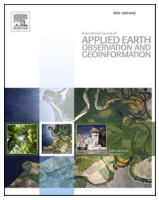
International Journal of Applied Earth Observation and Geoinformation 134 (2024) 104215

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Dual-branch multi-modal convergence network for crater detection using Chang'e image

Feng Lin [a,b], Xie Hu [a,*], Yiling Lin [a], Yao Li [c], Yang Liu [b], Dongmei Li [d]

[a] College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
[b] National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China
[c] State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China
[d] School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China

ARTICLE INFO

Keywords:
Moon
Crater detection
Multi-modal remote sensing
Deep learning
Data fusion
Attention

ABSTRACT

Knowledge about the impact craters on rocky planets is crucial for understanding the evolutionary history of the universe. Compared to traditional visual interpretation, deep learning approaches have improved the efficiency of crater detection. However, single-source data and divergent data quality limit the accuracy of crater detection. In this study, we focus on valuable features in multi-modal remote sensing data from Chang'e lunar exploration mission and propose an Attention-based Dual-branch Segmentation Network (ADSNet). First, we use ADSNet to extract the multi-modal features via a dual-branch encoder. Second, we introduce a novel attention for data fusion where the features from the auxiliary modality are weighted by a scoring function and then being fused with those from the primary modality. After fusion, the features are transferred to the decoder through skip connection. Lastly, high-accuracy crater detection is achieved based on the learned multi-modal data features through semantic segmentation. Our results demonstrate that ADSNet outperforms other baseline models in many metrics such as IoU and F1 score. ADSNet is an effective approach to leverage multi-modal remote sensing data in geomorphological feature detection on rocky planets in general.

1. Introduction

Craters formed by small impactors are typical landforms on rocky planets. The geographic locations and occurrence frequency of these impact craters offer valuable insights into the geological era of a celestial body (Neukum et al., 2001), the development of geological formations, and the chronological order of geological processes (Fassett et al., 2012; Wu et al., 2018; 2019). Crater inventory can also help plan for the landing sites in future space exploration missions (Bernard and Golombek, 2001; Wu et al., 2020).

Efforts have been taken towards compiling a global catalog of lunar impact craters. A global inventory of lunar impact craters was presented by Head et al. (2010), encompassing data on 5,185 craters with diameters of 20 km and larger. Other researchers (Barlow, 2017; Krüger et al., 2018; Salamunićcar et al., 2014) produced global catalogs including craters with diameters of 3 km and larger. Robbins (2019) compiled a lunar impact crater catalog with diameters more than 1–2 km. Wang and Wu (2020) initiated the development of a comprehensive catalog of lunar craters. Subsequently, Wang and Wu (2021) released a crater catalog, including 1,319,373 craters with diameters larger than 1 km.

Currently, records and chronologies of impact craters play a crucial role in theories regarding the evolution of celestial bodies. Additionally, the distribution of craters in variant dimensions can be used to infer the size of impactors (Strom et al., 2005). For example, Minton et al. (2015) contended, based on the distribution of craters, that the impactors comprise relatively fewer large-sized bodies compared to those found in the asteroid belt today.

In traditional approaches, craters are manually detected through the visual interpretation. However, inventories generated by this method are often limited to large impact craters or particular geographic areas (Bandeira et al., 2012). Additionally, visual interpretation may result in up to 40 % discrepancy among different human experts (Robbins et al., 2014). Consequently, researchers have designed crater detection algorithms based on feature extraction, machine learning, and neural networks to achieve the automation of crater classification. Emami et al. (2015) proposed a feature-based edge detection algorithm. Machine learning methods have been employed for crater detection by various

* Corresponding author.
E-mail address: hu.xie@pku.edu.cn (X. Hu).

https://doi.org/10.1016/j.jag.2024.104215
Received 1 May 2024; Received in revised form 25 August 2024; Accepted 10 October 2024
Available online 17 October 2024

researchers. Salamunićcar and Lončarić (2010) utilized the Hough Transform to detect craters. Wetzler et al. (2005) applied SVM, and Stepinski et al. (2012) employed Decision Trees. Additionally, some researchers employed neural networks to achieve automation in crater detection (Cohen et al., 2016; Palafox et al., 2017). Di et al. (2014) also attempt multi-step approaches. They utilized an enhancement algorithm to enclose regions containing craters, followed by the Hough Transform for delineating their edges. Similarly, Boukercha et al. (2014) employed a method, where crater candidates provided by initial detection algorithms were subsequently categorized as true positive or false positive using SVM or polynomial classifiers. Wang et al. (2022) further addressed the intricate challenges by proposing a method for detecting multi-scale lunar craters, accounting for diverse characteristics such as shape, lighting conditions, and size. In this method, they derived the feature in different light condition for detecting craters. Subsequently, they utilized the isolation forest algorithm to remove false craters. These models often rely on single data source, making them less effective when the dataset vary significantly in illumination across regions.

Researchers have started using multi-modal approaches to improve the performance. The primary remote sensing data used for planetary crater detection include Digital Elevation Models (DEM), near-infrared images, and Digital Orthophoto Maps (DOM). The data possess distinct characteristics: DOMs and infrared images are influenced by solar angles, resulting in oversaturation; DEMs are unaffected but lack spectral and physical information. If the unique features of different datasets are jointly utilized, it can achieve complementary information and lead to better performance. To leverage the complementary information of optical images and DEMs, Mao et al. (2022) proposed a CNN with dual-path construction that consolidates features. Similarly, the ordinary convolutional blocks can be replaced with residual blocks in the encoder to extract DEM and optical image features separately (Wang et al., 2020; Lee et al., 2021). The DEM features extracted in the bridging network are integrated with the features of optical images. The decoder network incorporates attention to expand features, enhancing feature information. However, these methods do not assign different attention levels to different modalities, which may lead to incomplete usage of the effective features of each modality and unwarranted noise in acquiring multi-modal features. This will result in a slower convergence rate during model optimization and will also lead to local optima, thus influencing the enhancement of model accuracy.

We propose the Attention-based Dual-branch Segmentation Network (ADSNet), which leverages multi-modal data by obtaining the importance of different features through an attention and utilizes skip connections to transfer the fused data features to the decoder. Finally, Concurrent Spatial and Channel 'Squeeze & Excitation' (scSE) is employed in the decoder to enhance data features by fully utilizing the different modal features on impact craters. The main contributions of our work are as follows:

(1) To the best of our knowledge, this is the first attempt to utilize the novel attention in the fusion step of multi-modal remote sensing data for impact crater detection. By allowing the model to assign varying degrees of attention to different data sources, this approach increases the available features and improves the accuracy of impact crater detection.
(2) Our proposed model fully considers multi-modal features by integrating the features of impact craters from multi-modal remote sensing data using attention and transferring them to the decoder. Subsequently, scSE is employed to enhance the crucial parts of the features, followed by decoding based on the new features. In other words, we make full use of the features from different modalities, thereby enhancing the model's generalizability and accuracy in impact crater detection.
(3) Systematic experiments are conducted on impact crater detection to verify the performance of the ADSNet model. Our results disclose that multi-modal data feature fusion contributes to a higher accuracy of impact crater detection than single-modal scenario. Furthermore, the ADSNet is compared with the baseline models, demonstrating its outstanding performance.

2. Related work

2.1. Semantic Segmentation-Based crater detection

The crater detection method based on semantic segmentation networks typically involves classifying each pixel in an image as either a crater or non-crater. Most semantic segmentation-based methods follow an encoder-bridge-decoder structure. The networks consist of multiple encoder and decoder blocks, respectively. They function as feature extractors through a series of encoder blocks ($[E_1, E_2, E_3, \ldots, E_i, \ldots, E_n]$) and gain knowledge about the abstract patterns from input images. The encoder block E_i receives downsampling output from E_{i-1} as an input, and its output is transferred to both decoder D_i and encoder E_{i+1}. Similarly, D_i receives input from E_i concatenated with the upsampling from D_{i+1}. The encoder network increases the number of feature channels while reducing the spatial dimensions of features to capture high-dimensional semantic information. Some researchers have adopted the idea of residual connections introduced by He et al. (2016), proposing residual convolution (Wang et al., 2020), residual blocks (Lee and Hogan, 2021), and a specialized residual block (Mao et al., 2022).

Ronneberger et al. (2015) proposed U-Net to segment biomedical images. U-Net has been widely used in various directions of semantic segmentation. Some researchers improved the U-Net model in terms of the filter quantity, kernel size, and depth, achieving advancements in impact crater detection (Silburt et al., 2019; Lee et al., 2019).

U-Net architecture exhibits inefficiency in multi-scale information fusion and fails to fully exploit high-resolution contextual information from multiple images. We still lack a learning mechanism to enhance features across various scales and to facilitate the transmission of features in the network for detecting small and overlapping craters. Consequently, researchers have begun to explore the creation of various U-Net variants aimed at achieving more precise segmentation outcomes. He et al. (2016) introduced a residual module through shortcut connections to streamline training and mitigate degradation issues in training deep networks. Wang et al. (2020) proposed an architecture called ERU-Net for crater detection, which replaces the convolutional blocks in U-Net with residual modules called "Residual Conv". Similarly, Lee et al. (2021) utilized the ResU-Net architecture (Zhang et al., 2018), which also employs residual units from U-Net to enhance the learning capacity of the network.

In a typical architecture, skip connections link encoder blocks to decoder blocks, mitigating the loss of high-level features induced by downsampling. In the work by Jia et al. (2021), skip connections were substituted with nested dense connections to enhance the preservation of high-level features. They introduced NAU-Net, which combines U-Net and attention (Oktay et al., 2018) with nested dense connections. The attention helps enhance feature extraction capabilities for detecting overlapping craters. Chen et al. (2021) utilized the HRNet (Sun et al., 2019) framework for crater detection on the lunar. HRNet preserves the high-resolution information and surpasses the constraints of the U-Net by learning comprehensive data representations through multi-scale fusion of input data. Juntao et al. (2024) proposed a knowledge-aware network for the detection of small-scale craters utilizing DEM. However, these methods rarely consider attention in fusion, which fails to highlight the contributions of significant features within multimodal feature layers, and thus limits the improvements in detection accuracy.

2.2. Multi-modal remote sensing

Numerous classic shallow multi-modal models are utilized for feature extraction of multi-modal remote sensing observations. For example, Liao et al. (2014) proposed a model with graph-based subspace

learning to fuse hyperspectral and Light Detection and Ranging (LiDAR)-derived morphological profiles. Hong et al. (2020a,b) presented five different fusion strategies for spatial information modeling tasks and pixel classification tasks. Hu et al. (2021) collaboratively employed 23 interdisciplinary datasets to characterize the structural, hydrological, and anthropogenic origins of surface deformation. Yokoya et al. (2018) combined multiple features extracted from multispectral and OpenStreetMap data and then inputted them into a classifier for local climate zone classification. Hu et al. (2019) proposed a mapper-based manifold alignment technique, achieving semi-supervised fusion of polarimetric SAR images and hyperspectral. Several subsequent studies (Gu et al., 2015; Liu et al., 2019; Hong et al., 2020; Hu et al., 2021; Liu et al, 2022; Yu et al., 2024) were conducted to improve the capacity of information fusion between multiple modalities using more advanced methods.

Due to its ability to capture finer and richer scene features, deep learning techniques have made significant advancements in multi-modal feature extraction. Researchers have been exploring the possibilities of developing multi-modal deep learning and its variants to effectively classify multi-modal remote sensing images. These models fall into two main directions. One category is the commonly used pixel-level multi-modal networks. Ghamisi et al. (2016) extracted extinction profiles from LiDAR data and hyperspectral. They fused them by CNNs in the deep feature space. Additionally, Xu et al. (2017) proposed the use of a CNN with cascading blocks for feature extraction and fusing multi-modality. Chen et al. (2017) proposed an end-to-end fusion network where feature was extracted by two CNNs and fused by a DNN. Benedetti et al. (2018) developed a DL-based generic framework to fuse multi-temporal and multi-modal satellite data. Another category is semantic segmentation, which involves assigning a semantic class to each pixel in an object-level manner. For example, Xu et al. (2019) achieved excellent semantic segmentation results through fusion fully convolutional networks on multispectral LiDAR data and hyperspectral data.

3. Data and method

Utilizing a complementary strategy between two types of images, we explore a dual-branch convolutional neural network framework based on encoder-bridge-decoder architecture for crater detection, termed as the ADSNet. Generally, the encoder-bridge-decoder structure excels at handling images with simple semantic structures, and lunar images possess intuitive semantics and representative structures. Therefore, we employ an encoder-bridge-decoder structure to construct a dual-branch CNN framework (Fig. 1). It consists of dual-branch encoding, bridging layer, and decoding. The dual-branch encoding extracts features from DEM and DOM data, and these features are fused with an attention. The bridge layer uses the skip-connection to pass fusion features from the encoders to the decoders. The decoding process combines the output from the previous decoder with the fusion result from the bridging layer, serving as the input for the current decoder. This approach utilizes successive convolutions of the feature map to effectively recover the image information.

3.1. Data

We utilize two modal images of lunar (DOM and DEM), both obtained from the CCD stereo camera carried by the Chang'e-2 mission. Fig. 2 showcases these two types of images. The Chang'e lunar exploration missions are lunar exploration projects initiated by the China National Space Administration (CNSA) in 2007. The objective of these missions is to achieve lunar exploration, sample return, and to explore scientific questions such as lunar surface geology and resource distribution. Chang'e 2 is extremely successful in China's lunar orbit exploration. Its scientific tasks, including high-resolution lunar image capture and lunar surface elevation mapping conducted in lunar orbit, provide crucial data for the advancement of lunar exploration missions. The DEM images used in the experiments have a resolution of 20 m/pixel in dimensions of 935 × 931 pixels. The DOM images have a resolution of 7 m/pixel in dimensions of 2,669 × 2,657 pixels. The DEM images were aligned to the DOM image, covering an area of approximately 18 × 18 km on the lunar surface. Most of the current lunar crater catalogs are at the kilometer scale, we aim to detect more and smaller craters (>50 m in diameter) that the DEM image cannot achieve alone. To better utilize the features of the DEM image in the model, we compute the slope of the DEM and used it as the input.

We crop the data into multiple patches, specifying the size of the dual-branch input images as 256 × 256 pixels. We focus on identifying craters in the images with a diameter greater than 7 pixels, allowing us to extract craters with a minimum diameter of 50 m. After data processing, we obtain 2,550 training samples, 500 validation samples, and 500 test samples. Each sample consists of a pair of DOM and DEM images

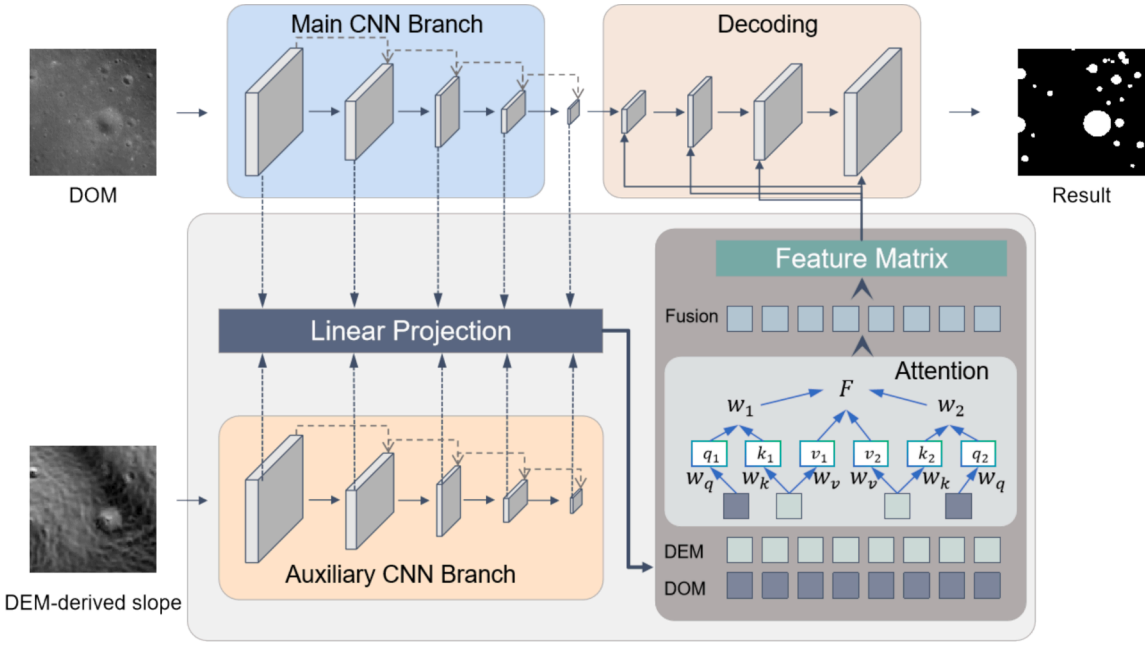


Fig. 1. The schematic diagram of the ADSNet.

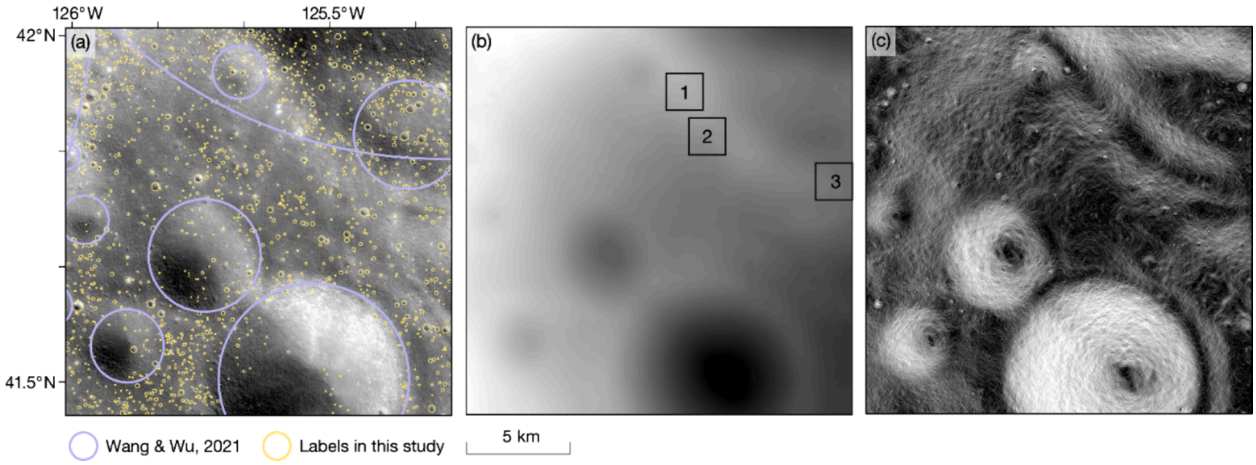


Fig. 2. (a) DOM, (b) DEM, and (c) the DEM-derived slope. Purple circles delineate craters from inventory generated by Wang and Wu (2021). Yellow polygons are labels used in this study.

of craters. Furthermore, we utilize existing crater catalog (Wang et al., 2021) to create semantic segmentation labels.

3.2. Dual-branch encoding

The encoding consists of two independent input branches (one main branch and one auxiliary branch) to address DOM and DEM images (Fig. 1). Each branch in the dual encoding comprises 5 encoder units and every unit consists of residual unit and max-pooling layers. Fig. 3 illustrates the structure of the residual unit, which consists of 2 convolutional layers, 2 batch normalization, and a rectified linear unit. The feature fusion of the two contracting paths is achieved using attention. It integrates the features of the main path and the features calculated through attention in the auxiliary path into the final features. Meanwhile, the features of the main branch are preserved.

3.2.1. Residual unit

We introduce residual units instead of regular units to further enhance the model performance. Deeper networks often degrade due to the vanishing gradient. To solve this problem, He et al. (2016) introduced residual neural networks. Fig. 3 illustrates the structure of the residual units. The residual unit consists of a combination of BN, ReLU activation, and convolutional layers. Based on this, we construct the encoding part of the ADSNet model, leading to two benefits: (1) Simplified training process due to the presence of residual units; (2) Facilitated information propagation without degradation between residual units and higher-level networks through skip connections. Consequently, a neural network with fewer parameters is designed with a better performance in semantic segmentation.

3.2.2. Fusion based on attention

In this paper, the data used for crater detection consist of DOM and DEM. We aim to fully leverage the diverse data features instead of directly concatenating them as previous studies. Therefore, we employ the scaled dot-product attention for computing scores. To meet the requirements of attention computation, the features obtained from the convolutional layers, namely the DOM and DEM feature matrices, must first be reshaped into patches. Next, these reshaped features are projected into a lower-dimensional embedding space through a linear transformation for fusion. Finally, after completing the fusion, the feature maps need to be reshaped back into their original matrix form to facilitate the decoder in recovering the image features. Specifically, we use the following dot product equation to compute the attention scores:

$$W = softmax\left(\frac{W_q \cdot F_O \cdot (W_k \cdot F_E)^T}{d_k}\right) \quad (1)$$

where W_q and W_k are the weight matrices used for the linear transformation, F_O and are F_E are the patches reshaped from the DOM and DEM feature matrices, respectively, d_k is the dimensionality of the embedding, and W represents the computed weight matrix. For our dataset, the resolution of the DOM data (7 m) is better than that of the DEM data (20 m). Therefore, we choose DOM as the primary modality to preserve the features while extracting effective features from the DEM through attention and reduced the influence of noise therein. The workflow of the attention in ADSNet is shown in Fig. 4. Thus, the formula for computing the final fused features is given by,

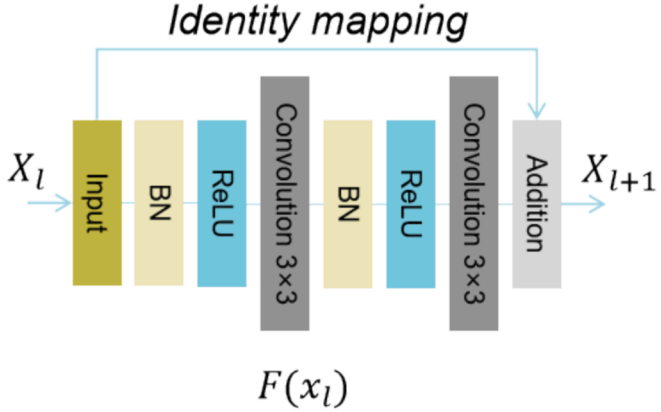


Fig. 3. The structure of residual unit.

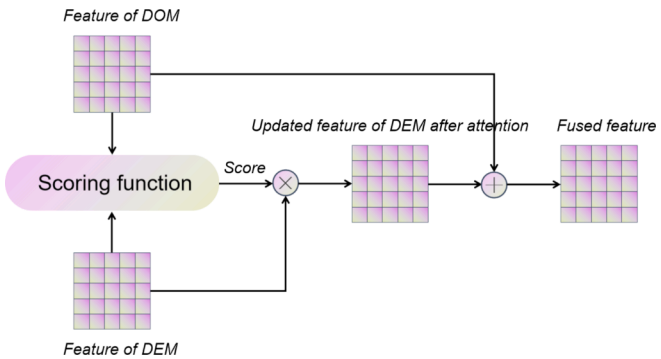


Fig. 4. The workflow of the attention in ADSNet.

$$Feature = F_1 + WF_2 \quad (2)$$

where F_1 is the feature matrix of the DOM, F_2 is the feature matrix of the DEM, and *Feature* is the feature matrix obtained after fusing.

3.3. Bridging layer

The bridging layer connects the encoding and decoding parts to facilitate information propagation. We perform concatenation operations for each layer of encoding and fuse the feature maps of the two residual convolution blocks involved in the dual branch (Fig. 1).

In the bridging pathway, we use 1×1 convolutional layers to mitigate aliasing effects during upsampling. As a result of the downsampling through pooling in encoding, semantic information is diminished in the low-level features, thereby achieving precise target localization. Conversely, high-level feature maps have stronger semantic information but coarse target localization. Therefore, the integrity of both positional and semantic information can be enhanced by connecting each layer of encoding through convolution.

3.4. Decoding

We restore the image to obtain precise segmentation of craters using decoding. The decoding consists of 4 decoder blocks, each containing two 3×3 convolutional layers and an scSE module (Fig. 5). Finally, we use a convolution to reduce the number of features to one. Subsequently, the sigmoid function is applied to obtain the output to be either 0 or 1.

In the decoding, we introduce "Feature recalibration module with spatial and channel 'squeeze & excitation' block" (scSE) to further optimize the feature space. Human visual attention involves focus on the areas of interest. Similarly, attention operations in deep learning enable the selection of critical information from vast feature spaces, thereby enhancing recognition performance. Specifically, the scSE module (Roy et al, 2018) consists of Spatial Squeeze & Excitation (sSE) and Channel Squeeze & Excitation (cSE) modules. The sSE module applies the attention weights to the spatial dimension of the input feature to highlight important spatial locations. Conversely, the cSE module applies the attention weighting to the channel dimension of the feature to capture correlations in channel-level. The combination of these two modules allows for a more effective utilization of the different information in input features, thereby improving the model performance.

4. Experiments

To objectively measure the performance of the model, we evaluate the model using multiple metrics. First, we validate the effectiveness of multi-modal image inputs and the impact of integrating the scSE module in the decoding phase. Second, we compare the performance of the ADSNet with several other models to assess its improvement. We provide detailed experimental setup in Section 4.1. Then, we compare the performance of the ADSNet and other baseline models in Section 4.2. Finally, in Section 4.3, we discuss the ablation experiments conducted on the ADSNet.

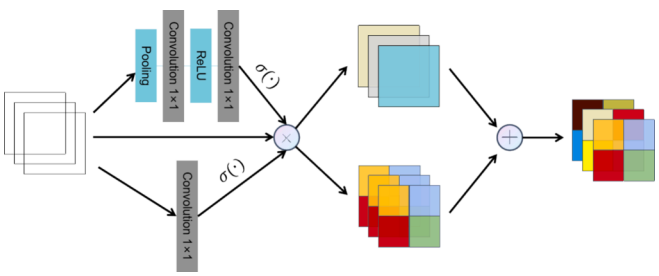


Fig. 5. The structure of scSE.

4.1. Experimental setup

4.1.1. Evaluation protocol

In the field of semantic segmentation, there are several commonly used evaluation metrics for assessing the model performance. For the task at hand, where pixels need to be classified into two categories (i.e., impact craters and non-impact craters), these metrics facilitate effective semantic segmentation and crater identification. Subsequently, model predictions are quantitatively compared with ground truth labels to evaluate model performance.

We choose four main evaluation metrics are often utilized in semantic segmentation tasks. *Mean Intersection over Union* (*mIoU*) assesses the degree of overlap between model predictions and the corresponding ground truth labels. *Pixel Accuracy* (*PA*) gauges the accuracy of pixel classification by determining the ratio of correctly classified pixels to the total number of pixels. *Mean Pixel Accuracy* (*mPA*) averages the pixel accuracy for each class, addressing the sensitivity of pixel accuracy to class imbalance. *F1 Score* is a comprehensive metric that considers both *Precision* and *Recall*, and it can also be regarded as a variant of *IoU*.

While the aforementioned evaluation metrics provide objective assessments of semantic segmentation models, for the specific task of crater detection, the prevalence of non-crater pixels in the image may lead to inflated evaluation metrics for the non-impact crater class. Consequently, the overall metrics such as *mIoU*, *mDice,* and *mPA* will be biased. To provide a more objective evaluation of our model's performance in impact crater detection, our experiments not only consider the overall metrics but also include metrics for the crater class such as *IoU*, *F1 Score*, and *cPA* (*Class Pixel Accuracy*).

4.1.2. Baselines

We establish two baselines to validate the effectiveness of the ADSNet. The first baseline consists of other semantic segmentation models: (1) DeeplabV3+ (Chen et al., 2018), (2) DMNet (He et al., 2019), (3) DANet (Fu et al., 2019), (4) Semantic FPN (Kirillov et al., 2019), and (5) SAN (Xu et al., 2023).

The second baseline involves ablation experiments. In the encoding phase, we employ a dual-branch strategy to encode both inputs simultaneously and utilize the scaled dot-product attention for fusion. In the decoding phase, we introduce the scSE module to enhance image features of critical areas of interest. Here our ablation experiments include three scenarios: (1) Exclude the dual-branch strategy and only input the 7-m-resolution DOM image; (2) Adopt the dual-branch strategy in traditional fusion methods without the scaled dot-product attention; (3) Employ the dual-branch strategy and the scaled dot-product attention for fusion, but omit the scSE module in decoding, where each encoding block only contains the convolutional layers.

4.1.3. Experiment settings

We set some parameters as constants throughout the entire experiment (e.g., number of epochs = 100, image size = 512, learning rate = 0.0001, patience = 5, and class weights = 1.0). Here, patience = 5 means that in the validation set, the model's effectiveness does not improve continuously by 5 times.

In the field of semantic segmentation, researchers often use *Cross-Entropy Loss* as a loss function. However, as discussed in Section 4.1.1, in the task of impact crater detection, there is a significant imbalance between impact craters and non-impact crater pixels, which challenges the model training and degrades the model performance. We opt to use the *Focal Loss* (Eq (3)), which addresses the issue of class imbalance problem by balancing the weights of easy and hard to classify samples.

$$FocalLoss = -(1 - p_t)^{\gamma} log(p_t) \quad (3)$$

where p_t represents the probability of predicting a certain sample as a positive sample, γ is a hyperparameter for tuning, a higher γ reduces the contribution of losses from easy samples, $(1 - p_t)^{\gamma}$ is used to adjust the

weights of easy and hard samples. When an impact crater is misclassified, p_t is small, $(1-p_t)^\gamma$ approaches 1, and its loss is hardly affected. Conversely, when p_t approaches 1 (suggesting good classification prediction for simple samples), $(1-p_t)^\gamma$ approaches 0, and its loss is reduced. After multiple validations, we choose $\gamma = 1$ as the hyperparameter.

4.2. Comparisons with other models

We compare ADSNet with other models in the field of semantic segmentation (Fig. 6). The results show that ADSNet outperforms most other models on many metrics (Table 1).

First, as a multi-modal model, ADSNet benefits from richer input features compared to the single-modal models and effectively leverages the multi-modal features via the scaled dot-product attention. Second, the residual units in ADSNet address the issue of vanishing and exploding gradients, thereby enhancing the model's performance. Lastly, the integration of scSE in the decoder part of ADSNet further enhances image features, resulting in superior accuracy. These reasons give ADSNet better performance.

Additionally, we note that in certain metrics, the performance of ADSNet is inferior to the comparison models. For example, Semantic FPN performs the best in terms of mIoU, and DANet performs the best in terms of Recall. However, it is evident that any individual metric can hardly fully characterize the performance of impact crater detection. In our analysis of the performance of ADSNet and DANet in a selected area, we note that misalignments at the edges of larger impact craters significantly affect detection accuracy more than the omission of smaller craters. It is evident that DANet fails to identify a number of small impact craters (Fig. 7). Nevertheless, DANet's delineation of large impact craters in this area aligns more closely with the actual data, resulting in similar overall metrics for both models.

4.3. Ablation experiment

Here we discuss the impact of various remote sensing images on our model. We note that the inputs of multiple remote sensing images and the fusion methods of image features both affect the model's performance. We also discuss the effects of residual units and scSE.

The ADSNet utilizes the scaled dot-product attention to fuse multi-modal data. Based on the results in Table 2, we affirm that the design of ADSNet is rational and effective. Specifically, we can draw the following four conclusions:

(I) When the residual units are not introduced, the model performance greatly declines due to vanishing gradient.

(II) When ADSNet uses single-modal inputs, there is a significant decrease in all performance metrics. This is attributed to the limited feature information compared to multi-modal inputs.

(III) When ADSNet relies on traditional fusion methods without the attention to fuse multi-modal features, its performance is notably better than that of the single-modal input ADSNet. However, it cannot focus on the area of interest. This is due to the presence of noise and irrelevant features in the directly fused information.

(IV) When ADSNet does not incorporate the scSE module, its performance slightly deteriorates compared to the original model. This is because the absence of spatial and channel attention affects the decoding results.

We also note that metrics such as *mIoU*, *mdice*, *PA*, and *mPA* show minimal variations across different experiments. This is because the inclusion of both crater and non-crater pixel results in these metrics. The dominance of non-crater pixels in the image leads to an overestimation of non-crater class metrics, resulting in relatively consistent metric changes across experiments. The effectiveness of ADSNet is undeniable, demonstrating that the fusion of multi-modal information provides richer information for crater detection. The use of attention enables the model to precisely detect craters.

Fig. 8 demonstrates the feature maps in ADSNet, visually explaining its decision-making process in completing the crater detection task. To determine the importance of different features, the scaled dot-product attention computes the attention scores. In the context of ADSNet, these features are derived from both DOM and DEM inputs. The attention assigns weights to these features based on their relevance to the task of crater detection. Higher weights indicate greater importance, allowing the model to focus on the most informative aspects of the data. The attention selectively highlights relevant features, guiding the network to focus on areas that are most likely to contain craters. This selective attention is crucial for distinguishing between craters and other similar surface features.

Early layers of the decoding, influenced by the attention scores, capture surface patterns and elevation changes. As the decoding proceeds, subsequent layers refine these maps by incorporating more detailed features, such as the precise contours of craters. The attention ensures that these detailed features are accurately integrated from both DOM and DEM data, leading to more precise crater detection.

4.4. Crater detection results

We compare the detection results of the model with the existing catalog of lunar impact craters. The existing catalog contains 7 impact craters with diameters on the order of kilometers in our study area, while ADSNet identifies 1537 impact craters with diameters greater than 50 m, two orders of magnitude smaller than the previous attempts.

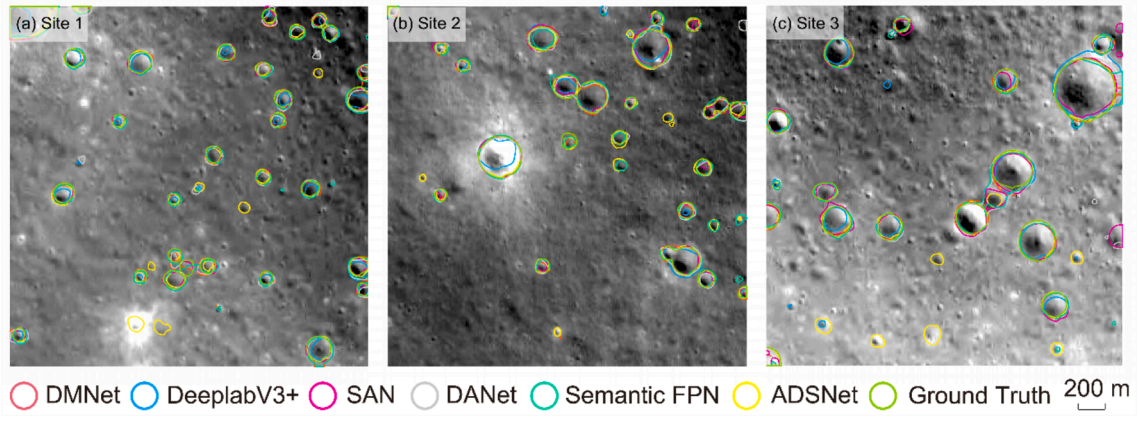


Fig. 6. Results of ADSNet and baseline models.

Table 1
Comparison of ADSNet and baseline. Bold digits refer to the best performance for the corresponding metrics.

	Precision	Recall	F1 score	mIoU	IoU	mDice	PA	mPA
DeeplabV3+	0.8270	0.5376	0.6516	0.7301	0.4832	0.8200	0.9775	0.7665
DMNet	0.8572	0.8323	0.8446	0.8593	0.7309	0.9192	0.9880	0.9133
DANet	0.8635	**0.8770**	0.8702	0.8798	0.7703	0.9324	0.9898	0.9357
Semantic FPN	0.8961	0.8099	0.8508	**0.8945**	0.7404	0.9225	0.9889	0.9030
SAN	0.9039	0.8413	0.8714	0.8811	0.7722	0.9332	0.9903	**0.9188**
ADSNet	**0.9265**	0.8265	**0.8736**	0.8830	**0.7756**	**0.9344**	**0.9906**	0.9119

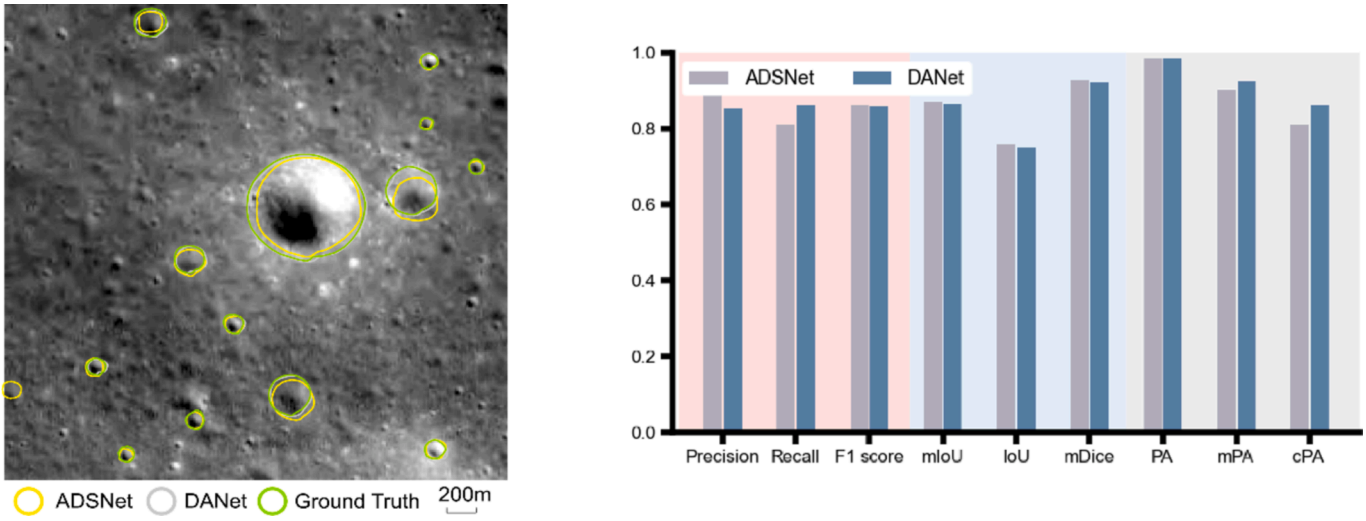


Fig. 7. Comparison between ADSNet and DANet.

Table 2
Ablation experiment on ADSNet.

	ADSNet w/o residual unit	ADSNet w/o multi-modal	ADSNet w/ o attention	ADSNet w/o scSE	ADSNet
Precision	0.8165	0.8574	0.8824	0.9209	**0.9265**
Recall	0.7965	0.7781	0.8003	0.8125	**0.8265**
F1 score	0.8064	0.8158	0.8394	0.8633	**0.8736**
mIoU	0.8300	0.8374	0.8553	0.8745	**0.8830**
IoU	0.6756	0.6889	0.7232	0.7594	**0.7756**
mDice	0.8993	0.9043	0.9166	0.9290	**0.9344**
PA	0.9850	0.9862	0.9880	0.9899	**0.9906**
mPA	0.8946	0.8864	0.8980	0.9048	**0.9119**
cPA	0.7965	0.7781	0.8003	0.8125	**0.8265**

ADSNet evidently complement our knowledge about craters in small dimensions.

5. Conclusion

To address the issue of previous automatic crater detection methods that overlooked multi-model data including surface elevation and optical textures, this study proposes the ADSNet method to utilize multi-modal remote sensing images and to integrate impact crater contextual information from different modalities. The incorporation of encoder residual units and decoder scSE enhances the convergence speed and accuracy of the model. By introducing attention to fuse multi-modal features, ADSNet extracts important features and reduces noise. By conducting experiments systematically, we demonstrate the significance of multi-modal features and attention for improving the accuracy of crater detection. In the future, we will complement multi-scale lunar craters through data resampling and conduct a comprehensive catalog of the entire lunar craters. Additionally, we will explore other optimization mechanisms and incorporate small craters to improve lunar chronology theory. Overall, our study provides a promising tool for geomorphological feature detection on rocky planets in general.

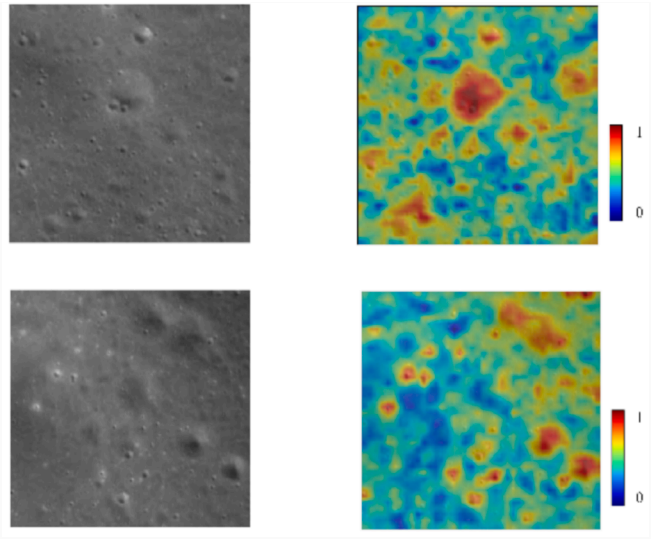


Fig. 8. Sample data and their feature maps in ADSNet.

CRediT authorship contribution statement

Feng Lin: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Xie Hu:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Yiling Lin:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Yao Li:**

Writing – review & editing, Supervision, Investigation, Formal analysis. **Yang Liu:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Dongmei Li:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (42241132), the Youth Open Grant of the National Space Science Center of the Chinese Academy of Sciences (NSSDC2302002), and the Distinguished Young Scholars of the National Natural Science Foundation of China (Overseas).

Data availability

Data will be made available on request.

References

Bandeira, L., Ding, W., Stepinski, T.F., 2012. Detection of subkilometer craters in high resolution planetary images using shape and texture features. Adv. Space Res. 49, 64–74.

Barlow, N.G., 2017. Status report on crater databases for Mercury, the Moon, Mars, and Ganymede. In Third Planetary Data Workshop and the Planetary Geologic Mappers Annual Meeting. 1986, 7027.

Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R., Dupuy, S., 2018. M3Fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (12), 4939–4949.

Boukercha, A., Al-Tameemi, A., Grumpe, A., Wöhler, C., 2014. Automatic crater recognition using machine learning with different features and their combination. In 45th Annual Lunar and Planetary Science Conference , No. 1777, 2842).

Chen, Y., Li, C., Ghamisi, P., Jia, X., Gu, Y., 2017. Deep fusion of remote sensing data for accurate classification. IEEE Geosci. Remote Sens. Lett. 14 (8), 1253–1257.

Chen, S., Li, Y., Zhang, T., Zhu, X., Sun, S., Gao, X., 2021. Lunar features detection for energy discovery via deep learning. Appl. Energy 296, 117085.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: In Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.

Cohen, J.P., Lo, H.Z., Lu, T., Ding, W., 2016. Crater detection via convolutional neural networks. In: In: Lunar and Planetary Science Conference. Vol. 47 of Lunar and Planetary Inst. Technical Report, p. 1143.

Di, K., Li, W., Yue, Z., Sun, Y., Liu, Y., 2014. A machine learning approach to crater detection from topographic data. Adv. Space Res. 54 (11), 2419–2429.

Emami, E., Bebis, G., Nefian, A., Fong, T., 2015. Automatic crater detection using convex grouping and convolutional neural networks. Springer, Cham, pp. 213–224.

C. I. Fassett et al., 2012. Lunar impact basins: Stratigraphy, sequence and ages from superposed impact crater populations measured from Lunar Orbiter Laser Altimeter (LOLA) data. J. Geophys. Res., 117(E12) 2011JE003951, Dec. 2012.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3146–3154.

Ghamisi, P., Höfle, B., Zhu, X.X., 2016. Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (6), 3011–3024.

Golombek, M., Bernard, D., 2001. Crater and rock hazard modeling for Mars landing. In: In AIAA Space 2001 Conference and Exposition, p. 4697.

Gu, Y., Wang, Q., Jia, X., Benediktsson, J.A., 2015. A novel MKL model of integrating LiDAR data and MSI for urban area classification. IEEE Trans. Geosci. Remote Sens. 53 (10), 5312–5326.

He, J., Deng, Z., Qiao, Y., 2019. Dynamic multi-scale filters for semantic segmentation. In: In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3562–3572.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern recognition,770–778.

Head III, J.W., Fassett, C.I., Kadish, S.J., Smith, D.E., Zuber, M.T., Neumann, G.A., Mazarico, E., 2010. Global distribution of large lunar craters: Implications for resurfacing and impactor populations. Science 329 (5998), 1504–1507.

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B., 2020a. More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. IEEE Trans. Geosci. Remote Sens. 59 (5), 4340–4354.

Hong, D., Yokoya, N., Xia, G.S., Chanussot, J., Zhu, X.X., 2020b. X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data. ISPRS J. Photogramm. Remote Sens. 167, 12–23.

Hu, X., Bürgmann, R., Xu, X., Fielding, E., Liu, Z., 2021. Machine-learning characterization of tectonic, hydrological and anthropogenic sources of active ground deformation in California. J. Geophys. Res. Solid Earth 126 (11) e2021JB022373.

Hu, J., Hong, D., Zhu, X.X., 2019. MIMA: MAPPER-induced manifold alignment for semi-supervised fusion of optical image and polarimetric SAR data. IEEE Trans. Geosci. Remote Sens. 57 (11), 9025–9040.

Jia, Y., Liu, L., Zhang, C., 2021. Moon impact crater detection using nested attention mechanism based UNet++. IEEE Access 9, 44107–44116.

Juntao, Y., Shuowei, Z., Lin, L., Zhizhong, K., Yuechao, M., 2024. Topographic knowledge-aware network for automatic small-scale impact crater detection from lunar digital elevation models. Int. J. Appl. Earth Obs. Geoinf. 129, 103831.

Kirillov, A., Girshick, R., He, K., Dollár, P., 2019. Panoptic feature pyramid networks. In: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6399–6408.

Krüger, T., Hergarten, S., Kenkmann, T., 2018. Deriving morphometric parameters and the simple-to-complex transition diameter from a high-resolution, global database of fresh lunar impact craters (D≥~3 km). J. Geophys. Res. Planets 123 (10), 2667–2690.

Lee, C., 2019. Automated crater detection on Mars using deep learning. Planet. Space Sci. 170, 16–28.

Lee, C., Hogan, J., 2021. Automated crater detection with human level performance. Computers Geosciences 147, 104645.

Liao, W., Pižurica, A., Bellens, R., Gautama, S., Philips, W., 2014. Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features. IEEE Geosci. Remote Sens. Lett. 12 (3), 552–556.

Liu, X., Deng, C., Chanussot, J., Hong, D., Zhao, B., 2019. StfNet: A two-stream convolutional neural network for spatiotemporal image fusion. IEEE Trans. Geosci. Remote Sens. 57 (9), 6552–6564.

Liu, S., Zhao, H., Du, Q., Bruzzone, L., Samat, A., Tong, X., 2022. Novel cross-resolution feature-level fusion for joint classification of multispectral and panchromatic remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 1–14, Art no. 5619314.

Mao, Y., Yuan, R., Li, W., Liu, Y., 2022. Coupling complementary strategy to u-net based convolution neural network for detecting lunar impact craters. Remote Sens. (Basel) 14(3):661, 2022.

Minton, D.A., Richardson, J.E., Fassett, C.I., 2015. Re-examining the main asteroid belt as the primary source of ancient lunar craters. Icarus 247, 172–190.

G. Neukum, B. A. Ivanov, W. K. Hartmann, "Cratering records in the inner solar system in relation to the lunar reference system," in Chronology and Evolution of Mars, vol. 12, R. Kallenbach, J. Geiss, and W. K. Hartmann, Eds., in Space Sciences Series of ISSI, vol. 12., Dordrecht: Springer Netherlands, 2001, 55–86.

Palafox, L.F., Hamilton, C.W., Scheidt, S.P., Alvarez, A.M., 2017. Automated detection of geological landforms on Mars using Convolutional Neural Networks. Comput. Geosci. 101, 48–56.

Robbins, S.J., 2019. A new global database of lunar impact craters >1–2 km: Crater locations and sizes, comparisons with published databases, and global analysis. J. Geophys. Res. Planets 124 (4), 871–892.

Robbins, S.J., Antonenko, I., Kirchoff, M.R., Chapman, C.R., Fassett, C.I., Herrick, R.R., Singer, K., Zanetti, M., Lehan, C., Huang, D., Gay, P.L., 2014. The variability of crater identification among expert and community crater analysts. Icarus 234, 109–131.

Ronneberger, O., Fischer, P., Brox, T., 2015. In: U-Net: Convolutional Networks for Biomedical Image Segmentation. Springer International Publishing, Cham, pp. 234–241.

Oktay, O., Schlemper, J, Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al. Attention U-Net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999, 2018.

Roy, A. G., Navab, N., Wachinger, C., 2018. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I (421–429). Springer International Publishing.

Salamunićcar, G., Lončarić, S., 2010. Method for crater detection from digital topography data: interpolation based improvement and application to Lunar SELENE LALT data. 38th COSPAR Scientific. Assembly 38, 3.

Salamunićcar, G., Lončarić, S., Grumpe, A., Wöhler, C., 2014. Hybrid method for crater detection based on topography reconstruction from optical images and the new LU78287GT catalogue of Lunar impact craters. Adv. Space Res. 53, 1783–1797.

Silburt, A., Ali-Dib, M., Zhu, C., Jackson, A., Valencia, C., Kissin, Y., Tamayo, D., Menou, K., 2019. Lunar crater identification via deep learning. Icarus 317, 27–38.

Stepinski, T., Ding, W., Vilalta, R., 2012. Detecting impact craters in planetary images using machine learning. Information Science Reference, Hershey, PA, pp. 146–159.

Strom, R.G., Malhotra, R., Ito, T., Yoshida, F., Kring, D.A., 2005. The origin of planetary impactors in the inner solar system. Science 309, 1847–1850.

Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. In: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5693–5703.

Wang, Y., Wu, B., 2020. A new global catalogue of lunar craters (≥1 km) with 3D information and preliminary results of global analysis. Paper presented at the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. B3-2020, 2020 XXIV ISPRS Congress (Vol. XLIII).

Wang, S., Fan, Z., Li, Z., Zhang, H., Wei, C., 2020. An effective lunar crater recognition algorithm based on convolutional neural network. Remote Sens. (Basel) 12 (17), 2694.

Wang, Y., Wu, B., Xue, H., Li, X., Ma, J., 2021. An improved global catalog of lunar impact craters (≥1 km) with 3D morphometric information and updates on global crater analysis. J. Geophys. Res. Planets 126, e2020JE006728.

Wang, Y., Xie, H., Huang, Q., Feng, Y., Yan, X., Tong, X., Liu, S., Xu, X., Wang, C., Jin, Y., 2022. A novel approach for multiscale lunar crater detection by the use of path-profile and isolation forest based on high-resolution planetary images. IEEE Trans. Geosci. Remote Sens. 60, 1–24.

Wetzler, P., Honda, R., Enke, B., Merline, W., Chapman, C., Burl, M., 2005. Learning to detect small impact craters. In: 7th IEEE Workshop on Application of Computer Vision. Vol. 1. 178–184.

Wu, B., Huang, J., Li, Y., Wang, Y., Peng, J., 2018. Rock abundance and crater density in the candidate Chang' E-5 landing region on the Moon. J. Geophys. Res. Planets 123, 3256–3272.

Wu, B., Wang, Y., Lin, T.J., Hu, H., Werner, S.C., 2019. Impact cratering in and around the Orientale Basin: Results from recent high-resolution remote sensing datasets. Icarus 333, 343–355.

Wu, B., Li, F., Hu, H., Zhao, Y., Wang, Y., Xiao, P., et al., 2020. Topographic and geomorphological mapping and analysis of the Chang' E-4 landing site on the far side of the Moon. Photogramm. Eng. Remote Sens. 86 (4), 247–258.

Xu, Y., Du, B., Zhang, L., Cerra, D., Pato, M., Carmona, E., Le Saux, B., 2019. Advanced multi-sensor optical remote sensing for urban land use and land cover classification: outcome of the 2018 IEEE GRSS data fusion contest. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (6), 1709–1724.

Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B., 2017. Multisource remote sensing data classification based on convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56 (2), 937–949.

Xu, M., Zhang, Z., Wei, F., Hu, H., Bai, X., 2023. Side adapter network for open-vocabulary semantic segmentation. In: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2945–2954.

Yokoya, N., Ghamisi, P., Xia, J., Sukhanov, S., Heremans, R., Tankoyeu, I., Tuia, D., 2018. Open data for global multimodal land use classification: outcome of the 2017 IEEE GRSS data fusion contest. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (5), 1363–1377.

Yu, X., Hu, X., Song, Y., Xu, S., Li, X., Song, X., Wang, F., 2024. Intelligent assessment of building damage of 2023 Turkey-Syria Earthquake by multiple remote sensing approaches. npj Natural Hazards 1 (1), 3.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual U-Net. IEEE Geosci. Remote Sens. Lett. 15 (5), 749–753.