# Digital Surface Model Super-Resolution by Integrating High-Resolution Remote Sensing Imagery Using Generative Adversarial Networks

Guihou Sun, Yuehong Chen, Jiamei Huang, Qiang Ma, and Yong Ge, *Senior Member, IEEE*

***Abstract*—Digital surface model (DSM) is the fundamental data in various geoscience applications, such as city 3-D modeling and urban environment analysis. The freely available DSM often suffers from limited spatial resolution. Super-resolution (SR) is a promising technique to increase the spatial resolution of DSM. However, most existing SR models struggle to reconstruct spatial details, such as buildings, valleys, and ridges. This article proposes a novel DSM super-resolution (DSMSR) model that integrates high-resolution remote sensing imagery using generative adversarial networks. The generator in DSMSR contains three modules. The first DSM feature extraction module uses the residual-in-residual dense block to extract features from low-resolution DSM. The second multiscale attention feature extraction module employs the pyramid convolutional residual dense blocks to capture the spatial details of ground objects at multiple scales from remote sensing imagery. The third DSM reconstruction module uses a squeeze-and-excitation block to fuse the extracted features from low-resolution DSM and high-resolution remote sensing imagery for generating SR DSM. The discriminator of DSMSR uses the relativistic average discriminator for adversarial learning. The slope loss is further introduced to ensure the accurate representation of topographic features. We evaluate DSMSR on four different terrain regions in the U.K. to downscale the 30-m AW3D30 DSM to 5-m DSM. The experimental results indicate that DSMSR outperforms the traditional interpolation algorithms and four existing deep-learning-based SR models. The DSMSR restores more spatial detail of topographic features and generates more accurate image quality, elevation, and terrain metrics.**

***Index Terms*—Digital surface model (DSM), generative adversarial networks (GANs), remote sensing imagery, slope loss, super-resolution (SR).**

Manuscript received 7 March 2024; revised 25 April 2024; accepted 7 May 2024. Date of publication 10 May 2024; date of current version 5 June 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFC3006701, in part by the National Natural Science Foundation of China under Grant 42071315, in part by Beijing Natural Science Foundation under Grant IS23117, and in part by Tianjin Science and Technology Program under Grant 23KPHDRC00320. *(Corresponding authors: Yuehong Chen; Qiang Ma.)*

Guihou Sun, Yuehong Chen, and Jiamei Huang are with the College of Geography and Remote Sensing, Hohai University, Nanjing 211100, China (e-mail: chenyh@lreis.ac.cn).

Qiang Ma is with the Research Center on Flood and Drought Disaster Reduction, China Institute of Water Resources and Hydropower Research, Beijing 100038, China (e-mail: maqiang@iwhr.com).

Yong Ge is with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.

Digital Object Identifier 10.1109/JSTARS.2024.3399544

## I. Introduction

TOPOGRAPHIC data play a crucial role in characterizing the elevation and location of Earth's surface and they are fundamental inputs for a wide range of geoscience applications [1], [2], [3], [4], [5], [6]. Digital terrain model (DTM), digital elevation model (DEM), and digital surface model (DSM) are three types of commonly used topographic data [3], [7]. Each of these models represents Earth's surface with distinct meanings and representations [3], [7], [8]. Generally, DTM is a representation of various topographic features, such as slope, gradient, valley lines, and ridges, organized in a 1-D or multidimensional feature vector space overlaid on a 2-D geographic space [1]. DEM only characterizes the elevation of bare-Earth surface by removing all natural and built objects [9], [10]. In contrast, DSM not only provides the elevation of bare-Earth surface but also includes the elevation of natural and built objects that exist above the terrain, such as buildings and vegetation [7], [8]. The natural and built information derived from DSM, such as building and tree height, is critical for a variety of applications, such as city 3-D modeling [11], [12], [13], population estimation [14], [15], urban environment analysis [6], [16], [17], [18], and flood impact assessment [2], [19], [20], [21].

Currently, there are several freely available topographic datasets [3], [7], [8], [22], including the shuttle radar topography mission (SRTM) DEM, the advanced spaceborne thermal emission and reflection radiometer global DEM (ASTER GDEM), and the advanced land observing satellite world 3-D DSM (AW3D30). SRTM is a 90-m DEM dataset created from radar data and it was enhanced to a 30-m version in 2014. ASTER GDEM is a 30-m DEM dataset created from stereopair images collected by advanced spaceborne thermal emission and reflection radiometer. Although SRTM and ASTER GDEM are primarily used as DEM datasets, they inherently contain both natural and built elevation information, which also qualifies them as DSM datasets [1], [3], [7], [8]. AW3D30 is a 30-m DSM dataset created from high-resolution (HR) phased array-type *L*-band synthetic aperture radar data [23]. These freely available DSM datasets are extensively utilized in numerous geoscience fields; however, their limited spatial resolutions ($\leq$30 m) are insufficient for regional scale and local scale applications [9], [19], [24], [25], [26], [27], [28].

Typically, there are two main ways to obtain HR DSM datasets. The first way is to improve the resolution of surveying

equipment (e.g., light detection and ranging laser scanner and stereo image pairs from an aerial camera), which includes the utilization of more sophisticated sensors, instruments, or techniques capable of capturing topographic data with higher resolution [2], [5], [29]. However, this way often requires a significant investment in labor, materials, and technical expertise. The second way is to use the super-resolution (SR) technique to increase the spatial resolution of DSM datasets, particularly those based on machine learning [9], [24], [25], [27], [28], [30], [31], [32]. SR refers to the process of enhancing the resolution of an image or dataset beyond its original level to achieve a more detailed representation of underlying features [5], [25], [33], [34], [35], [36], [37], [38]. With the rapid development of deep learning, many image SR algorithms have been proposed by leveraging the advantages of convolutional neural networks (CNNs) and generative adversarial networks (GANs), such as super-resolution convolutional neural network (SRCNN) [39], super-resolution generative adversarial networks (SRGANs) [40], and enhanced super-resolution generative adversarial networks (ESRGANs) [41]. Recent studies have demonstrated the effectiveness of attention and multiscale mechanisms in capturing features in both channel and spatial dimensions in image SR [42], [43], [44]. The attention mechanisms can emphasize important features and suppress unnecessary ones in the channel or spatial dimension, while the multiscale mechanism can capture the objects of different sizes in the spatial dimension [43]. For example, Xu et al. [45] introduced self-attention in GAN to capture global object features, while Jia et al. [46] proposed an advanced multiattention generative adversarial network (MA-GAN) for downscaling remote sensing images and it used a pyramid convolutional residual dense (PCRD) block with channel attention (CA) to enhance feature extraction at different scales. Hu et al. [47] proposed a squeeze-and-excitation (SE) block, which adaptively recalibrates channelwise feature responses by explicitly modeling interdependencies between channels. These SR algorithms have been successfully applied in downscaling DEM datasets. Chen et al. [24] used a similar architecture to SRCNN for DEM downscaling. A multiscale CNN has also been developed for DEM SR [48]. Both SRGAN and ESRGAN have been adopted to increase the spatial resolution of DEM [9], [25], [27], [49].

For downscaling DSM, Zhang et al. [28] adopted a CNN-based single-image SR algorithm to increase the spatial resolution of DSM, and Karatsiolis et al. [50] combined aerial imagery and DSM to estimate the height of buildings and vegetation using U-Net and residual networks. However, compared with DEM downscaling by various SR algorithms, little attention has been paid to DSM downscaling in recent years. One challenge in DSM downscaling using single-image SR models is the limited effectiveness in capturing spatial details due to the lack of auxiliary data [1], [26], [50]. Satellite images contain rich ground details that are useful auxiliary data to improve the spatial details of DSM, as stereo aerial image pairs are often used to create DSM datasets [7], [8], [13], [26], [50]. Meanwhile, the direct usage of image SR algorithms for DSM downscaling may be inappropriate because topographic details in DSM are significantly different from natural images [9], [25], [26], [28], [48]. As typical SR models take a loss function for natural images, topography-related loss functions have been proven to be more effective for topographic datasets [27], [51]. Moreover, advanced attention and multiscale mechanisms have seldom been used in downscaling DSM.

This article proposes a novel DSM super-resolution (DSMSR) model that integrates HR remote sensing imagery using GAN. The generator of the DSMSR comprises three modules: DSM feature extraction module, multiscale attention feature extraction module, and DSM reconstruction module. The residual-in-residual dense block (RRDB) from ESRGAN is used to extract features from low-resolution (LR) DSM in the DSM feature extraction module [41]. The PCRD blocks from MA-GAN are employed to capture the spatial details of ground objects at multiple scales from remote sensing imagery in the multiscale attention feature extraction module [46]. The SE block is employed to fuse the extracted features from LR DSM and HR remote sensing imagery and generate the final SR DSM in the DSM reconstruction module [47]. The relativistic average discriminator (RaD) is employed as the discriminator of DSMSR [52]. To ensure the accurate representation of topographic features during the training process, a slope loss is introduced and integrated into the loss functions in DSMSR. Compared with the existing DSM downscaling methods, DSMSR has several features and contributions.

1) The fusion of HR remote sensing imagery and LR DSM through GAN in DSMSR can fully exploit the HR remote sensing imagery to improve the spatial detail reconstruction in DSM.
2) The combination of RRDB and PCRD block in DSMSR is able to maximize the utilization of hierarchical features from LR DSM and HR remote sensing imagery, respectively.
3) The introduction of the slope loss facilitates the accurate representation of topographic features in DSM.

The rest of this article is organized as follows. Section II describes the proposed method. Section III illustrates the experimental results. Section IV discusses the experimental results. Finally, Section V concludes this article.

## II. Methodology

### A. DSMSR Architecture

The proposed DSMSR is built upon the foundation of GAN. A GAN model often consists of a generator and a discriminator, which are engaged in a zero-sum game where they compete with each other [40], [41]. DSMSR also has a generator $G$ and a discriminator $D$ to implement adversarial learning using a function Optimal$G, D$. The general process of DSMSR can be expressed by

$$\text{Optimal } G, D = \min_{G} \min_{D} (L_G + L_D) \tag{1}$$

where $G$ is the generator, $D$ is the discriminator, $L_G$ is the generator loss function, and $L_D$ is the discriminator loss function. In the competition, $G$ and $D$ obtain the capability to generate realistic data and distinguish the predicted data from the truth

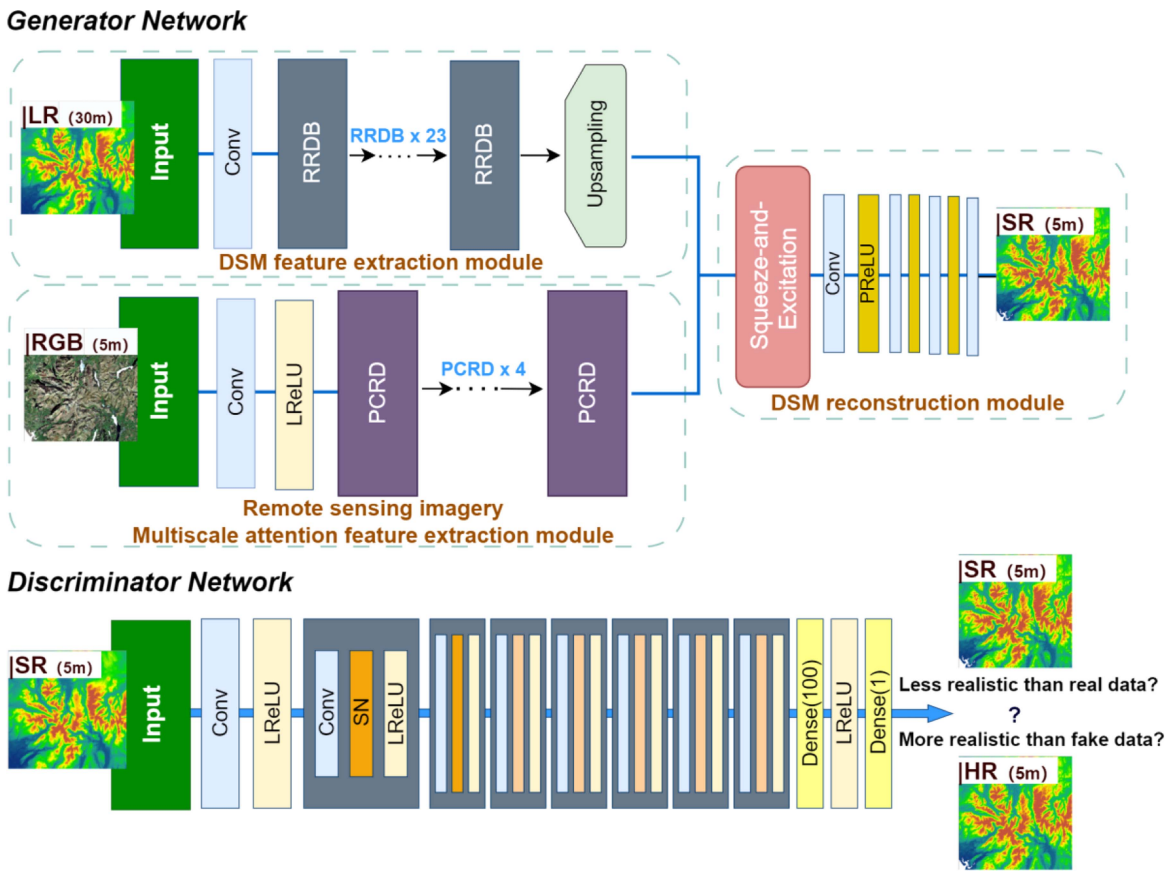


Fig. 1. Architecture of DSMSR.

data, respectively. The expected result is that $G$ will be able to generate data that cannot be discriminated by $D$.

The DSMSR is designed to generate HR DSM by fusing HR remote sensing imagery with LR DSM. The DSMSR consists of two components, including the generator and the discriminator, as illustrated in Fig. 1. The generator takes both LR DSM and HR remote sensing imagery as inputs to produce HR DSM results. The discriminator receives either a fake or a real HR DSM as the input to distinguish its source.

*1) Generator Network:* To take both LR DSM and HR remote sensing imagery as inputs for restoring HR DSM, the generator is designed to include three modules: the DSM feature extraction module, the multiscale attention feature extraction module, and the DSM reconstruction module.

*a) DSM feature extraction module:* ESRGAN, an enhanced version of SRGAN, is an advanced image SR model [41]. Compared with the traditional SRGAN, the generator network of ESRGAN adopts a different approach by removing the batch normalization layer and introducing a deeper and more complicated structure known as RRDB to replace the residual block in SRGAN [40], [41]. These modifications not only ensure stable training and consistent performance but also improve its generalization ability while reducing computational complexity and memory consumption [41]. Consequently, the DSM feature extraction module of the proposed DSMSR inherits the generator of ESRGAN to extract features from LR DSM and improve their resolution to a target resolution. The generator contains 23 RRDB blocks producing features with 64 channels.

*b) Multiscale attention feature extraction module:* To capture features in both spatial and channel dimensions, we design the multiscale attention feature extraction module. This module comprises a convolution-LReLU block and four PCRD blocks. The PCRD block incorporates an attention pyramid convolution (AttPConv) operator (see Fig. 2) [46]. The AttPConv operator conducts multiscale convolutions and then calculates CA on the resulting features, dynamically adjusting the weights of each feature [53]. Compared with ESRGAN [41], PCRD blocks can enhance the generalization ability in SR tasks [46]. The module can extract multiscale features from HR remote sensing images, which facilitates the recovery of spatial details for both natural and built objects in DSM reconstruction.

*c) DSM reconstruction module:* The DSM reconstruction module comprises an SE block [47], four convolutional layers, and three parametric rectified linear unit functions. The SE block is introduced to adaptively squeeze the 128 channelwise features from LR DSM and HR remote sensing imagery and adjust the weights for optimal multiscale feature fusion. The adjusted 128 features are then fused by four convolutional layers to generate the SR DSM at the target resolution.

*2) Discriminator Network:* The discriminator plays a crucial role in the adversarial training process. It acts as a critic to distinguish between the SR images produced by the generator

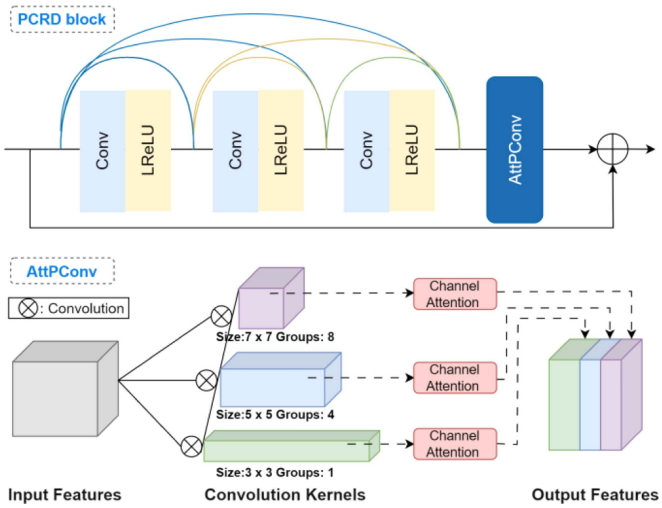


Fig. 2. PCRD block and AttPConv operator.

and real HR images. The RaD is proved to be more effective than the standard discriminator in SRGAN [40], [41]. It aims to learn to judge whether an image is "more real than another" rather than "whether an image is true or false" in a standard discriminator, guiding the generator to recover more detailed textures. Therefore, the RaD is used as the discriminator of DSMSR by

$$D_{\mathrm{Ra}}\left(X_{\mathrm{HR}}, X_{\mathrm{SR}}\right) = \sigma\left(C\left(X_{\mathrm{HR}}\right) - \mathbb{E}_{X_{\mathrm{SR}}}\left[C\left(X_{\mathrm{SR}}\right)\right]\right) \rightarrow 1 \quad \text{More realistic than fake data?} \tag{2}$$

$$D_{\mathrm{Ra}}\left(X_{\mathrm{SR}}, X_{\mathrm{HR}}\right) = \sigma\left(C\left(X_{\mathrm{SR}}\right) - \mathbb{E}_{X_{\mathrm{HR}}}\left[C\left(X_{\mathrm{HR}}\right)\right]\right) \rightarrow 0 \quad \text{Less realistic than real data?} \tag{3}$$

where $D_{\mathrm{Ra}}()$ is the RaD, $X_{\mathrm{HR}}$ is the HR DSM, $X_{\mathrm{SR}}$ is the SR DSM by the generator $G$ with the inputs of LR DSM and HR remote sensing imagery, and $\sigma()$, $C()$, and $\mathbb{E}[]$ represent the sigmoid function, discriminator output, and the mean of discriminator outputs for all generated images, respectively.

### B. Loss Function

The loss function is the key component used to quantify the error or discrepancy between the predicted output and the ground truth. Employing an appropriate loss function can effectively reduce training error and improve the model's performance. To train DSMSR, a comprehensive loss function is designed to evaluate the difference between the predicted SR DSM and the real HR DSM. The designed loss function not only includes the content loss, adversarial loss, and perceptual loss as in ESRGAN but also introduces a slope loss to recover terrain details. Slope is a critical attribute defining the topography of a geographic region [25], [27], [51]. Thus, slope loss is advantageous for restoring intricate terrain details and ensuring a more consistent alignment between the predicted DSM and the real DSMs terrain features. Further details about the slope loss will be explained in Section IV-A. The designed loss function of DSMSR is expressed as

$$L_G = L_{\mathrm{content}} + \alpha L_{\mathrm{adv}}^{G} + \beta L_{\mathrm{perception}} + \gamma L_{\mathrm{slope}} \tag{4}$$

where $L_{\mathrm{content}}$ represents the elevation loss, $L_{\mathrm{adv}}^{G}$ represents the adversarial loss, $L_{\mathrm{perception}}$ represents the perceptual loss, and $L_{\mathrm{slope}}$ represents the slope loss. $\alpha$, $\beta$, and $\gamma$ are the hyperparameters used to control the weights of each loss term.

*1) Content Loss:* The DSMSR chooses $L_1$ function as the content loss for the generator. $L_{\mathrm{content}}$ calculates the one-norm elevation distance between the SR DSM and the HR DSM and it can optimize the elevation accuracy of DSM at the pixel level as follows:

$$L_{\mathrm{content}} = \frac{1}{N}\sum_{i=1}^{N}\left|X_{\mathrm{HR},i} - X_{\mathrm{SR},i}\right| \tag{5}$$

where $X_{\mathrm{HR},i}$ and $X_{\mathrm{SR},i}$ are the HR and SR DSM elevation for pixel $i$, respectively.

*2) Adversarial Loss:* To increase the similarity between the predicted HR DSM and the real HR DSM, the RaD is employed as the discriminator in the proposed DSMSR [41], [52]. The RaD evaluates the performance of DSMSR by comparing the relative authenticity between the real and predicted DSM. The discriminator's adversarial loss function is defined as

$$L_{\mathrm{adv}}^{D} = -\mathbb{E}_{X_{\mathrm{HR}}}\left[\log\left(D_{\mathrm{Ra}}\left(X_{\mathrm{HR}}, X_{\mathrm{SR}}\right)\right)\right] - \mathbb{E}_{X_{\mathrm{SR}}}\left[\log\left(1 - D_{\mathrm{Ra}}\left(X_{\mathrm{SR}}, X_{\mathrm{HR}}\right)\right)\right]. \tag{6}$$

The adversarial loss function for the generator is defined as

$$L_{\mathrm{adv}}^{G} = -\mathbb{E}_{X_{\mathrm{HR}}}\left[\log\left(1 - D_{\mathrm{Ra}}\left(x_{\mathrm{HR}}, x_{\mathrm{SR}}\right)\right)\right] - \mathbb{E}_{X_{\mathrm{SR}}}\left[\log\left(D_{\mathrm{Ra}}\left(X_{\mathrm{HR}}, X_{\mathrm{SR}}\right)\right)\right]. \tag{7}$$

Equation (7) shows that the adversarial loss of the generator contains both $X_{\mathrm{HR}}$ and $X_{\mathrm{SR}}$; and hence, it benefits from the gradients of the generated and truth images during the training process.

*3) Perceptual Loss:* Perceptual loss was originally proposed by Johnson et al. [54] and later adapted and optimized in SRGAN [40]. ESRGAN further improved perceptual loss by using features before activation functions, leading to enhanced reconstruction results. The effectiveness of using features before activation functions lies in the fact that, after passing through very deep networks, the activated features tend to become sparse and can result in inconsistent brightness in the reconstructed images [41]. As DSM and natural images share similarities in containing rich object details, using perceptual loss before activation functions can aid in recovering object details. To achieve this, the predicted SR DSM and the real HR DSM are fed into a pretrained VGG19 network. The average $L_1$ distance between the features extracted from the fourth convolutional layer before the fifth max-pooling layer of VGG19 is then calculated. The perceptual loss function for the generator is defined as

$$L_{\mathrm{perception}} = \frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{VGG}_{19}(X_{\mathrm{HR},i}) - \mathrm{VGG}_{19}(X_{\mathrm{SR},i})\right| \tag{8}$$

where $\mathrm{VGG}_{19}()$ is the pretrained VGG 19 [55].

*4) Slope Loss:* Referring to Wang et al.'s article [41], the $L1$ loss [i.e., mean absolute error (MAE)] of slope between the predicted SR DSM and the real HR DSM is calculated as the slope loss. It is computed by

$$L_{\text{slope}} = \frac{1}{N}\sum_{i=1}^{N}|S_{\text{HR}} - S_{\text{SR}}| \tag{9}$$

$$S_{\text{HR}} = S(X_{\text{HR}}) \tag{10}$$

$$S_{\text{SR}} = S(S_{\text{SR}}) \tag{11}$$

$$S_{ij} = \arctan\left(\sqrt{dx^2 + dy^2}\right) \times \frac{\pi}{180} \tag{12}$$

where $S_{\text{HR}}$ is the slope of HR DSM, $S_{\text{SR}}$ is the slope of SR DSM, $S_{ij}$ is the slope at pixel (*i,j*), $dx$ is the gradient of the pixel in the horizontal direction, and $dy$ is the gradient of the pixel in the vertical direction.

### C. Accuracy Evaluation

Six metrics were selected to evaluate the performance of the DSMSR: peak signal-to-noise ratio (PSNR), mean structural similarity (MSSIM), root-mean-square error of elevation ($\text{RMSE}_{\text{E}}$), mean absolute error of elevation ($\text{MAE}_{\text{E}}$), mean absolute error of slope ($\text{MAE}_{\text{S}}$), and mean absolute error of aspect ($\text{MAE}_{\text{A}}$). These metrics were chosen to evaluate image quality, elevation accuracy, and terrain accuracy in SR DSM.

Visual perception can be used to evaluate the authenticity and level of detail recovery in the reconstructed images. The presence of detailed texture features in DSM, such as buildings, trees, and roads, is similar to remote sensing and natural images. Therefore, two widely used metrics (i.e., PSNR and MSSIM) for evaluating image quality in image SR are used to evaluate the image quality of reconstructed DSM. PSNR is calculated by the maximum pixel of the image and the mean square error between images, and it is defined as

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}}\right) \tag{13}$$

where MAX is the maximum elevation and MSE is the mean square error of elevation. PSNR does not have a specific range, and a higher PSNR indicates superior image quality.

MSSIM measures the structural similarity between images by considering three relatively independent components: luminance, contrast, and structure. It is defined as

$$\text{MSSIM} = \frac{1}{N}\text{SSIM}(x_i, y_i) \tag{14}$$

$$\text{SSIM}(x_i, y_i) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{15}$$

where $x_i$ and $y_i$ are the *i*th slices of $X_{\text{SR}}$ and $X_{\text{HR}}$, $\mu_x$ and $\mu_y$ are separately the means of $x_i$ and $y_i$, $\sigma_x^2$ and $\sigma_y^2$ are separately the variances of $x_i$ and $y_i$, $\sigma_{xy}$ is the covariances of $x_i$ and $y_i$, $c_1$ and $c_2$ are the constants used to avoid division by zero, and $N$ is the number of DSM slices. The range of MSSIM is [0,1], where a higher MSSIM indicates greater similarity between the reconstructed image and the real image, implying better structural similarity.

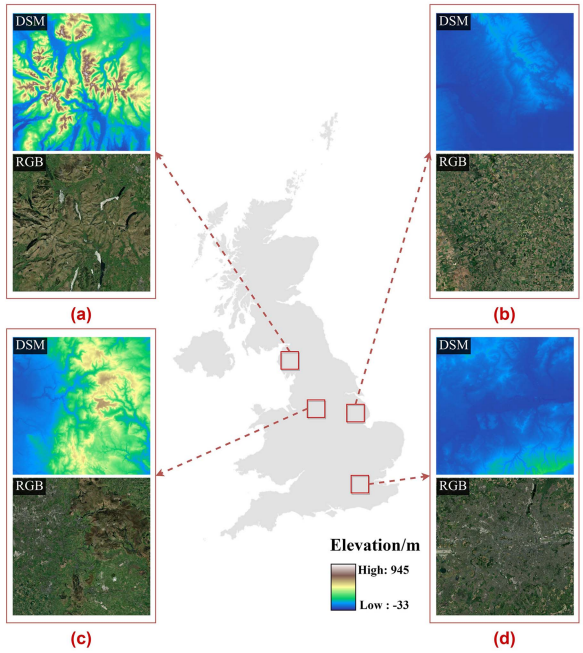


Fig. 3. Study area and data. (a) Mountainous region. (b) Suburban region. (c) Mixed region. (d) Urban region.

Elevation and terrain characteristics are crucial for DSM. Thus, we use four pixel-level metrics to quantitatively evaluate the accuracy of the reconstructed DSM. The $\text{RMSE}_{\text{E}}$ in (16) and the $\text{MAE}_{\text{E}}$ in (17) are adopted as the two metrics for assessing elevation accuracy. The $\text{MAE}_{\text{S}}$ in (18) and the $\text{MAE}_{\text{A}}$ in (19) are employed as the two metrics for evaluating terrain accuracy

$$\text{RMSE}_E = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_{\text{HR}} - X_{\text{SR}})^2} \tag{16}$$

$$\text{MAE}_E = \frac{1}{N}\sum_{i=1}^{N}|X_{\text{HR}} - X_{\text{SR}}| \tag{17}$$

$$\text{MAE}_S = \frac{1}{N}\sum_{i=1}^{N}|S_{\text{HR}} - S_{\text{SR}}| \tag{18}$$

$$\text{MAE}_A = \frac{1}{N}\sum_{i=1}^{N}|A_{\text{HR}} - A_{\text{SR}}| \tag{19}$$

where $S_{\text{HR}}$ and $S_{\text{SR}}$ are the slopes of HR DSM and SR DSM, respectively, and $A_{\text{HR}}$ and $A_{\text{SR}}$ are the aspects of HR DSM and SR DSM, respectively.

## III. Experiments

### A. Experimental Design

The study area is located in the United Kingdom and comprises four representative terrain regions: a mountainous region in Southern Cumbria [see Fig. 3(a)], a suburban region in Lincolnshire [see Fig. 3(b)], a mixed region in Manchester

[see Fig. 3(c)], and an urban region in London [see Fig. 3(d)]. Each region covers an area of 45 000 m × 45 000 m. This study employs three types of datasets. The first is the LR DSM, 30 m AW3D30 DSM, which is freely accessible.[1] The second is the HR DSM from the National LIDAR Programme DSM, which provides the accurate elevation data for the entire England on the U.K. Department for Environment Food and Rural Affairs Data Services Platform. The original resolution of the HR DSM is 1 m and was aggregated to 5 m as the HR DSM. The third is the 5-m HR remote sensing images with three bands (RGB) from the Mapbox platform. All datasets in this study use the British National Grid as the projection coordinate system, and the elevation coordinate system is Ordnance Datum Newlyn. The datasets in the mixed region are used as test data, while datasets in the other three regions are adopted to train the proposed DSMSR.

The scale factor is set to 6 to account for the resolution difference between LR and HR images. This means that the 30-m AW3D30 DSM is downscaled to the 5-m DSM using the proposed DSMSR. To obtain sufficient training samples, we created a total of 7200 patches from three different regions using both sequential and random cropping schemes. The LR DSM patches were cropped to a size of 32 × 32 pixels, while the HR DSM and remote sensing images were cropped to patches of 192 × 192 pixels. Before training, the mean elevation of each LR DSM patch and each HR DSM patch was subtracted to eliminate the impact of trend surfaces [30].

The training process was divided into two stages. During the first stage, the generator network of DSMSR was trained for 50 epochs using content loss and slope loss to obtain a pretrained generator. This pretrained generator was then used to initialize the generator of DSMSR during the second stage. The generator was then trained using the loss function with empirical weights $\alpha = 0.001$, $\beta = 0.01$, and $\gamma = 0.5$ in (4). The weights of $\alpha$ and $\beta$ were set following MA-GAN [46], and the weights of $\gamma$ were derived from experimental tests in Section IV-A.

The Adam algorithm was used for optimization during the training. The learning rate of the generator and discriminator in DSMSR was set to $1 \times 10^{-4}$. The generator and discriminator were updated alternately for 50 epochs. The DSMSR was implemented by the PyTorch package and executed on the NVIDIA RTX 4060 Ti GPU with 16 GB of RAM. A batch size of 16 was chosen to fit into the GPU memory.

To demonstrate the effectiveness of the proposed DSMSR, it was compared with the traditional bilinear (BL) interpolation algorithm and four widely used deep-learning-based SR models (i.e., SRCNN, SRGAN, ESRGAN, and MA-GAN). SRCNN, SRGAN, and ESRGAN have been employed in the DEM SR [9], [24], [25], [27], [56], and they were retrained using the DSM data from this experiment.

### B. Experimental Results

*1) Training Performance:* Fig. 4 illustrates the loss changes in each of the three training regions during the training process.

[1][Online]. Available: https://www.eorc.jaxa.jp/ALOS/en/aw3d30/data/index.htm

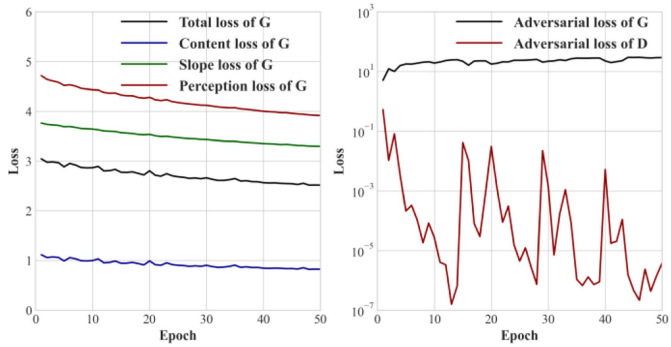


Fig. 4. Loss curves of DSMSR during the training process.

Content loss, perceptual loss, and slope loss fluctuate down and converge near 50 epochs. The adversarial losses of the generator and discriminator remain relatively stable, indicating that the training has achieved a favorable Nash equilibrium and resulted in a satisfactory trained DSMSR.

*2) SR Results in the Test Region:* The HR DSM and RGB remote sensing imagery in Fig. 5 reveal the diverse and complex topography of the test region, which is divided into two distinct areas. The western part of the test region is the urban area of Manchester, while the eastern part is the mountainous terrain. The SR 5-m DSM produced by the DSMSR exhibits more building details in the urban area and finer terrain features in the mountainous area compared with the original 30-m LR DSM images.

Fig. 6 shows the error maps of the six SR methods in the test region. The errors of all six methods are primarily concentrated in mountainous regions with significant topographic variations and urban areas with dense buildings. Specifically, for BL, SRCNN, ESRGAN, and MA-GAN methods, the errors are mostly negative in flat areas but mostly positive in mountain areas, while the errors of SRGAN are mostly negative. Moreover, for these five methods, the absolute errors in the urban are all large(>4 m). Compared with these methods, the majority of absolute errors in DSMSR are within 1 m in two types of regions. Meanwhile, the errors of DSMSR are more uniformly distributed than the other five methods. Additionally, DSMSR exhibits smaller errors in both densely built-up and mountainous regions, further confirming its excellent performance.

Table I presents the six accuracy metrics of different SR methods in the test region, including MSSIM, PSNR, $\text{RMSE}_\text{E}$, $\text{MAE}_\text{E}$, $\text{MAE}_\text{S}$, and $\text{MAE}_\text{A}$. As shown in Table I, all deep-learning-based SR models, including SRCNN, SRGAN, ESRGAN, MA-GAN, and DSMSR, show significant improvements over the traditional BL interpolation method. When comparing the SRCNN, SRGAN, ESRGAN, and MA-GAN, it can be observed that SRGAN performs slightly worse than the other three models, and ESRGAN and MA-GAN exhibit better accuracy levels. Notably, the proposed DSMSR outperforms all other models, achieving the best results across all six metrics.

Based on the two image quality metrics, the MSSIM and PSNR of the DSMSR are 0.17 and 2.04 higher than the best

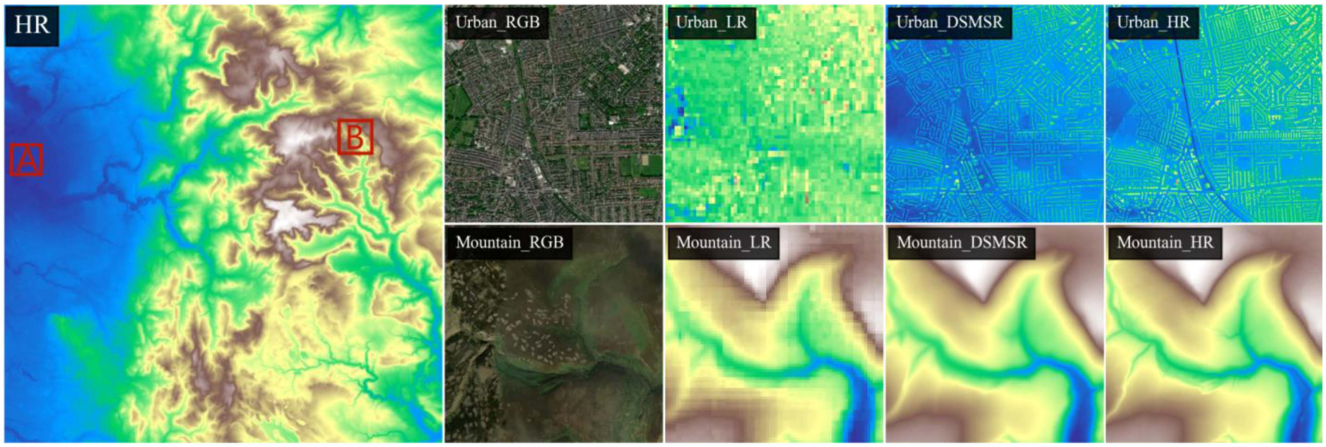


Fig. 5. SR results in the test region. (a) Urban area. (b) Mountain area.

TABLE I
ACCURACY OF DIFFERENT SR MODELS IN THE TEST REGION

| | BL | SRCNN | SRGAN | ESRGAN | MA-GAN | **DSMSR** |
|---|---|---|---|---|---|---|
| MSSIM | 0.45 | 0.52 | 0.49 | 0.52 | 0.53 | **0.70** |
| PSNR | 47.16 | 48.44 | 48.28 | 48.51 | 48.32 | **50.55** |
| $RMSE_E$ | 2.70 | 2.33 | 2.38 | 2.32 | 2.37 | **1.83** |
| $MAE_E$ | 1.79 | 1.52 | 1.53 | 1.47 | 1.50 | **1.07** |
| $MAE_S$ | 5.50 | 4.93 | 5.08 | 4.91 | 5.14 | **3.75** |
| $MAE_A$ | 70.47 | 63.28 | 68.57 | 65.31 | 63.49 | **49.96** |

Bold values indicate the best result in the same evaluation indicator.

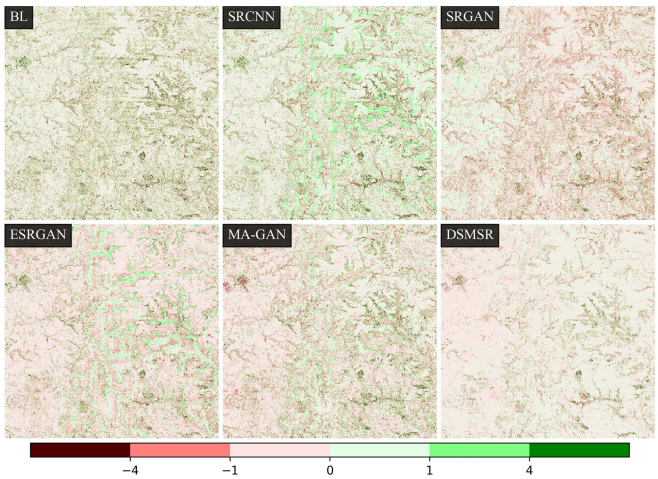


Fig. 6. Error maps of the reconstructed DSMs by different SR methods.

MSSIM and PSNR of other models, respectively. It is evident from the results that the reconstructed DSM produced by DSMSR has superior image quality with sharper edges and richer details of topographic features. Based on the two elevation accuracy metrics, DSMSR achieves a decrease of 0.49 m (21.1%) and 0.4 m (27.2%) in $RMSE_E$ and $MAE_E$, respectively, compared with the best values achieved by other models. When analyzing the two terrain-related metrics, it was found that the $MAE_S$ and $MAE_A$ of DSMSR have decreased by 1.16° (23.6%) and 13.32° (21%), respectively, compared with the best values of $MAE_S$ and $MAE_A$ achieved by other models. The findings indicate that DSMSR can recover more accurate pixel-level elevation and more significant topographic features in the reconstructed DSM.

*3) SR Results in a Zoom-in Urban Area:* Fig. 7 shows the result of the zoom-in urban area, which is marked with a red rectangle A in Fig. 5. The SR DSM of BL and SRCNN in the zoom-in urban area lack most of the building details and exhibit smoother results, and the SR DSM of MA-GAN is presented as flocculated. In contrast, SRGAN and ESRGAN show slightly improved performance in recovering details for some larger buildings, but most of the details are not true and accurate. Comparing with the other methods, the reconstructed DSM generated by DSMSR is significantly more consistent with the ground truth. DSMSR demonstrates an exceptional ability in capturing detailed surface texture information and achieve impressive restoration results for buildings and other objects. Additionally, the reconstructed DSM generated by DSMSR exhibits clear edge textures for buildings and various ground objects. This is because DSMSR integrates HR remote sensing images to restore these spatial details of objects.

According to further observation of the error maps in the zoom-in area, as shown in Fig. 7, it is evident that the five SR methods without HR remote sensing images exhibit higher errors in most artificial objects, such as buildings. In contrast,

TABLE II
ACCURACY OF DIFFERENT SR MODELS IN THE ZOOM-IN URBAN AREA

| | BL | SRCNN | SRGAN | ESRGAN | MA-GAN | **DSMSR** |
|---|---|---|---|---|---|---|
| MSSIM | 0.06 | 0.05 | 0.05 | 0.05 | 0.05 | **0.52** |
| PSNR | 20.46 | 20.74 | 20.46 | 20.68 | 20.71 | **22.82** |
| $RMSE_E$ | 2.75 | 2.67 | 2.75 | 2.69 | 2.68 | **2.10** |
| $MAE_E$ | 2.04 | 1.98 | 2.03 | 1.91 | 1.90 | **1.34** |
| $MAE_S$ | 12.99 | 13.70 | 12.45 | 13.67 | 14.14 | **7.55** |
| $MAE_A$ | 115.12 | 115.81 | 114.18 | 115.97 | 115.37 | **79.34** |

Bold values indicate the best result in the same evaluation indicator.

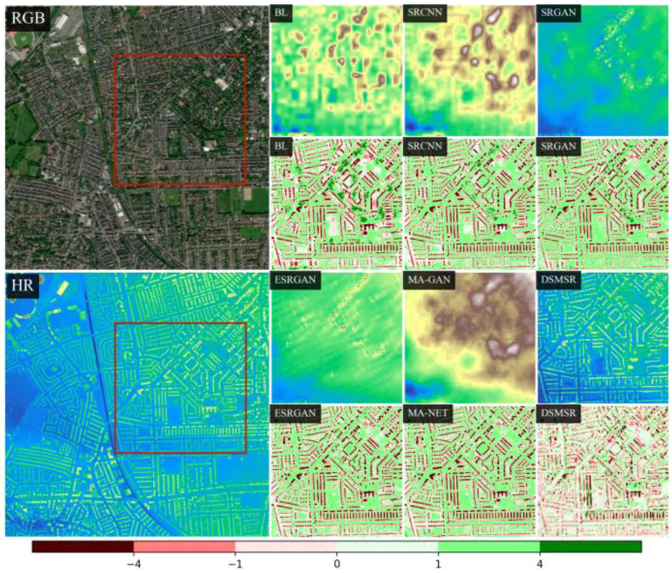


Fig. 7. SR results and error maps in the urban zoom-in area.

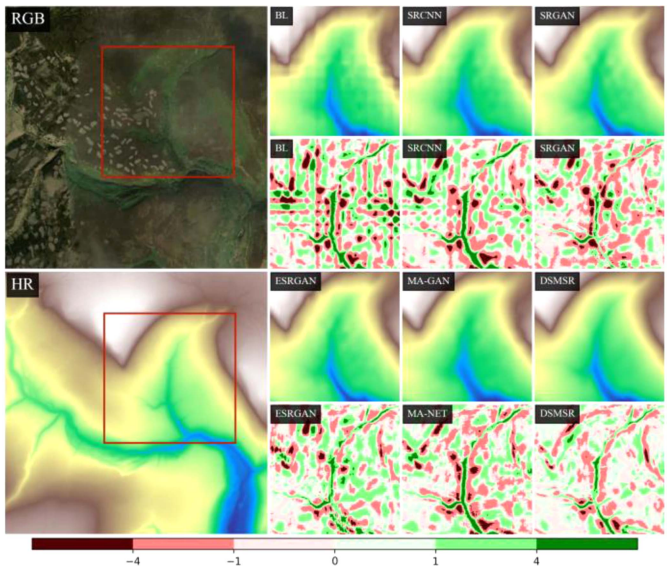


Fig. 8. SR results and error maps in the mountain zoom-in area.

the reconstructed DSM of DSMSR displays lower errors for most built and natural objects, which further highlights the ability of DSMSR to restore topographic features in urban areas, particularly buildings.

Table II presents that DSMSR achieves a significant improvement in accuracy and outperforms all other methods in the zoom-in urban area. The MSSIM and PSNR of DSMSR increase by 0.46 and 2.08, respectively, compared with the best result of the other five methods. Additionally, $RMSE_E$ and $MAE_E$ decrease by 0.57 m (21.3%) and 0.56 m (29.5%), respectively, while $MAE_S$ and $MAE_A$ decrease by 4.9° (39.4%) and 34.84° (30.5%), respectively.

*4) SR Results in a Zoom-in Mountain Area:* Fig. 8 shows the result of the zoom-in urban area, which is marked with a red rectangle B in Fig. 5. It can be seen that the BL and SRCNN produce smoother results. In contrast, SRGAN, ESRGAN, and MA-GAN show slightly better performance in recovering topographic lines, such as valleys and ridges, due to their use of GAN. In particular, ESRGAN, with its deeper networks and RaD, outperforms SRGAN and MA-GAN in terms of topographic recovery. The reconstructed terrain details produced by DSMSR are clearer and they are more realistic to the ground truth DSM. This improvement can be attributed to its ability to capture terrain texture from remote sensing imagery and the introduction of slope loss.

The error maps of the zoom-in mountain area in Fig. 8 show that the errors are mainly concentrated along the topographic lines (e.g., valleys) and their surroundings. The errors of both BL and SRCNN exhibit a grid pattern, while the errors of SRGAN, ESRGAN, MA-GAN, and DSMSR present a patchy distribution. In addition, the errors of DSMSR in the topographic lines are clearly lower than those in other methods, further suggesting the ability of DSMSR in restoring topographic features in mountain areas.

Table III presents accuracy metrics for the zoom-in mountain area, highlighting the effectiveness of the DSMSR model in accurately reconstructing detailed topographic features. DSMSR achieves the highest MSSIM and PSNR values when compared with the other five methods. Additionally, the $RMSE_E$ and $MAE_E$ of DSMSR decrease by 0.33 m (16.1%) and 0.29 m (21.6%), respectively, compared with the best results of the other five methods. Furthermore, DSMSR exhibits reductions of 0.61° (17.5%) and 7.43° (20.1%) in $MAE_S$ and $MAE_A$, respectively. These results highlight the effectiveness of the DSMSR model in accurately reconstructing detailed topographic features.

TABLE III
ACCURACY OF DIFFERENT SR MODELS IN THE ZOOM-IN MOUNTAIN AREA

| | BL | SRCNN | SRGAN | ESRGAN | MA-GAN | **DSMSR** |
|---|---|---|---|---|---|---|
| MSSIM | 0.70 | 0.79 | 0.80 | 0.79 | 0.79 | **0.90** |
| PSNR | 45.35 | 45.96 | 47.65 | 47.75 | 46.68 | **49.24** |
| $RMSE_E$ | 2.70 | 2.52 | 2.07 | 2.05 | 2.32 | **1.72** |
| $MAE_E$ | 1.80 | 1.61 | 1.35 | 1.34 | 1.48 | **1.05** |
| $MAE_S$ | 4.65 | 3.78 | 3.48 | 3.54 | 3.75 | **2.87** |
| $MAE_A$ | 45.41 | 36.99 | 41.16 | 39.67 | 39.04 | **29.56** |

Bold values indicate the best result in the same evaluation indicator.

TABLE IV
ACCURACY OF RECONSTRUCTED DSMS WITH DIFFERENT SLOPE LOSS COEFFICIENTS

| $\gamma$ | 0 | 0.1 | 0.5 | 1 | 5 |
|---|---|---|---|---|---|
| MSSIM | 0.67 | 0.69 | **0.7** | 0.69 | 0.68 |
| PSNR | 49.95 | 50.35 | **50.55** | 50.16 | 49.73 |
| $RMSE_E$ | 1.96 | 1.87 | **1.83** | 1.91 | 2.01 |
| $MAE_E$ | 1.18 | 1.11 | **1.07** | 1.13 | 1.2 |
| $MAE_S$ | 4.03 | 3.84 | **3.75** | 3.8 | 3.91 |
| $MAE_A$ | 51.93 | 50.41 | **49.96** | 50.72 | 51.89 |

Bold values indicate the best result in the same evaluation indicator.

TABLE V
ACCURACY OF RECONSTRUCTED DSMS WITH AND WITHOUT MULTISCALE AND ATTENTION MECHANISMS

| | RESNET18 | PCRD | **PCRD + SE** |
|---|---|---|---|
| MSSIM | 0.68 | 0.68 | **0.7** |
| PSNR | 49.87 | 50.14 | **50.55** |
| $RMSE_E$ | 1.98 | 1.92 | **1.83** |
| $MAE_E$ | 1.2 | 1.14 | **1.07** |
| $MAE_S$ | 3.93 | 3.84 | **3.75** |
| $MAE_A$ | 51.78 | 51.09 | **49.96** |

Bold values indicate the best result in the same evaluation indicator.

TABLE VI
ACCURACY OF DSMSR BY DIFFERENT SCALES

| Scale | MSSIM | PSNR | $RMSE_E$ | $MAE_E$ | $MAE_S$ | $MAE_A$ |
|---|---|---|---|---|---|---|
| 3 | **0.78** | **50.91** | **1.75** | **1.05** | 4.26 | **43.65** |
| 6 | 0.70 | 50.55 | 1.83 | 1.07 | **3.75** | 49.96 |

Bold values indicate the best result in the same evaluation indicator.

## IV. DISCUSSION

### A. Ablation Analysis for Slope Loss

To evaluate the impact of slope loss functions with different coefficients on DSMSR, we performed a comparative analysis of performance using different $\gamma$ values in the test region. Note that the value of 0 means that the slope loss is not used. Table IV presents that incorporating the slope loss leads to improvements in all accuracy metrics. As $\gamma$ increases, improvements in slope and aspect accuracy become more pronounced. However, an excessive emphasis on terrain precision may lead to a certain degree of decline in both scene quality and elevation accuracy. Notably, when $\gamma$ is set to 0.5, the model achieves an optimal balance in DSM image quality, elevation accuracy, and terrain precision.

### B. Ablation Analysis for Multiscale and Attention Mechanisms

The impact of multiscale and attention mechanisms is examined here. Traditional ResNet18 without multiscale and attention mechanisms is selected as a baseline for the ablation analysis. DSMSR employs the PCRD block with multiscale and attention mechanisms to extract features from HR remote sensing imagery and the SE block is integrated to fuse features from LR DSM and HR remote sensing imagery. The ablation result for the multiscale and attention mechanisms is shown in Table V. Compared with ResNet18, the introduction of the PCRD block improved PSNR to 50.14. It also significantly enhanced accuracy metrics, including a notable decrease in $RMSE_E$ (0.06 m) and $MAE_S$ (0.09°). Further refinement was obtained by the addition of the SE block for the fusion of features from LR DSM and HR remote sensing imagery. It achieved the highest MSSIM and PSNR, demonstrating superior accuracy in elevation, slope, and aspect. The combined use of PCRD blocks and SE blocks led to the best performance across all evaluated metrics. The importance of multiscale and attention mechanisms in enhancing feature extraction and fusing features within DSMSR is highlighted by these findings.

### C. Impact of Scale Factors on the Performance of DSMSR

The performance of DSMSR was demonstrated for a scale factor of 6. Its potential to other scale factors is further used to investigate. The metrics of DSMSR for two scales are shown in Table VI. It indicates that the smaller scale factor (i.e., 3) generated better results than the larger scale factor (i.e., 6). Specifically, the MSSIM and PSNR for the scale factor of 3 were higher than those for the scale factor of 6, while the $RMSE_E$, $MAE_E$, and $MAE_A$ for the scale factor of 3 were lower than those for the scale factor of 6. The MSSIM reveals that the scale factor of 6 is 11% lower than the scale factor of 3. The performance of DSMSR across different scales is basically consistent with many other SR models, as larger scale factors prove challenging for restoring more accurate details [27], [32], [33], [46].

### D. Performance Difference in Representative Areas

As observed from Tables II and III, all metrics of DSMSR in the mountain zoom-in area are better than those in the urban zoom-in area. The reason is mainly that the mountainous areas

TABLE VII
COMPUTATIONAL COMPLEXITY OF DIFFERENT MODELS

| | SRCNN | SRGAN | ESRGAN | MA-GAN | **DSMSR** |
|---|---|---|---|---|---|
| Parameters (Mb) | 0.06 | 2.04 | 16.7 | 0.7 | 17.44 |
| Throughput (image/s) | 221.28 | 24.96 | 15.84 | 27.84 | 5.99 |
| $RMSE_E$ | 2.33 | 2.38 | 2.32 | 2.37 | 1.83 |

have smoother terrain undulations, while urban areas have more intricate surface details (e.g., buildings). Despite these differences, the improved accuracy metrics in urban areas achieved by DSMSR are higher than those of mountainous areas. The reason for the significant improvement may be that remote sensing images can provide more detailed information for restoring urban objects in DSM.

### *E. Computational Complexity*

We conducted a comparative analysis of the computational complexities of different SR models for DSM in Table VII. DSMSR exhibits higher computational complexity compared with the other four models, primarily attributed to the incorporation of a multiscale attention feature extraction module tailored for remote sensing images. When comparing DSMSR with ESRGAN, the parameters of DSMSR are only 0.74 Mb more than ESRGAN, whereas there are significant accuracy improvements with 0.49-m decrease in $RMSE_E$ and 1.16° decrease in $MAE_S$. Therefore, the slight uptick in computational complexity for DSMSR remains relatively acceptable in light of these performance improvements.

### *F. Limitations*

While DSMSR demonstrates superior performance compared with the existing four SR models in downscaling DSM, it relies on high-quality remote sensing images to ensure the provision of accurate spatial details. Typically, cloud cover and building shadows are inevitable in remote sensing imagery and they may have a negative impact on the DSMSR outcomes. Meanwhile, the multiscale attention feature extraction module is designed to extract features from HR remote sensing imagery, which results in considerable computational complexity in DSMSR. Additionally, DSMSR mainly employs a convolutional architecture based on GAN and it is possible to leverage more sophisticated SR models (e.g., diffusion models) in the future.

## V. CONCLUSION

This study proposes a novel DSMSR model based on GANs to integrate HR remote sensing images. The DSMSR generator effectively learns multiscale features from both LR DSM and HR remote sensing imagery. The multiscale attention feature extraction module is designed and integrated to capture detailed spatial information of ground objects from remote sensing imagery for the combination with features from LR DSM. Furthermore, a slope loss is introduced to restore crucial topographic features. The experimental results on AW3D30 DSM datasets show that DSMSR outperforms the traditional interpolation method (BL) and four classic deep-learning-based models (SRCNN, SRGAN, ESRGAN, and MA-GAN) in terms of image quality, elevation accuracy, and terrain accuracy. When integrating HR remote sensing imagery, DSMSR is excellent at reconstructing ground objects, such as building contours and terrain lines across diverse landscapes, surpassing the performance of reference models. Additionally, the introduction of slope loss during training greatly improves the model's capacity to restore complex terrain features. The proposed DSMSR demonstrates its superior performance in DSMSR and its potential applications in geoscience and remote sensing.

## REFERENCES


[1] H. A. Amirkolaee, H. Arefi, M. Ahmadlou, and V. Raikwar, "DTM extraction from DSM using a multi-scale DTM fusion strategy based on deep learning," *Remote Sens. Environ.*, vol. 274, 2022, Art. no. 113014.

[2] J. M. Bodoque, E. Aroca-Jiménez, M. Á. Eguibar, and J. A. García, "Developing reliable urban flood hazard mapping from LiDAR data," *J. Hydrol.*, vol. 617, 2023, Art. no. 128975.

[3] L. Hawker et al., "A 30 m global map of elevation with forests and buildings removed," *Environ. Res. Lett.*, vol. 17, 2022, Art. no. 024016.

[4] H. Huang et al., "Estimating building height in China from ALOS AW3D30," *ISPRS J. Photogramm. Remote Sens.*, vol. 185, pp. 146–157, 2022.

[5] S. Karatsiolis, C. Padubidri, and A. Kamilaris, "Exploiting digital surface models for inferring super-resolution for remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 4414213, doi: 10.1109/TGRS.2022.3209340.

[6] C.-D. Wu, S.-C. C. Lung, and J.-F. Jan, "Development of a 3-D urbanization index using digital terrain models for surface urban heat island effects," *ISPRS J. Photogramm. Remote Sens.*, vol. 81, pp. 1–11, 2013.

[7] I. V. Florinsky, T. N. Skrypitsyna, and O. S. Luschikova, "Comparative accuracy of the AW3D30 DSM, ASTER GDEM, and SRTM1 DEM: A case study on the Zaoksky testing ground, Central European Russia," *Remote Sens. Lett.*, vol. 9, pp. 706–714, 2018.

[8] U. Alganci, B. Besol, and E. Sertel, "Accuracy assessment of different digital surface models," *ISPRS Int. J. Geo-Inf.*, vol. 7, 2018, Art. no. 114.

[9] B. Z. Demiray, M. Sit, and I. Demir, "D-SRGAN: DEM super-resolution with generative adversarial networks," *SN Comput. Sci.*, vol. 2, pp. 1–11, 2021.

[10] Q. M. Nguyen, T. T. H. Nguyen, P. H. La, H. G. Lewis, and P. M. Atkinson, "Downscaling gridded DEMs using the Hopfield neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4426–4437, Nov. 2019.

[11] D. Frantz et al., "National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112128.

[12] M. Li, E. Koks, H. Taubenböck, and J. van Vliet, "Continental-scale mapping and analysis of 3D building structure," *Remote Sens. Environ.*, vol. 245, 2020, Art. no. 111859.

[13] G. Priestnall, J. Jaafar, and A. Duncan, "Extracting urban features from LiDAR digital surface models," *Comput., Environ. Urban Syst.*, vol. 24, pp. 65–78, 2000.

[14] Y. Chen, R. Zhang, Y. Ge, Y. Jin, and Z. Xia, "Downscaling census data for gridded population mapping with geographically weighted area-to-point regression kriging," *IEEE Access*, vol. 7, pp. 149132–149141, 2019.

[15] B. Wu, C. Yang, Q. Wu, C. Wang, J. Wu, and B. Yu, "A building volume adjusted nighttime light index for characterizing the relationship between urban population and nighttime light intensity," *Comput., Environ. Urban Syst.*, vol. 99, 2023, Art. no. 101911.

[16] L. Feng, Y. Liu, Y. Zhou, and S. Yang, "A UAV-derived thermal infrared remote sensing three-temperature model and estimation of various vegetation evapotranspiration in urban micro-environments," *Urban Forestry Urban Greening*, vol. 69, 2022, Art. no. 127495.

[17] M. Aslani and S. Seipel, "Automatic identification of utilizable rooftop areas in digital surface models for photovoltaics potential assessment," *Appl. Energy*, vol. 306, 2022, Art. no. 118033.

[18] Y. Zhou, L. Feng, X. Zhang, Y. Wang, S. Wang, and T. Wu, "Spatiotemporal patterns of the COVID-19 control measures impact on industrial production in Wuhan using time-series earth observation data," *Sustain. Cities Soc.*, vol. 75, 2021, Art. no. 103388.
[19] N. Pradhan, Y. Tachikawa, and K. Takara, "A downscaling method of topographic index distribution for matching the scales of model application and parameter identification," *Hydrol. Processes*, vol. 20, pp. 1385–1405, 2006.
[20] C. Arrighi and L. Campo, "Effects of digital terrain model uncertainties on high-resolution urban flood damage assessment," *J. Flood Risk Manage.*, vol. 12, 2019, Art. no. e12530.
[21] R. Zhang, Y. Chen, X. Zhang, Q. Ma, and L. Ren, "Mapping homogeneous regions for flash floods using machine learning: A case study in Jiangxi province, China," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, 2022, Art. no. 102717.
[22] H. Li, J. Zhao, B. Yan, L. Yue, and L. Wang, "Global DEMs vary from one to another: An evaluation of newly released Copernicus, NASA and AW3D30 DEM on selected terrains of China using ICESat-2 altimetry data," *Int. J. Digit. Earth*, vol. 15, pp. 1149–1168, 2022.
[23] T. G. Farr et al., "The shuttle radar topography mission," *Rev. Geophys.*, vol. 45, 2007, Art. no. RG2004.
[24] Z. Chen, X. Wang, and Z. Xu, "Convolutional neural network based DEM super resolution," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 247–250, 2016.
[25] Y. Jiang, L. Xiong, X. Huang, S. Li, and W. Shen, "Super-resolution for terrain modeling using deep learning in high mountain Asia," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, 2023, Art. no. 103296.
[26] Z. Xu, Z. Chen, W. Yi, Q. Gui, W. Hou, and M. Ding, "Deep gradient prior network for DEM super-resolution: Transfer learning from image to DEM," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 80–90, 2019.
[27] Y. Zhang, W. Yu, and D. Zhu, "Terrain feature-aware deep learning network for digital elevation model superresolution," *ISPRS J. Photogramm. Remote Sens.*, vol. 189, pp. 143–162, 2022.
[28] Y. Zhang, Z. Zheng, Y. Luo, Yanfeng Zhang, J. Wu, and Z. Peng, "A CNN-based subpixel level DSM generation approach via single image super-resolution," *Photogramm. Eng. Remote Sens.*, vol. 85, pp. 765–775, 2019.
[29] A. Wang, D. Xue, H. Wu, and Y. Gu, "Efficient convolutional neural architecture search for LiDAR DSM classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 5703317, doi: 10.1109/TGRS.2022.3171520.
[30] O. Argudo, A. Chica, and C. Andujar, "Terrain super-resolution through aerial imagery and fully convolutional networks," in *Proc. Comput. Graph. Forum*, 2018, vol. 37, pp. 101–110.
[31] D. He, Q. Shi, X. Liu, Y. Zhong, and X. Zhang, "Deep subpixel mapping based on semantic information modulated network for urban land use mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10628–10646, Dec. 2021.
[32] Y. Chen, Y. Ge, and Y. Jia, "Integrating object boundary in super-resolution land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 219–230, Jan. 2017.
[33] Y. Chen, Y. Ge, G. B. Heuvelink, R. An, and Y. Chen, "Object-based superresolution land cover mapping from remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 328–340, Jan. 2018.
[34] Y. Chen, K. Shi, Y. Ge, and Y. Zhou, "Spatiotemporal remote sensing image fusion using multiscale two-stream convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4402112, doi: 10.1109/TGRS.2021.3069116.
[35] F. Ling et al., "Superresolution land cover mapping with multiscale information by fusing local smoothness prior and downscaled coarse fractions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5677–5692, Sep. 2014.
[36] Q. Wang, W. Shi, and P. M. Atkinson, "Spatiotemporal subpixel mapping of time-series images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5397–5411, Sep. 2016.
[37] Y. Zhang, P. M. Atkinson, X. Li, F. Ling, Q. Wang, and Y. Du, "Learning-based spatia-temporal superresolution mapping of forest cover with MODIS images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 600–614, Jan. 2017.
[38] X. Li et al., "SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111537.
[39] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.
[40] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.
[41] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 63–79.
[42] P. Wang, B. Bayram, and E. Sertel, "A comprehensive review on deep learning based remote sensing image super-resolution methods," *Earth-Sci. Rev.*, vol. 232, 2022, Art. no. 104110.
[43] Y. Chen and Y. Ge, "Spatiotemporal image fusion using multiscale attention-aware two-stream convolutional neural networks," *Sci. Remote Sens.*, vol. 6, 2022, Art. no. 100062.
[44] Y. Chen, J. Zhou, Y. Ge, and J. Dong, "Uncovering the rapid expansion of photovoltaic power plants in China from 2010 to 2022 using satellite data and deep learning," *Remote Sens. Environ.*, vol. 305, 2024, Art. no. 114100.
[45] Y. Xu, W. Luo, A. Hu, Z. Xie, X. Xie, and L. Tao, "TE-SAGAN: An improved generative adversarial network for remote sensing super-resolution images," *Remote Sens.*, vol. 14, 2022, Art. no. 2425.
[46] S. Jia, Z. Wang, Q. Li, X. Jia, and M. Xu, "Multiattention generative adversarial network for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 5624715.
[47] J. Hu, L. Shen, G. Sun, E. Wu, and S. Albanie, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
[48] L. Jiang, Y. Hu, X. Xia, Q. Liang, A. Soltoggio, and S. R. Kabir, "A multi-scale mapping approach based on a deep learning CNN model for reconstructing high-resolution urban DEMs," *Water*, vol. 12, 2020, Art. no. 1369.
[49] X. Deng, W. Hua, X. Liu, S. Chen, W. Zhang, and J. Duan, "D-SRCAGAN: DEM super-resolution generative adversarial network," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2022.3224296.
[50] S. Karatsiolis, A. Kamilaris, and I. Cole, "IMG2nDSM: Height estimation from single airborne RGB images with deep learning," *Remote Sens.*, vol. 13, 2021, Art. no. 2417.
[51] S. Li, G. Hu, X. Cheng, L. Xiong, G. Tang, and J. Strobl, "Integrating topographic knowledge into deep learning for the void-filling of digital elevation models," *Remote Sens. Environ.*, vol. 269, 2022, Art. no. 112818.
[52] A. Jolicoeur-Martineau, "The relativistic discriminator: A key element missing from standard GAN," 2018, *arXiv:1807.00734*.
[53] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, and A. Plaza, "Remote sensing image superresolution using deep residual channel attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9277–9289, Nov. 2019.
[54] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. 14th Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.
[55] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[56] Y. Zhang and W. Yu, "Comparison of DEM super-resolution methods based on interpolation and neural networks," *Sensors*, vol. 22, 2022, Art. no. 745.

**Guihou Sun** is currently working toward the B.S. degree in geographical information science with Hohai University, Nanjing, China.

His research interests include remote sensing image processing and super-resolution.

**Yuehong Chen** received the B.S. degree in geographical information system from Hohai University, Nanjing, China, in 2010, and the M.Sc. and Ph.D. degrees in cartography and geographical information system from the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, University of Chinese Academy of Sciences, Beijing, China, in 2013 and 2016, respectively.

He is an Associate Professor with the College of Geography and Remote Sensing, Hohai University. His current research interests include remote sensing image processing and geospatial Big Data analysis.

**Jiamei Huang** is currently working toward the B.S. degree in geographical information science with Hohai University, Nanjing, China.

Her research interest focuses on geospatial big data analysis.

**Qiang Ma** received the B.S. degree in geographical information science from Hohai University, Nanjing, China, in 2010, and the M.Sc. and Ph.D. degrees in hydroinformatics from the Université Nice Côte D'azur, Nice, France, in 2014 and 2018, respectively.

He is the Senior Engineer with the China Institute of Water Resources and Hydropower Research, Beijing, China. His current research interests include hydroinformatics and flood monitoring by remote sensing.

**Yong Ge** (Senior Member, IEEE) received the Ph.D. degree in cartography and geographical information system from the Chinese Academy of Sciences (CAS), Beijing, China, in 2001.

She is a Professor with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS. She has directed research in more than ten national projects. She is the author or coauthor of more than 80 scientific papers published in refereed journals, one book, and six chapters in books; she is an editor of one book, and she holds three granted patents in improving the accuracy of information extraction from remotely sensed imagery. Her research activity focuses on spatial data analysis and data quality assessment.

Dr. Ge has been involved in the organization of several international conferences and workshops. She is a member of the Theory and Methodology Committee of Cartography and Geographic Information Society, International Association of Mathematical Geosciences, and the Editorial Board of Spatial Statistics (Elsevier).