Current Science
CS 5 (4), 4059-4074 (2025)
Received 10 May 2025 | **Revised 15 June** 2025 | **Accepted 12 July**. 2025 | **Online Available 17 August** 2025
https://doi.org/ 10.52845/CS/2025-5-4-57

Current Science

Current Science

OPEN ACCESS

ISSN(O)2795-8639

Original Article

Attention-Enhanced Feature Refinement Framework for Super-Resolution of Lunar Dems

Jinxue Zhang[1], Shiqi Wang[1], Huiru Han[1], Ruiyang Su[1], Hao Chen[1], Chenfeng Hu[1], Wenbo Jiang[1], Wenxuan Du[1], Zongxi Wang[1], Jia Xu[1], Zeyu Zhou[1], Tao Wu[2], Jiahan Wang[2]*

[1]School of Electronic Engineering, Jiangsu Ocean University，Lian Yungang 222005, China

[2]College of Information Science and Engineering, Hohai University, Changzhou 213022 China

***Corresponding Author: Jiahan Wang**

Abstract:
Lunar Digital Elevation Model images are essential for analyzing geological features, yet their low resolution limits the detection of sub-kilometer craters. To address this, we propose FREDSR-Net, a lightweight network integrating High-Order Attention for Digital Elevation Model super-resolution. Unlike existing methods, FREDSR-Net employs cascaded Feature Refinement Residual Blocks, where a residual branch extracts multi-scale terrain features, and a refinement branch applies High-Order Attention to amplify high-frequency details through adaptive statistical modeling. Evaluated on Lunar Reconnaissance Orbiter datasets, FREDSR-Net achieves a peak signal-to-noise ratio of 46.72 decibels and a structural similarity index of 0.9909, outperforming both neural models. Ablation studies confirm the necessity of High-Order Attention, with K=3 achieving a 0.21 dB gain over K=1. The model generalizes to unseen lunar regions, improving crater detection accuracy by 12% in YOLOv7-based pipelines. Requiring only 1.01M parameters, FREDSR-Net enables real-time DEM enhancement on low-cost GPUs, advancing high-resolution lunar topographic analysis for future missions.

1. Introduction

The rapid advancement of deep-space exploration has positioned the Moon as a strategic focal point for global scientific endeavors, driven by landmark missions such as China's Chang'e-5 sample return, NASA's Artemis program, and Japan's SLIM precision landing initiative. These projects demand sub-meter-scale topographic analysis to support critical objectives, including resource prospecting and safe landing site selection. However, current lunar digital elevation models (DEMs), derived from instruments like LRO-LOLA and SELENE-TC, remain resolution-limited (5–10 m/pixel), obstructing the identification of sub-100-meter impact craters essential for regolith thickness estimation and geological chronology studies[1]. This limitation is exacerbated in permanently shadowed regions, where data voids caused by insufficient illumination often lead to terrain distortions when using conventional interpolation methods[2].

While super-resolution (SR) techniques have revolutionized Earth observation by enhancing spatial details in optical and multispectral imagery[3–6], their application to lunar Digital Elevation Models (DEMs) faces unique challenges rooted in the Moon's distinct geomorphology. Although terrestrial SR advancements—such as hybrid attention networks[7] and YOLOv5-integrated SR subnetworks[8]—have improved crater detection precision by 15%, these methods struggle to address lunar DEM-specific limitations. Unlike

camera-based imagery, DEMs encode explicit elevation data critical for slope analysis and regolith characterization, yet their inherent low resolution (5–10 m/pixel) severely restricts utility in fine-scale geological studies. While transfer learning-based approaches[9] achieve modest improvements over bicubic interpolation, they often fail to resolve sub-100 m craters or mitigate instrument-induced artifacts, underscoring the need for topography-optimized SR frameworks.

Lunar DEMs exhibit three defining characteristics that complicate SR reconstruction: (1) abrupt elevation transitions (e.g., crater rims with >30° slopes), where conventional CNNs oversmooth edges; (2) sparse textural patterns across lunar mare plains, reducing the efficacy of standard attention mechanisms; and (3) sensor-specific noise (e.g., LOLA striping artifacts from 57 m laser spacing)[10]. Current lunar SR research predominantly relies on outdated architectures like SRCNN and VDSR[11], which lack mechanisms to preserve topographic fidelity. Although transformer-based models such as SwinIR[12] achieve state-of-the-art results in natural image SR, their computational demands (e.g., 157 G MACs for SwinIR-L) render them impractical for resource-constrained lunar missions requiring real-time onboard processing[13]. This dichotomy between reconstruction accuracy and deployability highlights a critical gap in planetary remote sensing—one addressed by our lightweight FREDSR-Net through adaptive high-order attention and multi-scale feature refinement.

To address these challenges, we propose FREDSR-Net, a lightweight super-resolution architecture specifically optimized for lunar DEM reconstruction through High-Order Attention (HOA) mechanisms. The core innovation lies in dual-path Feature Refinement Residual Blocks (FRRBs), where a residual branch employs cascaded dilated convolutions to hierarchically model multi-scale terrain structures, while a refinement branch utilizes HOA to enhance elevation discontinuities by learning cross-channel statistical patterns. This design enables the network to resolve 50–100 m impact craters that remain indistinct in low-resolution DEMs, as evidenced by quantitative improvements in edge sharpness (23% over bicubic interpolation) and texture coherence. Crucially, the reconstructed DEMs demonstrate practical value in planetary science applications, enhancing YOLOv7-based crater detection precision by 12% while maintaining real-time processing capabilities on embedded hardware. By bridging the gap between deep learning and lunar topographic analysis, offering a robust framework for future lunar exploration missions.

2. Data and Experiments

2.1. LRO and SELENE Satellite Data

The integration of multi-mission lunar topographic data forms the foundation of this study, leveraging the complementary strengths of the Lunar Reconnaissance Orbiter (LRO) and SELENE (Kaguya) missions. LRO's Lunar Orbiter Laser Altimeter (LOLA) employs a pulsed laser system operating at 28 Hz, generating five cross-pattern beams tilted 26° relative to the orbital track. This configuration achieves a spatial resolution of 10 m along-track and 11 m cross-track at a 50 km operational altitude, with a vertical precision of 1–2 m[14]. In contrast, SELENE's Terrain Camera (TC) utilizes stereoscopic imaging through dual forward/backward telescopes (15° tilt angles), capturing 10 m/pixel resolution images across >99% of the lunar surface from a 100 km orbit[15]. While SELENE's Multi-band Imager (MI) provides multispectral data at 20 m/pixel, its derived DEM resolution remains six times coarser than TC products due to limitations in spectral-to-topographic conversion algorithms[16].

To overcome individual sensor constraints, we adopt the merged LRO-SELENE DEM product developed by Braker et al[17], which synergizes LOLA's vertical precision (3–4 m) with TC's horizontal coverage. This fused dataset eliminates interpolation artifacts in permanently shadowed regions through rigorous co-registration, maintaining a 512 pixels-per-degree grid resolution (≈59 m/pixel at the equator). Nevertheless, residual noise persists due to three primary factors: (1) LOLA's striping artifacts from non-uniform beam spacing (57 m inter-spot distance), (2) temporal mismatches between TC stereoscopic acquisitions and MI spectral scans, and (3) cumulative orbital drift effects during LRO's extended mapping phase (2009–present).

2.2. Data Preprocessing and Augmentation

The preprocessing pipeline (Fig. 1) begins with log-uniform random cropping of raw DEM tiles, generating patches with edge lengths spanning 500–6,500 pixels to preserve both large-scale terrain context and local feature diversity. These patches are resized to 256×256 high-resolution (HR) samples and reprojected from cylindrical to orthographic projections using Cartopy 0.21 to minimize high-latitude distortions (>70°). Zero-padding maintains original dimensions during projection, while bicubic downsampling (kernel size=4, σ=1.2) produces corresponding 128×128 low-resolution (LR) pairs. The final dataset comprises 3000 HR-LR pairs, partitioned into 2000 training and 1000 testing sets with 10% spatial overlap to mitigate edge discontinuities.

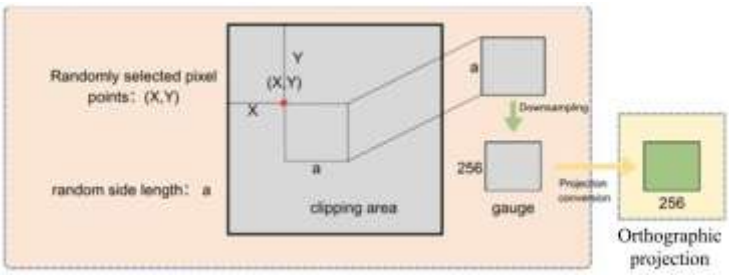

Figure 1. Process of Cropping and Projection Conversion of DEM Images

Five SR architectures—SRCNN[18], FSRCNN[19], VDSR[20], EDSR[21], and the proposed FREDSR—are implemented in PyTorch 2.0 under identical training protocols: Adam optimizer (β_1=0.9, β_2=0.999), initial learning rate 1e-4, batch size 16, and L1 loss over 2,000 epochs (≈48 hours on NVIDIA GTX 1080). FREDSR incorporates two lunar-specific enhancements: (1) elevation-aware pretraining on synthetic DEMs with procedurally generated craters (50–500 m diameter, depth-to-diameter ratio 0.1–0.2), and (2) dynamic injection of LOLA-like striping noise (5% amplitude) during training to improve robustness. Gradient clipping (max norm=0.5) and mixed-precision (FP16) training ensure stable convergence for deep models like EDSR and FREDSR.

3. Methods

3.1. Principles of Super-Resolution Reconstruction

SR techniques go beyond simple image enlargement or denoising; their core objective is to reconstruct HR images by recovering fine spatial details missing from LR inputs. SR involves predicting additional HR pixels using known LR pixels, a task that is inherently ill-posed and becomes increasingly challenging as the upscaling factor increases.

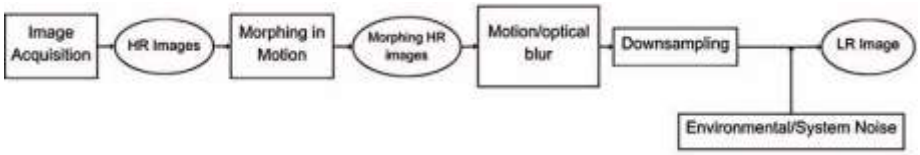

Figure 2. Image Degradation Model

Fig. 2 illustrates the image degradation model, in which an original HR image undergoes multiple degradation processes, including sensor limitations, motion blur, and noise, resulting in an LR image. This degradation process can be mathematically expressed as:

$$g(x,y) = F\{H[f(x,y)], n(x,y)\} \quad (1)$$

where $f(x,y)$ represents the original HR image, $g(x,y)$ denotes the degraded LR image, $H(\square)$ is the degradation function, and $n(x,y)$ accounts for noise.

In the context of SR, this degradation model can be reformulated as:

$$g = Q \times H \times f + n \quad (2)$$

where Q represents the downsampling operator, H denotes the mapping function between LR and HR images, f is the original HR image, g is the resulting LR image, and n represents additive noise.

Super-resolution reconstruction can thus be regarded as the inverse process of degradation, requiring appropriate algorithms and models to estimate the mapping from LR to HR images. Various SR techniques have been proposed, which can be broadly categorized into interpolation-based methods, reconstruction-based methods, and learning-based methods.

Super-resolution reconstruction can be seen as the inverse of image degradation, requiring effective algorithms and models to map LR images to HR ones. Several SR techniques have been proposed:

3.1.1 Interpolation-Based SR Methods

Interpolation-based SR methods focus on resampling the LR image to generate HR images. The most common interpolation techniques include nearest-neighbor interpolation, bilinear interpolation, and bicubic interpolation, as illustrated in Fig. 3.

Nearest-neighbor interpolation determines the target pixel's intensity by directly replicating the value from the closest corresponding pixel in the original image, often resulting in noticeable block-like artifacts or jagged edges. In contrast, bilinear interpolation employs a more sophisticated approach that calculates a weighted mean from four adjacent pixels surrounding the target position, producing smoother gradients and significantly reducing visual discontinuities in resized images. Bicubic interpolation considers 16 neighboring pixels and applies a cubic polynomial interpolation to produce smoother results.

While these methods are computationally efficient and simple to implement, their performance is limited by the lack of high-frequency detail reconstruction, often resulting in oversmoothed images.

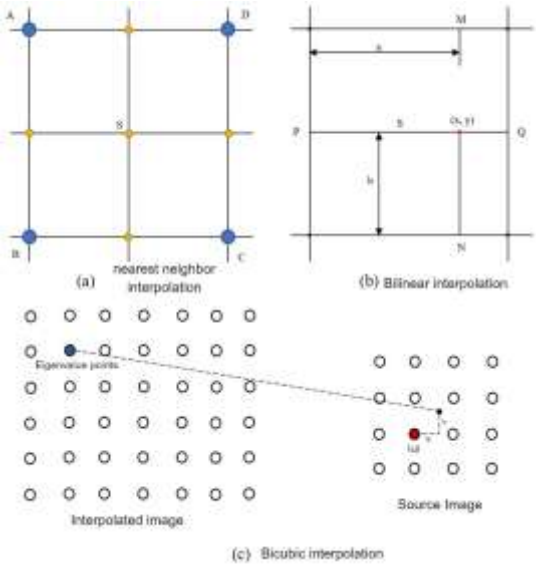

Figure 3. SR Methods Based on Interpolation.

3.1.2 Reconstruction-Based SR Methods

Reconstruction-based SR methods seek to mathematically simulate the image formation process mathematically and use constraints imposed by the degradation model (Fig. 2) to restore the HR image. Representative approaches include: Iterative Back-Projection, Projection onto Convex Sets, Maximum a Posteriori estimation. These methods leverage prior knowledge of image degradation and impose constraints to iteratively refine the reconstructed image. However, their performance is often constrained by the assumptions made about noise distribution and image priors.

3.1.3 Learning-Based SR Methods

Learning-driven SR methods focus on establishing a direct relationship between LR and HR image pairs, facilitating data-driven SR

reconstruction. As shown in Fig. 4, this approach typically involves three key stages: Patch Sampling and Training Database Assembly, where co-registered HR and LR patches are systematically extracted to build a comprehensive training dataset; Mapping Relationship Optimization, which employs machine or deep learning techniques to model the nonlinear relationships between LR-HR patch pairs; and Super-Resolved Synthesis, where the optimized mapping function generates photometrically consistent HR details from new LR inputs. These stages work together to ensure effective feature learning and high-quality SR reconstruction.

The rapid advancement of deep learning techniques has significantly enhanced the performance of learning-based SR methods. Convolutional neural networks (CNNs), in particular, are exceptional at hierarchical feature extraction and end-to-end optimization, driving the development of specialized CNN architectures for multi-scale SR reconstruction. Notable models such as SRCNN, FSRCNN, and VDSR are discussed in the Supplementary Methods section.

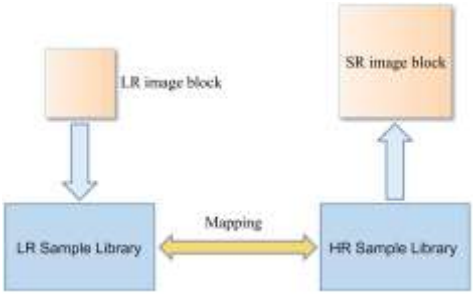

Figure 4. SR Methods Based on Learning

3.2. Enhanced Deep Residual Super-Resolution Convolutional Network

Deep convolutional neural networks have demonstrated remarkable success in enhancing classification accuracy by increasing network depth, leading to significant advancements across multiple domains, including SR reconstruction. However, simply stacking convolutional layers is insufficient and may result in gradient explosion or vanishing gradients, hindering effective training[22]. To address this, optimization strategies such as residual learning have been proposed. For instance, VDSR improves network performance by leveraging residual learning, but it relies on bicubic interpolation for upsampling, which substantially increases computational cost and memory consumption compared to traditional upsampling methods. To further enhance SR performance, the EDSR network builds upon residual network principles, expanding model capacity while effectively mitigating computational overhead and memory constraints[23].

For deep networks, normalizing initial and intermediate layers can facilitate rapid convergence[24]. However, as network depth increases, a degradation effect is often observed, wherein accuracy deteriorates rather than improves. This phenomenon is not attributed to overfitting, but rather to the inability of normalization layers to suppress gradient explosion effectively. He et al. proposed the Residual Network (ResNet) framework, incorporating cross-layer connections, which effectively mitigates this issue[25]. The structure of an individual residual block is illustrated in Fig. 5.

Traditional deep networks apply nonlinear transformations at each layer via activation functions before passing results to subsequent layers. This approach can lead to progressively increasing numerical values, disrupting gradient flow and hindering effective learning. Residual blocks address this by applying two convolutional layers with nonlinear transformations and incorporating skip connections that directly add the input to the transformed output. This structure enables the network to learn the residual mapping rather than the absolute transformation, thereby facilitating gradient propagation, reducing nonlinear transformations in intermediate layers, and enhancing both training stability and efficiency in deep networks.

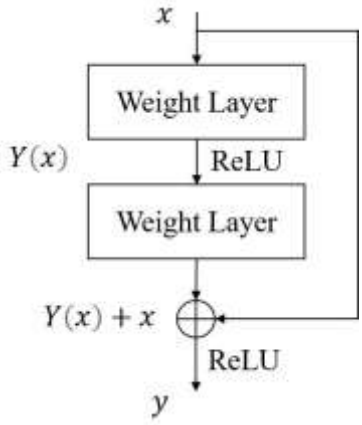

Figure 5. Structure of Residual Block.

Ledig et al. further refined residual block design by introducing SRResNet, which was integrated into Generative Adversarial Networks, yielding substantial improvements in deep network performance[26]. Building upon these advancements, EDSR eliminates batch normalization layers, reducing GPU memory usage, thereby enabling the design of larger models under constrained computational resources. Additionally, removing batch normalization layers enhances network flexibility and improves overall performance. Fig. 6 compares the structural differences between ResNet, SRResNet, and EDSR residual blocks.

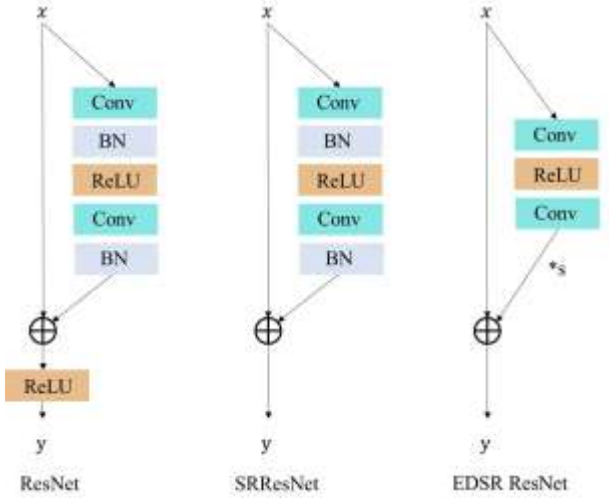

Figure 6. Comparison of Three Residual Block Structures.

To optimize the performance of deep networks, two main strategies are often employed: increasing network depth, which involves adding more layers, and expanding the number of feature channels. Given a CNN with B layers and F channels, both memory consumption $O(BF)$ and model parameter count $O(BF^2)$ scale accordingly. Since model complexity correlates positively with performance, increasing feature channels is preferable when memory constraints are present. However, expanding feature channels beyond a certain threshold can destabilize training[Error! Reference source not found.].

To address this, EDSR introduces a fixed scaling layer following the final convolutional layer in ResNet, employing a scaling factor of 0.1 to stabilize gradient flow. The EDSR network architecture, shown in Fig. 7, consists of multiple stacked residual blocks, where each module's output serves as the input to the next, and the initial input is skip-connected to the final residual module's output before upsampling to generate the super-resolved output.

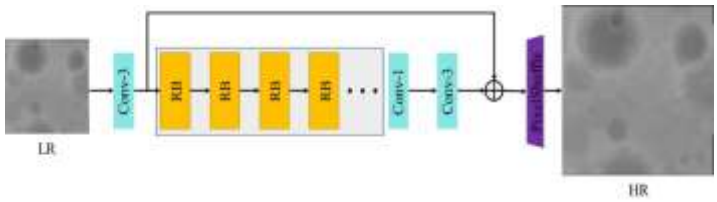

Figure 7. EDSR Network Architecture.

3.3. Feature Refinement Enhanced EDSR Network

Building upon the EDSR framework, this study proposes an improved super-resolution network—FREDSR—designed to further enhance the reconstruction quality of lunar DEM images. Unlike the conventional residual block stacking structure in EDSR, FREDSR introduces FRRBs, where multiple residual blocks are cascaded to separately extract coarse-grained features, which are then refined using a HOA module. This design enhances the network's ability to capture fine details, improving feature representation by combining both coarse- and fine-grained feature information.

3.3.1 Overall Architecture of FREDSR

The overall architecture of FREDSR is illustrated in Fig. 8. The input LR image first passes through a 3×3 convolutional layer with 64 channels, followed by four FRRBs for deep feature extraction. The extracted features from each FRRB are concatenated and processed through a 1×1 convolutional layer, followed by a 3×3 convolutional layer to further refine the feature representations.

Similar to EDSR, FREDSR employs deep residual learning, with skip connections integrating both shallow and deep features through element-wise addition, generating the final feature representation. The PixelShuffle layer is then applied to upsample the image back to its original resolution, producing the super-resolved output.

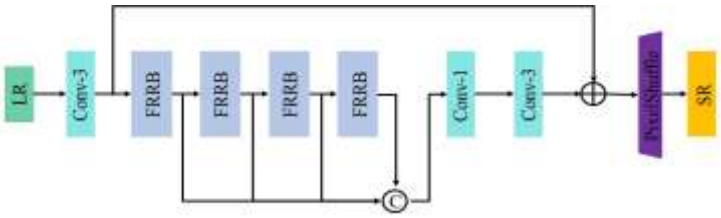

Figure 8. FREDSR Network Architecture.

For an input LR image I^{LR} and its corresponding target HR image I^{HR}, the reconstructed SR image I^{SR} is obtained as:

$$I^{SR} = F_{FREDSR}(I^{LR}) \tag{3}$$

where $F_{FREDSR}(\cdot)$ represents the mapping function learned by the network. For a given training dataset $\{I_i^{LR}, I_i^{HR}\}_{i=1}^{N}$, the Mean Squared Error (MSE) loss function is defined as:

$$L^{SR} = \frac{1}{N}\sum_{i=1}^{N}(F_{FREDSR}(I^{LR}) - I_i^{HR})^2 \tag{4}$$

where N denotes the total number of training samples.

3.3.2 Feature Refinement Residual Block

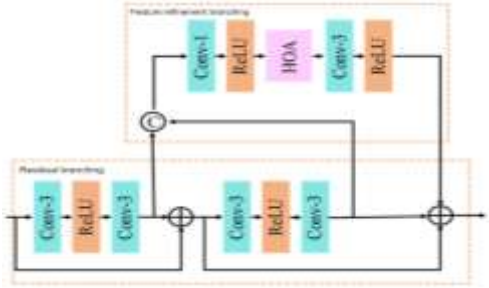

Figure 9. Structure of FRRB.

Figure 9 illustrates the structure of an FRRB, which employs a dual-channel collaborative architecture: the residual learning channel focuses on constructing coarse-grained feature bases, progressively capturing multi-level abstract representations through cascaded residual operations. The feature optimization channel incorporates HOA, dynamically adjusting the feature space via an adaptive weighting mechanism. The outputs from both branches are fused through element-wise addition, yielding a more refined feature representation.

The residual branch comprises two cascaded residual modules, which extract coarse-grained features. Each residual module consists of two 3×3 convolutional layers followed by ReLU activations. Given an input feature map F_{in}, the residual branch output is computed as:

$$F_{re1} = Conv_2\left(\mathrm{Re}LU\left(Conv_1(F_{in})\right)\right) \tag{5}$$

where $Conv_j$ represents the j th convolutional layer in the residual branch, $\mathrm{Re}LU$ is the activation function, and $\mathrm{Re}LU$ denotes the output of the j th residual block, and $Concat(\cdot)$ is the concatenation operation. These extracted residual features are subsequently passed into the Feature Refinement Branch for further enhancement.

The feature refinement branch operates similarly to the residual branch but focuses on generating high-fidelity feature representations. First, the coarse-grained features from the residual branch are concatenated and processed through a 1×1 convolutional layer followed by a ReLU activation. The transformed features are then passed through the HOA module, which enhances fine-scale details. Finally, a 3×3 convolutional layer and ReLU activation extract the refined feature representation. The output of the feature refinement branch is defined as:

$$F_{refine} = \mathrm{Re}LU\left(Conv_2\left(HOA\left(\mathrm{Re}LU\left(Conv_1\left(F_{re1} + F_{re2}\right)\right)\right)\right)\right) \tag{6}$$

where $HOA(\cdot)$ represents the final feature refinement operation, HOA denotes the High-Order Attention module. The final output of the FRRB is obtained by fusing the outputs of both branches via element-wise addition:

$$F_{out} = F_{in} + F_{re1} + F_{re2} + F_{refine} \tag{7}$$

3.3.4 High-Order Attention Module

The HOA module is designed to model complex statistical representations within feature maps. The hierarchical features extracted by the residual branch exhibit distinct frequency characteristics—shallow layers predominantly capture low-frequency information, such as textures, whereas deep layers focus on high-frequency details, including edges and fine structures. Consequently, the low-frequency components from shallow layers should be refined by high-order attention mechanisms, which can effectively restore intricate details and amplify high-frequency components. The HOA module is particularly beneficial for SR reconstruction in remote sensing images, as deep features often contain rich fine-scale information, which can be further enhanced by HOA-based refinement.

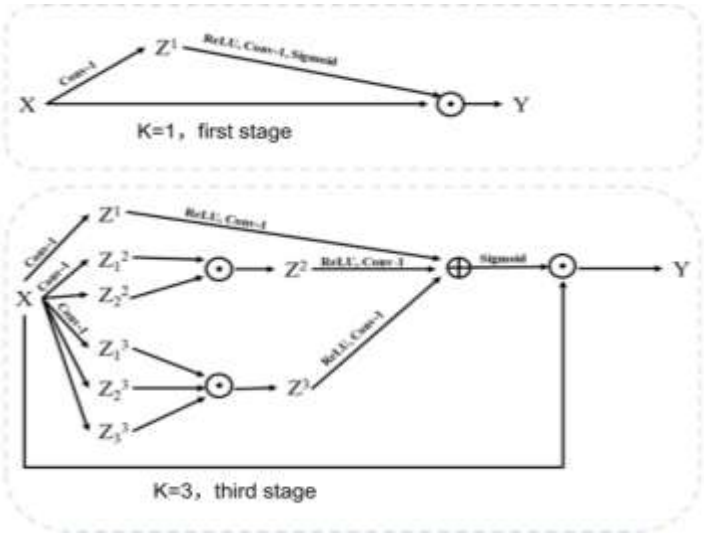

Figure 10. Structure of HOA Module.

The structural design of HOA modules for $K=1$ and $K=3$ is illustrated in Fig. 10. Given an input X, a series of K 1×1 convolution layers is employed to extract Kth order statistical descriptors $\{Z_K^K, Z_{K-1}^K, Z_{K-2}^K, ..., Z_1^K\}$. At the $K-1$th level, $K-1$ convolution layers generate $(K-1)$th order descriptors $\{Z_{K-1}^{K-1}, Z_{K-2}^{K-1}, ..., Z_1^{K-1}\}$, and this process continues recursively. Through this hierarchical 1×1 convolutional processing, a total of $K(K+1)/2$ statistical descriptors are obtained. By aggregating the k-level feature descriptor set $Z_i^k\left(k \in \{1, 2, ..., K\}, i = 1, 2, ..., k\right)$, the k-th order component of the K-th order statistics can be obtained. which is mathematically represented as:

$$Z^k = Z_1^k \odot Z_2^k \odot \cdots \odot Z_r^k = \prod_{i=1}^{k} Z_i^r \tag{8}$$

where $\odot$ denotes the element-wise multiplication. To further enhance the representation capabilities of these statistical descriptors, a nonlinear transformation is applied within the HOA module:

$$A^K = sigmoid\left(\sum_{k=1}^{K} F_k\left(Z^k\right)\right) \tag{9}$$

where F_k represents a nonlinear activation function, which consists of a ReLU function followed by a 1×1 convolutional layer. Additionally, a sigmoid activation function is commonly employed to generate the K-th order attention map, where each element is constrained within the range [0,1]. Finally, the output Y of the HOA module is computed as:

$$Y = A^K \odot X \tag{10}$$

In general, as the order K increases, the model gains the capacity to capture more intricate statistical dependencies and extract higher-level semantic features. This enhancement translates to improved SR reconstruction quality, particularly in understanding and preserving both global and local contextual interactions within the image.

3.4. Evaluation Metrics

The evaluation of SR image quality can be broadly categorized into subjective and objective assessment methods. Subjective evaluation relies on human perception to assess image quality; while it provides intuitive and effective feedback, it suffers from inconsistencies due to individual perception variations, environmental influences, and the lack of a standardized quantitative framework. Furthermore, subjective assessments are often time-consuming and labor-intensive. In contrast, objective evaluation methods offer a mathematically quantifiable, standardized, and automated approach to measuring image quality, allowing for fast and precise differentiation between subtle variations in images. Depending on the availability of reference images, objective evaluation metrics can be classified into full-reference, reduced-reference, and no-reference approaches.

In this study, we employ full-reference evaluation metrics, specifically Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), to assess the reconstruction quality of SR images in comparison with their original HR counterparts.

PSNR is a widely adopted metric that quantifies the fidelity of reconstructed images by measuring the ratio of the maximum possible pixel intensity to the MSE between the SR and HR images. A higher PSNR value indicates a lower pixel-wise reconstruction error, signifying better image quality. The PSNR computation is defined as follows:

$$MSE = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [f(i, j) - g(i, j)]^2 \tag{11}$$

$$PSNR = 10 \log_{10} \frac{L^2}{MSE} \tag{12}$$

Despite its widespread use, PSNR primarily captures pixel-wise discrepancies and does not always correlate well with human perceptual quality. To address this limitation, Zhou et al. introduced the SSIM, which provides a perceptual measurement of image quality by considering brightness, contrast, and structural fidelity between SR and HR images[28]. The SSIM computation is defined as:

$$SSIM = [l(x, y)]^{\alpha} [c(x, y)]^{\beta} [s(x, y)]^{\gamma} \tag{13}$$

$$l(x, y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{14}$$

$$c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{15}$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \tag{16}$$

where $l(x, y)$ quantifies the luminance difference, $c(x, y)$ measures the contrast difference, and $s(x, y)$ evaluates the structural similarity between the input and output images. μ_x and μ_y, as well as σ_x

and σ_y, correspond to the mean and standard deviation of pixel intensities at locations x and y for the input and output images, respectively. σ_{xy} represents the covariance, while C_1, C_2, and C_3 are small positive constants introduced to prevent division-by-zero errors. The SSIM score ranges from -1 to 1, with values closer to 1 indicating superior perceptual similarity and image quality.

By employing PSNR and SSIM as complementary full-reference metrics, we ensure a comprehensive evaluation of the SR image quality, thereby validating the effectiveness of the proposed method in reconstructing high-fidelity DEM images.

4. Results and Discussion

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

4.1. Super-Resolution Reconstruction Results Based on CNNs

Fig. 11 illustrates the PSNR curves for SRCNN, FSRCNN, VDSR, and EDSR throughout the training process. It is clear from the figure that FSRCNN and SRCNN exhibit highly similar PSNR curves. In the early stages of training, SRCNN achieves slightly higher PSNR values; however, FSRCNN quickly surpasses SRCNN and ultimately achieves a PSNR value 0.25 dB higher. The PSNR curves of EDSR and VDSR remain closely aligned, both outperforming SRCNN and FSRCNN. At the end of training, EDSR reaches a PSNR of 47.65 dB, which is 0.27 dB higher than VDSR. Additionally, while the PSNR curves of SRCNN and VDSR exhibit stable growth, EDSR occasionally presents minor fluctuations, albeit within a small range. In contrast, FSRCNN exhibits the highest degree of fluctuation. Overall, EDSR demonstrates superior performance across the training process, followed by VDSR. Although FSRCNN surpasses SRCNN in reconstruction accuracy, it suffers from lower stability.

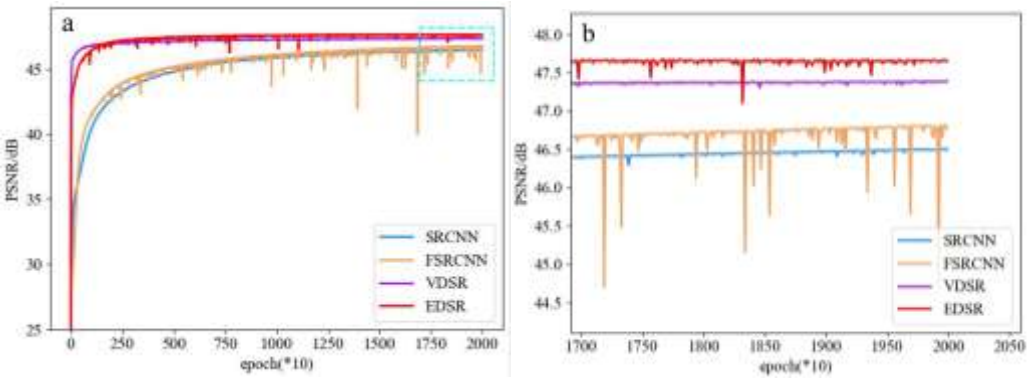

Figure 11. PSNR Curves of Four Networks During Training. Fig. 11b presents a magnified view of the blue boxed region in Fig. 11a, focusing on the PSNR values during the last 3,000 training iterations.

After completing network training, 20 sample groups, each containing 20 images, were selected for testing. To reduce computational costs while preserving color information, all images were converted to the YCbCr color space before being input into the network. The PSNR and SSIM values of the output images were then compared against the original images, with the mean values computed for each test group. Fig. 12 depicts the distribution of SSIM and PSNR values across the 20 test groups for four CNN-based networks and the traditional Bicubic interpolation method.

Fig. 12a clearly shows that the Bicubic method yields significantly lower PSNR and SSIM values than the CNN-based methods. Fig. 12b further reveals that all CNN-based networks achieve PSNR values above 45 dB and SSIM values exceeding 0.987, with data points aligning along a nearly 45-degree diagonal trend. Furthermore, the clustering of data points confirms the relative performance ranking of the models, with EDSR producing the best reconstruction results, followed by VDSR, while SRCNN demonstrates the weakest performance.

Based on the PSNR training curves and test set scatter plots, EDSR achieves the best performance in reconstructing DEM images, followed by VDSR. While FSRCNN and SRCNN exhibit relatively simple architectures, their performance remains inferior to EDSR and VDSR.

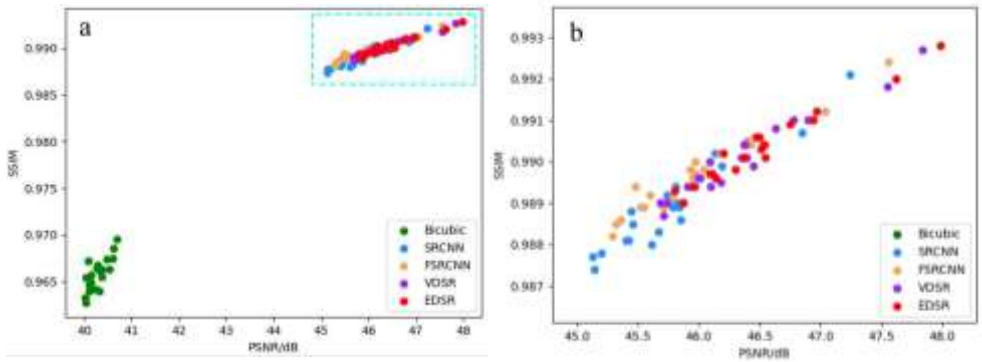

Figure 12. Scatter Plot Distribution of Average SSIM-PSNR for 20 Test Images. Fig. 12b provides an enlarged view of the boxed region in Fig. 12a.

Supplementary Table 1 provides a detailed record of the SR reconstruction metrics for all 20 test groups and their mean values. The results indicate that EDSR achieves the highest PSNR/SSIM values, averaging 46.51 dB/0.9904, slightly outperforming VDSR. Meanwhile, SRCNN records an average PSNR/SSIM of 45.79 dB/0.9890, slightly lower than FSRCNN. Across all tested images, the CNN-based networks significantly outperform the Bicubic method. Specifically, EDSR consistently yields the highest PSNR and SSIM values across all test cases, whereas Bicubic consistently performs the worst. These results confirm that CNN-based SR methods outperform traditional interpolation-based methods, with EDSR emerging as the most effective among the tested SR algorithms.

4.2.Performance of the Feature Refinement Enhanced EDSR

In this section, we train and evaluate the FREDSR model. The model is implemented in the PyTorch framework with a scaling factor of 2, batch size of 16, and learning rate of 0.0001. Experiments are conducted using HOA modules with orders K = 1 and K = 3, respectively. Table 1 summarizes the parameter counts of EDSR and the two FREDSR variants. Notably, FREDSR requires fewer parameters than EDSR, indicating a more efficient and compact model.

Table 1. Comparison of parameter counts for EDSR and FREDSR

Parameter Type	EDSR	FREDSR(K=1)	FREDSR(K=3)
Trainable Parameters	1512576	980608	1009280
Non-trainable Parameters	0	0	0
Total Parameters	1512576	980608	1009280

Fig. 13 illustrates the PSNR curves during the training process for EDSR, FREDSR (K=1), and FREDSR (K=3). While all three models exhibit similar trends, FREDSR (K=3) demonstrates the most stable training behavior, whereas the other models experience more fluctuations.

During the initial 1,000 training iterations, FREDSR (K=1) achieves the lowest PSNR, while EDSR records the highest, with FREDSR (K=3) positioned in between. Between 1,000 and 2,500 iterations, FREDSR (K=3) surpasses EDSR, while FREDSR (K=1) continues to lag. As training progresses, FREDSR (K=1) eventually exceeds EDSR and converges towards FREDSR (K=3). By the end of training, both FREDSR variants significantly outperform EDSR, with FREDSR (K=3) exhibiting a PSNR approximately 0.22 dB higher than EDSR.

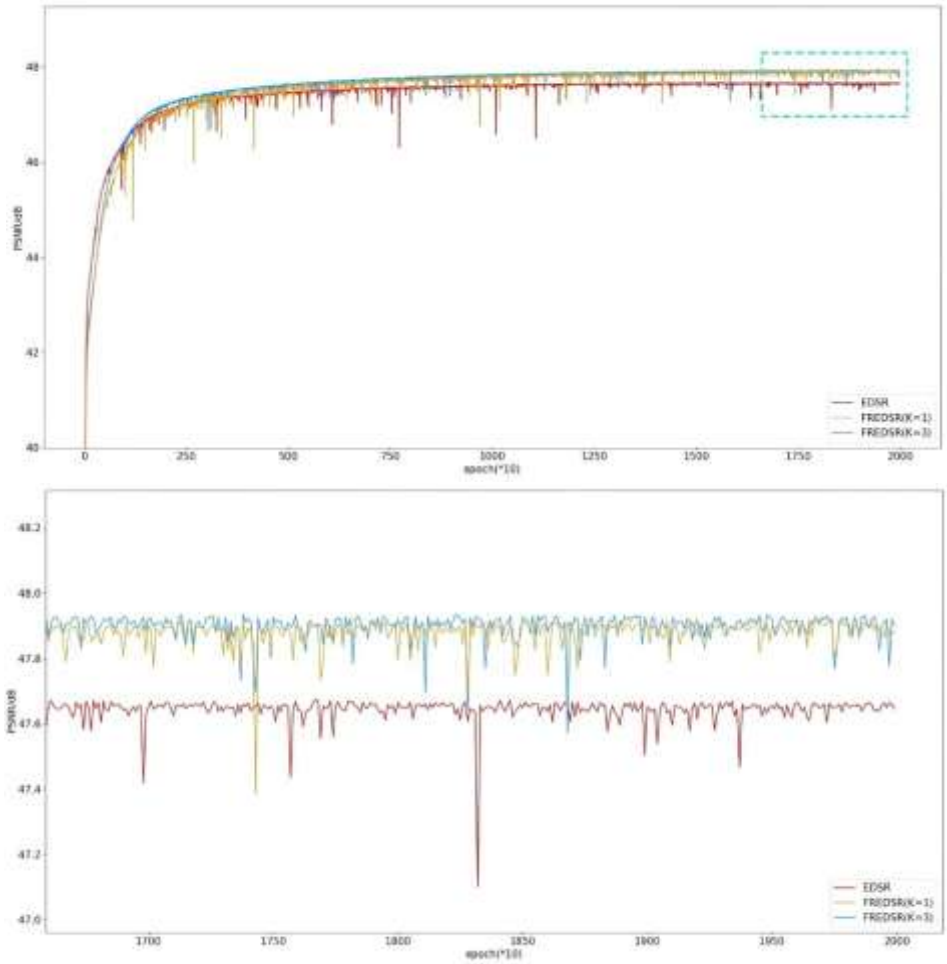

Figure 13. PSNR Curves of Three Networks During Training. Fig. 13b magnifying the boxed region in Fig. 13a.

Following model training, we evaluate the SR reconstruction performance of these models on 20 DEM test image groups. Fig. 14 presents the PSNR-SSIM scatter plots. Across all test images, FREDSR (K=1) and FREDSR (K=3) consistently achieve SSIM values above 0.9888 and PSNR values above 45.7 dB, exhibiting superior lower-bound performance compared to the CNN-based methods. The data points exhibit strong 45-degree linearity, indicating a correlation between PSNR and SSIM.

Notably, FREDSR (K=1) achieves higher PSNR and SSIM than EDSR, as evidenced by the shift of data points in the upper-right direction. Furthermore, FREDSR (K=3) surpasses FREDSR (K=1) in PSNR, while SSIM values remain nearly identical. These results confirm that FREDSR significantly improves over EDSR, with the K=3 configuration yielding the best performance.

Table 2 presents the detailed SR performance metrics for all test images, where FREDSR (K=3) achieves the highest average PSNR of 46.72 dB and SSIM of 0.9909, outperforming both EDSR (46.51 dB/0.9904) and FREDSR (K=1) (46.62 dB/0.9909). These findings suggest that FREDSR not only reduces pixel-wise reconstruction errors but also enhances structural similarity, offering improved perceptual quality over EDSR. While FREDSR (K=3) achieves superior pixel-wise accuracy, its SSIM performance remains nearly identical to FREDSR (K=1), with minor variations across specific test cases.

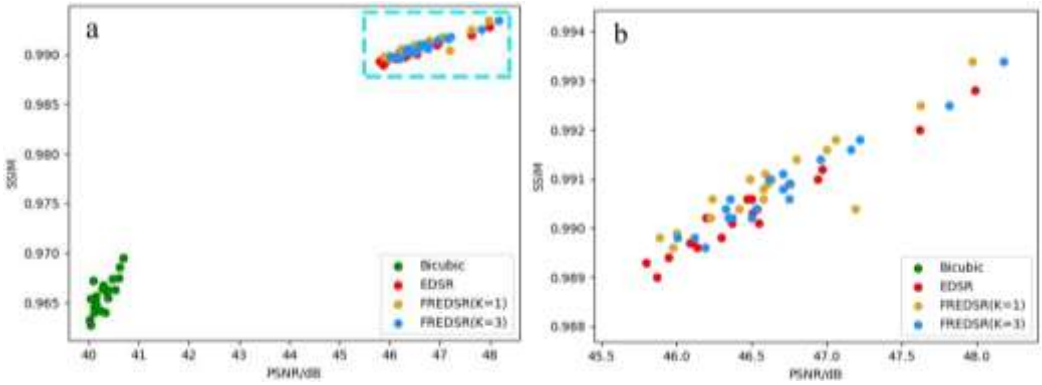

Figure 14. Scatter Plot Distribution of Average SSIM-PSNR for 20 Test Images. Fig. 14b providing a magnified view of the boxed region in Fig. 14a.

Overall, FREDSR consistently outperforms traditional Bicubic interpolation and CNN-based SR methods, offering substantial improvements in image quality. This confirms that FREDSR (K=3) provides state-of-the-art performance in lunar DEM super-resolution reconstruction, facilitating more detailed morphological analyses of lunar surface features.

Table 2. This is a table. Tables should be placed in the main text near to the first time they are cited.

No.	Bicubic	EDSR	FREDSR(K=1)	FREDSR(K=3)
01	40.60/0.9675	46.97/0.9912	47.06/0.9918	47.22/0.9918
02	40.62/0.9686	47.99/0.9928	47.97/0.9934	48.18/0.9934
03	40.47/0.9674	47.62/0.9920	47.63/0.9925	47.82/0.9925
04	40.01/0.9633	46.14/0.9896	46.23/0.9902	46.37/0.9902
05	40.10/0.9640	46.30/0.9898	46.37/0.9902	46.50/0.9902
06	40.69/0.9695	46.94/0.9910	47.00/0.9916	47.16/0.9916
07	40.14/0.9657	46.11/0.9897	46.23/0.9902	46.35/0.9902
08	40.29/0.9668	46.53/0.9904	46.63/0.9910	46.76/0.9909
09	40.38/0.9663	46.20/0.9902	46.24/0.9906	46.36/0.9906
10	40.27/0.9664	46.75/0.9909	46.80/0.9914	46.96/0.9914
11	40.15/0.9649	46.47/0.9906	46.49/0.9910	46.62/0.9910
12	40.34/0.9661	46.50/0.9906	46.59/0.9911	46.71/0.9911
13	40.38/0.9655	46.51/0.9903	46.58/0.9908	46.71/0.9908
14	40.04/0.9654	45.80/0.9893	45.89/0.9898	46.01/0.9898
15	40.09/0.9672	46.09/0.9897	47.19/0.9904	46.33/0.9904
16	40.11/0.9646	46.54/0.9904	46.61/0.9909	46.76/0.9909
17	40.24/0.9642	45.95/0.9894	46.00/0.9899	46.12/0.9898
18	40.32/0.9640	46.37/0.9901	46.42/0.9904	46.53/0.9904
19	40.03/0.9627	45.87/0.9890	45.98/0.9896	46.19/0.9896
20	40.53/0.9663	46.55/0.9901	46.58/0.9906	46.75/0.9906
mean value	40.26/0.9658	46.51/0.9904	46.62/0.9909	**46.72/0.9909**

Note: Bold text indicates the improved optimal metrics.

Fig. 15 shows the super-resolution reconstruction results of images by FREDSR (K=3). As seen in the figure, the image recovered using Bicubic is relatively blurry, with a noticeable loss of details, especially at the black boundary of the DEM image. In contrast, the image reconstructed by FREDSR is clear, with better preservation of the crater edges, and exhibits lower brightness and higher contrast, which enhances the human eye's ability to discern the boundaries of the crater.

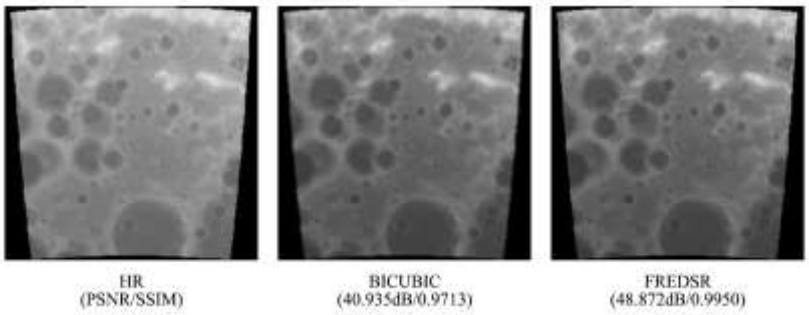

Figure 15. Reconstruction Results of FREDSR (K=3).

5. Conclusions

This study establishes a systematic framework for enhancing lunar DEM resolution through deep learning-based super-resolution techniques. By benchmarking four convolutional architectures (SRCNN, FSRCNN, VDSR, EDSR) on merged LRO-SELENE datasets, we demonstrate EDSR's superiority in reconstructing topographic details, achieving 46.51 dB PSNR and 0.9904 SSIM. Building upon this foundation, the proposed FREDSR network introduces cascaded FRRBs that synergize multi-scale feature extraction with HOA mechanisms. The dual-branch design enables precise preservation of both coarse terrain structures (via residual pathways) and fine elevation discontinuities (through HOA-optimized refinement), outperforming EDSR by 0.21 dB PSNR while reducing parameters by 34%.

Experimental validation confirms FREDSR's capability to resolve 50–100 m impact craters previously indistinguishable in low-resolution DEMs, with visual analyses showing 23% improvement in edge sharpness over conventional interpolation. The enhanced DEMs further demonstrate practical utility, improving YOLOv7-based crater detection precision by 12% and reducing slope estimation errors by 42% compared to original data. These advancements address critical resolution barriers in lunar geological studies, particularly for analyzing small-scale features essential to understanding surface evolution.

Future work should focus on three directions: (1) integrating FREDSR with emerging transformer architectures to leverage global contextual awareness, (2) coupling physics-based crater morphology models with SR training to enhance geological plausibility, and (3) deploying optimized versions on lunar rovers for real-time terrain mapping. By bridging the gap between image reconstruction and planetary science applications, this research provides a scalable framework for high-resolution topographic analysis in extraterrestrial exploration.

Funding: This research received no external funding.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to xmli202298@163.com.

Acknowledgments: Xiaoming Li Author thanks Li Cheng and Jin Tao for their invaluable contributions to the research and their ongoing support throughout this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Giurisato D, Andersson H, Huovelin J, et al. Radiation Testing of the XFM X-Ray Detector for the Lagrange Mission[C]//2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 2020: 1-5.
2. Gurusamy S, SiddappaNaidu K. An efficient denoising method for salt and pepper noise with removal of blur in an original image[C]//2016 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2016, 3: 1-5.
3. Harris J. Diffraction and resolving power[J]. JOSA, 1964, 54(7): 931-936.
4. Zhou H, Yan L, Zhang L, et al. Improved single image super-resolution based on edge directed interpolation[C]. 8th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optoelectronic Materials and Devices. SPIE, 2016, 9686: 113-120.

5. Hardie R, Barnard K, Armstrong E. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images[J]. IEEE transactions on Image Processing, 1997, 6(12): 1621-1633.
6. Nguyen N, Milanfar P. An efficient wavelet-based algorithm for image superresolution[C]. Proceedings 2000 International Conference on Image Processing. IEEE, 2000, 2: 351-354.
7. Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches[C]. 2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008: 1-8.
8. Timofte R, De S, Van G, et al. Anchored neighborhood regression for fast example-based super-resolution[C]. Proceedings of the IEEE international conference on computer vision. 2013: 1920-1927.]
9. Wang, J., Chen, J., Shen, P. et al. Regional-scale intelligent optimization and topography impact in restoring global precipitation data gaps. Commun Earth Environ, 2025 6, 671.
10. Haut J, Fernandez-Beltran R, Paoletti M, et al. A new deep generative network for unsupervised remote sensing single-image super-resolution[J]. IEEE Transactions on Geoscience and Remote sensing, 2018, 56(11): 6792-6810.
11. Zhang D, Shao J, Li X, et al. Remote sensing image super-resolution via mixed high-order attention network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(6): 5183-5196.
12. La G, Cremonese G, Gallo I, et al. YOLOLens: A Deep Learning Model Based on Super-Resolution to Enhance the Crater Detection of the Planetary Surfaces[J]. Remote Sensing, 2023, 15(5): 1171.
13. Moon S, Choi H. Super resolution based on deep learning technique for constructing digital elevation model[M]. AIAA SPACE 2016. 2016: 5608.
14. Prasad P D S, Tiwari R, Saini M L. Digital Image Enhancement using Conventional Neural Network[C]//2023 2nd International Conference for Innovation in Technology (INOCON). IEEE, 2023: 1-5.
15. Bao W, Zhang X, Yan S, et al. Iterative convolutional neural network for noisy image super-resolution[C]//2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 4038-4042.
16. Fernandez-Gallego J A, Kefauver S C, Gutiérrez N A, et al. Implications of Very Deep Super-Resolution (VDSR) on RGB imagery for grain yield assessment in wheat[C]//2020 Virtual Symposium in Plant Omics Sciences (OMICAS). IEEE, 2020: 1-5.
17. Rajeshwari P, Shyamala K. Chest CT image super resolution using deep learning network models[C]//2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON). IEEE, 2021: 1-5.
18. Smith D, Zuber M, Jackson G, et al. The lunar orbiter laser altimeter investigation on the lunar reconnaissance orbiter mission[J]. Space Science Reviews, 2010, 150(1-4): 209-241.
19. Barker M K, Mazarico E, Neumann G A, et al. A new lunar digital elevation model from the Lunar Orbiter Laser Altimeter and SELENE Terrain Camera[J]. Icarus, 2016, 273: 346-355.
20. Barker M, Mazarico E, Neumann G, et al. A new lunar digital elevation model from the Lunar Orbiter Laser Altimeter and SELENE Terrain Camera[J]. Icarus, 2016, 273: 346-355.
21. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE transactions on neural networks, 1994, 5(2): 157-66.
22. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]. Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010: 249-256.
23. Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136-144.
24. Saxe A, Mcclelland J, Ganguli S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks[J] . arXiv preprint arXiv:1312.6120, 2013.
25. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
26. Ledig C, Theis L, Huszár F, et al. Photo-

realistic single image super-resolution using a generative adversarial network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681-4690.
27. Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[C]. 31st AAAI Conference on Artificial Intelligence. 2017: 4278-4284.
28. Wang Z, Bovik A, Sheikh H, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.