Mitigating Class Imbalance and Enhancing Unlabeled Data Extraction in Semi-Supervised Deep Learning for Martian Terrain Segmentation

Tong Xiao, Qunming Wang, Peter M. Atkinson and Xiaohua Tong, *Fellow, IEEE*

***Abstract*—Semantic analysis of Martian terrain (MT) is essential for understanding the Martian surface and facilitating autonomous rover navigation. Currently, deep learning dominates MT semantic segmentation, but costly and time-consuming label generation hampers its application. To mitigate this, semi-supervised segmentation (SSS) methods have been developed to leverage a large number of unlabeled images to enhance training with a limited set of labeled images. Nevertheless, challenges persist regarding the quality of pseudo-labels for unlabeled images and class imbalance in labeled images. To alleviate these issues, we proposed a novel end-to-end and efficient SSS network, namely, Mars-Balance-Enhancer (MBE). Specifically, to increase the quality of pseudo-labels, a novel adaptive local data augmentation (ALDA) module was injected in MBE. This module identifies the most challenging regions adaptively to train within unlabeled images, thereby guiding the training of unlabeled images effectively. Meanwhile, to further mitigate the impact of class imbalance in labeled images, the Symmetrical-Cyclical Focal (SCF) loss was introduced, which emphasizes minority classes dynamically during key training stages, enhancing the model's sensitivity to these classes. We evaluated MBE on six open-source MT datasets: MER, MSL, S^{5}Mars, MarsScapes, TWMARS and SynMars. The results demonstrated the advantage of MBE over five advanced SSS methods across all MT datasets. Ablation studies validated the effectiveness of the ALDA module and SCF loss of MBE. Furthermore, the SCF loss was found to be more advantageous than five commonly adopted benchmark losses. Similarly, the ALDA module was more accurate compared to the widely adopted augmentation strategy, Adaptive Label-Aided (ALa) CutMix. MBE provides a robust and efficient framework that holds promise for advancing future applications in remote sensing and planetary exploration.**

***Index Terms*—Semi-supervised segmentation, Martian terrain, Class imbalance, Data augmentation, Deep learning, Remote sensing**

I. Introduction

As human exploration of Mars progresses, the detailed analysis of Martian terrain (MT) has become increasingly important for scientific research and exploration missions [1], [2]. Understanding MT not only provides deeper insights into the geological structure and environmental conditions of Mars [3], [4], but is also crucial for mission planning and rover navigation, as it enables the identification of safe landing sites [5], assessment of hazards [6] and mapping of potential resources [7], [8]. The collection of a wealth of MT images by rovers like Curiosity, Perseverance and Tianwen-1 has further facilitated such types of research focused on the detailed analysis of MT [9], [10]. To interpret these images efficiently and accurately, researchers have increasingly turned to deep learning techniques [11], [12], which have revolutionized the field by enabling automatic segmentation of MT images [13]–[16].

While deep learning techniques have advanced the analysis of MT images significantly, they often require large numbers of labeled data. This reliance on extensive labeling demands both specialized expertise and incurs high labor costs, making deep learning challenging and resource-intensive [17], [18]. In response to this challenge, semi-supervised segmentation (SSS) methods have been developed, that leverage a large volume of unlabeled data to assist in model training, even when only a few labeled data are available [19]–[21]. By reducing the dependency on labeled data, SSS methods offer a promising solution for achieving accurate segmentation in MT analysis [22], [23].

Consistency regularization is a common strategy in SSS methods, based on the assumption of smoothness [24]. Specifically, it ensures model predictions for unlabeled data remain consistent under various perturbations, including input, feature, network, and combined perturbations [25]. A notable example is the weak-to-strong consistency learning (WSCL) method [22], [26], [27], which increases model robustness by guiding the strongly augmented view of unlabeled data using the weakly augmented view, employing instant pseudo-labels in the process. However, pseudo-labels, especially generated by hard samples, may contain errors due to a lack of prior information, leading to inaccurate loss values that can affect the final results adversely [28]. To mitigate this issue, CutMix-related approaches have been proposed [29]–[31]. These methods identify low-confidence predictions among pseudo-labels generated from unlabeled images and augment them by replacing random areas with labeled regions. To further exploit the potential of unlabeled images, adaptive data augmentation methods [32], [33] have been presented, which decide dynamically whether to perform the CutMix operation during training. Even with these developments, a key challenge remains: even if a pseudo-label is of low confidence, the

This work was supported by the National Natural Science Foundation of China under grant nos. 42222108, 42221002 and 42171345. *(Corresponding author: Qunming Wang.)*

T. Xiao, Q. Wang and X. Tong are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: wqm11111@126.com).

P. Atkinson is with Faculty of Science and Technology, Lancaster University, LA1 4YR Lancaster, U.K., and also with Geography and Environment, University of Southampton, SO17 1BJ Southampton, U.K.

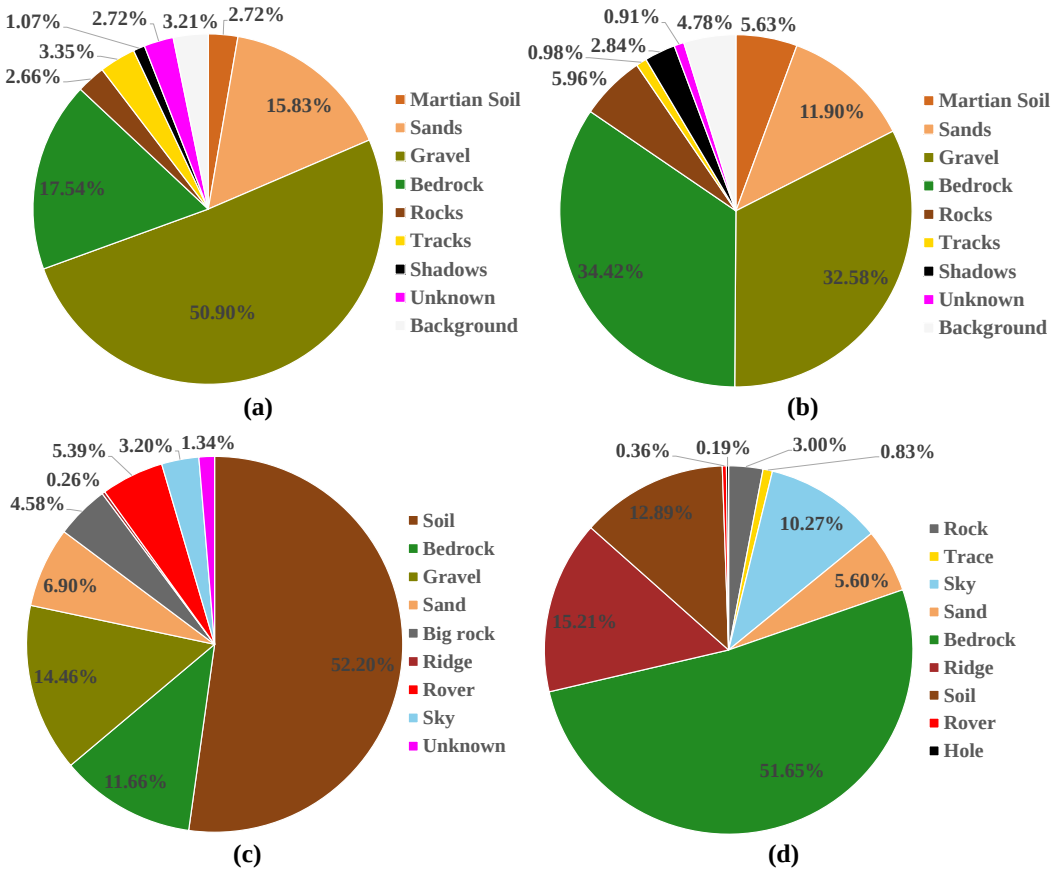


Fig. 1. Examples of class imbalance in typical MT datasets. (a) MER. (b) MSL. (c) MarsScapes. (d) S^{5}Mars.

dynamically cropped areas might still be of high confidence, making it difficult to train the model for the unlabeled image effectively.

Class imbalance is another prevalent issue in existing MT labeling data. As illustrated in Fig. 1, the class distribution is highly skewed across common MT datasets: MER [34], MSL [34], S^{5}Mars [35] and MarsScapes [36]. In these datasets, some categories, like Gravel in the MER, Bedrock in the MSL and S^{5}Mars and Soil in the MarsScapes, have obviously more samples than others. Conversely, the proportions of other categories, such as Shadows in the MER, Traces in the MSL, Ridge in the MarsScapes and Hole in the S^{5}Mars are much smaller. This imbalance can lead to biased model effectiveness, where the model becomes more adept at predicting the majority classes while struggling with the minority classes [37].

To alleviate class imbalance, two main strategies have been developed in recent years: resampling during data collection and weighting the loss function during model training. The former involves reconstructing the data distribution through sampling methods to balance the data [38], [39]. However, these methods can lead to data wastage and excessive resource consumption [40], which is particularly problematic for the scarce labeled data in MT, making resampling less practical. The latter approach aims at assigning larger weights to minority classes that are difficult to predict through reweighting [41]. For example, Focal loss [42] introduces a modulating factor γ to the standard Cross-Entropy (CE) loss, which reduces the relative loss for well-classified examples and focuses more on hard samples. Similarly, Online Hard Example Mining (OHEM) loss [43] selects dynamically the samples with the highest losses and computes the loss only on these samples during training. These methods, however, often overlook how class imbalance affects different stages of model training. To address this, a recent variant called Cyclical Focal (CF) loss [44] introduces cyclical adjustments to the loss weighting for high-confidence and low-confidence samples by applying a linear combination during different training stages. Despite the introduction of CF loss, the challenge remains of how to more effectively adjust the focus on different samples throughout the training process. Specifically, CF loss applies a fixed γ that emphasizes hard samples too early or too late, which may cause potentially instability or overfitting.

To mitigate the challenges of class imbalance and pseudo-label quality, we proposed an end-to-end and efficient SSS network, Mars-Balance-Enhancer (MBE), for MT semantic segmentation. MBE integrates a standard Teacher-Student framework [45] with the widely-used weak-to-strong consistent learning (WSCL) strategy. Unlike previous methods such as CutMix [46] and Adaptive Label-aided (ALa) CutMix [32], which apply augmentations across the entire image, MBE

introduces a novel adaptive local data augmentation (ALDA) module. This module identifies the lowest-confidence areas within a pseudo-label using a $h \times w$ box with a step size of δ—referred to as a confidence detector. By targeting specifically the most challenging areas for training, this approach avoids unnecessary distortions to regions with higher confidence. Moreover, to mitigate the effects of class imbalance, we designed a novel loss function, namely, Symmetrical-Cyclical Focal (SCF) loss. SCF loss adjusts dynamically the value of γ throughout the training process. This approach increases attention to the minority classes during the middle stages of training and stabilizes learning by reducing this attention in the early and late stages of training.

Overall, the main contributions of this paper are as follows:

I. A simple and end-to-end SSS method for MT segmentation, named MBE, was proposed. This method increases pseudo-label quality and mitigates the impact of class imbalance, enabling more effective MT analysis.
II. To increase pseudo-label quality, a novel ALDA module was introduced. This module identifies areas with the lowest pseudo-label confidence using a confidence detector and applies the CutMix method to replace these lowest confidence regions with labeled data.
III. To mitigate class imbalance on the MT datasets, a novel SCF loss was presented. SCF loss adjusts dynamically the value of γ throughout the training process, which increases attention to the minority classes during the middle stages of training and stabilizes learning by reducing this attention in the early and late stages of training.

II. Methodology

The proposed MBE method is illustrated in Fig. 2. Within a standard Teacher-Student framework, MBE trains the student model using both labeled and unlabeled data. The teacher model is trained on unlabeled data, with both models optimized by minimizing the supervised loss $\mathcal{L}_l$, where l represents labeled data, and the unsupervised loss $\mathcal{L}_u$, where u indicates unlabeled data. Firstly, the WSCL strategy is used to compute the unsupervised loss $\mathcal{L}_u$, which measures the consistency between the predictions of strongly augmented unlabeled data from the student model and the pseudo-labels generated from weakly augmented unlabeled data by the teacher model. To increase the quality of pseudo-labels, we proposed the ALDA module, integrated after the teacher model. Subsequently, the supervised loss $\mathcal{L}_l$ is calculated based on the predictions generated by the student model for labeled data and their corresponding ground truth. For this purpose, we introduced a novel class imbalance loss function, SCF loss, to refine the supervised loss $\mathcal{L}_l$ by adjusting dynamically the value of γ on minority classes throughout the training process.

A. Problem Definition

Consider a compact collection of labeled images, $\mathbf{D}_l = \left\{\left(\mathbf{x}_i^l, \mathbf{y}_i^l\right)\right\}_{i=1}^{N_l}$, and a substantially larger pool of unlabeled images, $\mathbf{D}_u = \{\mathbf{x}_i^u\}_{i=1}^{N_u}$. Each $\mathbf{x}_i^l$ is the i-th labeled image of dimensions $H \times W \times M$, and $\mathbf{y}_i^l$ presents its pixel-wise accurate set of labels across C categories, where C denotes the total number of classes. Meanwhile, $\mathbf{x}_i^u$ presents the i-th image without labels. Here, correspondingly, N_l and N_u denote the counts of labeled and unlabeled datasets, respectively.

To utilize the unlabeled data efficiently, we employed a simple yet effective SSS framework consisting of a student model and a teacher model, parameterized by θ_s and θ_t, respectively, where s represents the student model and t represents the teacher model. The teacher model generates pseudo-labels from the unlabeled data, while the student model produces predictions from both the labeled and unlabeled data. The primary objective of the SSS framework is to train the student model by minimizing a supervised loss $\mathcal{L}_l$ and an unsupervised consistency loss $\mathcal{L}_u$ simultaneously. Consistent with most SSS methods [47]–[49], we used a standard pixel-wise Cross-Entropy (CE) loss to obtain $\mathcal{L}_u$ on unlabeled data. To mitigate class imbalance in the labeled data, we proposed the SCF loss for the supervised loss $\mathcal{L}_l$, which will be detailed in Section II-D. The total training loss for the student model is defined as follows:

$$\mathcal{L} = \mathcal{L}_l + \lambda_u \mathcal{L}_u, \tag{1}$$

where λ_u is a scalar hyper-parameter to adjust the weight of the unsupervised loss. Following [22] and [32], it is set to 1 by default. For the teacher model, it is not trained directly; instead, the parameters θ_t are updated gradually through an exponential moving average (EMA) of the student weights θ_s, as follows:

$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s, \tag{2}$$

where α is a momentum parameter, set to 0.999 by default, as per [45].

B. Weak-to-Strong Consistency Learning

WSCL [26] is a robust strategy for increasing model generalization by leveraging unlabeled data. This method addresses the challenge of utilizing unlabeled samples by enforcing consistency between pseudo-labels from weakly augmented views and unlabeled predictions from strongly augmented versions of the same image. Specifically, WSCL uses pseudo-labels from weakly augmented images to supervise and refine the unlabeled predictions from strongly augmented images, thereby enhancing the model's ability to generalize from limited labeled data [27], [50], [51].

In our implementation, we applied three geometric augmentation methods—random scaling, random flipping, and random cropping—to create weakly augmented images, following [32]. For the strongly augmented images, we used a random intensity-based augmentation strategy that selects from a pool of 11 strong augmentation methods, such as identity, autocontrast and equalize, with a maximum of k augmentations applied. This approach ensures a diverse range of augmentations to challenge the model during training [32].

After generating the weakly and strongly augmented images, the weakly augmented images are input into the teacher model to produce pseudo-labels, while the strongly augmented images are fed into the student model to generate predictions

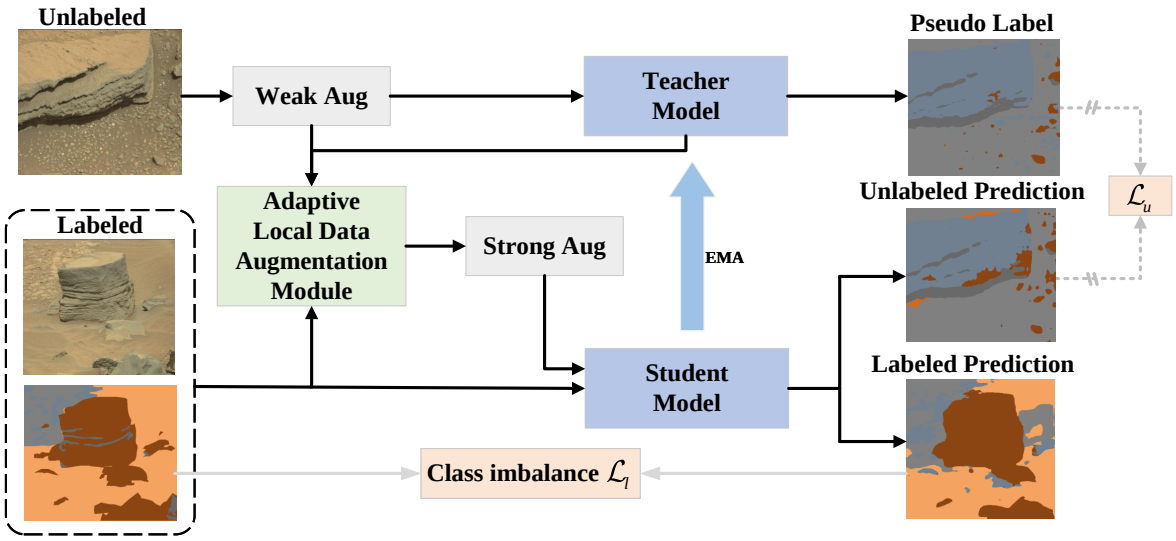


Fig. 2. Workflow of the proposed MBE method. Black arrows represent the primary data and operation flow; gray-white arrows indicate loss calculation paths, with solid lines representing losses involved in backpropagation and dashed lines representing those excluded from it.

for the unlabeled data. The unsupervised loss $\mathcal{L}_u$ is computed based on the consistency between the pseudo-labels and unlabeled predictions. Both the teacher and student models utilize encoder-decoder architectures, with ResNet101 [52] serving as the encoder and DeepLabV3 [53] as the decoder, to effectively capture features and obtain robust segmentation results.

C. Adaptive Local Data Augmentation Module

To increase the quality of pseudo-labels and maximize the use of unlabeled data, we introduced the ALDA module inspired by [31] and [32]. This module uses a confidence detector to identify areas with the lowest pseudo-label confidence and then applies the CutMix method to replace these areas with labels. The process begins by obtaining logit predictions from the teacher model. Then a random decision strategy is applied to determine whether to execute the CutMix operation during each epoch. If the decision is negative, the logit predictions remain unchanged, allowing for full utilization of the unlabeled data. If positive, the following steps are processed as illustrated in Fig. 3. In stage (a), a $h \times w$ box with a step size of δ, referred to as a confidence detector is designed to scan the logit predictions and compute the confidence score for each area. The confidence score is similar to that used in [32], which is calculated as follows:

$$\rho_i = \frac{1}{h \times w} \sum_{j=1}^{h \times w} \left(\max\left(p_i(j)\right) \left(1 - \frac{-\sum p_i(j) \log p_i(j)}{\log N}\right)\right) \tag{3}$$

where $p_i(j)$ represents the predicted probability for class j in the i-th logit within a confidence detector, and N is the total number of classes. This equation calculates the average confidence score within the detector, providing an estimate of the reliability of the pseudo-labels in that region. Areas with the lowest confidence scores are identified for augmentation.

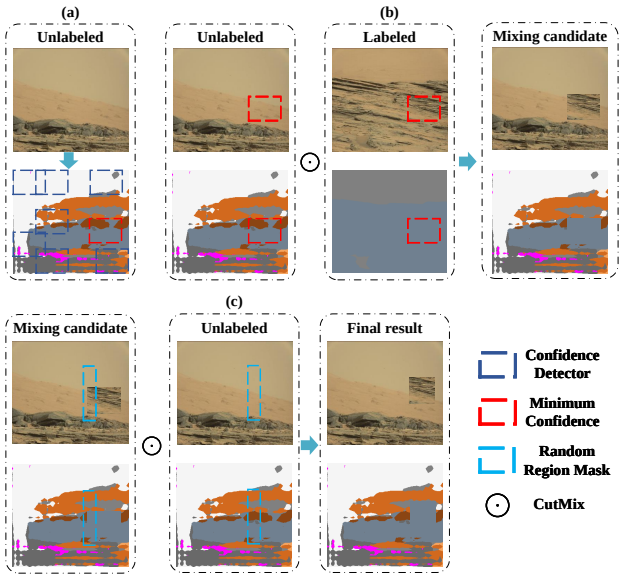


Fig. 3. Flowchart of the adaptive local data augmentation (ALDA) module. (a) A confidence detector identifies the lowest confidence region in a pseudo-label. (b) The CutMix method is then applied to create the initial mixing candidate. (c) CutMix is then applied again to produce the final augmented candidate.

In stage (b), a mixing operation between labeled and unlabeled regions is leveraged to generate mixing candidates $\{\mathbf{u}'_n\}$:

$$\mathbf{u}'_n \leftarrow \mathbf{M}_n \odot \mathbf{u}_n + (1 - \mathbf{M}_n) \odot \mathbf{x}^l_n, \tag{4}$$

where $\mathbf{M}_n$ denotes the mask for regions with the lowest confidence and $\odot$ refers to the CutMix operation. In stage (c), following [32], we performed the final mixing step between the unlabeled instance $\{\mathbf{u}_m\}$ and the permuted mixing candidates $\{\mathbf{u}'_n\}$:

$$\mathbf{u}'_m \leftarrow \mathbf{M}_m \odot \mathbf{u}_m + (1 - \mathbf{M}_m) \odot \mathbf{u}'_n \tag{5}$$

where $\mathbf{M}_m$ is a randomly generated binary mask for the mixing operation. This stage can generate prediction disagreement, which increases the model's robustness and helps prevent overfitting ultimately [32].

To ensure that the confidence detector can effectively identify small areas of low confidence, its side length is randomly selected within constraints: the maximum side length is no more than half the side length of the image, and the minimum side length is no less than 30 pixels.

D. Class Imbalance Loss

Class imbalance is a common problem in training datasets, leading to a decrease in model accuracy [54]–[56]. To alleviate this problem, we proposed a novel loss function, namely, SCF loss, which combines elements from Focal loss [42] and CF loss [44] to handle class imbalance more effectively. Focal loss was designed to mitigate class imbalance by focusing more on hard samples while down-weighting the loss for easy samples. It is defined as:

$$\mathrm{FL}\left(p_t\right) = -\alpha_t \left(1 - p_t\right)^{\gamma} \log\left(p_t\right) \tag{6}$$

where p_t is the predicted probability of the true class, α_t is a weighting factor for the class, and γ is a focusing parameter that adjusts the degree of down-weighting. To further advance the handling of class imbalance, SCF loss introduces a symmetrical-cyclical adjustment to the focusing parameter γ to increase attention to the minority classes during the middle stages of training and stabilize learning by reducing this attention during the early and late stages of training. Here, the early, middle and late stages refer to different learning phases. Specifically, the early stage is characterized by the model primarily learning simple patterns, such as dominant class features [57]; the middle stage involves learning more fine-grained and discriminative features [57]; in the late stage, the model enters the refinement stage, but it also faces the risk of overfitting [58]. SCF loss calculates the adjustment factor $\gamma_{\mathrm{cyclical}}$ using the following function of the epoch number:

$$\begin{aligned} \gamma_{\mathrm{cyclical}} &= \max\left(\gamma_{\min}, \min\left(\gamma_{\max}, g\right)\right) \\ g &= 1 - \omega \cdot \left|\frac{t}{T} - 0.5\right| \end{aligned} \tag{7}$$

here, $\gamma_{\min}$ and $\gamma_{\max}$ are the minimum and maximum values of the focusing parameter, set to 0 and 1, respectively. The parameter ω scales the cyclical adjustment, while t and T denote the current and total training epochs, respectively. This design provides three key advantages:

I. Early stage (stability). SCF loss initializes $\gamma_{\mathrm{cyclical}}$ to close to 0, making it in line with CE loss. This prevents overfitting to hard examples too soon and ensures balanced learning across classes.
II. Middle stage (focus on minority classes): As training progresses, SCF loss increases $\gamma_{\mathrm{cyclical}}$ gradually, shifting towards Focal Loss. This adaptive focus helps the model to enhance its understanding of minority classes after obtaining sufficient knowledge of majority classes.
III. Late stage (fine-tuning and generalization): SCF loss starts to reduce $\gamma_{\mathrm{cyclical}}$ gradually, mitigating excessive focus on hard examples. This shift back to CE helps stabilize training, avoiding overfitting while improving generalization ability.

Unlike CF loss, which applies a fixed γ throughout training and continuously emphasizes difficult samples, potentially causing instability in the early stage and overfitting in the later stage. By adjusting $\gamma_{\mathrm{cyclical}}$, SCF loss enables the model to first learn general patterns, then refine minority class handling, and finally stabilize its learned representations. Theoretically, this cyclical adjustment improves training robustness by aligning the emphasis of the loss function with different learning phases.

To ensure that SCF loss appropriately emphasizes challenging examples during the most effective training periods, careful control of the $\gamma_{\mathrm{cyclical}}$ curve is necessary. If ω was set too large, the curve will be too steep, causing the model to focus intensely on minority classes for only a short period, which limits its ability to handle complex samples consistently. Setting $\omega = 2$ ensures a smoother and more balanced adjustment of the $\gamma_{\mathrm{cyclical}}$ curve, allowing for gradual and sustained attention to minority classes during the middle stages of training when the model is most capable of learning from them.

III. Experiments

In this section, we detail our experimental data and setup. We then compare the MBE method with recent state-of-the-art techniques on SSS benchmarks. Additionally, we conduct a series of ablation studies to further validate the effectiveness and stability of the ALDA module and SCF loss.

A. Data Description

In this research, six heterogeneous MT datasets[1], namely, MER [34], MSL [34], S^{5}Mars [35], MarsScapes [36], TWMARS [59] and SynMars [60], were used to evaluate the performance of the proposed MBE. These MT datasets contain both multi-class and single-class samples, allowing for testing the ability of the proposed model to handle various classification scenarios. Moreover, each dataset was partitioned randomly into training and testing sets with a 4/5 and 1/5 ratio, as described in [22]. Simultaneously, the training data were partitioned randomly for semi-supervised learning using 1/4 and 1/8 labeled data protocols.

The MER dataset consists of single-channel grayscale images sourced from the planetary data system. It contains 1024 high-definition images, each with dimensions of 512 × 512 pixels, captured by the navigation camera and panoramic camera of the Spirit and Opportunity Mars rovers. The semantic information within the MER dataset includes nine categories: Martian Soil, Sands, Gravel, Bedrock, Rocks, Tracks, Shadows, Unknown and Background.

[1] https://github.com/Fluestech/Awesome-Martian-Terrain-Segmentation-Dataset

The MSL dataset, the same as MER in its semantic categories, comprises RGB-channel images obtained from the Mars 32k, totaling 4155 images. These images, captured by the mast camera of the Curiosity Rover, have been resized from 1024 × 1024 pixels to 560 × 500 pixels using bilinear interpolation [34].

The S^{5}Mars dataset comprises 6000 fine-resolution images captured on the surface of Mars by the color mast camera aboard Curiosity. The RGB images in this dataset have a spatial resolution of 1200 × 1200 pixels, and they are categorized into nine label categories: Sky, Ridge, Soil, Sand, Bedrock, Rock, Rover, Trace and Hole.

The MarsScapes dataset, provided by NASA/JPL-Caltech, is the first panorama dataset for Martian terrain understanding. It contains 195 panoramic RGB images with widths ranging from 1230 to 12062 pixels and heights from 472 to 1649 pixels. To facilitate input into the deep learning model, these images were cropped into 512 × 512 pixel segments. The semantic information includes nine different categories: Big Rock, Bedrock, Sand, Soil, Gravel, Sky, Ridge, Rover and Unknown.

Both the TWMARS and SynMars datasets consist of RGB-channel images and include two categories: Rock and Background. The TWMARS dataset, captured by NaTeCam, consists of 336 images, each with dimensions of 1024 × 1024 pixels. The SynMars dataset, totaling 10000 images, is a synthetic dataset simulated from Tianwen-1. Each image in the SynMars dataset is generated by Blender software[2] with a resolution of 1920 × 1080 pixels.

B. Experimental Setup

To validate the effectiveness and reliability of MBE, two comparative experimental sections were designed.

Section III-C focused on comparing the outcome of the proposed MBE method with five state-of-the-art SSS methods: s4GAN [61], ClassMix [62], LSST [63], WSCL [22] and AugSeg [32]. To make a fair comparison, the backbones of these methods are ResNet and DeepLab series, which are consistent with MBE. In this section, the training datasets were randomly cropped to 321 × 321 pixels, and the batch size of each approach was set to 8. For s4GAN and ClassMix, the iterations were set to 40000. For other methods, the training epochs were set to 50. The backbone was optimized using stochastic gradient descent (SGD) with a base learning rate of 0.01, a weight decay of 0.0005 and a momentum of 0.9. The learning rate was decayed during training using a polynomial scheduling strategy: $l_r = \text{base\_}l_r \times (1 - (\text{ iter })/(\max\_\text{iter }))^{0.9}$.

Section III-D concentrated on validating the effectiveness of the ALDA module and SCF loss through three perspectives. Specifically, Section III-D1 aimed to determine the influence of the ALDA module and SCF loss through a basic ablation experiment. In this experiment, the baseline was the MBE method, with the ALDA module removed and SCF loss replaced by Cross-Entropy (CE) loss. Section III-D2 further tested the effectiveness of SCF by comparing it with five benchmark loss functions, specifically CE, Focal [42], Label Smoothing (LS) [64], Cyclical Focal (CF) [44] and Online Hard Example Mining (OHEM) losses [43], across six datasets. The f_c parameter of CF loss was set to 2 to ensure a fair comparison with SCF loss, maintaining the same cyclical symmetry structure, and γ was set to 1 in the Focal and CF losses to ensure a consistent comparison. All modules (except for the loss function) and parameters were identical to those in Section III-C. Additionally, we evaluated the performance of the proposed CutMix method against the popular method, Adaptive Label-aided (ALa) CutMix [32], for the same datasets in Section III-D3. The supervised loss functions used were the SCF and CE losses, and other modules and parameter settings were consistent with Section III-C.

The evaluation metric for all experiments was Intersection over Union (IoU). All operations were performed on a Windows 10 workstation running Python 3.8.10. The workstation was equipped with a 13th Gen Core i7-13700 CPU and a 12 GB NVIDIA GeForce RTX 4070 Ti GPU. All semi-supervised segmentation methods were implemented using PyTorch 1.21.

C. Experimental Results

As shown in Fig. 4, the proposed method outperformed other advanced SSS models in terms of Mean IoU for all classes across the MER [34], MSL [34], S^{5}Mars [35], MarsScapes [36], TWMARS [60] and SynMars [59] datasets under both 1/4 and 1/8 labeled partition protocols. Notably, the proposed method excelled in handling scenarios with limited labeled data, not only increasing segmentation IoU y across various terrain datasets but also demonstrating higher generalization ability and robustness in diverse and complex environments.

For the MER dataset, the visual results (to fair comparison, all visual results of SSS models were shown under the 1/4 labeled protocol), as illustrated in Fig. 5, demonstrated that the boundaries of objects in the MBE results were smoother compared to those of other methods. This was particularly evident in categories such as Tracks, Shadows and Unknown, where the MBE results aligned more closely with the characteristics of the reference image. Furthermore, other methods exhibited greater mis-segmentation rates when identifying small objects, such as rocks, than the MBE method. Due to these advantages, MBE exhibited high-quality visual results. This observation was supported by the quantitative analysis results, as shown in Table I, where MBE outperformed existing state-of-the-art methods across all partition protocols, surpassing the second-best method by 0.99% and 1.44% under the 1/8 and 1/4 partition protocols, respectively. Additionally, MBE demonstrated effective segmentation for the imbalanced categories, such as Martian Soil, Rock, Tracks and Shadows, across both partition protocols. It is important to note that MBE achieved the greatest IoU for Rocks and Tracks, further demonstrating the effectiveness of SCF loss.

Regarding the MSL dataset, the visual results, as shown in Fig. 6, revealed that each method produces smooth object boundaries, with differences arising primarily from mis-segmentation. The s4GAN and ClassMix methods exhibited

[2]https://www.blender.org

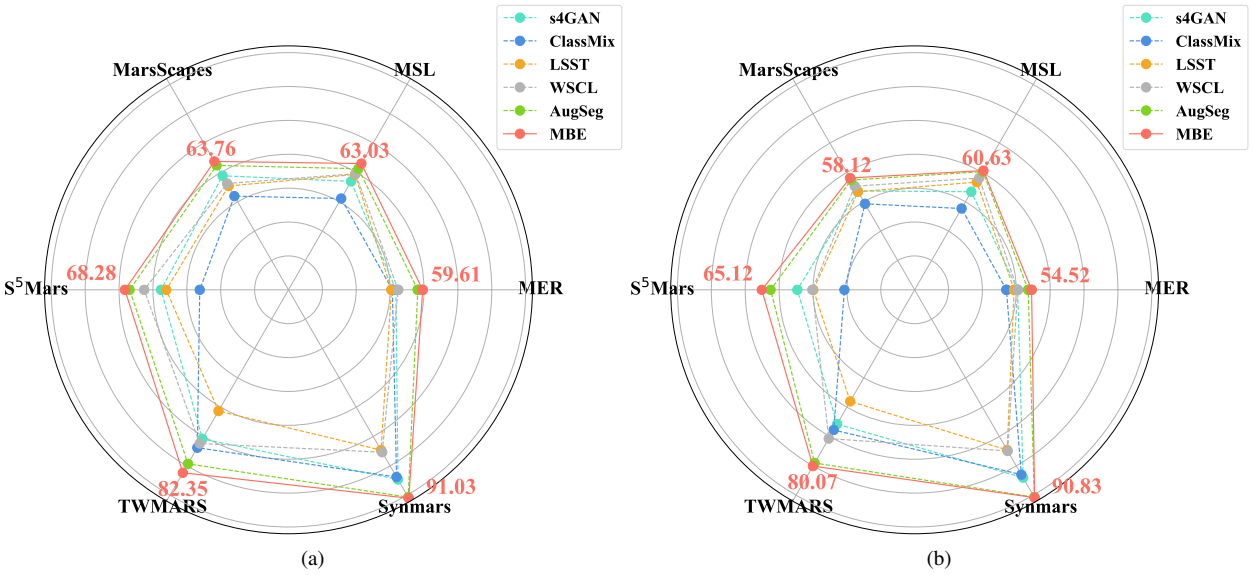


Fig. 4. Mean IoU (%) results of various semi-supervised segmentation methods for all classes across six datasets. (a) Under the 1/4 labeled partition protocol. (b) Under the 1/8 labeled partition protocol.

TABLE I
PER-CLASS IOU(%) FOR THE MER TEST SET.

Labeled	Method	Martian Soil	Sands	Gravel	Bedrock	Rocks	Tracks	Shadows	Unknown	Background	Mean IoU
1/4	s4GAN [61]	11.30	56.02	73.21	65.09	33.52	53.23	22.72	68.71	81.52	51.70
	ClassMix [62]	15.39	52.21	69.56	61.14	33.57	47.38	28.59	65.08	83.03	50.66
	LSST [63]	5.86	51.67	71.61	64.04	31.95	48.74	20.70	71.44	87.52	50.39
	WSCL [22]	6.82	53.40	70.89	63.71	34.93	49.93	26.76	77.52	**87.38**	52.37
	AugSeg [32]	**22.26**	58.15	**74.65**	69.19	49.27	55.55	34.12	74.57	85.77	58.17
	MBE	21.70	**58.37**	73.75	**69.26**	**49.97**	**61.80**	**37.68**	**77.00**	86.96	**59.61**
1/8	s4GAN [61]	**4.82**	49.23	71.71	58.83	33.10	49.39	26.58	69.91	**90.13**	50.41
	ClassMix [62]	3.78	44.55	67.66	57.40	34.87	48.11	**39.89**	65.51	62.05	47.09
	LSST [63]	3.39	53.24	71.30	61.74	29.41	52.64	28.80	64.39	79.30	49.36
	WSCL [22]	2.33	**55.93**	69.80	61.20	32.07	53.50	28.18	69.99	80.60	50.40
	AugSeg [32]	3.53	52.77	**72.91**	68.29	47.63	58.92	37.49	71.70	68.50	53.53
	MBE	4.03	52.65	71.81	**69.09**	**50.25**	**60.92**	34.97	**75.18**	71.81	**54.52**

notable mis-segmentations in the background category, while the LSST and WSCL methods struggled to recognize the Tracks and Rock categories accurately. AugSeg failed to correctly identify gravel types, and its rock classification was also inaccurate. In contrast, MBE produced higher-quality visual results compared to the other methods. The quantitative analysis, presented in Table II, showed that the MBE method achieved significant increases in IoU over other methods, with gains of 1.77% and 1.45% over the second-best methods under the 1/4 and 1/8 partition protocols, respectively (p <0.05, McNemar's test [65]). MBE also excelled in handling unbalanced classes; for example, under the 1/4 partition protocol, it delivered the highest IoU in categories such as Martian Soil, Rocks, Tracks and Unknown. More importantly, it achieved a high IoU value of 70.75% in the Tracks category, despite this class comprising only 0.98% of the labeled data. Furthermore, under the 1/8 partition protocol, MBE continued to yield accurate results in the unbalanced categories, achieving nearly the same IoU as the most accurate method in the Tracks category, with only a 0.45% difference.

The visual results for the MarsScapes dataset are presented in Fig. 7. Similar to previous datasets, the results from s4GAN, ClassMix and LSST showed more jagged object boundaries, particularly in the Soil and Redrock categories, compared to WSCL, AugSeg and MBE. The WSCL method showed inaccuracy in segmentation for the Sands category, while AugSeg showed discontinuity in the Soil category. In comparison, MBE achieved more accurate and comprehensive segmentation results. The quantitative analysis, as shown in Table III, highlighted the exceptional IoU of MBE, which surpassed the second-best methods by 1.35% and 0.72% under the 1/4 and 1/8 partition protocols, respectively. These high IoU scores extended to the minority classes such as Ridge and Sky, where MBE achieved the highest IoU scores. Furthermore, the MBE method maintained its effectiveness even with reduced numbers of labeled data, underscoring its

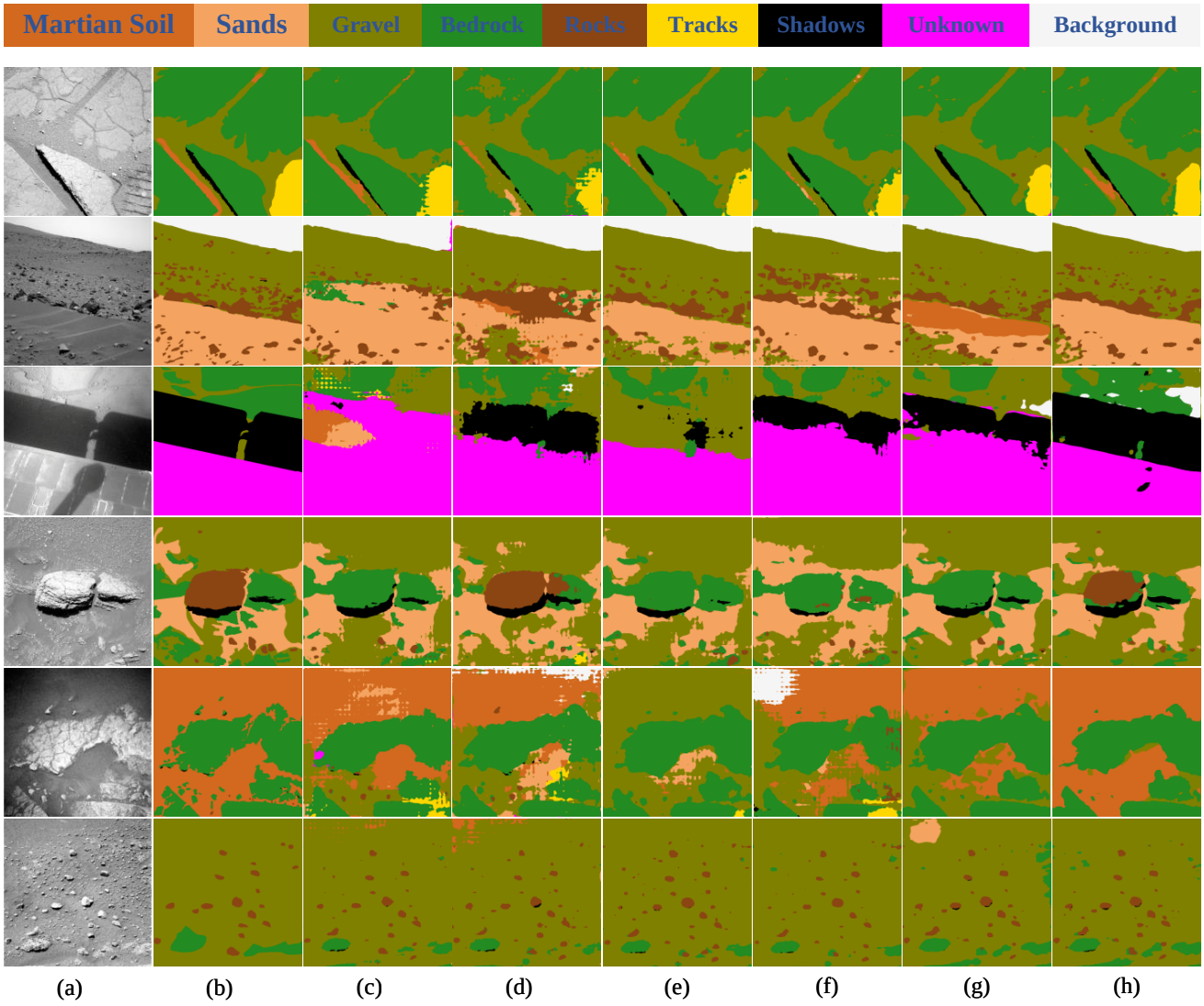


Fig. 5. Visual comparisons of different semi-supervised segmentation methods for the MER dataset. (a) Raw image. (b) Reference. (c) s4GAN. (d) ClassMix. (e) LSST. (f) WSCL. (g) AugSeg. (h) MBE.

TABLE II
PER-CLASS IOU(%) FOR THE MSL TEST SET.

Labeled	Method	Martian Soil	Sands	Gravel	Bedrock	Rocks	Tracks	Shadows	Unknown	Background	Mean IoU
1/4	s4GAN [61]	28.95	63.32	69.75	71.31	41.84	51.43	41.92	53.52	87.48	56.95
	ClassMix [62]	26.74	57.80	65.42	67.04	32.48	41.10	36.52	52.38	80.50	51.11
	LSST [63]	29.04	64.79	69.61	74.87	39.33	62.85	43.53	**61.83**	88.13	59.33
	WSCL [22]	31.36	64.44	70.48	75.60	39.89	66.02	39.59	60.93	87.91	59.58
	AugSeg [32]	28.73	68.67	**71.99**	76.40	45.13	62.01	**51.25**	57.97	89.24	61.27
	MBE	**31.94**	**69.36**	71.86	**76.63**	**47.21**	**70.75**	49.61	59.41	**90.61**	**63.04**
1/8	s4GAN [61]	24.58	62.35	67.43	72.08	37.30	42.43	37.28	54.83	82.67	53.44
	ClassMix [62]	23.71	55.11	63.55	63.82	36.43	36.91	37.02	41.35	68.79	47.71
	LSST [63]	**28.66**	63.67	67.87	74.02	38.82	55.48	39.73	56.17	85.96	56.71
	WSCL [22]	25.46	64.61	67.43	74.19	37.68	**69.24**	38.95	**57.82**	87.47	58.09
	AugSeg [32]	22.97	**66.91**	70.43	**76.86**	**47.18**	62.47	45.78	54.66	85.34	59.18
	MBE	27.02	66.80	**70.84**	76.46	45.35	68.79	**46.11**	53.44	**90.83**	**60.63**

Fig. 6. Visual comparisons of different semi-supervised segmentation methods for the MSL dataset. (a) Raw image. (b) Reference. (c) s4GAN. (d) ClassMix. (e) LSST. (f) WSCL. (g) AugSeg. (h) MBE.

efficiency in scenarios with limited labeled data.

In terms of the S^5Mars dataset, the visual results of the MBE method were more accurate than those of other methods, as shown in Fig. 8. For example, for the Trace category, other methods exhibited severe mis-segmentation, whereas the MBE method provided relatively accurate segmentation. Additionally, MBE achieved notably greater performance in segmenting the Rock and Gravel categories compared to other methods. The advances in Martian terrain classification achieved by MBE were highlighted by its higher IoU scores compared to other methods, as evidenced in Table IV. With only 1/4 of the labeled data, MBE achieved the highest IoU for most classes, including a 0.79% increase over AugSeg for the Sky category, a 1.45% increase for Ridge and a 0.87% increase for Soil. Its ability to maintain high IoU with reduced labeled data was particularly impressive, as demonstrated by a 14.33% advantage over ClassMix for Sky with just 1/8 of the labeled data. Furthermore, the MBE method excelled in challenging classes such as Trace, producing an increase in IoU of 9.69% over AugSeg.

For both TWMARS and Synmars datasets, the visual results are presented in Fig. 9. These datasets were used primarily for detecting the position and shape of rocks. For the TWMARS dataset, as highlighted in the red box, the MBE method demonstrated greater performance in detecting both the shape and location of rocks compared to other methods. For the Synmars dataset, which contained more supervised information, all methods achieved great performance. However, distinctions arose in the detection of small rocks. Methods such as ClassMix, LSST and WSCL showed limited success in identifying small rocks, as shown in the red box, while s4GAN and AugSeg missed or misclassified rocks in the yellow box. In contrast, the MBE method excelled in these areas. Quantitative analysis supported this conclusion, as shown in Tables V and VI. For example, the MBE method advanced rock detection on the TWMARS and Synmars datasets by 5.87% and 0.66%, respectively, compared to the second-best methods when using 1/4 of the labeled data. With 1/8 of the labeled data, MBE achieved increases of 2.06% and 0.34% on the TWMARS and Synmars datasets, respectively. Although the increase over other methods on the Synmars dataset was less pronounced, the MBE method achieved an IoU exceeding 90%. This result

Fig. 7. Visual comparisons of different semi-supervised segmentation methods for the MarsScapes dataset. (a) Raw image. (b) Reference. (c) s4GAN. (d) ClassMix. (e) LSST. (f) WSCL. (g) AugSeg. (h) MBE.

TABLE III
PER-CLASS IOU(%) FOR THE MARSSCAPES TEST SET.

Labeled	Method	Soil	Bedrock	Gravel	Sand	Big Rock	Ridge	Rover	Sky	Unknown	Mean IoU
1/4	s4GAN [61]	80.74	67.51	42.50	57.51	43.33	15.90	80.81	71.84	**69.71**	58.87
	ClassMix [62]	77.36	60.47	36.34	50.41	42.00	1.43	82.04	63.09	54.19	51.93
	LSST [63]	77.66	65.88	44.89	53.77	45.55	3.14	85.86	69.14	52.50	55.38
	WSCL [22]	77.76	68.11	47.17	57.00	42.33	3.37	81.91	67.56	60.85	56.23
	AugSeg [32]	80.93	71.88	48.02	62.69	**51.78**	14.04	**93.57**	74.21	64.52	62.41
	MBE	**81.05**	**73.19**	**48.45**	**64.40**	50.29	**20.49**	**93.57**	**79.31**	63.06	**63.76**
1/8	s4GAN [61]	76.46	51.83	31.81	49.83	39.01	**21.06**	83.87	**72.71**	**56.69**	53.70
	ClassMix [62]	76.22	57.51	31.91	46.70	44.70	9.10	76.94	60.29	41.14	49.32
	LSST [63]	78.83	59.36	40.68	**68.33**	40.22	17.44	81.27	65.36	49.40	53.43
	WSCL [22]	79.03	62.09	43.75	49.75	41.58	17.28	81.38	66.61	56.42	55.32
	AugSeg [32]	**80.24**	64.00	**46.41**	51.10	**48.34**	18.24	84.09	71.29	52.85	57.40
	MBE	78.49	**69.04**	45.76	54.31	48.12	13.37	**86.28**	72.43	55.28	**58.12**

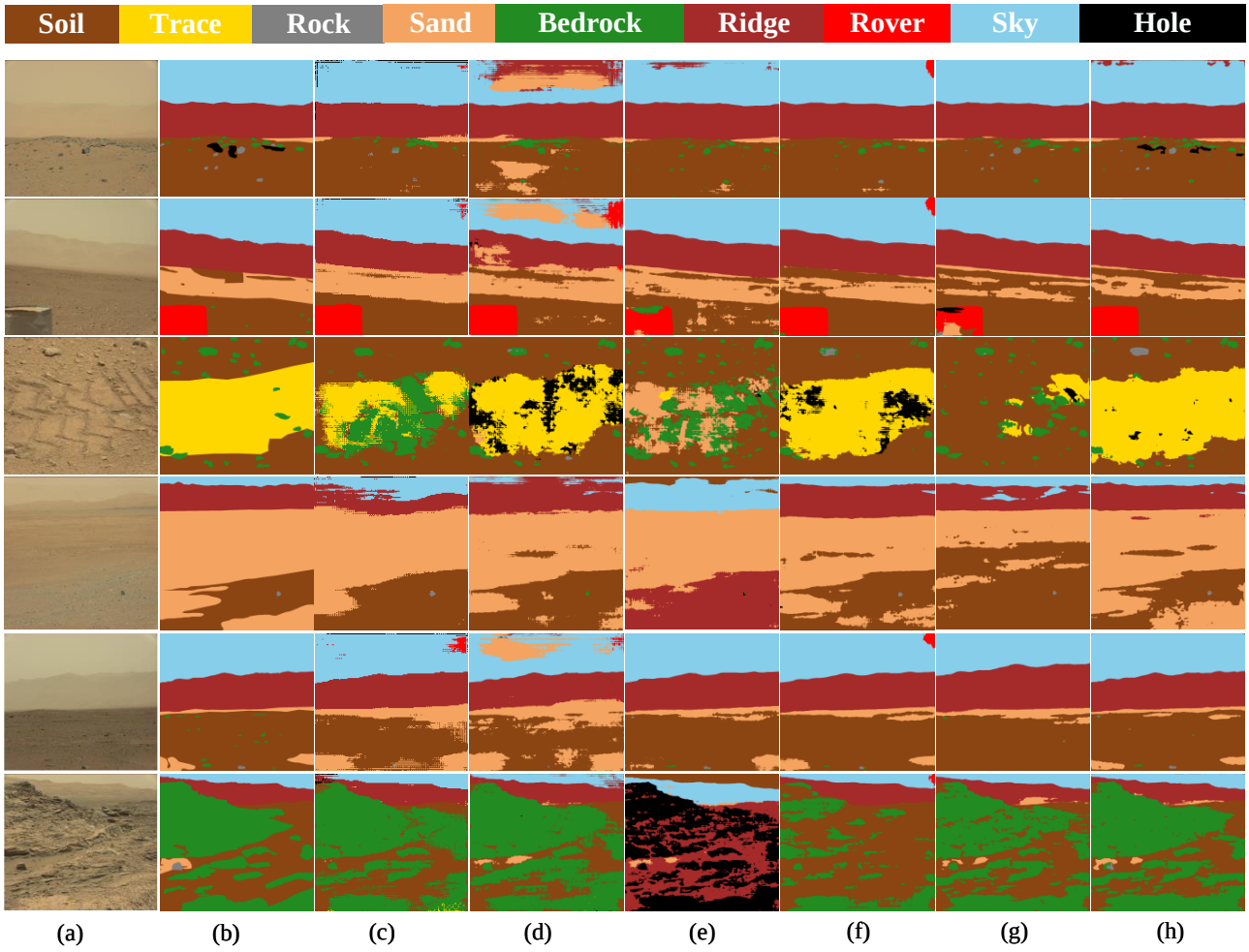


Fig. 8. Visual comparisons of different semi-supervised segmentation methods for the S^5Mars dataset. (a) Raw image. (b) Reference. (c) s4GAN. (d) ClassMix. (e) LSST. (f) WSCL. (g) AugSeg. (h) MBE.

TABLE IV
PER-CLASS IOU(%) FOR THE S^5MARS TEST SET.

Labeled	Method	Sky	Ridge	Soil	Sand	Bedrock	Rock	Rover	Trace	Hole	Mean IoU
1/4	s4GAN [61]	79.09	93.17	91.71	40.14	54.02	73.90	15.17	43.19	28.43	57.64
	ClassMix [62]	75.41	60.55	65.94	29.89	30.71	70.03	14.04	35.43	33.70	46.19
	LSST [63]	73.76	91.63	90.46	31.00	49.98	69.42	13.80	62.70	22.01	56.08
	WSCL [22]	79.93	94.82	92.99	39.65	60.63	70.78	17.16	64.36	43.06	62.60
	AugSeg [32]	80.33	**96.91**	**95.13**	39.74	62.00	73.28	19.95	68.83	**65.74**	66.88
	MBE	**81.12**	95.46	94.13	**40.95**	**62.01**	**75.23**	**22.34**	**78.64**	64.62	**68.28**
1/8	s4GAN [61]	78.88	**96.06**	94.35	**35.73**	42.09	72.64	15.90	46.28	9.71	54.63
	ClassMix [62]	64.84	86.10	62.20	24.86	19.30	62.20	17.86	14.91	14.34	40.73
	LSST [63]	76.34	93.58	91.40	31.42	45.65	69.52	14.37	37.55	19.31	53.23
	WSCL [22]	78.43	93.31	91.99	31.89	36.16	69.82	15.76	46.97	27.35	54.34
	AugSeg [32]	**80.02**	93.70	92.78	26.39	**59.74**	73.69	19.18	58.25	58.94	62.52
	MBE	79.17	94.80	**94.44**	32.45	51.58	**74.49**	**20.22**	**67.94**	**70.96**	**65.12**

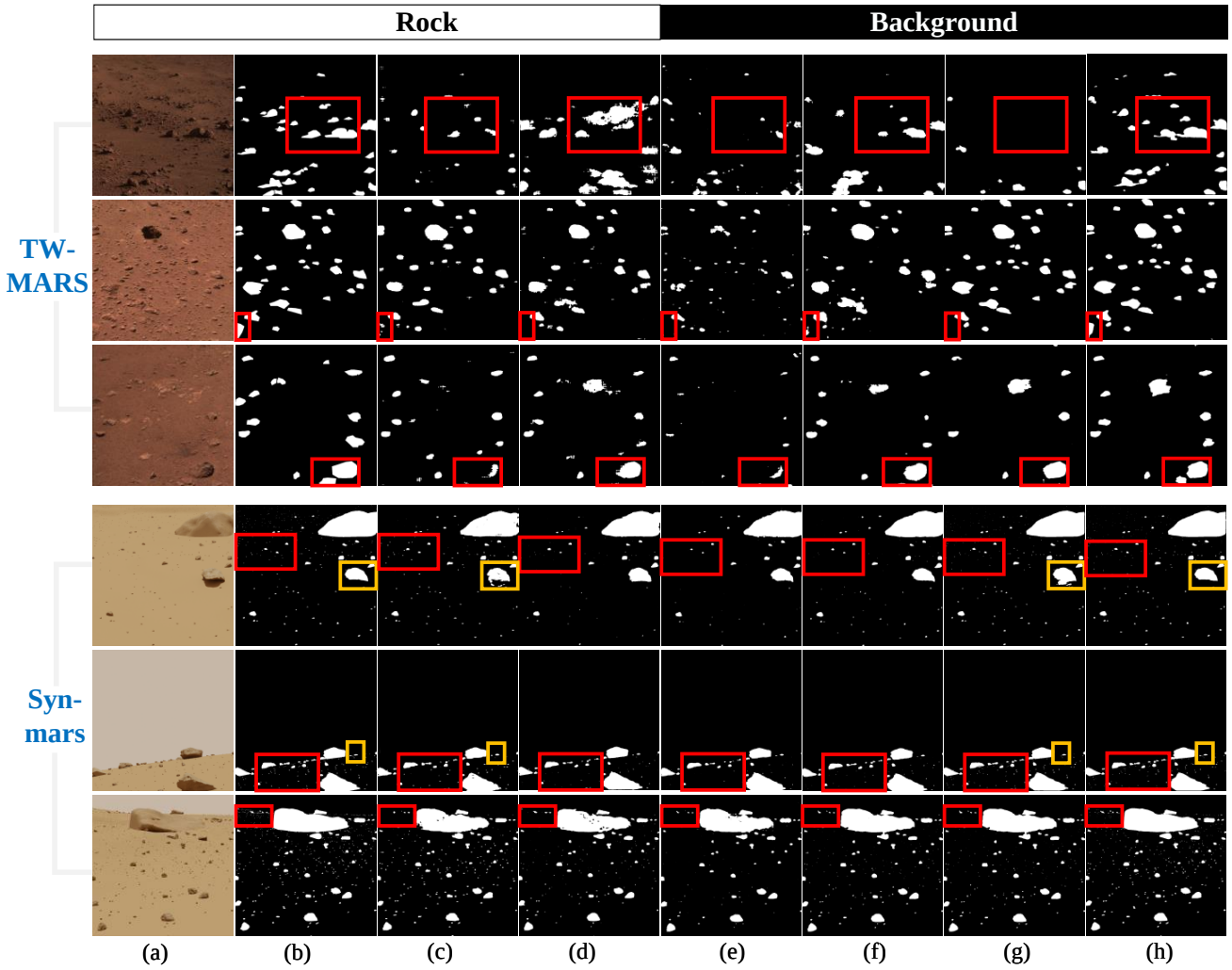


Fig. 9. Visual comparisons of different semi-supervised segmentation methods for the TWMARS and Synmars datasets. (a) Raw image. (b) Reference. (c) s4GAN. (d) ClassMix. (e) LSST. (f) WSCL. (g) AugSeg. (h) MBE.

highlighted the MBE method's stability and reliability in high-Mean IoU scenarios.

TABLE V
PER-CLASS IOU(%) FOR THE TWMARS TEST SET.

Labeled	Method	Rock	Background	Mean IoU
1/4	s4GAN [61]	43.62	98.10	70.86
	ClassMix [62]	49.08	98.56	73.82
	LSST [63]	25.01	97.61	61.31
	WSCL [22]	46.57	98.19	72.38
	AugSeg [32]	59.97	98.64	79.31
	MBE	**65.84**	**98.86**	**82.35**
1/8	s4GAN [61]	33.57	97.92	65.75
	ClassMix [62]	37.43	98.12	67.77
	LSST [63]	18.65	74.40	58.02
	WSCL [22]	43.40	98.00	70.70
	AugSeg [32]	59.39	98.65	79.02
	MBE	**61.45**	**98.69**	**80.07**

Despite the overall effectiveness of MBE, it still has some limitations. For example, on the MER dataset with the 1/8 labeled protocol, s4GAN achieved a notably higher IoU of 90.13% for the Background category compared to MBE's 71.81%. Similarly, on the Marscapes dataset with the 1/4 labeled protocol, s4GAN outperformed MBE in the Unknown

TABLE VI
PER-CLASS IOU(%) FOR THE SYNMARS TEST SET.

Labeled	Method	Rock	Background	Mean IoU
1/4	s4GAN [61]	70.43	98.86	84.64
	ClassMix [62]	68.90	98.70	83.80
	LSST [63]	52.34	96.92	74.63
	WSCL [22]	52.68	98.01	75.35
	AugSeg [32]	82.09	99.28	90.68
	MBE	**82.75**	**99.31**	**91.03**
1/8	s4GAN [61]	69.30	98.76	84.03
	ClassMix [62]	67.28	98.72	83.00
	LSST [63]	51.34	97.87	74.60
	WSCL [22]	51.98	97.88	74.93
	AugSeg [32]	82.03	99.28	90.65
	MBE	**82.37**	**99.29**	**90.83**

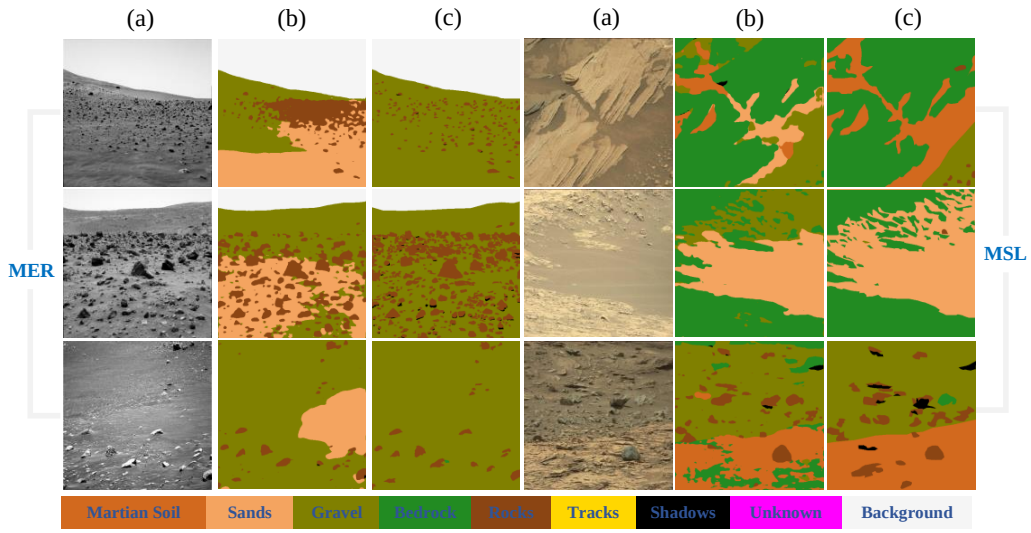


Fig. 10. Visual examples illustrating limitations of MBE on the MER and MSL datasets. (a) Raw image. (b) MBE prediction. (c) Reference.

category, achieving 69.71% versus 63.76%. This suggests that the adversarial training strategy of s4GAN may better capture complex background textures when labeled data are scarce. As illustrated in Fig. 10, MBE segments occasionally an entire object into scattered fragments. This can be observed in several categories, such as Sands and Gravel. This issue arises mainly from the use of the CutMix operation in MBE, which, while effective for uncertain regions, can disrupt the structural integrity of large or homogeneous areas, leading to decreased segmentation accuracy ultimately. While MBE could not achieve the highest IoU in every category across all six datasets, its integration of the ALDA module and SCF loss could enable the refinement of class boundary delineation, partially mitigating these limitations. Future work could focus on enhancing MBE's ability to model background textures and fuzzy regions, possibly by improving adversarial learning strategies or adaptive feature extraction techniques.

Overall, MBE's consistent top-tier IoU and notable gains across multiple classes highlighted its effectiveness in handling sparse labeled data. Its ability to achieve high IoU across various datasets underscored its robustness and adaptability, making it a valuable tool for enhancing remote sensing applications in planetary exploration. This high IoU suggested that MBE could advance terrain classification and data analysis in challenging environments, offering potential advances in the field of planetary science.

D. Ablation Experiments

1) Basic Ablation Experiments: This section explores the contributions of the ALDA module and SCF loss to the proposed method. To ensure the robustness of our findings, we conducted basic ablation studies across all six MT datasets. The baseline model was based on the MBE model, which removed the CutMix augmentation and used standard CE loss, serving as a point of comparison. All experiments were carried out under the 1/4 and 1/8 labeled data partition protocols.

The results, presented in Table VII, revealed several key insights. For example, incorporating the ALDA module alone consistently achieved higher IoU over the baseline model. However, the combination of the ALDA module with the SCF loss demonstrated even more significant gains (p <0.05, McNemar's test [65]). On average, this combined approach increased the Mean IoU by approximately 1-2% compared to the ALDA-only method and 2-3% compared to the baseline model across different datasets and labeled data fractions.

Furthermore, with reduced labeled data (1/8 partition), the boost in IoU achieved by combining the ALDA module and the SCF loss became more evident. The combined method showed variable improvements across datasets, with the most substantial gain of 4.06% observed on MSL (1/8 partition). The mean improvement across all datasets was 2.28% (1/8 partition) and 1.93% (1/4 partition), indicating enhanced efficacy under greater label scarcity. However, a slight drop was observed on MER (1/8 partition), likely due to noisy pseudo-labels under extremely limited supervision. In this case, SCF loss may have overemphasized incorrectly labeled hard examples, leading to degradation of the performance. Nonetheless, SCF increases consistently segmentation performance on other datasets, validating its overall robustness.

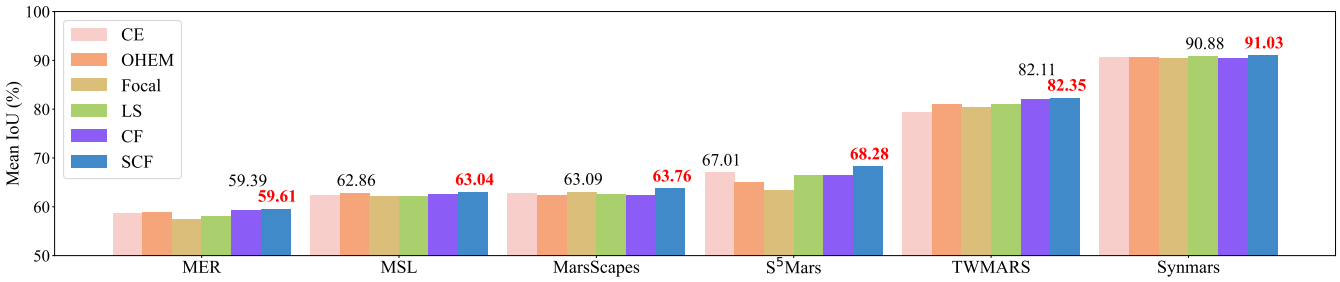
The above findings underscored that integrating the ALDA module and SCF loss not only provided notable increases in scenarios with ample labeled data but also delivered substantial benefits when labeled data were scarce. The combined approach was particularly effective in handling class imbalance and increasing segmentation IoU across various datasets and data conditions.

2) Comparisons of SCF Loss with Other Losses: This section compares the SCF loss with the CE, Focal, LS, CF and OHEM losses. As illustrated in Fig. 11, SCF loss obtained high Mean IoU across all datasets and labeled data fractions.

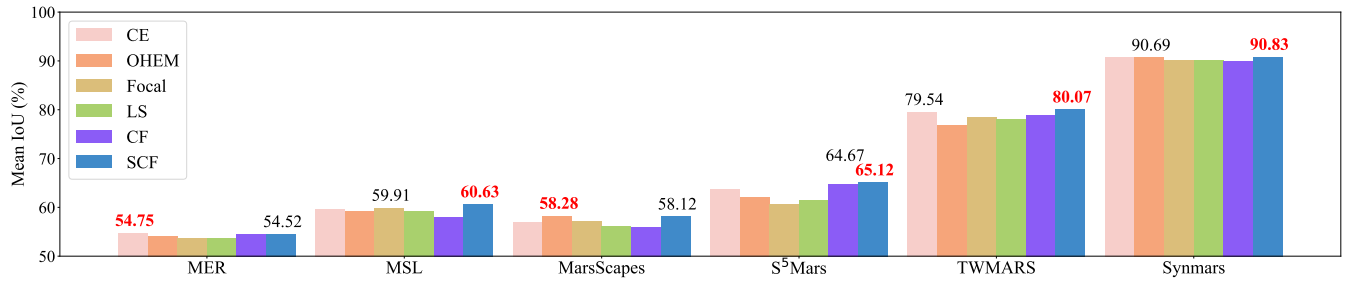
Specifically, with 1/4 of the labeled data, SCF loss achieved the highest Mean IoU consistently, surpassing the second-best method by 0.64%, 0.18%, 0.67%, 1.27%, 0.24%, and

TABLE VII
MEAN IOU (%) RESULTS OF ABLATION EXPERIMENTS FOR ALL SIX DATASETS UNDER DIFFERENT LABELED DATA PARTITION PROTOCOLS.

Labeled	Method	MER	MSL	MarsScapes	S^5Mars	TWMARS	Synmars
1/4	Baseline	57.64	60.63	62.18	65.60	80.03	90.44
	Baseline + ALDA module	58.74	62.45	62.83	67.01	79.32	90.68
	Baseline + ALDA module + SCF loss	**59.61**	**63.04**	**63.76**	**68.28**	**82.35**	**91.03**
1/8	Baseline	52.46	56.57	55.79	62.26	78.06	90.46
	Baseline + ALDA module	**54.75**	59.51	56.94	63.64	79.54	90.65
	Baseline + ALDA module + SCF loss	54.52	**60.63**	**58.12**	**65.12**	**80.07**	**90.83**



(a)



(b)

Fig. 11. Comparisons of semi-supervised segmentation methods using different loss functions. (a) under the 1/4 labeled partition protocol. (b) under the 1/8 labeled partition protocol. The red and black numbers represent the largest and second-largest IoU values, respectively.

0.15% on the MER, MSL, MarsScapes, S^5Mars, TWMARS and Synmars datasets, respectively. Furthermore, SCF loss, which is based on Focal loss, demonstrated a notable increase over the standard Focal loss. For example, while Focal loss achieved the lowest Mean IoU on the MER, MSL and S^5Mars datasets, SCF loss achieved the highest Mean IoU across these datasets consistently.

Similarly, with 1/8 of the labeled data, SCF loss demonstrated the highest Mean IoU on all datasets except MER and MarsScapes. For example, SCF loss surpassed the second-best method by 0.45% and 0.53% on the S^5Mars and TWMARS datasets, respectively. Although SCF loss did not achieve the top result on the MER and MarsScapes dataset, it iwas only 0.23% and 0.16% lower than the most accurate methods, respectively.

Additionally, SCF loss outperformed CE loss across all datasets and labeled data fractions consistently. Notably, the performance using CF loss exhibited instability, particularly on the MSL, MarsScapes and Synmars datasets under the 1/8 labeled partition protocol. This instability might stem from the inherent nature of CF loss, where the fixed γ amplifies consistently the loss of misclassified samples throughout training. Specifically, this characteristic causes the model to focus heavily on hard samples early in the training process, which might impede the model's ability to build a solid foundation on simpler patterns. As a result, the model's overall learning process became less stable, causing fluctuations in performance.

Overall, SCF loss demonstrated robust and excellent Mean IoU across most datasets and varying numbers of labeled data in this semi-supervised segmentation model. The symmetric and cyclical adjustment of SCF loss allows the model to progressively refine its ability to handle both simple and complex samples without prioritizing challenging examples prematurely. This approach helps to alleviate class imbalance effectively, consistently yielding high-IoU results even in scenarios with limited labeled data.

3) ALDA Module vs. Ala Module: This section compares the ALDA module with the popular Adaptive Label-aided (ALa) [32] module on the six MT datasets. For a fair comparison, we used two loss functions: CE and SCF losses, and other conditions were the same as in Section III-C.

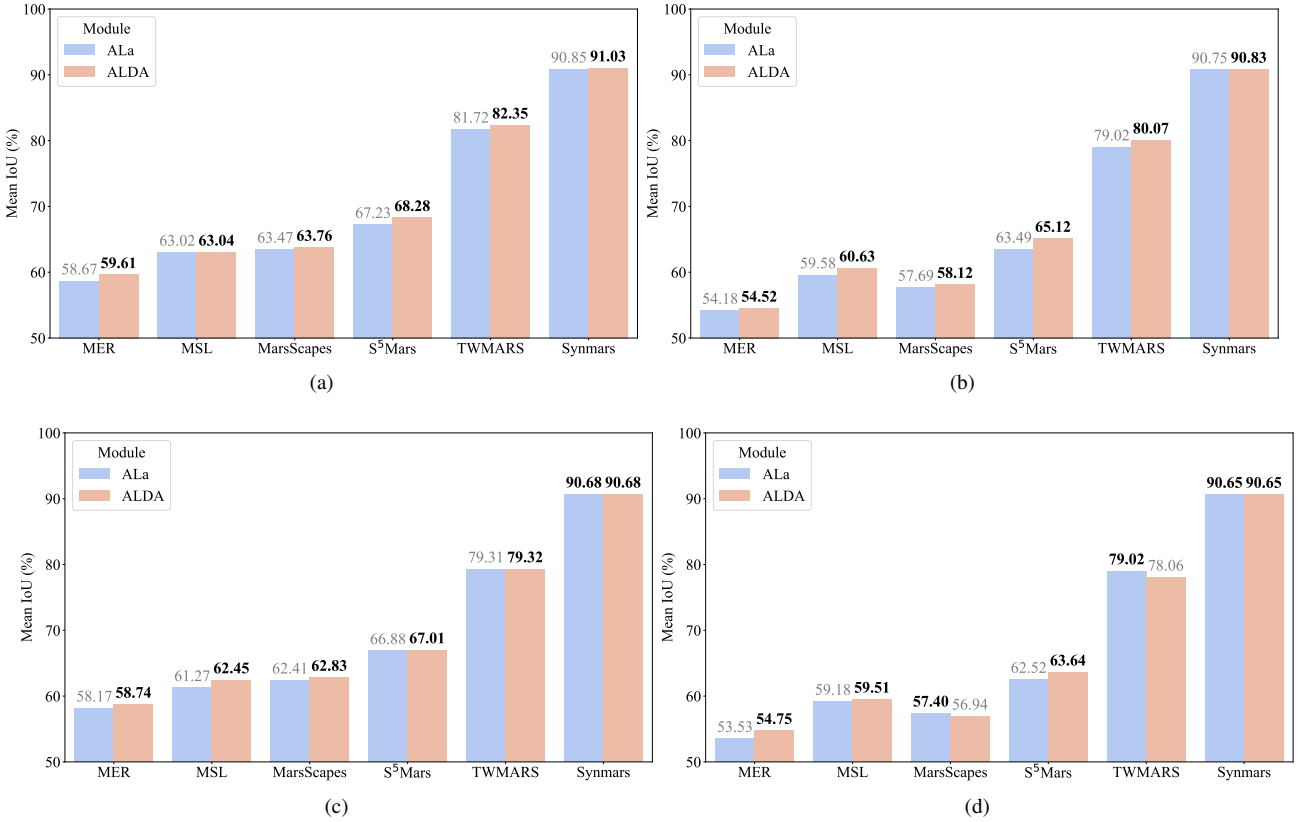


Fig. 12. Comparisons of Ala and ALDA modules under different labeled partition protocols and loss functions. (a) 1/4 labeled partition protocol with SCF loss. (b) 1/8 labeled partition protocol with SCF loss. (c) 1/4 labeled partition protocol with CE loss. (d) 1/8 labeled partition protocol with CE loss.

The results are shown in Fig. 12. The ALDA module demonstrated notable increases over the ALa method across these datasets. Specifically, for the 1/4 partition protocol using SCF loss, ALDA surpassed ALa consistently, achieving a 0.94% higher Mean IoU on the MER dataset. The MarsScapes and S^{5}Mars datasets also saw an increase in Mean IoU, with ALDA outperforming ALa by 0.29% and 1.05%, respectively. Even on the Synmars dataset, ALDA maintained its advantage, delivering a 0.18% higher Mean IoU compared to ALa.

In most cases, similar trends were observed across other partition protocols and loss functions. ALDA outperformed ALa generally or at least maintained parity. However, there was one notable exception: on the TWMARS dataset when using CE loss (Fig. 12 (d)), ALDA's Mean IoU was 0.96% lower than that of ALa. This discrepancy might be due to the CE loss's equal treatment of all classes, which could be less effective in handling the class imbalance in datasets.

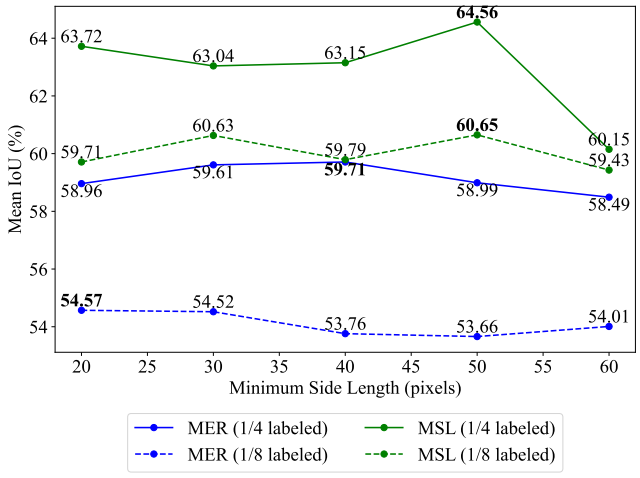


Fig. 13. Impact of minimum side length on Mean IoU (%) for the MER and MSL datasets with different labeled partition protocols.

IV. Discussion

A. Sensitivity Analysis of ALDA Module Parameters

The key contribution of the ALDA module is its use of a confidence detector to identify the lowest confidence regions in pseudo-labels. This confidence detector is characterized as an $h \times w$ box, where the box size influences directly the final segmentation results. Among various parameters, the minimum side length has an impact on the detection and refinement of the uncertain region. To investigate its impact, we conducted a sensitivity analysis by testing different minimum side lengths (20, 30, 40, 50, and 60 pixels) and observing their effect on the segmentation results. To ensure efficiency and focused analysis, this sensitivity analysis was conducted on two representative datasets: MER and MSL.

The result (Fig. 13) indicated that setting the minimum side length between 20 and 50 pixels generally yields high Mean IoU results across both datasets. For example, for the

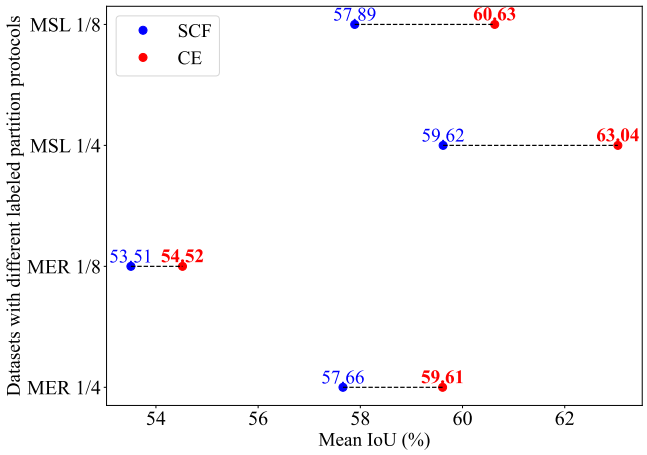


Fig. 14. Comparisons of unsupervised loss: SCF loss vs. CE loss across the MER and MSL Datasets.

MSL dataset, the IoU values for 1/4 of labeled images ranged from 63.04% to 64.56% consistently. Similarly, for the MER dataset, the IoU values for 1/4 of labeled images remained stable within this range.

Increasing the minimum side length to 60 pixels led to a notable decline in Mean IoU. For example, on the MSL dataset, the IoU for 1/4 labeled images decreased by 4.41% when increasing from 50 to 60 pixels, and for 1/8 labeled images, it decreased by 1.22%. This decline was unexpected because, theoretically, a larger minimum side length should capture more significant regions. The observed decline might be due to the inclusion of excessive background or non-target information, which can disrupt the object's semantic integrity and lead to incomplete segmentation. Especially the CutMix method might exacerbate this issue by fragmenting object information, thus affecting the model's learning effectiveness.

In summary, while larger minimum side lengths covered a greater area, they compromised the complete semantic information of objects, resulting in decreased overall performance. Consequently, despite comparable Mean IoU values at 40-50 pixels, we selected 30 pixels as the optimal threshold, which can maintain precise uncertainty localization while avoiding the performance collapse observed at 60 pixels.

B. Application of SCF Loss to Unsupervised Loss

In this research, we utilized the SCF loss in the student model primarily for two key reasons: first, the student model is the main network responsible for learning and adaptation; second, the teacher model provides guidance and generates pseudo-labels, but its role is less central in the loss calculation process. However, we also aimed to test the effectiveness of SCF loss when applied to the teacher model to explore its potential benefits and impact in that context.

The experiment was tested on the MER and MSL datasets, and the results are shown in Fig. 14. The results indicated a notable decrease in performance early when the SCF loss was applied to the teacher model instead of the standard CE loss. On the MER dataset, the Mean IoU dropped 1.95% and 1.01% from CE to SCF loss when using 1/4 and 1/8 of the labeled images. More importantly, on the MSL datasets, the Mean IoU experienced a larger decline of 3.42% and 2.74%, respectively when using 1/4 and 1/8 of the labeled images, showing a more pronounced decrease in performance.

The above results suggested that applying SCF loss to the teacher model did not yield the same level of performance as it did in the student model. One potential reason for this decline is that the teacher model focuses primarily on generating pseudo-labels rather than direct adaptation and learning from the data. The introduction of SCF loss may disrupt this process, leading to less accurate pseudo-labels, which negatively impacts the student model's effectiveness.

C. Applicability of MBE

In this research, the effectiveness of MBE was evaluated on MT images. However, the applicability of MBE is expected to extend well beyond MT segmentation. This model can be employed effectively for similar tasks on other celestial bodies, such as the Moon and Earth, as well as in various remote sensing applications.

MBE alleviates two key challenges in MT segmentation: the generation of high-quality pseudo-labels and the mitigation of class imbalance, which are also prevalent issues in other remote sensing contexts. By introducing the ALDA module, MBE can learn from diverse unlabeled data distributions, which is critical when applying it to different planetary terrains that may exhibit varying characteristics. Notably, ALDA introduces only a modest computational overhead ($\sim$2.4 GFLOPs, +27 MB memory during training, and $\sim$3.6 ms/image in inference), making it feasible for practical applications, including real-time scenarios. Additionally, the proposed SCF loss not only increases the model's ability to segment minority classes but also maintains overall segmentation performance across all classes.

By tackling the above fundamental issues in segmentation, we expect that MBE has the potential to advance the state of the art in various fields, including agricultural land mapping, urban planning, environmental monitoring on Earth and rock mapping on the Moon. For example, in Lunar surface analysis, where manual labeling is extremely costly, ALDA can help to increase the segmentation performance of small or ambiguous objects such as craters or rocks.

D. Limitations of MBE

The mechanism underlying the MBE method has two primary limitations. First, the structure of objects in the raw image is destroyed when employing the CutMix operation. This may make it difficult for the model to capture the correct object features during learning, especially in categories such as Rock and Bedrock, that are sensitive to structure, which will introduce uncertainty. Further research is needed to consider object completeness when using the CutMix method. Second, MBE needs to be more lightweight. MBE utilizes the teacher and student models simultaneously, which requires substantial computational power and memory, posing challenges for deployment on Mars rovers. Therefore, it is essential to develop a more lightweight model that maintains the performance of

the semi-supervised segmentation process, thereby facilitating effective terrain segmentation and analysis.

E. The Challenge of Classes Being Scale-Dependent in MT Images

Unlike Earth-based satellite images, which are captured at a standardized distance from the lens to the object, MT images are acquired typically by the navigation camera and panoramic camera mounted on rovers. Thus, MT images include oblique perspectives frequently, resulting in significant variations in the scale of measurement (i.e., spatial resolution relative to the scene). This introduces a unique challenge in MT image segmentation, specifically, the scale dependency of class definitions. For example, what is classified as Rock at a distance might be identified as Gravel when examined closely. This phenomenon leads to ambiguity of category labeling, which has the potential to affect the consistency and performance of segmentation results negatively. For this challenge, future research should be undertaken from two perspectives. One is labeling based on continuous images. Since MT images are not isolated from each other, they are taken according to the direction of the rover, resulting in inherent spatial and temporal continuity between images. Therefore, when marking an object within the distant view of an image, it is possible to refer to other MT images where this object might be in the close view. In this way, the labeling error can be reduced. The second perspective concerns the exploration of scale-invariant features and multi-scale learning techniques to alleviate this challenge more effectively, thereby increasing the segmentation robustness to the planetary terrain.

F. The Implications of Class Definitions for Generalization

In this research, the mean segmentation IoU for multi-class MT images is approximately 60%, which is relatively low for segmentation tasks on Earth. While the proposed method is semi-supervised, prior studies have shown that semi-supervised methods have reached basically the level of fully supervised methods. This suggests that the challenge may lie not only in the method itself but also in the way MT classes are defined, as unclear or overlapping class definitions could lead to reduced performance. Unlike Earth-based satellite images, where class hierarchies and definitions are well-established, MT classes are often less clear and may not follow traditional ontological frameworks. For example, regions where sand dunes meet rocky outcrops may display both sandy and rocky features, with no clear-cut boundary between the two. The surface may contain both loose Sand and embedded Rocks, making it challenging to classify the area strictly as either Sand or Rock. It is difficult to distinguish these unclear or overlapping classes due to overlapping characteristics or ambiguous boundaries. Regarding this issue, future research should focus on developing more standardized and precise class definitions. This could involve refining current categories, exploring hierarchical classification models to reflect potential class relationships, or even adopting fuzzy classification techniques to account more effectively for overlapping or ambiguous terrain features.

V. Conclusion

In this paper, we presented an end-to-end, efficient SSS method, namely, MBE, for MT semantic segmentation. MBE aims to alleviate the challenging issues of pseudo-label quality for unlabeled data and class imbalance for labeled data in existing SSS methods. Experimental results demonstrated that the introduction of the ALDA module and SCF loss in MBE could increase semantic segmentation IoU. Extensive comparisons between MBE and five benchmark methods (s4GAN, ClassMix, LSST, WSCL and AugSeg) across six MT datasets—MER, MSL, MarsScapes, S^5Mars, TWMARS and Synmars—demonstrated the robustness and efficiency of MBE for MT semantic segmentation. The key findings of this paper are summarized as follows.

I. The proposed MBE method outperformed five advanced SSS methods on six open-sourced MT datasets consistently. For example, for the S^5Mars dataset, MBE achieved the highest IoU of 68.28% , exceeding the second-best method by 1.40% under the 1/4 partition protocol.
II. The ALDA module generated higher-quality pseudo-labels than other methods by detecting the lowest confidence areas in pseudo-labels and subsequently increasing segmentation Mean IoU. For example, for the S^5Mars dataset, the IoU of MBE was 1.05% larger than for the second-best method under the 1/4 partition protocol.
III. The SCF loss provided a more effective solution for alleviating class imbalance compared to classical and class imbalance-sensitive loss functions, such as the CE, Focal, LS, CF and OHEM losses. For example, for the S^5Mars dataset, the IoU of MBE surpasses the second-best method by 1.17% under the 1/4 partition protocol.

Overall, MBE offers an effective solution for MT semantic segmentation. Future research will focus on generating higher-quality pseudo-labels with refined object completeness and developing more refined model architectures.

References

[1] S. J. Hoffman, *Human exploration of Mars: the reference mission of the NASA Mars exploration study team*. National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, 1997, vol. 6107.
[2] J. P. Grotzinger, J. Crisp, A. R. Vasavada, R. C. Anderson, C. J. Baker, R. Barry, D. F. Blake, P. Conrad, K. S. Edgett, B. Ferdowski *et al.*, "Mars science laboratory mission and science investigation," *Space science reviews*, vol. 170, pp. 5–56, 2012.
[3] D. Liu and W. Cheng, "Progress and prospects for research on martian topographic features and typical landform identification," *Frontiers in Astronomy and Space Sciences*, vol. 10, p. 1275516, 2023.
[4] F. Neukart, "Towards sustainable horizons: A comprehensive blueprint for mars colonization," *Heliyon*, 2024.
[5] J. A. Grant, M. P. Golombek, S. A. Wilson, K. A. Farley, K. H. Williford, and A. Chen, "The science process for selecting the landing site for the 2020 mars rover," *Planetary and Space Science*, vol. 164, pp. 106–126, 2018.
[6] F. A. Cucinotta, "Review of nasa approach to space radiation risk assessments for mars exploration," *Health physics*, vol. 108, no. 2, pp. 131–142, 2015.
[7] N. E. Putzig, G. A. Morgan, H. G. Sizemore, D. M. Hollibaugh Baker, E. I. Petersen, A. V. Pathare, C. M. Dundas, A. M. Bramson, S. W. Courville, M. R. Perry *et al.*, "Ice resource mapping on mars," in *Handbook of space resources*. Springer, 2023, pp. 583–616.

[8] T. Ke, Y. Zhong, M. Song, X. Wang, and L. Zhang, "Mineral detection based on hyperspectral remote sensing imagery on mars: From detection methods to fine mapping," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 218, pp. 761–780, 2024.

[9] J. Zhao, Z. Xiao, J. Huang, J. W. Head, J. Wang, Y. Shi, B. Wu, and L. Wang, "Geological characteristics and targets of high scientific interest in the zhurong landing region on mars," *Geophysical Research Letters*, vol. 48, no. 20, p. e2021GL094903, 2021.

[10] J.-F. Yang, D.-W. Liu, B. Xue, J. Lyu, J.-J. Liu, F. Li, X. Ren, W. Ge, B. Liu, X.-L. Ma *et al.*, "Design and ground verification for multispectral camera on the mars tianwen-1 rover," *Space Science Reviews*, vol. 218, no. 3, p. 19, 2022.

[11] Z. Zeng, J. Zhou, B. Li, Y. Tang, and M. Yan, "Lfea-net: semantic segmentation for urban point cloud scene via local feature extraction and aggregation," *Urban Informatics*, vol. 2, no. 1, p. 8, 2023.

[12] M. Hao, S. Chen, H. Lin, H. Zhang, and N. Zheng, "A prior knowledge guided deep learning method for building extraction from high-resolution remote sensing images," *Urban Informatics*, vol. 3, no. 1, p. 6, 2024.

[13] Y. Dai, T. Zheng, C. Xue, and L. Zhou, "Segmarsvit: Lightweight mars terrain segmentation network for autonomous driving in planetary exploration," *Remote Sensing*, vol. 14, no. 24, p. 6297, 2022.

[14] T. Panambur, D. Chakraborty, M. Meyer, R. Milliken, E. Learned-Miller, and M. Parente, "Self-supervised learning to guide scientifically relevant categorization of martian terrain images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1322–1332.

[15] L. Yang, G. Huang, and Y. Xing, "Improved mars terrain segmentation in terms of style transfer and network architecture," *International Journal of Aeronautical and Space Sciences*, pp. 1–14, 2024.

[16] Y. Xiong, X. Xiao, M. Yao, H. Cui, and Y. Fu, "Light4mars: A lightweight transformer model for semantic segmentation on unstructured environment like mars," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 214, pp. 167–178, 2024.

[17] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[18] T. S. Kim, G. Jang, S. Lee, and T. Kooi, "Did you get what you paid for? rethinking annotation cost of deep learning based computer aided detection in chest radiographs," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 261–270.

[19] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," *arXiv preprint arXiv:1802.07934*, 2018.

[20] H. Kervadec, J. Dolz, É. Granger, and I. Ben Ayed, "Curriculum semi-supervised segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019, pp. 568–576.

[21] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2613–2622.

[22] X. Lu, L. Jiao, L. Li, F. Liu, X. Liu, S. Yang, Z. Feng, and P. Chen, "Weak-to-strong consistency learning for semisupervised image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[23] W. Wang, L. Lin, Z. Fan, and J. Liu, "Semi-supervised learning for mars imagery classification and segmentation," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 4, pp. 1–23, 2023.

[24] O. Chapelle, B. Scholkopf, and A. Zien, Eds., "Semi-supervised learning (chapelle, o. et al., eds.; 2006) [book reviews]," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 542–542, 2009.

[25] A. Peláez-Vegas, P. Mesejo, and J. Luengo, "A survey on semi-supervised semantic segmentation," *arXiv preprint arXiv:2302.09899*, 2023.

[26] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.

[27] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7236–7246.

[28] C. Liang, G. Yang, L. Qiao, Z. Huang, H. Yan, Y. Wei, and W. Zuo, "Layermatch: Do pseudo-labels benefit all layers?" *arXiv preprint arXiv:2406.14207*, 2024.

[29] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *Advances in neural information processing systems*, vol. 32, 2019.

[30] X. Wei, X. Wei, X. Kong, S. Lu, W. Xing, and W. Lu, "Fmixcutmatch for semi-supervised deep learning," *Neural Networks*, vol. 133, pp. 166–176, 2021.

[31] L. Yang, X. Zhang, J. Li, L. Wang, M. Zhu, C. Zhang, and H. Liu, "Mix-teaching: A simple, unified and effective semi-supervised learning framework for monocular 3d object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 11, pp. 6832–6844, 2023.

[32] Z. Zhao, L. Yang, S. Long, J. Pi, L. Zhou, and J. Wang, "Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 350–11 359.

[33] L. Bai, H. Wang, X. Zhang, W. Qin, B. Liu, and S. Du, "Ap-semi: Improving the semi-supervised semantic segmentation for vhr images through adaptive data augmentation and prototypical sample guidance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[34] J. Li, S. Zi, R. Song, Y. Li, Y. Hu, and Q. Du, "A stepwise domain adaptive segmentation network with covariate shift alleviation for remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[35] J. Zhang, L. Lin, Z. Fan, W. Wang, and J. Liu, "S^5Mars: Semi-supervised learning for mars semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.

[36] H. Liu, M. Yao, X. Xiao, and H. Cui, "A hybrid attention semantic segmentation network for unstructured terrain on mars," *Acta Astronautica*, vol. 204, pp. 492–499, 2023.

[37] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5375–5384.

[38] M. Peng, Q. Zhang, X. Xing, T. Gui, X. Huang, Y.-G. Jiang, K. Ding, and Z. Chen, "Trainable undersampling for class-imbalance learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4707–4714.

[39] E. Rendon, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutierrez, "Data sampling methods to deal with the big data multi-class imbalance problem," *Applied Sciences*, vol. 10, no. 4, p. 1276, 2020.

[40] A. Katharopoulos and F. Fleuret, "Not all samples are created equal: Deep learning with importance sampling," in *International conference on machine learning*. PMLR, 2018, pp. 2525–2534.

[41] S. Shi, Y. Zhong, Y. Liu, L. Zhang, and D. Li, "Cross-temporal high spatial resolution urban scene classification and change detection based on a class-weighted deep adaptation network," *Urban Informatics*, vol. 3, no. 1, p. 3, 2024.

[42] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[43] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 761–769.

[44] L. N. Smith, "Cyclical focal loss," *arXiv preprint arXiv:2202.08978*, 2022.

[45] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol. 30, 2017.

[46] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

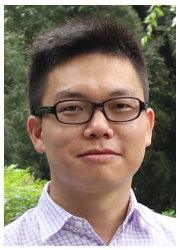
[47] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 674–12 684.

[48] I. Alonso, A. Sabater, D. Ferstl, L. Montesano, and A. C. Murillo, "Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8219–8228.

[49] I. Grubišić, M. Oršić, and S. Šegvić, "Revisiting consistency for semi-supervised semantic segmentation," *Sensors*, vol. 23, no. 2, p. 940, 2023.
[50] Y. Li, X. Wang, L. Yang, L. Feng, W. Zhang, and Y. Gao, "Diverse cotraining makes strong semi-supervised segmentor," *arXiv preprint arXiv:2308.09281*, 2023.
[51] X. Wang, H. Bai, L. Yu, Y. Zhao, and J. Xiao, "Towards the uncharted: Density-descending feature perturbation for semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 3303–3312.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[53] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
[54] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the class imbalance problem," in *2008 Fourth International Conference on Natural Computation*, vol. 4, 2008, pp. 192–201.
[55] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem," *Int. J. Advance Soft Compu. Appl*, vol. 5, no. 3, pp. 176–204, 2013.
[56] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of big data*, vol. 6, no. 1, pp. 1–54, 2019.
[57] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 41–48.
[58] L. Rice, E. Wong, and Z. Kolter, "Overfitting in adversarially robust deep learning," in *International conference on machine learning*. PMLR, 2020, pp. 8093–8104.
[59] W. Lv, L. Wei, D. Zheng, Y. Liu, and Y. Wang, "Marsnet: Automated rock segmentation with transformers for tianwen-1 mission," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.
[60] H. Liu, M. Yao, X. Xiao, and Y. Xiong, "Rockformer: A u-shaped transformer network for martian rock segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.
[61] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high- and low-level consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 4, pp. 1369–1379, 2021.
[62] V. Olsson, W. Tranheden, J. Pinto, and L. Svensson, "Classmix: Segmentation-based data augmentation for semi-supervised learning," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 1368–1377.
[63] X. Lu, L. Jiao, F. Liu, S. Yang, X. Liu, Z. Feng, L. Li, and P. Chen, "Simple and efficient: A semisupervised learning framework for remote sensing image semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.
[64] R. Müller, S. Kornblith, and G. E. Hinton, "When does label smoothing help?" *Advances in neural information processing systems*, vol. 32, 2019.
[65] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

Tong Xiao received the B.S. degree in Geographic Information Science (GIS) and M.S. degree in Cartography and Geographic Information from Hunan Normal University, Changsha, China in 2019 and 2023, respectively. He is currently pursuing a Ph.D. in the College of Surveying and Geo-Informatics at Tongji University.

His research focuses on Martian terrain recognition and small object detection.



Qunming Wang received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2015.

He was a Lecturer (an Assistant Professor) with the Lancaster Environment Centre, Lancaster University, Lancaster, U.K., from 2017 to 2018, where he is currently a Visiting Professor. He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He has authored or coauthored over 100 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, *IEEE Transactions on Geoscience and Remote Sensing*, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Dr. Wang serves as Associate Editor for *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering and Remote Sensing*. He was an Associate Editor of *Computers and Geosciences* from 2017 to 2020.

Peter M. Atkinson received the Ph.D. degree from the University of Sheffield (NERC CASE Award with Rothamsted Experimental Station), Sheffield, U.K., in 1990, and the M.B.A. degree from the University of Southampton, Southampton, U.K., in 2012.

He was a Professor of geography at the University of Southampton, where he is currently a Visiting Professor. He is also a Distinguished Professor of spatial data science and the Dean of the Faculty of Science and Technology, Lancaster University, Lancaster, U.K., and a Visiting Professor at Tongji University, Shanghai, China. He has published over 400 peer-reviewed articles in international scientific journals and around 50 refereed book chapters. He has also edited 13 journal special issues and eight books. His research interests include remote sensing, geographical information science, and spatial (and space–time) statistics applied to a range of environmental science and socioeconomic problems.

Prof. Atkinson is a Fellow of the Learned Society of Wales. He was a recipient of the Peter Burrough Award of the International Spatial Accuracy Research Association. He held the Belle van Zuylen Chair at Utrecht University, Utrecht, The Netherlands. He is an Editor-in-Chief of *Science of Remote Sensing*, a sister journal of *Remote Sensing of Environment*.

Xiaohua Tong (Fellow, IEEE) received the Ph.D. degree in geographic information system from Tongji University, Shanghai, China, in 1999.

From 2001 to 2003, he was a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. In 2006, he was a Research Fellow with Hong Kong Polytechnic University, Hong Kong. From 2008 to 2009, he was a Visiting Scholar with the University of California, Santa Barbara, CA, USA. He is currently a Professor with the College of Surveying and Geoinformatics, Tongji University. He is also an Academician at the Chinese Academy of Engineering, Beijing, China. His research interests include remote sensing, geographic information systems, uncertainty and spatial data quality, and image processing for high-resolution and hyperspectral images.