terms of overall accuracy, as indicated by its higher RMSE compared to DSRT. This underscores the exceptional performance of the DSRT method, which manages to strike a balance between maintaining overall accuracy and preserving specific features of the terrain. The remaining three methods – Tfasr, SRGAN, and ESRGAN – all exhibit certain shortcomings in terms of RMSE and preservation of geographical features. These deficiencies become particularly pronounced when dealing with high-resolution DEMs with significant elevation differences, such as those found in canyon regions.

6.1.4. *Performance evaluation in plateau regions*

It is evident in Table 3 that the Tfasr method, while showing a slight increase in RMSE compared to DSRT, still manages to maintain a relatively low error rate across all resolutions. This suggests that, while not as accurate as DSRT, Tfasr is still capable of producing reasonably accurate super-resolved DEMs of plateau regions. Interestingly, the Bicubic method, despite having a slightly higher RMSE than DSRT at the 64 × 64 resolution, manages to achieve a lower RMSE at the 512 × 512 resolution. This suggests that the Bicubic method may be more effective at capturing the topographical features of plateau regions at higher resolutions.

In Fig. 20, visually, the DSRT method produces a DEM that is strikingly similar to the original, capturing the subtle topographical features of the plateau region with remarkable accuracy. This visual impression is supported by the RMSE values presented in the accompanying Table 3, where DSRT consistently achieves the lowest error across all resolutions. The Bicubic method also performs well in terms of visual quality, particularly in its ability to accurately reproduce the contours of the peaks. However, its slightly higher RMSE values, as compared to DSRT, indicate a somewhat lower overall accuracy. The other methods – Tfasr, SRGAN, and ESRGAN – while capable of generating visually similar DEMs, struggle to capture the topographical features of the plateau region accurately. This is reflected in their higher RMSE values and is particularly noticeable in the case of ESRGAN, which exhibits the highest error rates across all resolutions.

6.2. *The effect of the collaborative loss function on the model*

To reveal the effect of the collaborative loss function on topographic feature recovery, we conduct additional experiments. We only keep $L1$ loss as the loss function, named DSRT$_{L1}$. In the second experiment, we retained L_1 loss and $RMSE$ loss, named DSRT$_{L_1,RMSE}$. The third experiment is a model with a full collaborative loss function, namely DSRT. Tfasr, which has the second-best performance in Section 5, is selected as the baseline in this experiment.

Table 4
RMSE assessment results of collaborative loss.

Method	RMSE			
	Elevation (m)	Slope (°)	Aspect (°)	Curvature
DSRT	1.29	1.96	79.55	0.00026
DSRT$_{L1}$	1.39	2.82	88.22	0.00048
DSRT$_{L1,RMSE}$	1.25	2.42	83.26	0.00038
Tfasr	1.49	2.32	82.67	0.00036

Table 4 lists the results of additional experiments of collaborative loss. If only the L_1 loss were kept, the obtained elevation error of DSRT$_{L1}$ is similar to DSRT, but without the feature loss function ℓ_{feat}, the RMSE-slope, RMSE-Aspect, and RMSE-Curvature are the highest. After adding $RMSE$ loss on DSRT$_{L_1}$, DSRT$_{L_1,RMSE}$ achieves the best RMSE-Elevation result, and the RMSE-slope, RMSE-Aspect, and RMSE-Curvature are further reduced. Since Tfasr used slope and aspect as a collaborative loss function in order to restore geographical features, the RMSE-slope, RMSE-Aspect, and RMSE-Curvature of DSRT$_{L_1,RMSE}$ are still inferior to Tfasr. This proves that RMSE can be used as a loss function that constrains elevation accuracy while also constraining the preservation of topographic features globally. Although the elevation accuracy results are slightly lower than DSRT$_{L1,RMSE}$, DSRT achieves the best results in RMSE-slope, RMSE-Aspect, and RMSE-Curvature, which indicates that a complete collaborative loss function can optimally restore terrain features while preserving elevation accuracy as much as possible.

6.3. *The impact of the weight coefficients on the model*

To better analyse the impact of different weight coefficients in the collaborative loss function, we further tested the impact of α and β in Eq. (12) on the model. As shown in Table 5, different weight coefficients bring great differences to the final results of the model. When the weight coefficient α of $RMSE$ loss is set to 1, as the coefficient β of ℓ_{feat} decreases, the model can gradually retain more complete terrain features (decrease trend of RMSE-Elevation, RMSE-Slope, RMSE-Aspect, and RMSE-Curvature). For the weight coefficient α of $RMSE$ loss, if α is set to 10, the final model generates the worst RMSE-Elevation result. If α is reduced to 0.1 and 0.001, although better results can be obtained compared to the value of 10, the model will generate a larger error when $\alpha = 0.001$. This shows that for $RMSE$ loss, the value of α that is too high or too low will affect the final performance of the model, and the feature loss ℓ_{feat} cannot well constrain the collaborative loss of the model.

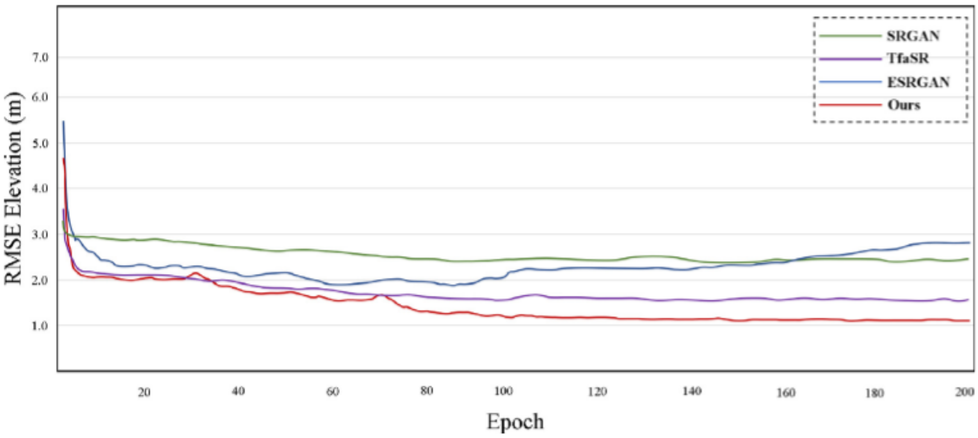


Fig. 21. Elevation loss (RMSE) of four models.

Table 5
RMSE assessment results of weight coefficients (baseline: $\alpha - 1$, $\beta - 0.001$).

α	β	RMSE			
		Elevation (m)	Slope (°)	Aspect (°)	Curvature
1	1	2.589	3.058	89.552	0.458
1	0.1	1.896	2.422	86.332	0.356
1	**0.001**	**1.286**	**1.958**	**79.552**	**0.255**
10	0.001	5.386	3.902	90.673	0.504
0.1	0.001	2.474	2.774	85.233	0.385
0.001	0.001	2.993	2.988	87.066	0.399

6.4. *Effect of additional training epochs*

To understand whether the performance can be further improved if the training epoch model is increased, we continued to train the four networks for 100 epochs. We set the learning rate to decrease in steps, reducing it to half every 20 epochs. As shown in Fig. 21, our model has achieved good results at 100 epochs. Although the final results are slightly improved, the difference is insignificant. SRGAN and Tfasr have slight loss fluctuations but have remained in a stable range without significant improvement. The loss of ESRGAN has a clear upward trend, further proving that ESRGAN is unsuitable for DEM super-resolution tasks.

6.5. *Evaluation based on 2 × SR (upscale = 2)*

In order to explore the performance of DSRT in dealing with the super-resolution of different scales, we trained the model based on 2 × SR (upscale = 2). The training parameter settings are the same as in Section ??.

In Fig. 22, when setting the scale of SR to 2 (upscale = 2), the generated high-resolution DEM is very close to the original picture. As the upscale value increases, the elevation information retained by the low-resolution DEM will also decrease. As shown in Fig. 22, when upscale = 4, we need to downsample the 64 × 64 resolution DEM to 16 × 16. The resulting quadruple high-resolution DEM retains much less elevation information than the low-resolution DEM downsampled to 32 × 32 when upscale = 2. Table 6 shows the results of the different upscale factors of DSRT. Similar to the visualization results in Fig. 22, better results can be obtained using upscale = 2 as the sampling factor, especially in slope and aspect. This indicates that DSRT can generate more realistic HR DEMs with sufficient elevation and topographic features.

6.6. *Evaluation on higher-resolution DEMs*

To further validate the generalization capability of our model, we conducted additional experiments, directly comparing the model-generated HR DEM with the existing 8 m resolution DEM.

6.6.1. *Data preparation and processing*

The High Mountain Asia (HMA) dataset primarily focuses on mountainous regions and boasts an 8 m spatial resolution. This dataset was chosen for its higher precision in challenging terrains, making it an ideal candidate for comparison against our derived models. Given the inherent nature of high-altitude regions, the raw HMA DEM contained numerous data voids. To address this, an elevation-contour-based interpolation method was employed to fill these gaps. The rationale behind contour-based interpolation is that it can effectively maintain the topographical features while filling in the missing values, ensuring minimal distortion. Once the voids were addressed, the HMA DEM was systematically cropped into multiple tiles with a resolution of 512 × 512 pixels. This dimension was selected to ensure our experiments' consistency and facilitate efficient data processing (Jiang et al., 2023). For this study and to simulate a real-world application where higher-resolution DEMs might not always be available, the extracted regions from the ASTER GDEM were downsampled using a Cubic interpolation method. The resultant DEMs had a resolution of 128 × 128 cells and were used as test data in subsequent experiments.

6.6.2. *Results*

Due to the high-resolution and low-resolution DEMs originating from two distinct data sources, inherent discrepancies in elevation values are to be expected. However, the experimental results Table 7 shows DSRT maintains a leading performance. Fig. 23 show that the elevation values of the low-resolution and high-resolution DEMs differ in certain areas. Nonetheless, the HR DEM generated by DSRT manages to restore the geographical features as closely as possible.

6.7. *Limitations*

In the proposed DSRT model, we use a self-attention mechanism-based deep learning model for DEM super-resolution. Despite its significant advancements, the application of our model could potentially be bounded by several limitations.

Firstly, the constraints related to the training data should be considered. Our model might exhibit biases resulting from the data it was trained on. For instance, if the training dataset encompasses only a limited variety of terrain types, complexities, or geographical extents, the model's performance might be less reliable when extrapolated to terrains or regions outside of these parameters. This limitation could

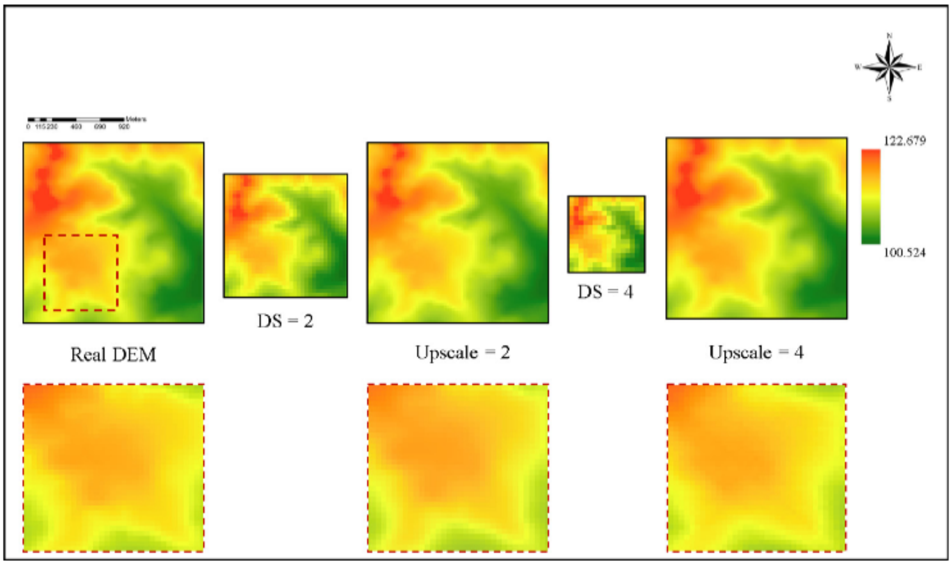


Fig. 22. Generated DEM from different upscale factors (red dotted frames are the details).

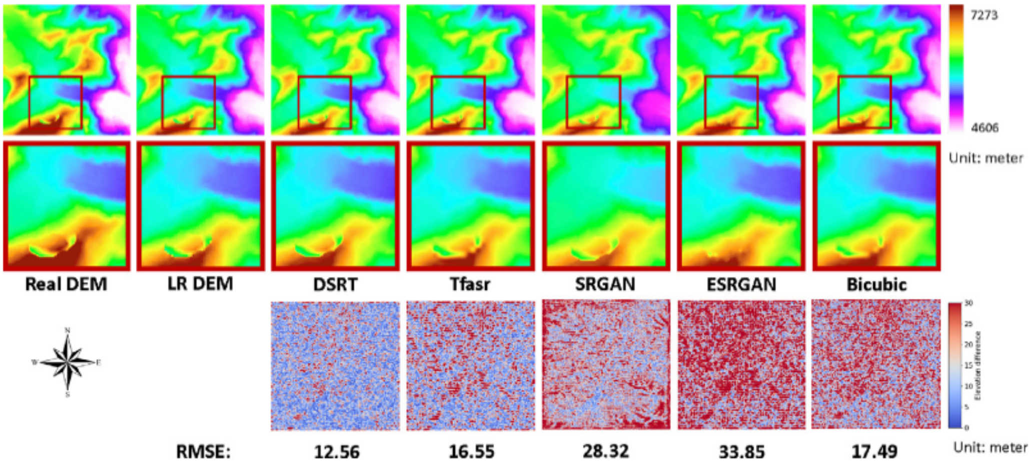


Fig. 23. Performance evaluation on HMA DEM. 512 × 512 resolution DEMs generated by different methods and their elevation differences.

Table 6
RMSE results of different upscale factors.

	RMSE-elevation (m)	RMSE-slope (°)	RMSE-aspect (°)	RMSE-curvature (m^{-1})
DSRT_upscale − 4	1.286	1.958	79.552	0.000255
DSRT_upscale − 2	1.268	1.799	77.004	0.000237

Table 7
RMSE results of different upscale factors.

	RMSE-elevation (m)	RMSE-slope (°)	RMSE-aspect (°)	RMSE-curvature
DSRT	**9.738**	**7.734**	**89.463**	**3.96**
Tfasr	10.268	8.005	92.004	4.98
SRGAN	19.779	11.045	98.233	8.38
ESRGAN	28.443	18.271	99.001	11.99
Bicbuic	13.003	10.280	89.821	9.99

affect the model's ability to generalize effectively to new and diverse geographic areas or to significantly different terrain types. Future research could focus on enhancing the diversity of the training data to improve the model's versatility.

Furthermore, training and optimization challenges inherent in models with a large number of parameters, like Vision Transformers, should be acknowledged. These models necessitate vast amounts of data and meticulous fine-tuning for optimal performance. In scenarios with insufficient data or improper optimization, the overall performance of the model could be compromised.

Lastly, while the power of Transformer models in capturing complex patterns is well acknowledged, their black-box nature and lack of interpretability could be a potential drawback. Although we used the LAM method to demonstrate our model in using elevation points, understanding the internal workings of these models can be difficult, which may hinder applications that require comprehensible model decisions. Future work could explore incorporating methods to increase the interpretability of these models.

By acknowledging these limitations, we aim to guide future research in this field, providing a realistic perspective on the challenges that may be encountered and outlining potential areas for further exploration and development

7. Conclusion

In this paper, we propose a self-attention-based deep learning network to increase the resolution of digital elevation models. In order to improve DEM resolution and take spatial information continuity into account, our network (DSRT) employs a unique Transformer-based technique. We carry out in-depth tests to confirm the efficacy of DSRT, and the findings demonstrate that our approach beats image-based SR and DEM-oriented SR approaches. In comparison to bicubic, SRGAN, ESRGAN, and Tfasr, our network further reduces the RMSE-Elevation, RMSE-Slope, RMSE-Aspect, and RMSE-Curvature of the generated HR DEMs, demonstrating the superiority of the deep learning model based on the Transformer in learning DEM features over interpolation, CNNs, and GANs.

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

The authors do not have permission to share data.

Acknowledgements

All authors approved the final version of the manuscript.

Funding

This research was funded by the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (SZS2022004), the Key Programme Special Fund of XJTLU (KSF-A-19), the Suzhou Science and Technology Project (SYG202122), the Research Development Fund of XJTLU (RDF-19-02-23) and Jiangsu Engineering Research Center for Data Science and Cognitive Computing.

References

Agterberg, F.P., 1984. Trend surface analysis. Spat. Statist. Models 147–171.

Anagun, Y., Isik, S., Seke, E., 2019. SRLibrary: Comparing different loss functions for super-resolution over various convolutional architectures. J. Vis. Commun. Image Represent. 61, 178–187.

Ba, J.L., Kiros, J.R., Hinton, G.E., 2016. Layer normalization. arXiv preprint arXiv:1607.06450.

Chen, Z., Wang, X., Xu, Z., et al., 2016. Convolutional neural network based DEM super resolution. Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci. 41.

Da Wang, Y., Armstrong, R.T., Mostaghimi, P., 2019. Enhancing resolution of digital rock images with super resolution convolutional neural networks. J. Pet. Sci. Eng. 182, 106261.

Daihong, J., Sai, Z., Lei, D., Yueming, D., 2022. Multi-scale generative adversarial network for image super-resolution. Soft Comput. 26 (8), 3631–3641.

Daly, C., 2006. Guidelines for assessing the suitability of spatial climate data sets. Int. J. Climatol.: J. R. Meteorol. Soc. 26 (6), 707–721.

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

Dong, C., Loy, C.C., Tang, X., 2016. Accelerating the super-resolution convolutional neural network. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, pp. 391–407.

Fran, J.S., Martin, Y., 2006. High spatial resolution satellite imagery, DEM derivatives, and image segmentation for the detection of mass wasting processes. Photogramm. Eng. 72 (6), 687–692.

Fu, K., Peng, J., Zhang, H., Wang, X., Jiang, F., 2020. Image super-resolution based on generative adversarial networks: a brief review.

Ge, Y., Jiang, S., Xu, Q., Jiang, C., Ye, F., 2018. Exploiting representations from pre-trained convolutional neural networks for high-resolution remote sensing image retrieval. Multimedia Tools Appl. 77, 17489–17515.

Gu, J., Dong, C., 2021. Interpreting super-resolution networks with local attribution maps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9199–9208.

Han, D., 2013. Comparison of commonly used image interpolation methods. In: Conference of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013). Atlantis Press, pp. 1556–1559.

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.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al., 2022. A survey on vision transformer. IEEE Trans. Pattern Anal. Mach. Intell. 45 (1), 87–110.

Hayat, K., 2018. Multimedia super-resolution via deep learning: A survey. Digit. Signal Process. 81, 198–217.

He, X., Cheng, J., 2022. Revisiting L1 loss in super-resolution: a probabilistic view and beyond. arXiv preprint arXiv:2201.10084.

Hu, Y., Li, J., Huang, Y., Gao, X., 2019. Channel-wise and spatial feature modulation network for single image super-resolution. IEEE Trans. Circuits Syst. Video Technol. 30 (11), 3911–3927.

Hu, F., Xia, G.-S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens. 7 (11), 14680–14707.

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. http://dx.doi.org/10.1016/j.jag.2023.103296, URL https://www.sciencedirect.com/science/article/pii/S1569843223001188.

Jo, Y., Yang, S., Kim, S.J., 2020. Investigating loss functions for extreme super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 424–425.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, pp. 694–711.

Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M., 2022. Transformers in vision: A survey. ACM Comput. Surv. (CSUR) 54 (10s), 1–41.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4681–4690.

Li, J., Fang, F., Mei, K., Zhang, G., 2018a. Multi-scale residual network for image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 517–532.

Li, K., Yang, S., Dong, R., Wang, X., Huang, J., 2020. Survey of single image super-resolution reconstruction. IET Image Process. 14 (11), 2273–2290.

Li, Y., Zhang, L., Dingl, C., Wei, W., Zhang, Y., 2018b. Single hyperspectral image super-resolution with grouped deep recursive residual network. In: 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM). IEEE, pp. 1–4.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R., 2021. Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1833–1844.

Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144.

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), http://dx.doi.org/10.3390/rs14092181, URL https://www.mdpi.com/2072-4292/14/9/2181.

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

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10012–10022.

Lu, G.Y., Wong, D.W., 2008. An adaptive inverse-distance weighting spatial interpolation technique. Comput. Geosci. 34 (9), 1044–1055.

Ma, C., Rao, Y., Cheng, Y., Chen, C., Lu, J., Zhou, J., 2020. Structure-preserving super resolution with gradient guidance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7769–7778.

Ma, C., Yu, P., Lu, J., Zhou, J., 2022. Recovering realistic details for magnification-arbitrary image super-resolution. IEEE Trans. Image Process. 31, 3669–3683.

Mitas, L., Mitasova, H., 1999. Spatial interpolation. Geogr. Inf. Syst.: Princ. Tech. Manage. Appl. 1 (2).

Moore, I.D., Grayson, R., Ladson, A., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5 (1), 3–30.

Nagaraj, P., Muthamilsudar, K., Naga, S., Mohammed, R., Sujith, K., 2020. Perceptual image super resolution using deep learning and super resolution convolution neural networks (SRCNN). Intell. Syst. Comput. Technol. 37 (3).

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp. 807–814.

Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., Shen, H., 2020. Single image super-resolution via a holistic attention network. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16. Springer, pp. 191–207.

Oliver, M.A., Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. Int. J. Geogr. Inf. Syst. 4 (3), 313–332.

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.

Park, S.C., Park, M.K., Kang, M.G., 2003. Super-resolution image reconstruction: a technical overview. IEEE Signal Process. Mag. 20 (3), 21–36.

Rad, M.S., Bozorgtabar, B., Marti, U.-V., Basler, M., Ekenel, H.K., Thiran, J.-P., 2019. Srobb: Targeted perceptual loss for single image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2710–2719.

Rahman, H.S., Alireza, K., Reza, G., et al., 2010. Application of artificial neural network, kriging, and inverse distance weighting models for estimation of scour depth around bridge pier with bed sill. J. Softw. Eng. Appl. 3 (10), 944.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Rukundo, O., Cao, H., 2012. Nearest neighbor value interpolation. arXiv preprint arXiv:1211.1768.

Santurkar, S., Tsipras, D., Ilyas, A., Madry, A., 2018. How does batch normalization help optimization? Adv. Neural Inf. Process. Syst. 31.

Schuler, C.J., Hirsch, M., Harmeling, S., Schölkopf, B., 2015. Learning to deblur. IEEE Trans. Pattern Anal. Mach. Intell. 38 (7), 1439–1451.

Shaw, P., Uszkoreit, J., Vaswani, A., 2018. Self-attention with relative position representations. arXiv preprint arXiv:1803.02155.

Shean, D.E., Alexandrov, O., Moratto, Z.M., Smith, B.E., Joughin, I.R., Porter, C., Morin, P., 2016. An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. ISPRS J. Photogramm. Remote Sens. 116, 101–117.

Simonyan, K., Vedaldi, A., Zisserman, A., 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

Singla, K., Pandey, R., Ghanekar, U., 2022. A review on single image super resolution techniques using generative adversarial network. Optik 169607.

Smith, M.J., Clark, C.D., 2005. Methods for the visualization of digital elevation models for landform mapping. Earth Surf. Process. Landf. 30 (7), 885–900.

Su, H., Tang, L., Wu, Y., Tretter, D., Zhou, J., 2011. Spatially adaptive block-based super-resolution. IEEE Trans. Image Process. 21 (3), 1031–1045.

Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks. In: International Conference on Machine Learning. PMLR, pp. 3319–3328.

Taud, H., Mas, J., 2018. Multilayer perceptron (MLP). Geomat. Approaches Model. Land Change Scen. 451–455.

Tun, N.L., Gavrilov, A., Tun, N.M., Aung, H., et al., 2021. Remote sensing data classification using a hybrid pre-trained VGG16 CNN-SVM classifier. In: 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). IEEE, pp. 2171–2175.

Uysal, M., Toprak, A.S., Polat, N., 2015. DEM generation with UAV photogrammetry and accuracy analysis in Sahitler hill. Measurement 73, 539–543.

Voita, E., Talbot, D., Moiseev, F., Sennrich, R., Titov, I., 2019. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. arXiv preprint arXiv:1905.09418.

Wan, Z., Zhang, J., Chen, D., Liao, J., 2021. High-fidelity pluralistic image completion with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4692–4701.

Wang, Z., Chen, J., Hoi, S.C., 2020. Deep learning for image super-resolution: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (10), 3365–3387.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C., 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

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.

Yang, D., Li, Z., Xia, Y., Chen, Z., 2015. Remote sensing image super-resolution: Challenges and approaches. In: 2015 IEEE International Conference on Digital Signal Processing (DSP). IEEE, pp. 196–200.

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., Liao, Q., 2019. Deep learning for single image super-resolution: A brief review. IEEE Trans. Multimed. 21 (12), 3106–3121.

Zagoruyko, S., Komodakis, N., 2016. Wide residual networks. arXiv preprint arXiv:1605.07146.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13. Springer, pp. 818–833.

Zeiler, M.D., Taylor, G.W., Fergus, R., 2011. Adaptive deconvolutional networks for mid and high level feature learning. In: 2011 International Conference on Computer Vision. IEEE, pp. 2018–2025.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., 2018. Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2472–2481.

Zhang, S., Wu, R., Xu, K., Wang, J., Sun, W., 2019. R-CNN-based ship detection from high resolution remote sensing imagery. Remote Sens. 11 (6), 631.

Zhang, Y., Yu, W., 2022. Comparison of DEM super-resolution methods based on interpolation and neural networks. Sensors 22 (3), http://dx.doi.org/10.3390/s22030745, URL https://www.mdpi.com/1424-8220/22/3/745.

Zhang, Y., Yu, W., Zhu, D., 2022a. Terrain feature-aware deep learning network for digital elevation model superresolution. ISPRS J. Photogramm. Remote Sens. 189, 143–162. http://dx.doi.org/10.1016/j.isprsjprs.2022.04.028, URL https://www.sciencedirect.com/science/article/pii/S0924271622001332.

Zhang, X., Zeng, H., Guo, S., Zhang, L., 2022b. Efficient long-range attention network for image super-resolution. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII. Springer, pp. 649–667.

Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4 (2), 22–40.

Zhu, D., Cheng, X., Zhang, F., Yao, X., Gao, Y., Liu, Y., 2020. Spatial interpolation using conditional generative adversarial neural networks. Int. J. Geogr. Inf. Sci. 34 (4), 735–758.

Zhu, X., Hu, H., Lin, S., Dai, J., 2019. Deformable convnets v2: More deformable, better results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9308–9316.