An Encoder-Decoder Network Trained with Multi-Branch Auxiliary Learning for Extracting Transverse Aeolian Ridge Morphological Parameters from High-Resolution Mars Imagery

Jiahui Sun[1,3], Zhen Cao[1,3,*], Zhizhong Kang[1,2,3], Jingwen Li[1],Jianwu Jiang[1], Bo Song[1], Xing Zhang[1]

[1] College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China, sunjiahui@glut.edu.cn, caozhen@glut.edu.cn, zzkang@cugb.edu.cn, lijw@glut.edu.cn, fengbuxi@glut.edu.cn, bosong@glut.edu.cn, zhangxing@glut.edu.cn
[2] School of Land Science and Technology, China University of Geosciences, Xueyuan Road, Beijing, 100083, zzkang@cugb.edu.cn
[3] Subcenter of International Cooperation and Research on Lunar and Planetary Exploration, Center of Space Exploration, Ministry of Education of The People’s Republic of China, No. 29 Xueyuan Road, Haidian District, Beijing 100083 China, sunjiahui@glut.edu.cn, caozhen@glut.edu.cn, zzkang@cugb.edu.cn

Keywords: Mars exploration, Transverse aeolian ridges, Deep learning, Morphological parameters.

Abstract

Transverse aeolian ridges (TARs) are the most widely distributed and enigmatic aeolian landforms on the surface of Mars, holding significant research value and implications for interpreting ancient wind fields and environments, searching for water and life, and selecting landing sites. However, accurately interpreting the morphological parameters of TARs, including their edge contours and ridge lines, remains a challenge. To tackle this issue, this paper proposes a Multi-branch Auxiliary Training Encoder-Decoder Network (MATED-Net) for detecting the edge contours and ridge lines of TARs on Mars. Built upon the Unet architecture, MATED-Net incorporates four auxiliary training losses to perceive features at different scales. Then, We introduce a lightweight attention mechanism to guide the fusion of multi-scale features. Finally, an edge tracing loss is introduced to enhance the distinction between edge pixels and surrounding confusing pixels, thereby accurately tracking the true positions of edges. To verify the effectiveness of the MATED-Net in detecting TARs’ contours and ridge lines, this paper constructs a dataset of TAR ridge lines based on HiRISE and HiRIC imagery. To facilitate subsequent training and testing, all images were clipped to a size of $512 \times 512$and converted to the VOC dataset format, resulting in a total of 1000 images and corresponding label data. The experimental results demonstrate a precision of 0.72, a recall of 0.67, and a mean Intersection over Union (MIoU) of 0.57 for edge extraction.

1. Introduction

TARs are among the most widespread landforms on the Martian surface (Liu et al., 2023, Hugenholtz et al., 2017). Their morphology resembles both sand dunes (Lu et al., 2021) and ripples (Zimbelman et al., 2012), typically appearing as linear or crescent-shaped features perpendicular to the prevailing wind direction. However, the scale of TARs lies between sand dunes and ripples, and their near-global distribution makes TARs a unique landform. Consequently, TARs serve as a critical proxy for understanding ancient Martian climate and environmental changes.

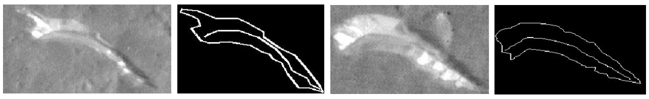

Figure 1. TARs and their ridges

With the advancement of Mars exploration missions, research on the morphology and spatial distribution of TARs has significantly accelerated, aiming to unravel their formation and evolutionary mechanisms. Wilson et al. (Wilson and Zimbelman, 2004) conducted the first systematic survey of the latitudinal distribution of TARs within the range of 180°E to 240°E, spanning from the South Pole to the North Pole, using High-Resolution Stereo Camera (HRSC) imagery. Their study provided crucial data for an initial understanding of TAR spatial distribution. Gou et al. (Gou et al., 2022) extracted TARs within a $1.9\,\mathrm{km}^2$region surrounding the Zhurong rover landing site, indicating that regional sediment transport is primarily driven by north-to-south winds, with a slightly greater contribution from northward-transported sediments. However, manually extracting TARs over large areas is both labor-intensive and time-consuming. With the rapid progress of artificial intelligence in computer vision, an increasing number of studies have begun to leverage deep learning techniques to enhance the automated extraction and interpretation of TARs. Palafox et al. (Palafox et al., 2017) were the first to propose Mars-Net for identifying the distribution areas of TARs on HiRISE images, defining the detection of TARs as a classification problem, and delineating the span of TARs areas by determining whether image blocks contain transverse aeolian ridge targets. Furthermore, Cao et al. (Cao et al., 2024) defined the extraction of TARs as a rotated object detection task, using directional bounding boxes to enclose the targets, thus not only locating the positions of TARs but also obtaining their directions. Zhang et al. (Zhang et al., 2024) employed the Mask R-CNN (He et al., 2017) model to extract TARs.This method initially extracts contours and subsequently estimates edges and ridges, identifying the centerline as the ridge line of the TARs based on the binary mask image. While this approach is effective for symmetric TARs, it only estimates the centerline for asymmetric TARs, which can compromise extraction accuracy. Overall, current deep learning methods can accurately locate the positions of , but interpreting their morphological characteristics remains a significant challenge.

The main challenges encountered in the extraction of TARs morphological parameters based on high-resolution images are:

- The morphological features of TARs are delicate, narrow,

[*] Corresponding author

and information-poor, which makes it challenging to perceive and extract these features accurately.

- Ten-meter-scale TARs often exhibit mixed and blurry areas between their contours and the background, as well as between the contours and ridge lines in high-resolution images. This blending of features makes it difficult to accurately extract mixed features.

- There is currently a lack of datasets for model training and testing, resulting in research in this field being still in its infancy.

To address these challenges, this paper proposes a multi-branch auxiliary training encoder-decoder network, termed MATED-Net, for detecting the edge contours and ridge lines of TARs in an end-to-end manner. The specific contributions of this paper are as follows:

- we introduce four auxiliary training losses from shallow to deep features on the basis of Unet, with shallow features used to perceive low-level features such as lighting and texture, and deep features used to perceive high-level semantic features.

- A feature fusion mechanism guided by attention is proposed, which first fuses the encoding and decoding stage features to produce features with edge discriminative power. Second, the multi-scale features produced during the encoding and decoding stages are fused to produce features with rich edge information;

- Finally, an edge tracking loss function is introduced to expand the difference between edge pixels and surrounding ambiguous pixels, tracking the actual positions of true edges.

2. Method

Figure 2 illustrates the overall architecture of the MATED-Net, which consists of three main modules: a feature extraction network based on the ”encoder-decoder” structure, a feature optimization module, and a multi-scale feature fusion module.

The encoder-decoder module is based on the Unet architecture and is used to extract multi-scale features of the target. We first employs a 3 × 3 convolution layer to expand the input image into a 32-dimensional feature space, followed by five stages of feature extraction in a top-down manner. The number of channels in each feature extraction layer is doubled, while the spatial dimensions are halved, aiming to aggregate contextual features and improve the overall feature representation. The feature scales outputted by the first four feature layers are $(512, 512, 64)$, $(256, 256, 128)$, $(128, 128, 256)$, and $(64, 64, 512)$, with the $(64, 64, 1024)$ network shallow features used to express low-level features such as edge texture features, and deep features expressing semantic features. Thus, the aggregation of low-level and high-level features can produce a more comprehensive and accurate feature representation to improve subsequent detection accuracy. To mix shallow and deep features to obtain a powerful feature representation, the decoding phase first upsamples the image, and then the upsampled features are added to the corresponding encoded features. To avoid merging irrelevant background information during feature addition, this paper introduces a lightweight attention mechanism called Attention Gate (LAG) as the feature optimization module. The LAG attention mechanism can assign weight information to each pixel, focusing on key features and avoiding irrelevant features such as the background. Finally, the multi-scale features obtained from the decoding are fused to form a complete application, using contour and ridge features for expression, and a 3×3 convolution is used as the target regression branch to output the classification result for each pixel.

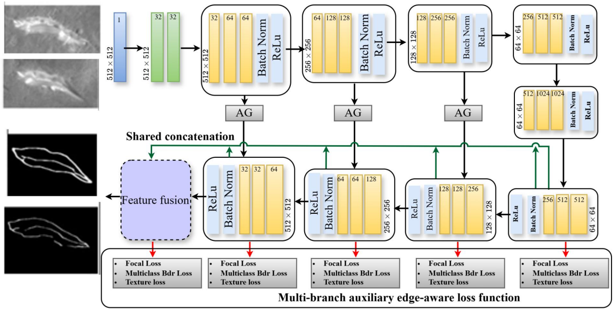

Figure 2. The pipline of MATED

2.1 Feature Optimization Network

The feature optimization network mainly employs an LAG attention to add weight information to pixels, prompting the model to focus on key edge and weak texture features, thereby guiding feature fusion to focus on the features that should be attended to. The calculation formula for the LAG attention is as follows:

$$\alpha_i = \psi^T \left(\sigma_1(W_x^T x_i + W_g^T g_i + b_g)\right) + b_\psi \tag{1}$$

$$Q_{Ag} = \sigma_2(\alpha_i(x_i, g_i)) \tag{2}$$

Where σ_1, σ_2 represent the activation functions. W,ψ represent the convolution weights. b_g, $b_\psi$are the convolution bias terms.

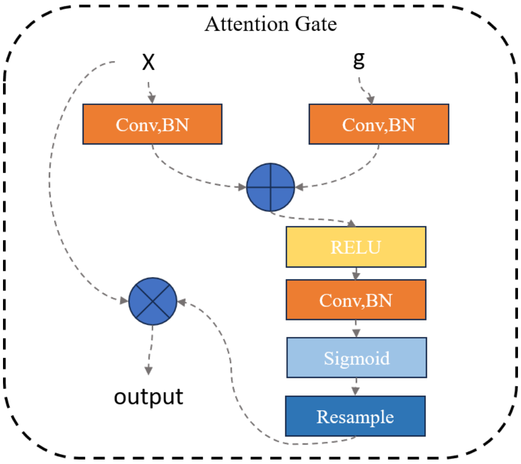

Figure 3. The pipline of LAG attention

The encoding feature is denoted as x, and the decoding feature is denoted as g. Both x and g are mapped to a single-channel image through a $1 \times 1 \times 1$ convolution, then they are fused by addition. Subsequently, the fused features are activated using the ReLU function, and the features after fusion continue to be enhanced through interaction with a $1 \times 1 \times 1$ convolution. Finally, the features pass through a Sigmoid function to filter them, and then they are resampled to obtain attention weights. The final decoded features are multiplied by the attention weights to obtain the enhanced features. The purpose of the $1 \times 1 \times 1$ convolution is to compress the dimensions in a low dimension, aggregate information, and make the fusion more complete, while reducing the number of parameters to improve computational efficiency. Compared to the CBAM, LAG attention has a smaller number of parameters and is more computationally efficient.

2.2 Multi-levle feature fusion

The MATED-Net adopts the Unet framework as the basic framework for auxiliary training. The model outputs four layers of features from shallow to deep, thus enhancing the expression of contour edge features through the method of fused features. The shallow features can provide low-level features, such as texture and edge information, while the deep features can provide high-level semantic features. The fusion of low-level and high-level features will produce features with more edge expression. Traditional edge contour extraction methods based on deep learning often use weighted methods to fuse multi-layer features, but they cannot guarantee that weights are assigned to key layers, which leads to ineffective feature fusion in the model. To this end, this article designs an attention-guided context-aware fusion module. Through the attention mechanism, it discriminates pixel information from the background and edges, then aggregates multi-layer features to enhance the expression of edge information, thereby avoiding manual weighting.

Let $Z = \{Z_1, Z_2, Z_3, Z_4\}$ be the set of decoded output features. When fusing features, we first use the LAG attention module to learn attention weights $Q_{Z_l} \in R^{H \times w \times L}$ from the edge feature map Z. Then, the attention weights are multiplied with the edge map to obtain attention features, which can highlight the target and suppress the background, thereby enhancing the fused feature representation. In the actual input process, the LAG attention module requires two feature maps as input. In this section, the two input feature maps are the same, both being Z. This process can be expressed as:

$$Q_{Z_1} = Q_{Ag}(Z_l, Z_l) \tag{3}$$

$$\text{CoFusion}(Z) = \sum_{i=1}^{L} Q_{z_i} \otimes Z \tag{4}$$

$$P_{\text{final}} = \text{sigmoid}\big(\text{CoFusion}(Z)\big) \tag{5}$$

Where the P_{final} is denoted as the edge prediction result, where the predicted edge values are 1, and non-edge values are 0.

2.3 Edge-tracking loss

The loss function assesses the discrepancy between the model’s predicts and ground truth. Therefore, the quality of the loss function directly affects the accuracy of the model detection. The limitations of directly applying classification or semantic segmentation loss to contour feature extraction are that contour features are narrow and elongated, with a pixel distribution that significantly differs from that of the background, making them easily overshadowed by background information. To address these challenges, this paper introduces a multi-level edge tracing loss function, which is designed to capture the edge contours and ridge lines of TARs effectively. Edge-tracking loss include, Focal Loss, edge-tracking loss function, and texture suppression loss.

2.4 Focal Loss

Focal Loss is used to address the issue of imbalanced background and edge pixel numbers. The main idea is to reduce the weight of easily classified samples, prompting the training model to focus on difficult samples, thereby improving the model’s performance under class imbalance conditions. It introduces a balancing parameter γ to adjust the weight difference between easily and difficultly classified samples. In edge detection tasks, due to the extremely imbalanced number of linear feature pixels compared to background pixels, Focal loss can reduce the impact of severely imbalanced samples on linear feature detection to some extent. Thus, it pays more attention to the target categories that are difficult to classify. The formula for multi-class Focal loss is as follows:

$$L_f(\widehat{Y}_i) = -\alpha_t (1 - \widehat{Y}_i)^{\gamma} \log\left(\widehat{Y}_i\right) \tag{6}$$

Where P_t represents the model’s prediction for each category, while α_t is the category weight that can be tuned based on the importance of the categories. If there are fewer target categories, a higher weight can be assigned to them. γ is used to adjust the focus of the loss, and γ is usually set to a value greater than 0 to increase attention to samples that are difficult to classify.

2.5 Tracking Loss Function

The tracking loss function is used for classifying target and background mixed pixels. The loss function was proposed by Huan et al. (Huan et al., 2021), which highlights the values of boundary points by calculating the difference between the boundary points and the pixel matrix of their respective neighborhoods, thereby reducing mixed pixels. The calculation formula for the tracking loss function is as follows:

$$L_{bdr}\left(\widehat{Y}_i, Y_i\right) = -\sum_{P \in E} \log\left(\frac{\sum_{i \in 1,} \widehat{y}_i}{\sum_{i \in R_P^e} \widehat{y}_i + \sum_{i \in L} \widehat{y}_i}\right) \tag{7}$$

Where E represents the set of boundary points, R^{e} represents the neighborhood matrix, and L represents the set of boundary points within the neighborhood matrix. When the loss function reaches its minimum value, $\sum_{i \in R_P^e} \widehat{y}_i$will tend to 0, while $\sum_{i \in L} \widehat{y}_i$ will tend to the maximum. Therefore, the boundary points are highlighted, and the background in the mixed pixel values is suppressed.

2.6 Texture Suppression Loss Function

After the edge tracing function has processed and identified the confusing pixels, the unnecessary texture areas remaining in the predicted map can be handled by the texture suppression function, as shown in the formula below:

$$L_{tex}\left(\widehat{Y}_i, Y_i\right) = -\sum_{P \in E} \log\left(1 - \sum_{i \in R_p^t} j\right) \tag{8}$$

Where R_p^t represents the neighborhood pixel matrix, with non-boundary points at the center. Since images contain more unnecessary texture information than edge-confusing pixels, the essence of the texture suppression function is to suppress the same texture regions of non-edge pixels, rather than operating on individual pixels.

In summary, the edge tracking loss function formula in this paper is as follows:

$$L = L_f(\widehat{Y_i}) + \lambda_1 L_{bdr}(\widehat{Y_i}, Y_i) + \lambda_2 L_{tex}(\widehat{Y_i}, Y_i) \qquad (9)$$

Where $\widehat{Y_i}$and $Y_i$represent the model predicted and label values, and λ_1 and λ_2 represent two hyperparameters to regulate the importance of the edge tracking function and the texture suppression function in the loss function.

3. Experiments

3.1 Experimental data

To verify the effectiveness of the MATED-Net network in detecting the contours and ridge lines of TARs, this paper randomly selected 600 TARs in the HiRISE images (0.25 m/pixel) and annotated their contours and ridge lines, with the contour and ridge line assigned a value of 1 and the background assigned a value of 0. To facilitate subsequent training and testing, all images were clipped to a size of 512 × 512 and then converted to VOC dataset format, resulting in a total of 1000 images and label data.

3.2 Experimental details and metrics

This experimental platform uses Ubuntu 20.04, with hardware configuration consisting of a Core i7-6700 3.30 GHz CPU and 2 NVIDIA GeForce RTX 3090 GPUs (each with 24GB memory). During the experiments, the Adam optimizer is used with an initial learning rate (Learning rate, Lr) set to 0.01, a learning momentum of 0.7, and a weight decay rate of 0.01. Distributed training is conducted across 2 GPUs simultaneously, with a batch size (Bs) of 4 for each GPU.

The evaluation metrics in this article use precision (P), recall (R), and mean intersection over union ($MIoU$). The calculation formulas for P, R, and $MIoU$ are as follows:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \qquad (10)$$

$$MIoU = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \qquad (11)$$

3.3 Extraction results

The MATED-Net extraction results are shown in Table 1. The extraction accuracy P is 0.72, the extraction rate R is 0.67, and MIoU is 0.57. Figure 4 (a) shows the extraction results of our method. It shows that our method accurately extracts the contours and ridge lines of TARs. Figure 4 (b) displays the multilayer feature extraction results of the MATED-Net. It shows that features from the first to the fourth layer are present, with a significant enhancement in feature perception capabilities. The fifth layer, which is the fused feature layer, encompasses the features of the previous four layers, resulting in more accurate feature extraction outcomes after the fusion.

	P	R	MIOU
MATED-Net	0.72	0.67	0.57

Table 1. Performance metrics of MATED-Net.

3.4 Comparison experiments

Table 2 shows the extraction results of our method compared with classical methods, including the CAT, HED, and Canny edge detectors. Our method achieved the best detection results, with a precision P higher than that of the Canny operator by 0.37 and a recall R of 0.16. However, it can be observed that the mIoU of our method is not the highest. This is because the predicted contours by our method are finer, making it difficult to match every pixel point with the manually annotated contours. Nevertheless, from the visualization results, our method successfully extracted the contours of each TARs.

Methods	P	R	MIoU
CAT	0.72	0.51	0.50
HED	0.73	0.49	0.48
Canny	0.35	0.51	0.26
MATED-Net-SE	0.70	0.63	0.59
MATED-Net with CBAM	0.70	0.60	0.59
MATED-Net with ECA	0.71	0.61	0.59
MATED-Net	0.72	0.67	0.57

Table 2. Comparison of different methods.

3.5 Ablation study

Tables 3 presents the ablation study results. The first line of the table shows the results of the LAG ablation experiment. Compared with the Unet model, the LAG module increased the accuracy, the extraction rate increased by 11%, and the MIoU increased by 13%. The second behavior increases the results after multiscale feature fusion, with P increasing by 2% and R by 5%. The final model in this paper, showed a significant 7% increase in accuracy, 1% increase in extraction rate and 4% more in MIoU compared to Unet + LAG + Cofusion. This experiment shows that this module is effective for improving the contour and ridge detection accuracy. To further verify the influence of the number of feature fusion layers on the detection progress of the model, the number of feature fusion layers was gradually increased from 1 to 4 to verify the influence of this module. As can be seen from Table 3, when the model is extracted after 4 layers of fusion, the accuracy is higher, the extraction rate is the highest, and the matching degree with the model is also the highest.

Methods	P	R	MIoU
Unet	0.60	0.50	0.44
Unet+LAG	0.63	0.61	0.53
Unet+LAG+Cofusion	0.65	0.66	0.53
Unet+LAG+Cofusion+Tracing Loss	0.72	0.67	0.57

Table 3. Ablation study.

4. Conclusion

This study has developed a novel MATED-Net to address the challenge of accurately interpreting the morphological para-

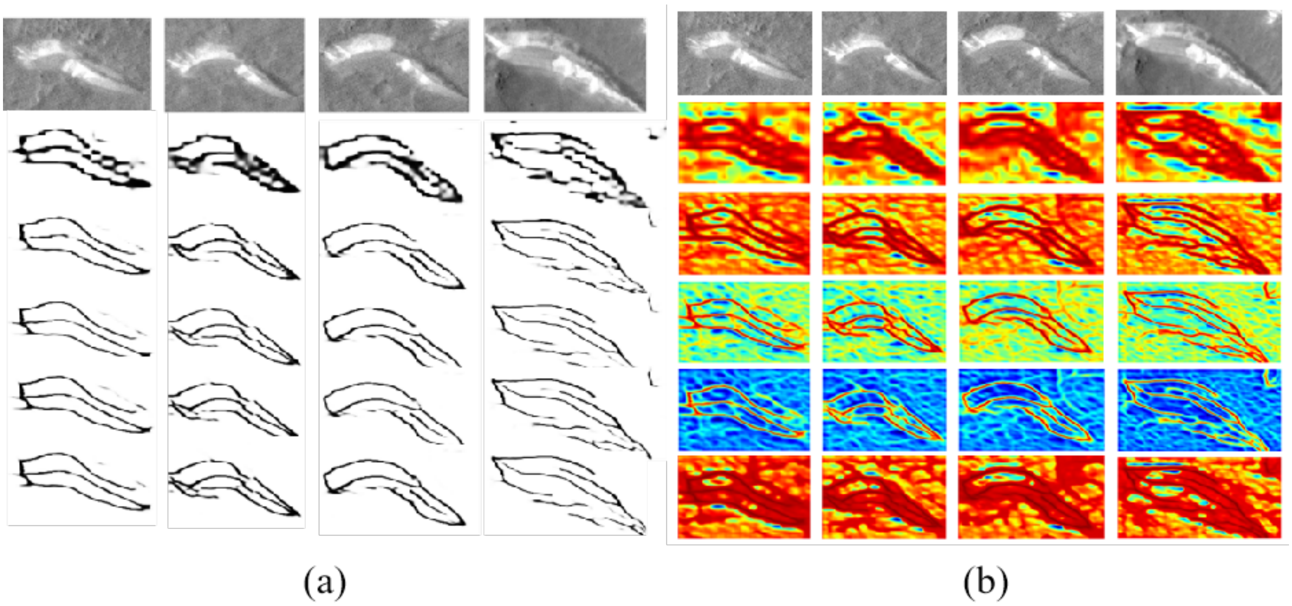

Figure 4. The extracting results of TARs. (a) the contours and ridge lines of TARs extracted by the four feature layers of the mated network, (b) the features of TARs perceived by the four feature layers of the mated network.

meters of TARs on Mars, specifically their edge contours and ridge lines. By integrating four auxiliary training losses, a lightweight attention mechanism, and an edge tracing loss, MATED-Net effectively enhances the detection of TAR features across different scales and complex backgrounds. The dataset constructed from HiRISE and HiRIC imagery, comprising 1000 images and corresponding labels in the VOC format, provided a robust foundation for training and testing MATED-Net. The experimental results are promising, with a precision of 0.72, a recall of 0.67, and a MIoU of 0.57 for edge extraction. These findings demonstrate the effectiveness of MATED-Net in accurately identifying TARs' contours and ridge lines. Future work will focus on further improving the robustness and accuracy of MATED-Net by incorporating additional data sources and exploring advanced training techniques. We also plan to extend the application of MATED-Net to the analysis of other aeolian landform features, aiming to contribute to a broader understanding of aeolian processes and landforms in diverse environments.

Acknowledgements

This work was supported by the China Postdoctoral Science Foundation (Certificate Number: 2024M760677)and the research startup funding from Guilin University of Technology.

References

Cao, Z., Kang, Z., Hu, T., Yang, Z., Chen, D., Ren, X., Meng, Q., Wang, D., 2024. AiTARs-Net: A novel network for detecting arbitrary-oriented transverse aeolian ridges from Tianwen-1 HiRIC images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 135–155.

Gou, S., Yue, Z., Di, K., Zhao, C., Bugiolacchi, R., Xiao, J., Cai, Z., Jin, S., 2022. Transverse aeolian ridges in the landing area of the Tianwen-1 Zhurong rover on Utopia Planitia, Mars. *Earth and Planetary Science Letters*, 595, 117764.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Huan, L., Xue, N., Zheng, X., He, W., Gong, J., Xia, G.-S., 2021. Unmixing convolutional features for crisp edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10), 6602–6609.

Hugenholtz, C. H., Barchyn, T. E., Boulding, A., 2017. Morphology of transverse aeolian ridges (TARs) on Mars from a large sample: Further evidence of a megaripple origin? *Icarus*, 286, 193–201.

Liu, J., Qin, X., Ren, X., Wang, X., Sun, Y., Zeng, X., Wu, H., Chen, Z., Chen, W., Chen, Y. et al., 2023. Martian dunes indicative of wind regime shift in line with end of ice age. *Nature*, 620(7973), 303–309.

Lu, Y., Edgett, K., Wu, Y., 2021. Ripples, Transverse Aeolian Ridges, and Dark-Toned Sand Dunes on Mars: A Case Study in Terra Sabaea. *Journal of Geophysical Research: Planets*, 126(10), e2021JE006953.

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

Wilson, S. A., Zimbelman, J. R., 2004. Latitude-dependent nature and physical characteristics of transverse aeolian ridges on Mars. *Journal of Geophysical Research: Planets*, 109(E10).

Zhang, J., Liu, S., Du, K., Tong, X., Xie, H., Feng, Y., Jin, Y., Lin, Y., Wan, B., 2024. Automatic extraction of Transverse Aeolian Ridges (TARs) and analysis of landform influence for the Zhurong landing area on Mars. *Geomorphology*, 467, 109489.

Zimbelman, J. R., Williams, S. H., Johnston, A. K., 2012. Cross-sectional profiles of sand ripples, megaripples, and dunes: a method for discriminating between formational mechanisms. *Earth Surface Processes and Landforms*, 37(10), 1120–1125.