JARS Journal of Applied Remote Sensing

RESEARCH PAPER

STLDF-Net: a semantic segmentation network for lunar surface linear structure detection—a case study of lobate scarps

Chenya Li,[a,b,c] Puyan Xu,[a,b,c] Xin Lu,[a,b,c] Zhiyuan Guo,[a,b,c] Ning Li,[a,b,c] and Gaofeng Shu[a,b,c,*]

[a]Henan University, School of Computer and Information Engineering, Kaifeng, China
[b]Henan Province Engineering Research Center of Spatial Information Processing, Kaifeng, China
[c]Henan Key Laboratory of Big Data Analysis and Processing, Kaifeng, China

ABSTRACT. The lunar surface is characterized by numerous linear structures. Investigating these linear features contributes to our understanding of the Moon's cooling processes and the evolutionary history of its crust. Currently, most methods for extracting linear structures from remote sensing images of the lunar surface rely on manual visual interpretation and semi-supervised learning. This leads to inefficient extraction of these structures from the vast amount of lunar remote sensing data. We take the typical lunar linear structure—lobate scarps—as a representative case and propose a semantic segmentation-based automatic detection algorithm. The proposed model, based on Swin Transformer for semantic segmentation, integrates three modules— long-connection Swin Transformer residual block, deformable pyramid pooling module, and feature pyramid and aggregation network—to significantly enhance the network's capability in extracting features of lobate scarps. The model is named STLDF-Net. Compared with other networks, STLDF-Net achieved the highest accuracy on our custom dataset, with an intersection over union of 95.71% and an $F1$-score of 97.81%. We applied the trained model to detect lobate scarps in the Aitken crater region and the Ansgarius crater region, successfully mapping their spatial distribution in these areas. In addition, we transferred the model to detect lobate scarps on Mars, obtaining favorable results and demonstrating the model's strong generalization capabilities. Finally, we conducted experiments and discussions on the model complexity of STLDF-Net, verifying its applicability for lunar lobate scarp segmentation tasks.

 [DOI: 10.1117/1.JRS.19.024511]

Keywords: lobate scarp; linear structure; transformer; semantic segmentation; narrow angle camera

Paper 250150G received Mar. 4, 2025; revised May 9, 2025; accepted May 16, 2025; published Jun. 4, 2025.

1 Introduction

The Moon is the closest extraterrestrial body to the Earth. According to their geometrical features, lunar structures can be broadly classified into linear or circular structures.[1] Linear structures are widely distributed on the lunar surface, with prominent linear features primarily located on the near side of the Moon. A common type of linear tectonic landform on the Moon is lobate scarps, which are characterized by horizontal compressional stress exceeding vertical stress.[2] The

*Address all correspondence to Gaofeng Shu, gaofeng.shu@henu.edu.cn

Handling Editor: Sicong Liu, Associate Editor

1931-3195/2025/$28.00 © 2025 SPIE

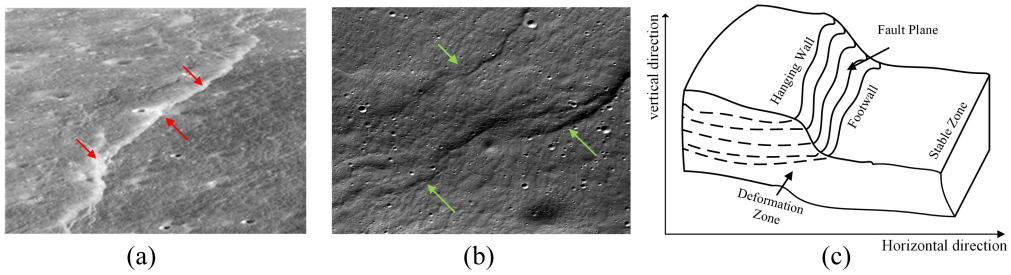

Fig. 1 (a) The red arrow points to the lobate scarp Slipher (48°14′N, 160°32′E). Source of the image.[3] (b) The green arrow points to the lobate scarp Simpelius (73.6°S, 8.76°E). Source of the image.[4] (c) Simple profile map of the lobate scarp.

lunar surface exhibits typical and distinct lobate scarps, as illustrated in Fig. 1. These landforms generally display asymmetric ridge-like topography, often linear or arcuate in shape, with steeply inclined cliff faces and gently sloping rear limbs. Their lengths range from tens of meters to several kilometers, widths span tens to hundreds of meters, and elevations can reach ∼150 m.[5,6] Lobate scarps represent some of the youngest landforms on the Moon. According to van der Bogert et al.[7], their formation ages support a late Copernican age (<800 ± 15 Ma). They typically exhibit sharp morphological expressions, lack superimposed large-diameter impact craters (>400 m), and crosscut smaller diameter craters. These thrust faults exhibit a global distribution, spanning all latitudes on the nearside and farside, which establishes them as the most ubiquitous tectonic features on the Moon. Lobate scarps are predominantly located in anorthosite highlands.[8–10] Unlike nearside wrinkle ridges and graben, they generally occur outside mare-filled basins in highland regions and are most extensively developed on the farside.[2]

Regarding the formation of lunar lobate scarps, previous studies suggest that the Moon originated from a collision between a Mars-sized body and the Earth.[11–14] Following the magma ocean stage, the Moon's interior gradually cooled, leading to solidification and contraction of its liquid outer core. This process generated horizontal compressional stresses (∼400 MPa) that acted on the lunar crust. The crust underwent displacement along low-angle thrust faults (average dip angle: 22.95 deg), with the upper crustal blocks overriding lower blocks, forming asymmetric step-like scarps.[15,16] Lobate scarps serve as direct evidence of global contraction driven by lunar interior cooling. Their dominant orientations (e.g., NE–SW trends) reflect directional patterns of regional stress fields, whereas geometric parameters (e.g., relief amplitude and displacement magnitude) reveal mechanical properties of the lunar crust. In-depth study of lobate scarps can unveil global or regional structural characteristics and stress states, providing critical insights into the Moon's internal geological evolution. Research on lunar linear structures, particularly lobate scarps, holds significant implications for understanding the Moon's tectonic history.[15,17]

Lunar remote sensing optical images are the primary data source for studying lunar tectonic activities and evolution. Early studies on lobate scarps utilized images from the Apollo Panoramic Camera; however, due to their limited spatial coverage, the research was confined to equatorial regions.[18] In 2010, the Lunar Reconnaissance Orbiter Camera (LROC) provided new global high-resolution images, leading to the discovery of previously unidentified lobate scarps in high-latitude areas (≥ 60 deg) and mapping their global spatial distribution.[2,5,19] By 2015, over 3200 scarps had been identified.[8]

Previously, the detection of linear structures on the lunar surface primarily relied on manual interpretation[1,20,21] and semi-supervised learning methods.[17,22]

Lu et al.[1] utilized dataset products from the China Lunar Exploration Program and high-quality dataset products from international exploration missions to manually map and annotate 227 lobate scarps, 474 rilles, and 11,046 wrinkle ridges longer than 2.5 km using the Mollweide projection on the ArcMap platform. Hurwitz et al.[20] employed the latest acquired imagery and topographical data to manually draw and analyze observed lunar channels using ArcMap. Nelson et al.[23] used Narrow Angle Camera (NAC) data from the LROC and ArcGIS software to digitally map lobate scarps. The aforementioned manual mapping processes are time-consuming, labor-intensive, inefficient, and prone to errors, largely depending on the mapper's experience and judgment. Typically, such studies require the mapper to have a profound understanding of

geology and planetary dynamics, resulting in a high barrier to entry. In most cases, this method is difficult to generalize. Considering the vastness of the lunar surface, manually detecting all linear structures from a large number of remote sensing images poses a significant challenge for experts, and different professionals may identify varying characteristics of lunar linear structures.

Previous studies have employed semi-supervised learning methods to detect linear structures on the lunar surface. Ke et al.[24] utilized multiscale parameters and multiresolution digital terrain models (DTM), introducing terrain curvature to automatically extract lunar surface linear structures (including rilles and wrinkle ridges). The extraction accuracy reached a maximum value of 0.1652 with a window size of 825 m. Peng et al.[22] proposed a method combining phase symmetry and morphological operations to detect wrinkle ridges, achieving a detection percentage of 90.7% at the best test points. Lou and Kang[17] used a DEM slope averaging filter method to extract lunar surface linear structures, attaining a completeness rate of 87.26%. Micheal et al.[25] introduced an automatic detection method for grabens based on Hessian techniques applied to lunar DTM, automatically identifying grabens through gradient changes and achieving a detection rate of 90%. These algorithms typically extract linear structures based on features of linear entities, such as shape, brightness differences, and texture variations. Compared with manual detection, these semi-automatic detection algorithms have significantly improved efficiency. However, due to their specialized design, they often perform poorly when processing images with complex geological information. The extraction of linear structures using these methods is generally limited by their specialized capabilities, requiring substantial knowledge in mathematics, physics, geology, and planetary science. In addition, when multiple geomorphological types coexist in the same region, the detection performance of these algorithms tends to be relatively poor.

In recent years, artificial intelligence technology has been rapidly advancing in the field of deep space exploration. Moghe et al.[26] proposed a deep learning method utilizing semantic segmentation to classify hazardous and safe areas from light detection and ranging (LIDAR) scans during lunar landing phases. Leveraging semantic segmentation algorithms in deep learning enables end-to-end, pixel-wise matching to extract linear structures on the lunar surface. Yan et al.[27] proposed a linear feature extraction method combining an improved UNet++ and You Only Look Once v5 (YOLOv5) to achieve object detection and semantic segmentation of linear structures, attaining an intersection over union (IoU) of 0.69 on a custom dataset. Zhang et al.[28] introduced a multimodal semantic segmentation method based on DeepLabv3+ to automatically identify and detect rilles, achieving a mean intersection over union (MIoU) of 93.90% on a custom dataset. These deep learning–based algorithms for extracting lunar surface linear structures demonstrate the clear advantages of deep learning over traditional methods in this task. However, due to the inherent limitations of convolutional operations, convolutional neural network (CNN)-based methods struggle to learn explicit global and long-term semantic information interactions. Some classical algorithms, lacking the introduction of new concepts and improvements and focusing solely on local information, have been unable to achieve satisfactory segmentation results.

With the rapid development of the deep learning field, vision transformers (ViTs), introduced by a Google team in 2020, represent models that apply transformers to image classification.[29] Semantic segmentation of images using global self-attention mechanisms has been widely adopted. A representative model in the ViT domain is the Swin Transformer. Swin Transformers utilize hierarchical feature maps and shifted windows to capture both local and global contextual information, significantly improving segmentation accuracy.[30] The ability of the Swin Transformer to handle varying scales and its efficiency in processing high-resolution images make it particularly suitable for complex segmentation tasks. Inspired by the success of the Swin Transformer,[31] researchers from the Technical University of Munich, Fudan University, and Huawei proposed Swin-UNet—the first purely transformer-based U-shaped architecture.[32] This model fully leverages the powerful capability of transformers to extract global information, demonstrating great potential in image semantic segmentation. Li et al.[33] proposed a U-Net and transformer-based semantic segmentation network for deep space rock images, achieving MIoU scores of 79.32% and 93.43% on two public datasets. However, directly applying Swin-UNet in its original form may not be optimally suited for the specific task of semantic segmentation of

lunar lobate scarps as it lacks targeted adaptations and specialized design considerations for this unique application.

In summary, manual annotation of lunar linear structures exhibits critical limitations: (1) heavy reliance on expert knowledge leading to subjective biases, (2) extreme time-consumption and labor-intensiveness, and (3) prohibitively high specialization barriers. Semi-automatic detection methods, while improving efficiency, demonstrate unsatisfactory performance and poor generalization when handling complex geological scenarios. Furthermore, existing automated approaches lack novel conceptual improvements and dedicated optimizations specifically for lobate scarp identification. These unresolved challenges underscore the urgent need to develop specialized, fully automated techniques for lunar lobate scarps detection. This study focuses on the automatic detection of lunar linear structures using lobate scarps as representative features. We designed STLDF-Net, a deep learning network for the automatic extraction of lobate scarps based on an improved Swin-UNet.

Our main contributions are as follows:

1. Creation of a lunar lobate scarp dataset: We have developed a lunar lobate scarp dataset, one of the few of its kind. This dataset is based on high-resolution NAC data captured by LROC using charge-coupled devices. It comprises 1000 manually collected contour samples of lunar lobate scarps, providing a unique and valuable resource for studying linear structures on the lunar surface. The scarcity of such datasets is attributed to the complexity and time-consuming nature of data collection and processing.
2. Proposal of STLDF-Net: We propose STLDF-Net, a semantic segmentation algorithm based on a global self-attention mechanism, designed to extract linear structures represented by lunar lobate scarps. This model demonstrates improved edge-matching capabilities, enabling more accurate extraction of edge details in linear structures, particularly lobate scarps.

The remainder of this paper is organized as follows. Section 2 introduces the selection and creation of the dataset. Section 3 details the STLDF-Net algorithm and the significance of each module's design. Section 4 presents the experimental evaluation metrics, compares the proposed algorithm with existing classical algorithms through comparative experiments, and conducts ablation experiments, model application experiments, and model transfer experiments. Section 5 concludes our findings.

2 Study Data and Area

2.1 LRO NAC Introduction

High-resolution optical images from LROC have revealed previously undetected lobate thrust fault scarps and associated meter-scale secondary tectonic landforms.[9] This is because it is equipped with NACs that are designed to provide 0.5 m-scale panchromatic images over a 5 km swath. LROC NAC anaglyphs are made from geometric stereo pairs (two images of the same area on the ground, taken from different view angles under nearly the same illumination).[34] These images are panchromatic (400 to 760 nm) with a pixel scale of 0.5 to 2 m, an image display width of 5 km, and a length of 25 km.[19] Due to the small size of lobate scarps, this study utilizes NAC images for scarp detection and confirmation.

2.2 Study Area

A typical NAC image measures ~10,000 pixels in width (east–west direction) and 52,000 pixels in length (north–south direction), corresponding to a normal coverage area ranging from $\sim 5 \times 26$ km to 20×100 km. The NAC optical image data used in this study were downloaded from NASA's Planetary Data System, comprising a total of 39 typical NAC Raw Data Record Products (geometrically corrected and georeferenced NAC data). These include 25 regional images (21 named regions and 6 unnamed regions), as shown in Table 1. The resolution of the obtained NAC images ranges from 0.8 to 5 m (the majority being 5 m per pixel), covering the middle, high, and low latitudes of the Moon (with a greater number in the mid and low latitudes and a few in the high latitudes).

Table 1 Data used for creating the lobate scarp detection dataset.

Region	Product name	Latitude	Longitude	Pixel/m
Aitken	NAC_ANAGLYPH_M1137772118_M1137765006	16°47′S	174°24′E	5
Aldrovandi	NAC_ANAGLYPH_M1197583152_M1197569085	25°12′N	28°59′E	5
De Vries	NAC_ROI_DEVRIES_LOA	17°53′S	178°23′W	1
Extension	NAC_ANAGLYPH_M154169223_M154162437	48°9′N	163°40′E	5
Galvani	NAC_ANAGLYPH_M1119442203_M1119420888	50°37′N	81°12′W	5
Galvani	NAC_ANAGLYPH_M1104094593_M1104073160	50°37′N	81°12′W	5
Horseshoe	NAC_ROI_CRSHORSELOA	18°55′N	61°30′E	1.1
Jules Verne	NAC_ANAGLYPH_M103539841_M103532684	36°33′S	148°29′E	5
Jules Verne	NAC_ANAGLYPH_M1100246028_M1100238883	36°40S	148°22′E	5
Jules Verne	NAC_ANAGLYPH_M182552124_M182544976	35°33′S	148°47′E	5
Jules Verne	NAC_ANAGLYPH_M1137943109_M1137935999	35°56′S	149°7′E	5
Jules Verne	NAC_ANAGLYPH_M1212094957_M1212087924	34°53′S	148°36′E	5
Korolev	NAC_ANAGLYPH_M182259815_M182245522	1°33′N	164°20′W	5
Lebedev	NAC_ANAGLYPH_M1151095979_M1151088860	45°3′S	115°23′E	5
Madler	NAC_ANAGLYPH_M1108082617_M1108075470	10°45′S	31°33′E	5
Mandelshtam	NAC_ANAGLYPH_M191909925_M191895630	6°53′N	161°1′E	5
Mandelshtam	NAC_ANAGLYPH_M161252379_M161245596	5°53′N	161°28′E	5
Mendel	NAC_ANAGLYPH_M1134950382_M1134943272	49°1′S	111°14′W	5
Morse	NAC_ANAGLYPH_M143425323_M143418540	18°1′N	176°42′W	5
Moscoviense	NAC_ANAGLYPH_M1205084460_M1205070393	25°43′N	144°47′E	5
Oken	NAC_ROI_OKENCTR_LOB	46°53′S	76°24′E	0.8
Oppenheimer	NAC_ANAGLYPH_M1220047500_M1220040467	37°5′S	164°32′W	5
Oppenheimer	NAC_ANAGLYPH_M151575728_M151568945	34°11′S	160°57′W	5
Racah	NAC_ANAGLYPH_M1189528636_M1189521607	11°18′S	178°7′W	5
Racah	NAC_ANAGLYPH_M1189528636_M1189521607	8°18′S	178°45′E	5
Seares	NAC_ANAGLYPH_M169600722_M169587158	73°35′N	146°52′E	5
Serenitatis	NAC_ANAGLYPH_M1190645305_M1190631250	30°56′N	10°35′E	5
Serenitatis	NAC_ROI_SERENITALOA_E251N0253	25°12′N	25°26′E	1.1
Serenitatis	NAC_ROI_SPACEIL_LOB_E326N0194	33°6′N	19°21′E	0.9
Tsiolkovskiy	NAC_ROI_TSIOLKOVLOH_E199S1286	19°15′S	128°31′E	0.9
Tsiolkovskiy	NAC_ANAGLYPH_M167370048_M167363261	19°19′S	128°36E	5
Tsiolkovskiy	NAC_ANAGLYPH_M143799104_M143792320	21°30′S	126°4′E	5
Tsiolkovskiy	NAC_ANAGLYPH_M1122743851_M1122736739	19°8′S	130°19′E	5
Virtanen	NAC_ANAGLYPH_M158795840_M158789053	15°54′N	177°13′E	5
Unnamed	NAC_ANAGLYPH_M151603560_M151596778	2°43′N	164°37′W	5
Unnamed	NAC_ANAGLYPH_M134272668_M134265884	6°1′N	140°33′E	5

Table 1 (*Continued*).

Region	Product name	Latitude	Longitude	Pixel/m
Unnamed	NAC_ANAGLYPH_M1191769650_M1191755597	36°6′N	162°41′W	5
Unnamed	NAC_ANAGLYPH_M189073069_M189044470	53°26′N	122°18′W	5
Unnamed	NAC_ANAGLYPH_M141456962_M141443389	72°9′N	121°8′E	5
Unnamed	NAC_ANAGLYPH_M186714150_M186685552	53°27′N	122°16′W	5

2.3 Dataset Creation

After obtaining the aforementioned NAC images, they were first converted into grayscale images. To ensure the accuracy of sample labels, we manually interpreted and annotated the labels. Following the initial labeling, multiple manual verifications were conducted to correct erroneous annotations, particularly around the edges, and the labeled images were then integrated into the dataset. To prevent memory overflow and increase the number of samples, this study employed a sliding window with a 50% overlap rate to crop the NAC images and their corresponding labels, with a crop size of 512×512 pixels. Images without label data were automatically deleted, followed by manual inspection to exclude low-quality images. The final sample consisted of a total of 1124 patches, which were randomly divided into training and testing sets in a ratio of 8:2.

3 Methodology

The lobate scarp detection method in this study primarily consists of the following components. First, we introduce an overall architecture of the proposed STLDF-Net. Then, we describe the modules designed within the model, which mainly include the long-connection Swin Transformer residual block (LCSRB) is a feature extraction module based on the Swin Transformer, designed to enhance the network's ability to extract features from lunar lobate scarps; the deformable pyramid pooling module (DPPM), which enhances the network's feature representation capabilities for lunar lobate scarps.; and the feature pyramid and aggregation network (FPAN) module, which integrates the feature pyramid network (FPN) and path aggregation network (PAN) to fuse multilevel features from different hierarchical layers, thereby improving the network's detection accuracy for lunar lobate scarps. Finally, we introduce our decoder component.

3.1 STLDF-Net Construction

To effectively address the problem of feature extraction and recognition of lobate scarps on the lunar surface, we designed STLDF-Net based on an encoder–decoder architecture, as shown in Fig. 2. This is an end-to-end network model. The encoder part of STLDF-Net utilizes the encoder from Swin-UNet as the backbone network.[32] The latter employs the Swin Transformer to replace the original U-Net structure, inheriting the advantages of both U-Net and transformers. It uses skip connections to link the encoder and decoder, fully extracting the semantic features of the images. There are various network architectures of the Swin Transformer, and STLDF-Net adopts the Swin-B (base) architecture, which has a model size and computational complexity similar to ViT-B/Dei-B.[30] First, we embed the LCSRB module with residual connections, which we designed, in each stage to prevent feature extraction loss that may be caused by directly connected features, thereby enhancing the model's expressive capacity and training stability. The encoder consists of four stages, with each stage containing 2, 2, 18, and 2 Swin Transformer blocks, respectively. In our network model, the input image size is $H \times W \times 3$, which is divided into 4×4 patches through the patch partition layer, generating patch tokens with a shape of (H/4, W/4, 48). The generated patch tokens go through the linear embedding in stage 1. They are then input into the LCSRB module, which contains two consecutive Swin Transformer V2 blocks, to generate tokens of (H/8, W/8, 256). Stages 2 and 3 consist of patch merging and the LCSRB module, respectively. In patch merging, adjacent 2×2 patches are

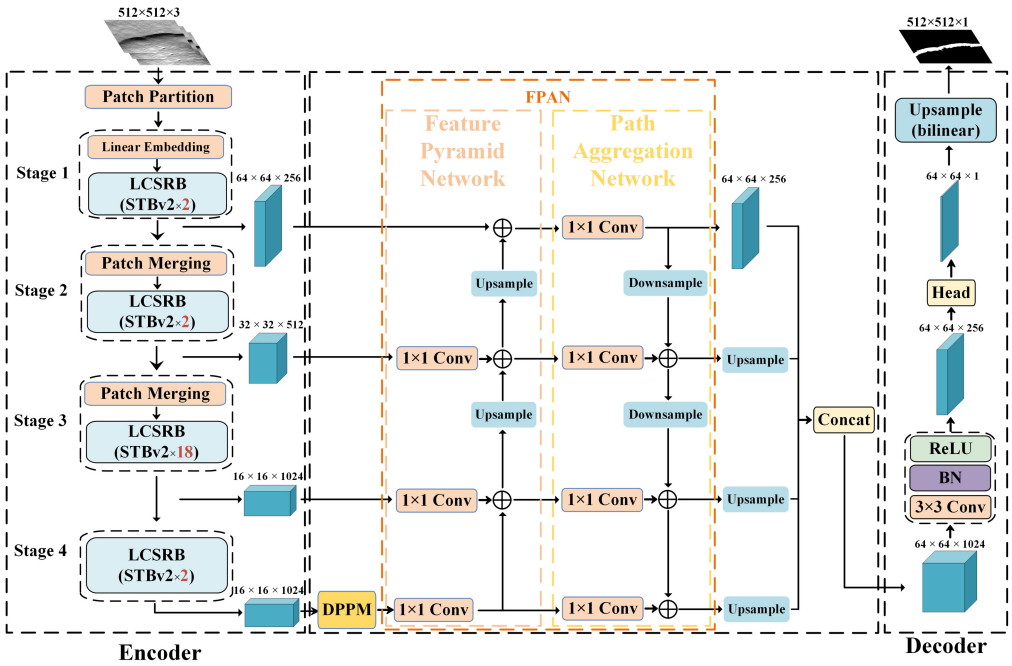

Fig. 2 Structure of STLDF-Net.

merged into one patch, and the tokens are downsampled by a factor of ½, whereas the channel dimension C is doubled.[35] In stages 2 and 3, the shape of the tokens is (H/16, W/16, 512) and (H/32, W/32, 1024). Stage 4 no longer performs patch merging, meaning its output feature map size and dimensions are the same as those of stage 3. This architectural choice preserves finer spatial details that prove particularly crucial for accurate localization and semantic segmentation tasks, especially when dealing with small-scale linear features such as lunar lobate scarps. Although patch merging in shallow layers helps retain detailed information for micro-terrain analysis, its application in deeper layers captures more global contextual information suitable for macro-structure interpretation. However, continued downsampling in deep networks risks losing critical spatial resolution, which would significantly compromise segmentation accuracy for precision-demanding geological features. It also ensures that high-level features remain stable, facilitating special processing of the final layer features and benefiting the input of the final layer features into the DPPM module for better performance. Next, we designed the DPPM module to process the image features obtained from the fourth stage of the decoder, thereby capturing key information in the images more precisely. Then, we incorporated the FPAN module to fuse feature maps from different hierarchical levels, enhancing the network's image detection capability. Finally, the decoder module fuses semantic features at different scales and restores resolution through convolution, batch normalization (BN), and upsampling, resulting in the predicted output.

3.2 Long-Connection Swin Transformer Residual Block

A classic Swin-UNet successfully integrates the transformer into the UNet architecture through the use of the Swin Transformer block.[32] Building upon this, we designed the LCSRB module. The Swin Transformer blocks employed in the LCSRB module are composed of two consecutive Swin Transformer blocks with different structures, one utilizing W-MSA and the other using SW-MSA. W-MSA and SW-MSA are multihead self-attention modules with regular and shifted windowing configurations, respectively. These two modules are used simultaneously and alternately, with an even number of iterations (to ensure paired usage), thereby enhancing the cross-window information connection. We utilize the Swin Transformer V2 architecture, which incorporates a technique called residual post-normalization (res-post-norm), replacing the previously used pre-normalization (pre-norm) structure,[36] as shown in Fig. 3. This method relocates the layer normalization (LN) layer from the beginning of each residual unit to the end of the multilayer perceptron (MLP), applies the LN layer between each multi-head self-attention (MSA) and MLP module, and introduces residual connections after each module. This approach improves the

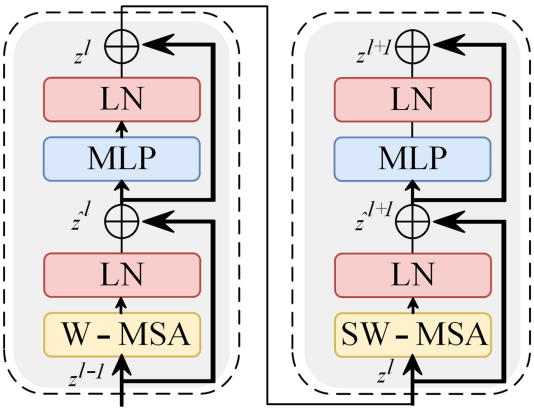

Fig. 3 Two successive Swin Transformer V2 blocks (post-norm Swin Transformer block).

stability of the training process and enhances the training accuracy, making the model more efficient in transferring information between window resolutions. Thus, the computational representation of the Swin Transformer V2 blocks is expressed as follows:

$$\hat{z}^l = \mathrm{LN}(W-\mathrm{MSA}(z^{l-1})) + z^{l-1}, \tag{1}$$

$$z^l = \mathrm{LN}(\mathrm{MLP}(\hat{z}^l)) + \hat{z}^l, \tag{2}$$

$$\hat{z}^{l+1} = \mathrm{LN}(\mathrm{SW}-\mathrm{MSA}(z^l)) + z^l, \tag{3}$$

$$z^{l+1} = \mathrm{LN}(\mathrm{MLP}(\hat{z}^{l+1})) + \hat{z}^{l+1}, \tag{4}$$

where $\hat{z}^l$ and z^l represent the input and output of the l'th block, respectively, and LN stands for layer normalization. W-MSA and SW-MSA perform self-attention within the window but ignore tokens outside the window, where each window only covers $M \times M$ patches. In the experiment, for the convenience of calculation, we set M to 4. The computation formulas for self-attention and multihead attention are as follows:

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

$$\mathrm{MultiHead}(Q,K,V) = \mathrm{Concat}_i(\mathrm{head}_1)W^O, \tag{6}$$

where

$$\mathrm{head}_i = \mathrm{Attention}(QW_i^Q, KW_i^k, VW_i^V), \tag{7}$$

where $Q, K, V \in \mathbb{R}^{M^2 \times d}$ denote the query, key, and value matrices, d is the query/key dimension, and M^2 is the number of patches in a window. As the relative position along each axis lies in the range $[-M+1, M-1]$, we parameterize a smaller sized bias matrix $\hat{B} \in \mathbb{R}^{(2M-1)\times(2M-1)}$, and values in B are taken from $\hat{B}$.[26] T is the transpose. W_i and W^O are parameter matrices.[37]

Building upon Swin Transformer V2 blocks, our improvement involves adding residual connections with three convolutional layers, forming the LCSRB module, as shown in Fig. 4. The residual learning framework was introduced in 2016,[38] and it is a classic approach in the field of deep learning that effectively alleviates the training burden of networks. The residual connections

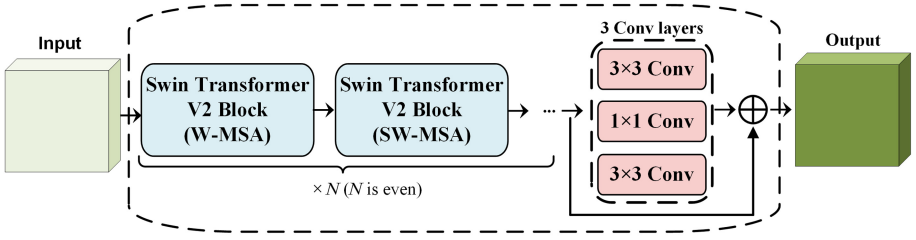

Fig. 4 Structure of the LCSRB.

added in the LCSRB module not only preserve the advantages of the Swin Transformer but also enhance the model's expressive capacity and training stability. In addition, they allow for parameter and memory savings by processing feature maps progressively, thereby maintaining computational efficiency. The introduction of three convolutional operations enables the model to better capture local spatial features, enhancing its ability to capture fine-grained features, which is particularly crucial for handling the complex textures and edge information of linear structures such as lunar lobate scarps. Moreover, through residual connections, the LCSRB module can more effectively fuse features at different scales, improving the model's performance in multiscale scenarios. As our backbone network based on the Swin-B architecture has more layers and greater depth, introducing residual connections effectively improves the model's gradient flow, mitigating the problems of gradient vanishing or exploding in deep networks without causing model performance degradation due to the vanishing gradient problem.

We designed a sequence consisting of three convolutional layers aimed at reducing the number of parameters and the computational burden of the model while enhancing its representational capacity. The first convolutional layer uses a 3×3 kernel to reduce the number of channels from C to $C/4$, where C denotes the number of channels, with the purpose of decreasing computational load while preserving important feature information. The second convolutional layer employs a 1×1 kernel to maintain the number of channels at $C/4$, facilitating the integration and reorganization of feature information. The third convolutional layer utilizes a 3×3 kernel to increase the number of channels from $C/4$ back to C, achieving feature dimensionality expansion and reintegrating the processed features to meet the requirements of subsequent network layers. Between each pair of these three convolutional layers, a LeakyReLU activation function is inserted. Unlike the standard ReLU, LeakyReLU does not completely zero out negative inputs but instead outputs them with a small slope (0.2).[39] This helps prevent neurons from permanently deactivating during training, thereby enhancing the model's expressive capacity. The computation formula for the LeakyReLU function is as follows:

$$\text{LeakyRelu}(x) = \max(0, x) + \alpha \cdot \min(0, x). \tag{8}$$

The processing procedure of the LCSRB module can be represented as follows:

$$F_{\text{stageout}} = F_L + \text{Conv}_{3\times3}(\text{LeakyRelu}(\text{Conv}_{1\times1}(\text{LeakyRelu}(\text{Conv}_{3\times3}(F))))), \tag{9}$$

where F_{stageout} is the output after an entire stage of the encoder, F_L is the feature map after being processed by N Swin Transformer V2 blocks, $\text{Conv}_{3\times3}$ is the convolution operation with a 3×3 kernel size, and F is the input feature of each stage.

3.3 Deformable Pyramid Pooling Module

In this section, we introduce the designed DPPM module, which is applied to process the feature maps output by the final stage of the encoder. In the fourth stage, we obtained the feature map $\mathrm{F} \in \mathbb{R}^{\frac{H}{32}\times\frac{W}{32}\times1024}$, which is subsequently sent to the DPPM module for processing. The DPPM module is an improved version of the pyramid pooling module,[40] which is a technique used in the field of computer vision to enhance the feature expression capabilities of CNNs. An insufficient number of branches would compromise multiscale feature capture, whereas excessive branches may introduce redundancy. In addition, even-sized pooling kernels demonstrate superior uniformity when processing feature maps with even dimensions. Although maintaining this four-level design paradigm in our DPPM module, we strategically adjusted the pooling kernel sizes to $\{1 \times 1, 2 \times 2, 4 \times 4, 6 \times 6\}$ and innovatively incorporated deformable convolution to enhance geometric deformation modeling capabilities. The architecture of the DPPM is shown in Fig. 5.

The purpose of designing this module is to enhance the feature representation in the final stage, enabling the model to more accurately segment targets with varying geometric shapes and sizes, thereby improving the precision and detail of segmentation. This is particularly beneficial for identifying linear structures such as lunar lobate scarps. The module first divides the input features of size $\mathrm{H} \times \mathrm{W} \times \mathrm{C}$ into four branches. Each branch undergoes average pooling of different sizes to obtain features at four distinct pyramid scales, with average pooling layer sizes of $1 \times 1, 2 \times 2, 4 \times 4$, and 6×6, respectively. These are then input into deformable convolutions. Deformable convolution[41] introduces learnable offsets into the receptive field, allowing the

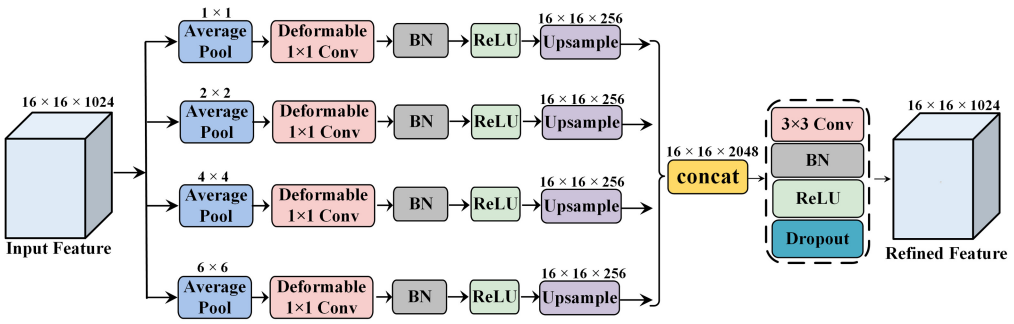

Fig. 5 Structure of the DPPM.

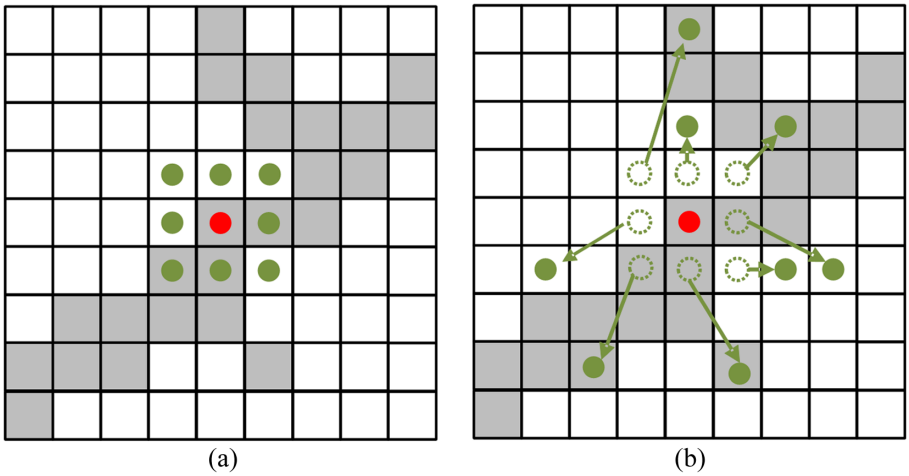

Fig. 6 (a) Standard convolution. (b)The deformable convolution.

convolution kernels to adapt beyond rigid square shapes to more closely match the actual shapes of objects, as illustrated in Fig. 6.

This approach is highly effective for extracting fine branches at the ends of linear structures. Regardless of their shape variations, the convolutional regions can consistently cover the periphery of the object shapes. Next, the features processed by the 1×1 deformable convolution undergo BN and ReLU operations, followed by upsampling to unify their size and dimensions. Subsequently, features obtained from all branches are fused. Finally, a set of 3×3 convolutions, BN, ReLU, and Dropout are applied, resulting in output feature maps of size $H \times W \times C$. The processing procedure of the DPPM can be represented as follows:

$$F_i = \text{Upsample}(\text{Relu}(\text{BN}(\text{Deformable Conv}_{1\times1}(\text{Avgpool}_{j\times j}(F_{\text{in}}))))), i = 1,2,3,4, j = 1,2,4,6, \tag{10}$$

$$F_{\text{out}} = \text{Dropout}(\text{Relu}(\text{BN}(\text{Conv}_{3\times3}(\text{Cat}(F_1, F_2, F_3, F_4))))), \tag{11}$$

where F_{in} is the input to the DPPM module, F_i corresponds to the outputs of the four branches processed by the DPPM, and F_{out} is the output after the entire DPPM module processing. Avgpool refers to average pooling, BN stands for batch normalization, and Dropout refers to the dropout layer.

3.4 Feature Pyramid Network and Path Aggregation Network Module

STLDF-Net introduces the FPN and PAN modules to construct the FPAN module, as shown in Fig. 7, aiming to enhance the performance and efficiency of recognizing linear structures such as lunar lobate scarps. FPN utilizes a top-down pathway and lateral connections to focus on information transfer from higher to lower levels, integrating high-level semantic information with low-level detailed information.[42] Conversely, PAN employs a bottom-up pathway to complement

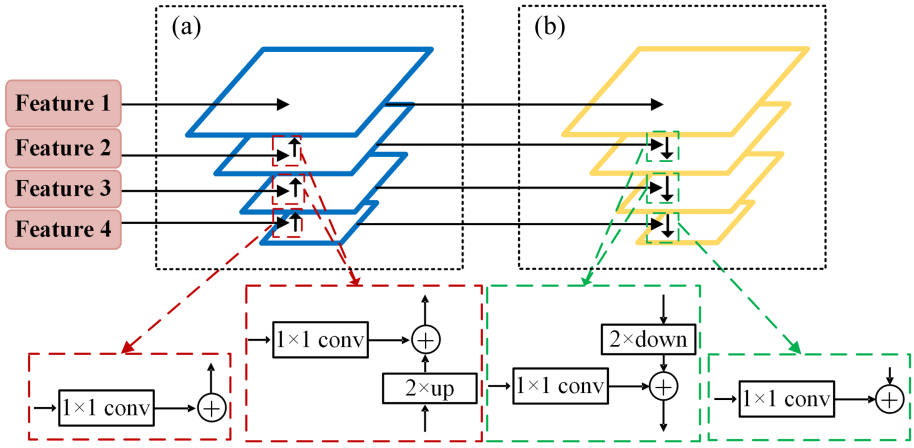

Fig. 7 Structure of the FPAN. (a) FPN backbone. Feature maps are indicated by blue outlines, and thicker outlines denote semantically stronger features. (b) PAN backbone. The yellow outlines denote the bottom-up path augmentation.

information transfer from lower to higher levels, further strengthening the connections between features of different scales.[43] This bidirectional information flow ensures that the model can effectively capture various target features ranging from large to small. The FPAN module combines the strengths of both FPN and PAN, enabling feature extraction and fusion at multiple scales. This bidirectional information flow helps avoid potential information bottlenecks caused by unidirectional paths, achieving more comprehensive information flow and more thorough feature fusion. Particularly for small target detection, such as lobate scarps, the network model with the FPAN module demonstrates superior performance by capturing more detailed information on high-resolution feature maps, making the features of small targets more distinct. Although the inclusion of the FPAN module adds additional feature fusion pathways, the shared convolution layers and modular design make it efficient, avoiding redundant computations and not significantly increasing computational costs or model parameters.

The processing procedure of the FPAN module is as follows: First, 1×1 convolutions are applied to the feature maps of the second, third, and fourth layers to uniformly adjust their channel numbers to a common value, reducing computational load and facilitating subsequent fusion. Next, a bottom-up feature fusion is performed to obtain P_1, P_2, P_3, P_4. Subsequently, 1×1 convolutions are applied to the feature maps obtained from FPN to generate lateral connection feature maps L_1, L_2, L_3, L_4. These convolution operations ensure that the channel numbers of the feature maps at each layer remain consistent, facilitating subsequent fusion. Finally, top-down feature fusion is performed, applying bilinear upsampling on the lower three feature maps to match the height and width of the first layer for subsequent task processing. The computation process of FPAN can be represented as follows:

$$C_i = \mathrm{Conv}_{1\times1}(F_i), i = 2,3,4, \tag{12}$$

$$P_4 = C4, \tag{13}$$

$$P_i = U(P_{i+1}) + C_i, i = 2,3, \tag{14}$$

$$P_1 = F_1, \tag{15}$$

$$P_i' = S_i(P_i), i = 2,3,4, \tag{16}$$

$$L_i = \mathrm{Conv}_{1\times1}(P_i'), i = 1,2,3,4, \tag{17}$$

$$Q_i = L_i, i = 1,4, \tag{18}$$

$$Q_i = L_i + D_{i-1}(Q_{i-1}), i = 2,3, \tag{19}$$

$$Q_i' = \begin{cases} Q_i, i = 1 \\ U(Q_i), i = 2,3,4, \end{cases} \tag{20}$$

where i denotes the level of the feature map, C_i is used to adjust the number of channels, F_i represents the input feature map, "+" denotes feature fusion, U represents bilinear upsampling, S_i represents smoothing convolution operations, P_i' represents the feature map after FPN fusion, L_i represents lateral connection convolution, and Q_i represents the feature map after PAN fusion. The output of the FPAN module is a set of feature maps with uniform sizes $\{Q_1', Q_2', Q_3', Q_4'\}$. For our inputs, each feature map output by this module has dimensions of $64 \times 64 \times 256$.

3.5 Decoder

After the feature maps are processed by the FPAN module, the four output feature maps are concatenated along the channel dimension to generate a comprehensive feature map containing multiple features. Then, the concatenated feature map undergoes 3×3 convolution, BN, and ReLU operations to further fuse and compress feature information, producing a unified fused feature map. Next, the fused feature map is passed through the head (which is essentially a 3×3 convolution layer) to generate the final output feature map. Finally, the feature map output by the head is upsampled using bilinear interpolation to match the spatial dimensions of the original input image, resulting in the network-processed prediction. This process can be represented as follows:

$$Y' = \text{Head}(\text{ReLU}(\text{BN}(\text{Conv}_{3\times3}(\text{Concat}(Q_1', Q_{21}', Q_3', Q_4'))))), \tag{21}$$

$$Y = \text{Upsample}(Y', \text{size} = (H, W), \text{mode} = 'bilinear', \tag{22}$$

where Q_1', Q_2', Q_3', Q_4' are the feature maps output by the FPAN module and Y is the output prediction with the same $H \times W$ size as the input image to STLDF-Net.

4 Experiment and Result

4.1 Implementation Details

This study utilizes the PyTorch deep learning framework to construct network models, which are executed on workstations equipped with 24 GB random access memory (RAM) and NVIDIA RTX A5000 graphics processing unit (GPUs). STLDF-Net is trained using the AdamW optimizer. Starting from the 30th epoch, the learning rate is adjusted to 10% of its previous value. The hardware and software configurations of the network are presented in Table 2. During training, the input image size is set to 512×512pixels.

4.2 Evaluation Metrics

To comprehensively analyze the performance of the proposed STLDF-Net, four commonly used evaluation metrics in deep learning semantic segmentation were employed: precision, recall, IoU, and $F1$-score as objective quantitative analysis indicators to assess our model's performance in automatically detecting lunar lobate scarps. Their computation formulas are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{23}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{24}$$

Table 2 Hardware and software configurations of the experiments.

Configuration	Version
Central processing unit	Intel(R) Xeon(R) Gold 6240R
GPU	NVIDIA RTX A5000
GPU RAM	24GB
PyTorch	1.9.1+cuda11.1
Language	Python 3.8.18

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{25}$$

$$F1 - \mathrm{score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \tag{26}$$

where true positives (TP) denote the number of pixels correctly classified as lunar lobate scarps, false positives (FP) denote the number of pixels incorrectly classified as lunar lobate scarps, and false negatives (FN) denote the number of pixels incorrectly classified as nonlunar lobate scarps. Precision quantifies the proportion of correctly predicted positive instances out of all instances predicted as positive by the model. It is a measure of the model's accuracy in identifying positive classes. Recall, also known as sensitivity, measures the proportion of actual positive instances that are correctly identified by the model. It reflects the model's ability to capture all relevant positive instances. IoU is defined as the ratio of the intersection to the union of the predicted and ground truth values. IoU is frequently used to evaluate the alignment between segmentation models and the actual object boundaries, with higher IoU values indicating better segmentation accuracy. The $F1$-score is a key statistic in the field of image segmentation and is widely used in deep learning as an important indicator of model performance, especially in classification tasks, where a higher $F1$-score signifies stronger model performance. As there is often a trade-off between precision and recall, both IoU and $F1$-score are included in the evaluation of lunar lobate scarp detection to ensure a more comprehensive analysis.

4.3 Manual Hyperparameter Tuning for STLDF-Net

In this section, we focus on the experimental part of hyperparameter optimization for STLDF-Net on the lunar lobate scarp dataset to determine the final hyperparameter selection. Due to the limited number of parameters, we manually tuned the parameters and conducted the following experiments: We performed three sets of experiments, as shown in Fig. 8. The first set controlled the size of the epoch and batch size while varying the learning rate, as illustrated in Fig. 8(a); the second set controlled the size of the epoch and learning rate while changing the batch size, as shown in Fig. 8(b); the third set controlled the learning rate and batch size while varying the epoch, as depicted in Fig. 8(c).

Learning rate: The choice of learning rate is a critical parameter that significantly impacts the training outcomes of deep learning models. In the context of lunar lobate scarp semantic segmentation, the learning rate directly influences the speed and magnitude of weight updates in the network. Theoretically, an excessively small learning rate may prolong the training process and potentially trap the model in a local optimum, preventing it from finding the global optimal solution. Conversely, an overly large learning rate, while accelerating weight updates, may introduce oscillations during training, hindering stable convergence or even causing complete failure to converge. In our experiments, Fig. 8(a) compares model performance under five different learning rates. To ensure a fair comparison, other hyperparameters were held constant. When the learning rate was set to 1×10^{-6}, the IoU value was the lowest (86.86%), suggesting that the weight update steps were too small, preventing the model from reaching the optimal solution. Increasing the learning rate to 5×10^{-6} slightly improved IoU, but the performance remained suboptimal. At learning rates of 5×10^{-5} or 1×10^{-4}, the larger update steps likely caused

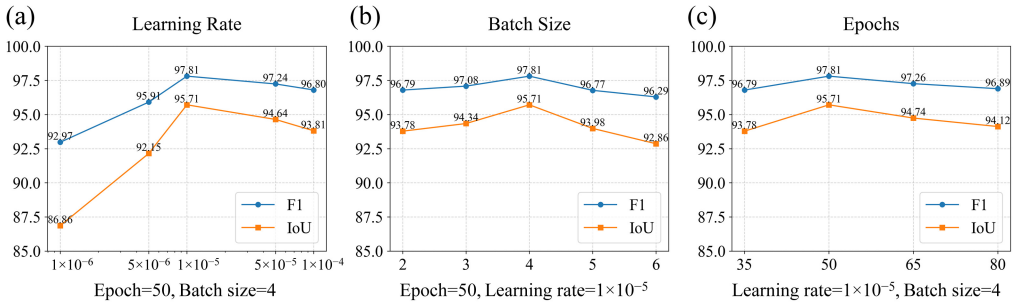

Fig. 8 Comparison of hyperparameter optimization experiments. (a) Learning rate of segmentation model. (b) Batch size of segmentation model. (c) Training epoch of segmentation model.

gradient oscillations around the global optimum, hindering precise convergence. Finally, a learning rate of 1×10^{-5} achieved the highest IoU (95.71%), striking an effective balance between stable weight updates and avoiding excessively slow training. Thus, a learning rate of 1×10^{-5} enabled the model to optimally explore the parameter space while mitigating overfitting, delivering peak performance.

Batch size: Batch size is another crucial hyperparameter that directly affects training efficiency and stability. It determines the number of samples used in each weight update. Too small a batch size may introduce instability during training and increase iteration counts, prolonging training time. Conversely, although larger batches reduce training duration and provide more accurate gradient estimates, they may limit the model's ability to explore the parameter space, potentially trapping it in local optima. Our experiments investigated the impact of different batch sizes on model performance [Fig. 8(b)]. With a batch size of 2, the IoU was lowest (93.78%), likely due to unstable gradient updates. Increasing the batch size to 3 improved IoU to 94.34%, and a further increase to 4 yielded the highest IoU (95.71%), indicating an optimal balance between computational efficiency and gradient estimation accuracy. However, larger batch sizes (5 and 6) led to significant IoU degradation, possibly due to reduced model generalization from limited exposure to data diversity. These results confirm that a batch size of 4 delivers the best performance for STLDF-Net in lunar lobate scarp segmentation.

Training epochs: The number of training epochs is another key hyperparameter requiring careful consideration. Each epoch represents a full forward and backward pass over the training dataset. Insufficient epochs may prevent the model from learning high-level features, whereas excessive epochs can lead to overfitting, degrading performance on unseen data. Determining the optimal epoch count involves balancing training duration, model complexity, and generalization capability. In Fig. 8(c), we observed the effects of varying epoch counts. At epoch = 50, STLDF-Net achieved peak performance for lunar lobate scarp segmentation. With only 35 epochs, IoU was lowest (93.78%), likely due to inadequate learning opportunities. Increasing epochs to 50 maximized IoU (95.71%), striking an ideal balance between training efficiency and convergence. However, further increasing epochs to 65 and 80 reduced IoU, likely due to overfitting. Thus, 50 epochs proved optimal for maintaining generalization while avoiding overfitting.

Based on the comprehensive experimental results, we have determined the following optimal hyperparameter configuration for subsequent lunar lobate scarp semantic segmentation experiments: learning rate of 1×10^{-5}, batch size of 4, and 50 training epochs.

4.4 Comparative Evaluation of STLDF-Net and Other Methods

In this study, we compared STLDF-Net with five classical and two state-of-the-art semantic segmentation models: FCN,[44] U-Net,[45] PSPNet,[40] DeepLabV3+[46] (using ResNet18[38] as the backbone), DeepLabV3+ (using ResNet101 as the backbone), TransUNet,[47] and Swin-UNet.[32] To ensure a fair comparison, all models were trained on the specially annotated dataset for this study. All models used the same hyperparameters as those used in our proposed model. Table 3 presents the quantitative analysis of STLDF-Net and the seven comparison methods for lunar lobate scarp detection, with the best results highlighted in bold and the second-best results in italics. The experimental results indicate that STLDF-Net delivers superior performance, achieving a precision of 97.93%, a recall of 97.69%, a $F1$-score of 97.81%, and an IoU of 95.71%. STLDF-Net outperforms the second-ranked Swin-UNet in both IoU and $F1$-score by 2.14% and 4.01%, respectively, surpasses the second-ranked TransUNet in precision by 1.95%, and exceeds DeepLabv3+ (ResNet101) in recall by 1.78%. In summary, STLDF-Net exhibits the highest accuracy, followed by Swin-UNet, TransUNet, DeepLabv3+ (ResNet101), DeepLabv3+ (ResNet18), PSPNet, U-Net, and FCN.

Figure 9 qualitatively presents the detection results for eight regions. These regions include the midlatitude area on the front side of the Moon's northern hemisphere (regions A–C), the equatorial area on the back side of the Moon's northern hemisphere (regions D–F), the midlatitude area on the back side of the Moon's northern hemisphere (region G), and the midlatitude area on the back side of the Moon's southern hemisphere (region H). The aim is to demonstrate the model's detection capability across different lunar regions. FCN is capable of detecting lunar lobate scarps; however, its detection results exhibit significant issues. When the contrast between the lobate scarps and the lunar background is insufficient, FCN makes notable errors, identifying

Table 3 Results of quantitative evaluation by different methods.

Method	Precision (%)	Recall (%)	*F*1-score (%)	IoU (%)
FCN	82.79	89.94	86.22	75.78
U-Net	86.60	92.67	89.54	81.06
PSPNet	91.59	94.39	92.97	86.87
DeepLabv3+(ResNet18)	92.68	94.30	93.49	87.78
DeepLabv3+(ResNet101)	93.59	95.91	94.74	90.00
TransUNet	*95.98*	94.96	95.47	91.33
Swin-UNet	95.56	*95.77*	*95.67*	*91.70*
STLDF-Net	**97.93**	**97.69**	**97.81**	**95.71**

The values in **bold** highlight the best evaluation metrics in the comparative experiments, whereas the values in *italics* indicate the second-best evaluation metrics.

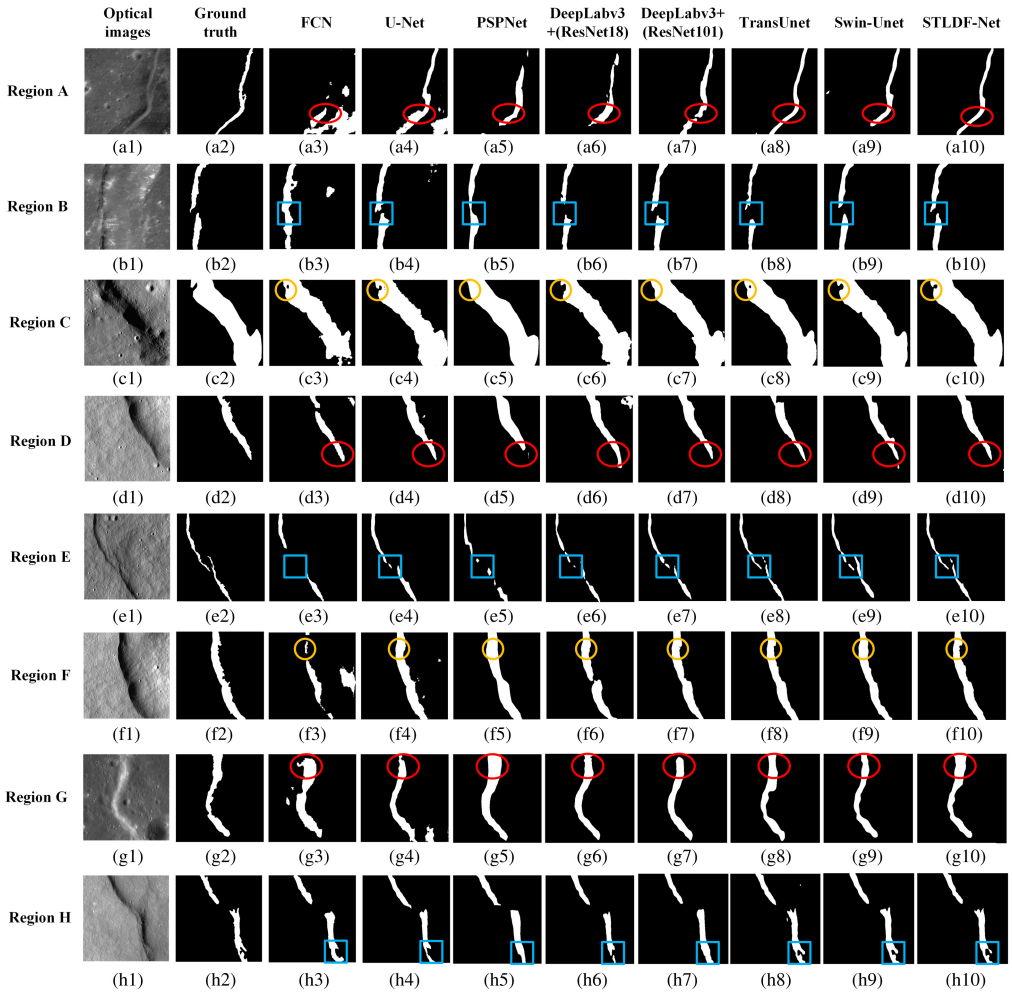

Fig. 9 Detection results of different methods. (a1)–(h1) Optical images. (a2)–(h2) Ground truth. (a3)–(h3) FCN. (a4)–(h4) U-Net. (a5)–(h5) PSPNet. (a6)–(h6) DeepLabv3+(resnet18). (a7)–(h7) DeepLabv3+(resnet101). (a8)–(h8) TransUNet. (a9)–(h9) Swin-UNet. (a10)–(h10) STLDF-Net.

the lunar background as lobate scarps, as shown in the red circle in Fig. 9(a3) and the blue box in Fig. 9(b3). In addition, the model overlooks correct pixels, as seen in the blue box area in Fig. 9(e3) and the yellow circle in Fig. 9(f3). U-Net also erroneously identifies the lunar background as lobate scarps, as shown in the red circle in Fig. 9(a4) and the blue box in Fig. 9(b4). Furthermore, the model struggles to detect the fine tips of the lobate scarps, as seen in the blue box area in Fig. 9(e4) and the red circle in Fig. 9(g4). DeepLabv3+(ResNet18) occasionally overidentifies the lunar background as lobate scarps, as shown in the red circle areas in Figs. 9(a6) and 9(d6). In addition, it performs poorly in detecting the tips of the lobate scarps, as seen in the blue box areas in Figs. 9(b6), 9(c6), and 9(h6). DeepLabv3+(ResNet101) shows poor smoothness in detecting lobate scarps and suffers from fragmentation, as seen in the red circle area in Fig. 9(a7). Moreover, it struggles to distinguish the fine gaps between lobate scarps and the lunar background, as illustrated in the yellow circle in Fig. 9(c7). Observations show that both TransUNet and Swin-UNet avoid many of the aforementioned issues and result in fewer misclassifications. The lobate scarps detected by these two models are relatively smooth and complete compared with the models mentioned earlier, as shown in the red circle in Fig. 9(a8) and the blue box areas in Figs. 9(h8) and 9(h9). In region H, these two models, along with our STLDF-Net, achieve the best segmentation results. However, they still exhibit some potential issues: TransUNet can detect small gaps but erroneously classifies the lunar background adjacent to the gaps as lobate scarps, as shown in the yellow circle area in Fig. 9(c8). Swin-UNet still suffers from detection interruptions, as seen in the red circle area in Fig. 9(d9) and the blue box area in Fig. 9(e9). After overall observation, our designed STLDF-Net achieves the fewest misclassifications. This model produces the best results for detecting lunar lobate scarps, with smooth and continuous images, as seen in the red circle areas in Figs. 9(a10) and 9(d10). STLDF-Net detects gaps, tips of lobate scarps, and fractures with the highest accuracy, as shown in the blue box area in Fig. 9(b10), the yellow circle in Fig. 9(c10), and the red circle in Fig. 9(g10).

4.5 Evaluation of Ablation Experiments

In this study, STLDF-Net integrates three key modules: LCSRB, DPPM, and FPAN, designed to enhance the network's feature extraction capabilities for lobate scarps. To validate the effectiveness of these three modules, we conducted ablation experiments. The different configurations of the ablation experiments and their quantitative evaluations are presented in Table 4.

The quantitative results in Table 4 show that, compared with case 1, the proposed STLDF-Net, which integrates LCSRB, DPPM, and FPAN, demonstrates superior performance metrics. When none of the three modules are incorporated, i.e., using the baseline Swin-UNet, the performance metrics are relatively low. When the DPPM, FPAN, and LCSRB modules are added individually, the model's performance improves compared with the baseline, with IoU increasing by 1.15%, 1.42%, and 1.60%, and $F1$-score increasing by 0.62%, 0.76%, and 0.86%,

Table 4 Different cases and the quantitative evaluation for ablation experiments.

Case	LCSRB	DPPM	FPAN	Precision (%)	Recall (%)	$F1$-score(%)	IoU (%)
1	×	×	×	95.56	95.77	95.67	91.70
2	×	√	×	95.47	97.13	96.29	92.85
3	×	×	√	95.87	97.01	96.43	93.12
4	√	×	×	96.20	96.86	96.53	93.30
5	√	√	×	95.88	96.93	96.40	93.06
6	×	√	√	96.01	*97.39*	96.69	93.60
7	√	×	√	*96.97*	96.98	*96.98*	*94.14*
STLDF-Net	√	√	√	**97.93**	**97.69**	**97.81**	**95.71**

The values in **bold** highlight the best evaluation metrics in the ablation experiments, whereas the values in *italics* indicate the second-best evaluation metrics. A check mark (√) indicates the module is present, whereas a cross (×) indicates the module is removed.

respectively. However, when compared with the complete model, the performance metrics decline: IoU decreases by 2.86%, 2.59%, and 2.41%, and $F1$-score decreases by 1.52%, 1.38%, and 1.28%, respectively. When two of the designed modules are included simultaneously, the performance metrics generally outperform the case where only one module is added. Compared with the complete model, when FPAN is independently removed, IoU decreases by 2.65% and $F1$-score decreases by 1.41%; when LCSRB is independently removed, IoU decreases by 2.11% and $F1$-score decreases by 1.12%; and when DPPM is independently removed, IoU decreases by 1.57% and $F1$-score decreases by 0.83%. The experimental results confirm the effectiveness of the LCSRB, DPPM, and FPAN modules incorporated into this study, indicating that the proposed model provides a robust method for effective detection of lunar lobate scarps.

Figure 10 qualitatively presents the lobate scarp detection results from the ablation experiments across five regions: the low-latitude area on the back side of the Moon's southern hemisphere (regions A-C), the equatorial area on the back side of the Moon's northern hemisphere (region D), and the midlatitude area on the front side of the Moon's northern hemisphere (region E). It is evident that in case 1, significant detection errors occur in areas with darker backgrounds, such as the blue box region in Fig. 10(b3). This is due to the absence of the modules we designed. When any of our three modules are incorporated, such severe errors are no longer present. Compared with case 1, adding the DPPM or FPAN modules enables the model to capture more complete information from the images. The results in case 2 and case 3, compared with case 1, show more complete detections, but still result in minor missed detections, such as in the blue box region of Fig. 10(b5) and the red box regions in Figs. 10(d4)–10(d5). When only the LCSRB module is added, the model improves its ability to capture fine-grained features, allowing for deeper detail extraction, as clearly seen in the blue box area in Fig. 10(e6). However, it suffers from interruptions in detecting linear structures, as shown in the yellow circle area in Fig. 10(c6). When any two of the aforementioned modules are incorporated, the network model's recognition of lobate scarps becomes more comprehensive, as seen in the red circle regions in Figs. 10(a8)–10(a9). Nevertheless, even with two modules, the model still lacks certain feature detection capabilities. This results in some missing details and an increased likelihood of detection omissions, such as in the blue box areas of Figs. 10(b8), 10(b9), and the red box regions in Figs. 10(d7) and 10(d9). STLDF-Net demonstrates the best detection performance, effectively capturing fine details of the lobate scarp tips. The model achieves near-perfect detection with minimal omission and redundant detections, and the detected images are highly smooth. This represents the closest

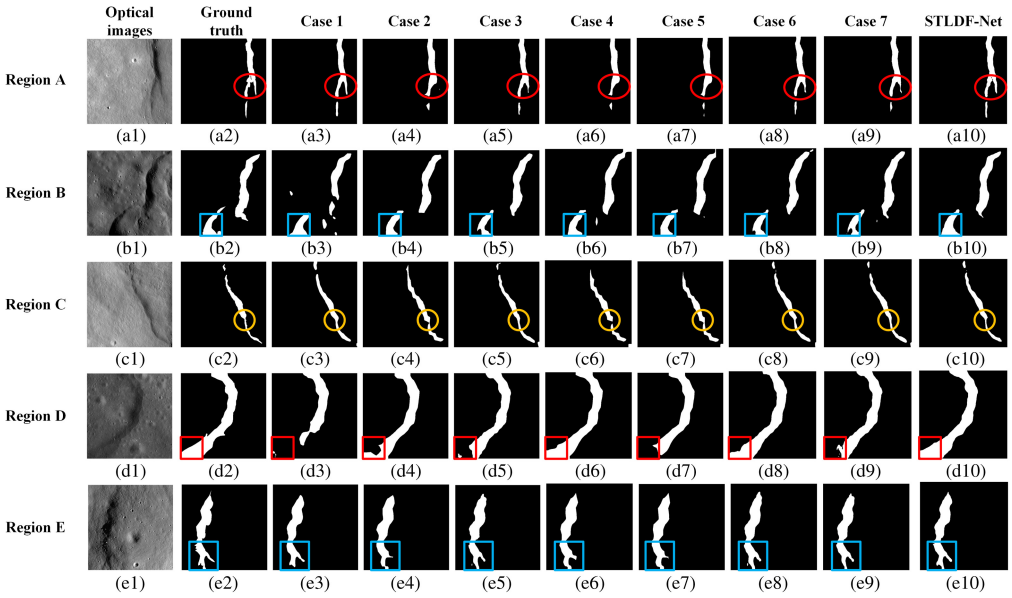

Fig. 10 Qualitative results of ablation experiments. The optical images, ground truth, and detection results based on case 1 to STLDF-Net are shown in (a1)–(e1), (a2)–(e2), (a3)–(e3), (a4)–(e4), (a5)–(e5), (a6)–(e6), (a7)–(e7), (a8)–(e8), and (a9)–(e9), respectively.

alignment with the ground truth in the experiment, as seen in the red circle region in Fig. 10(a10), the red box in Fig. 10(d10), and the blue box in Fig. 10(e10). This exceptional performance is largely due to the incorporation of the three modules, which enable the model to fully learn and achieve both comprehensive global context detection and fine pixel-level classification.

The LCSRB module demonstrates superior capability in preserving feature details, whereas the DPPM module excels at expanding the receptive field. As evidenced by cases 2 and 5 in Table 4, when DPPM is used alone, the IoU reaches 92.85%, whereas the combined use of LCSRB and DPPM yields an additional 0.11% improvement in IoU. This confirms that the residual structure of LCSRB can optimize DPPM's multiscale feature representation capability. Qualitative evidence from region E in Fig. 10 further supports this finding: although a noticeable gap appears in the blue box area of Fig. 10(e4), the detection of lobate scarp termini becomes more refined after incorporating LCSRB [Fig. 10(e7)]. The FPAN module enables multiscale feature extraction and fusion, facilitating more comprehensive information flow and enhancing detection performance. Table 4 data reveals that without any modules (case 1), the IoU is 91.70%. This increases to 93.06% when both LCSRB and DPPM are added (case 5), and further improves by 1.42% and 2.65%, respectively, with FPAN integration. These results demonstrate FPAN's significant contribution to overall detection accuracy. Qualitative validation can be observed in region D of Fig. 10: although noticeable omissions occur in the red circle areas of Fig. 10(d3) and Fig. 10(d7), FPAN incorporation enables complete lobate scarp detection in Fig. 10(d5) and perfect extraction in Fig. 10(d10). In conclusion, our designed modules exhibit mutually reinforcing capabilities, demonstrating quantifiable synergistic effects.

4.6 Model Application

Aitken crater is located on the far side of the Moon, with a diameter of ~135 km. The crater and its surrounding areas exhibit diverse geomorphological features and preserve relatively intact structural characteristics. Due to the compressive and extensional stresses experienced during early impacts and subsequent tectonic evolution, the crater's rim and inner walls have developed prominent fracture structures, including lobate scarps. The lobate scarps in this region vary in scale and morphology, forming scattered band-like distributions around the inner periphery of the crater. Aitken crater has undergone relatively limited modification since its formation, thus retaining more pristine impact and tectonic information, making it of significant research value. Similarly, Ansgarius, a large impact crater located on the eastern limb of the Moon's near side with a diameter of ~91.42 km, has also developed lobate scarps due to stress-induced tectonic evolution, rendering it equally valuable for research.

In this study, we utilized LRO NAC data from Aitken crater, which includes lobate scarps. The image covers a longitudinal range of 15.59°S to 17.10°S and a latitudinal range of 174.14°E to 174.40°E, as shown in Fig. 11(a), with a pixel resolution of 2493×7108. In addition, we employed LRO NAC data from Ansgarius crater, which also includes lobate scarps. The image spans a longitudinal range of 12.57°S to 14.80°S and a latitudinal range of 79.39°E to 80.20°E, as depicted in Fig. 11(c), with a pixel resolution of 2487×7159.

Both datasets were cropped and used as input for the STLDF-Net. After processing through STLDF-Net, the outputs were mosaicked, successfully detecting the overall spatial distribution of lobate scarps in these two regions, as illustrated in Figs. 11(b) and 11(d). Numerous scattered band-like lobate scarps were successfully identified. With the increasing number of lunar far-side exploration missions, such as the Chang'e-4 landing in Von Kármán Crater, the analysis of typical lunar geomorphological and geological units has become increasingly important. Compared with other lunar structural features, lobate scarps within impact craters have received relatively less attention, leaving many aspects unexplored. Our detection and analysis of lobate scarps in these regions significantly contribute to the study of linear structures on the Moon. This research helps us understand the stress field variations, tectonic activity frequency, and deformation mechanisms of the lunar crust over its long-term evolution. Furthermore, it provides new observational evidence for future studies on lunar tectonic models.

4.7 Model Migration

Mars also features lobate scarps similar to those on the Moon. These scarps were discovered earlier on Mars, and the cliffs observed on Mars are generally an order of magnitude larger than those on

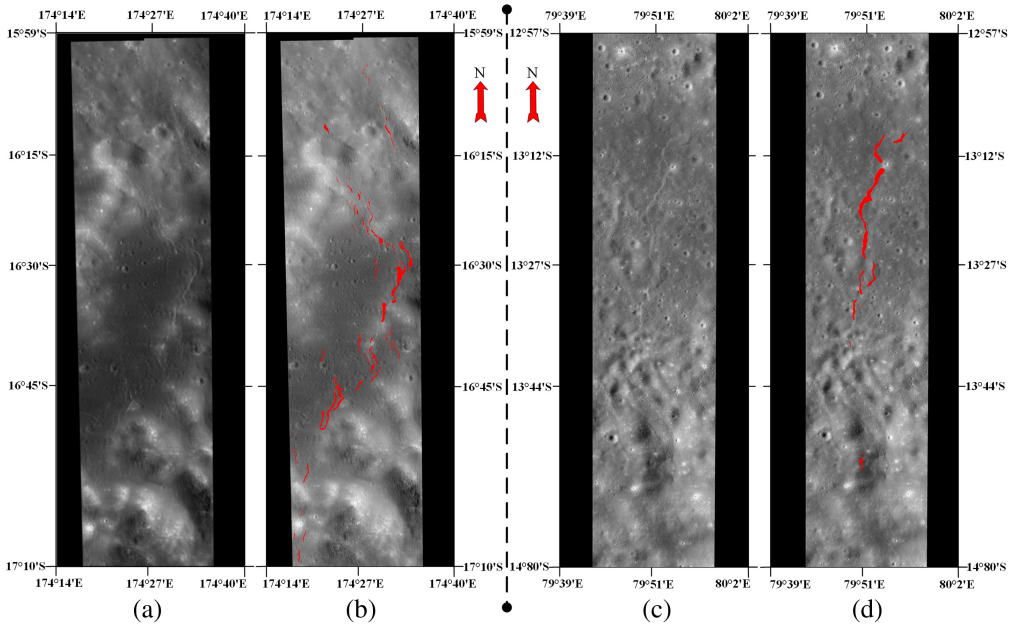

Fig. 11 (a) Optical image of the LRO NAC Aitken (NAC_ANAGLYPH_M1137772118_M1137765006) region. (b) Extraction results of lobate scarps in the Aitken region. (c) Optical image of the LRO NAC Ansgarius (NAC_ANAGLYPH_M1190196360_M1190189329) region. (d) Extraction results of lobate scarps in the Ansgarius region.

the Moon.[2,19] They are typically considered to result from reverse faulting or cliff formation caused by the cooling and contraction of Mars' interior. These features are widely distributed across the Martian surface, particularly concentrated in the highland regions of Mars' northern hemisphere, areas with frequent geological activity and significant tectonic stress. The asymmetric profiles and maximum slopes of lobate scarps on Mars are morphologically similar to those on the Moon.[5] Despite the differences in environmental and geological backgrounds between Mars and the Moon, lobate scarps on different celestial bodies exhibit certain similarities in their basic morphological characteristics, such as clear edges and linear or arcuate structures. This similarity provides a basis for model transferability, motivating the use of Martian lobate scarps to evaluate the generalization ability of STLDF-Net. Moreover, the complex geological background of Mars offers a rich testing ground for the newly designed detection model, allowing for a thorough evaluation of the algorithm's recognition accuracy and adaptability at various scales and complexities.

This study uses high-resolution Mars imagery from the Mars Reconnaissance Orbiter (MRO) HiRISE product (product ID: ESP_017171_2190) to assess the generalization performance of STLDF-Net. The image was captured in 2010, with a central latitude of 38.795°N and a central longitude of 2.061°E. The solar incidence angle is 42 deg, with the Sun about 48 deg above the horizon. The image was preprocessed with image enhancement and noise removal, and the lobate scarps in the image were manually annotated to serve as the validation dataset. The image was then cropped to 512×512pixels to meet the input requirements of the model. After removing images without lobate scarps and those with poor quality, a total of 45 images of Martian lobate scarps were used for testing the model. A small subset of regions was selected to display both the original Martian image and the STLDF-Net predicted lobate scarp results. A comparison between the predicted lobate scarps and the annotated images was performed, with qualitative evaluation metrics, as shown in Fig. 12. The quantitative evaluation metrics for the test images obtained by STLDF-Net were IoU = 86.59% and $F1 = 92.81\%$, indicating that even under the complex geological conditions of Mars, our model maintains high detection performance. These qualitative and quantitative experiments demonstrate that STLDF-Net exhibits strong generalization ability for identifying lobate scarps, showing high robustness and adaptability, and highlighting the model's practical applicability in planetary science.

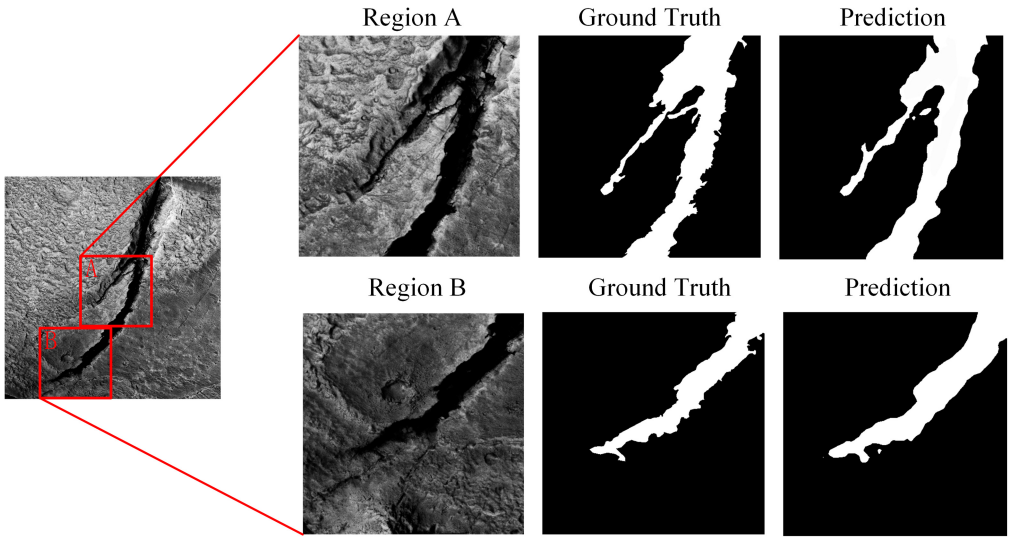

Fig. 12 Detection results of STLDF-Net in Mars testing regions.

5 Discussion

In this section, we conducted experiments on model complexity and computational efficiency between STLDF-Net and other baseline models to provide a more comprehensive evaluation of the proposed method. We performed model complexity experiments on eight models using the lunar lobate scarps dataset, measuring Params (M), FLOPs (G), and computing the efficiency-performance ratio. The experimental results are presented in Table 5 and Fig. 13.

The experimental data from the aforementioned tables and figures demonstrate that although STLDF-Net achieves FLOPs of 87.08G, significantly lower than U-Net (218.64G) and TransUNet (129.25G). However, its FLOPs remain higher than lightweight models such as DeepLabv3+ Res-18 (10.57G), indicating room for optimization in extremely low-resource scenarios. Although DeepLabv3+ (Res-18) attains the highest efficiency-performance ratio (2.16) due to its lightweight design, STLDF-Net achieves a slightly lower efficiency-performance ratio (1.13) but delivers substantially superior absolute performance ($F1 = 97.81\%$, IoU = 95.71%), making it particularly suitable for semantic segmentation tasks requiring stringent accuracy, such as lunar lobate scarps detection. The parameter count of STLDF-Net is 102.41M, exceeding most baseline models, primarily due to the introduction of the LSCRB module. However, parameter sharing strategies mitigate exponential growth in model size while achieving significant performance gains. The

Table 5 Comparison of model efficiency (Params/FLOPs) and segmentation performance between STLDF-Net and other methods on lunar lobate scarps.

Method	Params (M)	FLOPs (G)	$F1$ (%)	IoU (%)	Efficiency-performance ratio
FCN	134.26	160.68	86.22	75.78	0.50
UNet	31.04	218.64	89.54	81.06	0.68
PSPNet	46.57	26.47	92.97	86.87	1.50
DeepLabv3+(Res-18)	16.68	10.57	93.49	87.78	2.16
DeepLabv3+(Res-101)	62.55	45.79	94.74	90.00	1.34
TransUNet	93.23	129.25	95.47	91.33	0.84
Swin-UNet	81.52	71.33	95.67	91.70	1.22
STLDF-Net	102.41	87.08	97.81	95.71	1.13

Efficiency-performance ratio = ($F1$ + IoU)/(Params + FLOPs), scaled by 100 for readability.

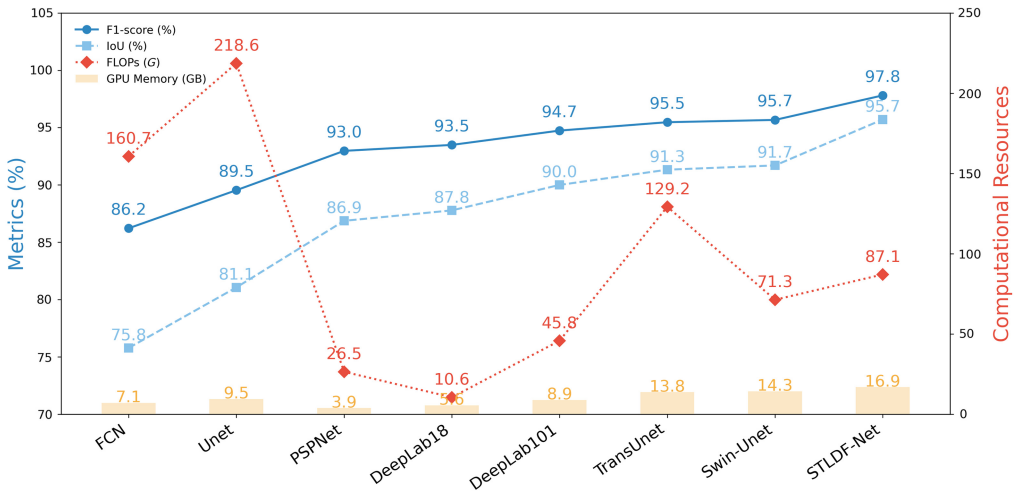

Fig. 13 Comprehensive analysis of computational complexity and segmentation accuracy for lunar Lobate scarps detection.

efficiency-performance ratio of 1.13 for STLDF-Net reflects its trade-off of higher computational costs for enhanced accuracy. Given its substantial performance improvements, this resource overhead is justifiable for precision-critical tasks, especially in the semantic segmentation of lunar lobate scarps.

In summary, STLDF-Net exhibits exceptional performance on the lunar lobate scarps dataset, highlighting its capability to capture complex terrain edges and fine-grained features. This validates its applicability to analogous planetary geological segmentation tasks.

6 Conclusion

This paper proposes the STLDF-Net model for detecting lunar lobate scarps, which effectively extracts semantic information from high-resolution optical images for semantic segmentation. Two innovative modules, LCSRB and DPPM, are introduced, and the FPAN module is successfully incorporated. The residual connections in the LCSRB module effectively enrich the gradient flow during the semantic feature extraction process, mitigating the vanishing gradient problem, enhancing the stability of training deep network models, and enhancing the detection performance of lobate scarps extraction. The DPPM module focuses on processing the semantic features in the final stage of the encoder, enhancing the feature representation at this stage. This module enables the model to more accurately segment targets of various shapes and sizes, improving the network's segmentation precision and detail representation. The introduced FPAN module connects FPN and PAN and integrates them into the network. This module allows for the mutual fusion and transfer of high- and low-level semantic information, and the bidirectional information flow ensures that the model can effectively capture features of targets at various scales, from large to small.

In this study, STLDF-Net is compared with seven classical or advanced semantic segmentation models. The quantitative and qualitative experimental results show that STLDF-Net outperforms the selected comparison algorithms, achieving an IoU of 95.71% and an $F1$-score of 97.81%. The subsequent ablation experiments provide strong evidence of the rationality and effectiveness of incorporating the three modules. The STLDF-Net model is then applied to detect lobate scarps in the Aitken crater region and the Ansgarius crater region, successfully identifying the overall spatial distribution of lobate scarps in these areas. In addition, a transfer experiment is conducted using STLDF-Net on Mars, and the results demonstrate that our model has good generalization ability. Finally, we conducted experiments and discussions on the model complexity of STLDF-Net, verifying its applicability for lunar lobate scarp segmentation tasks.

In summary, the automatic detection algorithm designed for lunar lobate scarps in this study has been successful, offering an insight for the future development of more advanced deep learning–based lunar exploration methods.

Disclosures

No potential conflict of interest was reported by the authors.

Code and Data Availability

The code and data that support the findings of this study are available from the authors upon reasonable request. The data that support the findings of this paper can be requested from the author at lichenya@henu.edu.cn

Acknowledgments

We acknowledge the use of publicly available data from the following sources:

1. LRO NAC products.[48]
2. MRO products.[49]

References

1. T. Lu et al, "The 1: 2,500,000-scale global tectonic map of the Moon," *Sci. Bull.* **67**(19), 1962–1966 (2022).
2. T. R. Watters et al., "Lunar tectonics," *Planet. Tecton.* **11**, 121 (2010).
3. NASA/GSFC/Arizona State University, "Slipher crater: fractured moon in 3-D," 2010, https://www.lroc.asu.edu/images/244.
4. NASA/GSFC/Arizona State University, "Simpelius scarp," 2010, https://www.lroc.asu.edu/images/455.
5. M. E. Banks et al., "Morphometric analysis of small-scale lobate scarps on the Moon using data from the Lunar Reconnaissance Orbiter," *J. Geophys. Res.: Planets* **117**(E12), E00H11 (2012).
6. A. B. Binder and H.-C. Gunga, "Young thrust-fault scarps in the highlands: evidence for an initially totally molten Moon," *Icarus* **63**(3), 421–441 (1985).
7. C. H. Van Der Bogert et al., "How old are lunar lobate scarps? 1. Seismic resetting of crater size-frequency distributions," *Icarus* **306**, 225–242 (2018).
8. T. R. Watters et al., "Global thrust faulting on the Moon and the influence of tidal stresses," *Geology* **43**(10), 851–854 (2015).
9. T. R. Watters et al., "Evidence of recent thrust faulting on the Moon revealed by the Lunar Reconnaissance Orbiter Camera," *Science* **329**(5994), 936–940 (2010).
10. T. R. Watters et al., "Recent tectonic activity on Mercury revealed by small thrust fault scarps," *Nat. Geosci.* **9**(10), 743–747 (2016).
11. W. K. Hartmann and D. R. Davis, "Satellite-sized planetesimals and lunar origin," *Icarus* **24**(4), 504–515 (1975).
12. A. G. W. Cameron and W. R. Ward. "The origin of the Moon," *Abstr. Lunar Planet. Sci. Conf.* **7**, 120 (1976).
13. M. Ćuk and S. T. Stewart, "Making the Moon from a fast-spinning Earth: a giant impact followed by resonant despinning," *Science* **338**(6110), 1047–1052 (2012).
14. R. M. Canup, "Forming a Moon with an Earth-like composition via a giant impact," *Science* **338**, 1052–1055 (2012).
15. J. D. Clark et al. "Investigation of newly discovered lobate scarps: implications for the tectonic and thermal evolution of the Moon," *Icarus* **298**, 78–88 (2017).
16. T. R. Watters et al., "Tectonics and seismicity of the lunar south polar region," *Planet. Sci. J.* **5**(1), 22 (2024).
17. Y. Lou and Z. Kang, "Extract the lunar linear structure information by average filtering method based on DEM data," *Sci. Surv. Mapp.* **43**, 155–160 (2018).
18. A. B. Binder, "Post-imbrium global tectonism: evidence for an initially totally molten Moon," *Moon Planets* **26**, 117–133 (1982).
19. M. S. Robinson et al., "Lunar Reconnaissance Orbiter Camera (LROC) instrument overview," *Space Sci. Rev.* **150**, 81–124 (2010).
20. D. M. Hurwitz et al. "Lunar sinuous rilles: distribution, characteristics, and implications for their origin," *Planet. Space Sci.* **79**, 1–38 (2013).
21. J. Ji et al., "The 1: 2,500,000-scale geologic map of the global Moon," *Sci. Bull.* **67**(15), 1544–1548 (2022).
22. M. Peng et al., "Automated detection of lunar ridges based on DEM data," *Int. Arch. Photogramm. Remote Sens. and Spatial Inf. Sci.* **XLII-2/W13**, 1431–1435 (2019).
23. D. M. Nelson et al., "Mapping lunar Maria extents and lobate scarps using LROC image products," in *45th Annu. Lunar and Planet. Sci. Conf. No. 1777* (2014).
24. L. I. Ke et al., "Geomorphometric multi-scale analysis for the automatic detection of linear structures on the lunar surface," *DIXUE QIANYUAN* **21**, 212–222 (2014).

25. A. A. Micheal et al. "Automatic Graben detection in lunar images using Hessian technique," *J. Indian Soc. Remote Sens.* **42**, 445–451 (2014).
26. R. Moghe and R. Zanetti, "A deep learning approach to hazard detection for autonomous lunar landing," *J. Astronaut. Sci.* **67**, 1811–1830 (2020).
27. P. Yan et al. "A new lunar lineament extraction method based on improved UNet++ and YOLOv5," *Sensors* **24**(7), 2256 (2024).
28. S. Zhang et al., "Detecting lunar linear structures based on multimodal semantic segmentation: the case of sinuous Rilles," *Remote Sens.* **16**(9), 1602 (2024).
29. A. Dosovitskiy, "An image is worth 16x16 words: transformers for image recognition at scale," arXiv:2010.11929 (2020).
30. Z. Liu et al. "Swin transformer: hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vision* (2021).
31. Y. He et al., "MoMFormer: mixture of modality transformer model for vegetation extraction under shadow conditions," *Ecol. Inf.* **83**, 102818 (2024).
32. H. Cao et al. "Swin-unet: Unet-like pure transformer for medical image segmentation," *Lect. Notes Comput. Sci.* **13803**, 205–218 (2022).
33. G. Li et al., "Diamond-Unet: a novel semantic segmentation network based on U-net network and transformer for deep space rock images," *IEEE Geosci. Remote Sens. Lett.* **21**, 8002205 (2024).
34. S. Mattson et al. "Exploring the Moon with LROC-NAC stereo anaglyphs," in *Eur. Planet. Sci. Congr.* (2012).
35. Y. Gu et al., "STHarDNet: Swin transformer with HarDNet for MRI segmentation," *Appl. Sci.* **12**(1), 468 (2022).
36. Z. Liu et al., "Swin transformer v2: scaling up capacity and resolution," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit*. (2022).
37. E. Voita et al., "Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned," arXiv:1905.09418 (2019).
38. K. He et al. "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit*. (2016).
39. J. Xu et al., "Reluplex made more practical: Leaky ReLU," in *IEEE Symp. Comput. and Commun. (ISCC)*, IEEE (2020).
40. H. Zhao et al. "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit*. (2017).
41. J. Dai et al., "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vision* (2017).
42. T.-Y. Lin et al., "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit*. (2017).
43. S. Liu et al., "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit*. (2018).
44. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit*. (2015).
45. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
46. L.-C. Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vision (ECCV)* (2018).
47. J. Chen et al., "Transunet: transformers make strong encoders for medical image segmentation," arXiv:2102.04306 (2021).
48. Lunar Reconnaissance Orbiter Camera, "RDR products," https://wms.lroc.asu.edu/lroc/rdr_product_select.
49. High Resolution Imaging Science Experiment HiRISE Operations Center, "Anaglyphs," https://www.uahirise.org/.

Chenya Li received his BE degree in safety engineering from Zhengzhou University of Aeronautics, Zhengzhou, China, in 2022, and he is working toward the ME degree in computer science and technology at Henan University, Kaifeng, China. His current research interests include deep space exploration and deep learning.

Puyan Xu received his BE degree in software engineering from Henan University of Science and Technology, Luoyang, China, in 2022, and he is currently working toward the MS degree in computer technology at Henan University, Kaifeng, China. His current research interests include SAR image processing and deep learning.

Xin Lu received his BE degree in water supply and drainage science and engineering from Zhongyuan University of Technology, Zhengzhou, China, in 2022, and he is working toward

the ME degree in computer science andtechnology at Henan University, Kaifeng, China. His current research interests include deepspace exploration and deep learning.

Zhiyuan Guo received his bachelor's degree in materials forming and control engineering from Henan University of Technology, China, in 2021 and is currently pursuing a master's degree in computer technology at Henan University, Kaifeng, China. His current research interests include deep space exploration and image processing.

Ning Li received his PhD in communications and information systems from the Institute of Electronics, Chinese Academy of Sciences (IECAS), Beijing, China, in 2015. Since December 2017, he has been a full professor at the School of Computer and Information Engineering, Henan University, Kaifeng, China. His research interests include SAR and inverse SAR imaging algorithms and autofocusing techniques, SAR polarimetric theory, and SAR image processing.

Gaofeng Shu received his BS degree from Wuhan University, Wuhan, Hubei, China, in 2015, and his PhD from the Department of Space Microwave Remote Sensing Systems, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2021. Since 2021, he joined the School of Computer and Information Engineering, Henan University, Kaifeng, China. His research interests include synthetic aperture radar (SAR) imaging, orbital angular momentum (OAM), and electromagnetic vortex technology in SAR systems.