C-PLES: Contextual Progressive Layer Expansion with Self-attention for Multi-class Landslide Segmentation on Mars using Multimodal Satellite Imagery

Abel A. Reyes[1], Sidike Paheding[1], A. Rajaneesh[2], K.S. Sajinkumar[1], Thomas Oommen[1]
[1]Michigan Technological University [2]University of Kerala
{areyesan, spahedin, skochapp, toomme}@mtu.edu rajaneesh@keralauniversity.ac.in

Abstract

Landslide segmentation on Earth has been a challenging computer vision task, in which the lack of annotated data or limitation on computational resources has been a major obstacle in the development of accurate and scalable artificial intelligence-based models. However, the accelerated progress in deep learning techniques and the availability of data-sharing initiatives have enabled significant achievements in landslide segmentation on Earth. With the current capabilities in technology and data availability, replicating a similar task on other planets, such as Mars, does not seem an impossible task anymore. In this research, we present C-PLES (Contextual Progressive Layer Expansion with Self-attention), a deep learning architecture for multi-class landslide segmentation in the Valles Marineris (VM) on Mars. Even though the challenges could be different from on-Earth landslide segmentation, due to the nature of the environment and data characteristics, the outcomes of this research lead to a better understanding of the geology and terrain of the planet, in addition, to providing valuable insights regarding the importance of image modality for this task. The proposed architecture combines the merits of the progressive neuron expansion with attention mechanisms in an encoder-decoder-based framework, delivering competitive performance in comparison with state-of-the-art deep learning architectures for landslide segmentation. In addition to the new multi-class segmentation architecture, we introduce a new multi-modal multi-class Martian landslide segmentation dataset for the first time. The dataset will be available at https://github.com/MAIN-Lab/C-PLES

1. Introduction

Landslides are a movement of terrain caused by different factors, including earthquakes, volcanic activity, and heavy rainfalls [17]. Therefore, understanding the cause that triggers their formations help us to unravel the morphological changes in a planet. The study of extraterrestrial terrains has been limited. The current availability of satellite imagery and cutting-edge deep learning (DL) algorithms opens the opportunity to exploit their capabilities for landslide detection outside of Earth. Since Mars has shown evidence that presumes a similar past that the current conditions on Earth, the study of the Martian terrain could have a potential contribution to estimating the transitions related to climate change that could be faced on Earth in a near future [13, 31]. The vast presence of landslides on the surface of Mars brings us the possibility to obtain valuable insights from their analysis. With the use of aerial imagery, a typical method to identify landslides on Mars requires the visual interpretation of optical images by geomorphologist analysis [5], a time-consuming procedure that could highly get bias the expertise of the analyst expert. In addition, this procedure based the criteria on similarities with terrestrial landslide [6], which includes shape, size, tone, mottling, texture, pattern of objects, and site topography [3, 21]. However, new studies suggested the inclusion of an additional set of criteria to recognize landslides in areas such as the Valles Marineris (VM) region [12], which is considered as a museum of landslides on Mars [38]. Characteristics such as length, width, relative relief, slope statistics, slope of the scarps, and geology are mentioned in addition to an empirical relationship between the volume and the area of the slides. The aforementioned

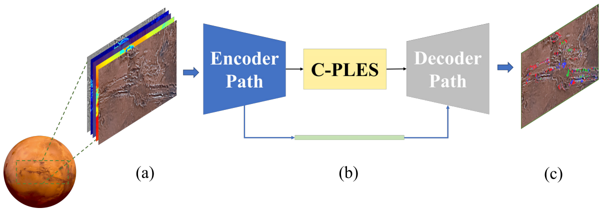


Figure 1. Location of Valles Marineris (VM) on Mars used for landslide mapping in this study. (a) The multi-modality imagery used to train the (b) proposed segmentation model. (c) The output landslide segmentation map.

methods rely upon the use of traditional machine learning algorithms, (i.e. logistic regression), to develop a landslide classifier based on the extracted features. With recent advancements in DL algorithms, several studies have achieved state-of-the-art results in computer vision tasks such as image classification [14, 49], object detection [18, 33], and semantic segmentation [11, 39]. However, there is still a gap between the implementation of cutting-edge techniques and the adoption of those techniques for specific tasks such as landslide segmentation. The use of automated feature extraction, provided by a DL architecture, eliminates any human bias during the feature extraction process and grants an end-to-end framework for landslide analysis, as it is shown in the adoption of Vision Transformers [15] for landslide classification task [27].

In this research work, as shown in Figure 1, we propose an end-to-end DL architecture to facilitate the task of identifying different types of landslides. To this end, the VM region is selected to build a dataset for landslide segmentation and to develop a model to automate this task efficiently. The proposed end-to-end Martian multi-class landslide segmentation model fuses heterogeneous multi-modality imagery data in a contextually aware manner. The main contributions of this research are summarized as follows:

- A new multi-modal multi-class dataset for the Martian landslide segmentation is introduced. The dataset is composed of three different versions, each version with a different image size, which will be publicly available to the research community.

- To the best of our knowledge, this study is the first attempt to propose an end-to-end DL framework for multi-class Martian landslide segmentation from satellite images. We named our proposed model C-PLES (Contextual Progressive Layer Expansion with Self-attention). Experimental results reveal effectiveness in the performance of C-PLES when compared to state-of-the-art DL architectures. In addition, the ablation study is conducted to support the significance of different components of C-PLES to reach this performance.

- The impact of different image modalities on the contribution of landslide segmentation accuracy is systematically analyzed. Experimental results indicate that imagery modalities, such as the Thermal Emission Imaging System (THEMIS) and Digital Elevation Model (DEM), have the most significant contribution to the segmentation accuracy in terms of Intersection over Union (IoU).

The rest of this paper is organized as follows. Section 2 reviews the landslide detection algorithms. Section 3 describes the proposed C-PLES architecture in detail. Section 4 introduces newly built datasets used in this study. Section 5 provides the experimental results. Section 6 summarizes the component contribution of the model throughout the ablation study. Section 7 concludes our findings in this study along with future research.

2. Related Work

To the best of our knowledge, we have not come across any published research that uses an end-to-end DL framework for multi-class Martian landslide segmentation. As a result, in this section, we discuss the most pertinent machine learning techniques for on-Earth landslides segmentation.

Traditional machine learning algorithms have been widely used in combination with DL and object-based image analysis (OBIA) techniques for landslide segmentation tasks [1, 28, 44]. For instance, Tavakkoli et al. [44] experimented OBIA with a set of machine learning methods to perform landslide detection, within a stacking machine learning framework. This method utilized random forest, logistic regression, and multilayer perceptron neural network as the set of selected machine learning algorithms (Level 0 of the stacking model), and logistic regression as the meta-learner (Level 1 of the stacking model) to make the final prediction. Overall, they reported the approach resulted in appropriate landslide detection. Keyport et al. [28] reported a comparative analysis for pixel-based landslide detection with the use of very high-resolution (VHR) remote sensing aerial imagery. They performed both a pixel-based and object-oriented analysis (OOA) for landslide mapping, in which the OOA method yielded better results with the presence of less number of false positives. Achariyaviriya et al. [1] utilized a DL approach by combining three versions of the ResNet [22] models, in which each model take a single data modality as an input, such as RGB color image, normalized difference vegetation index (NDVI), and slope factor (SP). Initially, Each ResNet model and the corresponding data modality are trained to generate a set of features per modality. Then, those sets of features are combined to train a final decision tree classifier model. As a result, the use of different modalities, rather than just RGB or gray-scale, helped to improve the performance of their proposed classifier.

The study conducted by Prakash et al. [37] used CNN as a semantic segmentation problem to map landslides. The researchers utilized high-resolution Lidar DEM and cloud-free optical images from Sentinel-2 for mapping. Pixel-based, object-based, and DL methods were used for generating landslide susceptibility maps. The study introduced CNN-based U-Net [40] and ResNet architectures for mapping landslides. The U-Net architecture was used for semantic segmentation, and ResNet was used for feature identification. The authors demonstrated that the U-Net with ResNet strategy outperforms pixel-based and object-based machine learning algorithms on a regional scale for map-

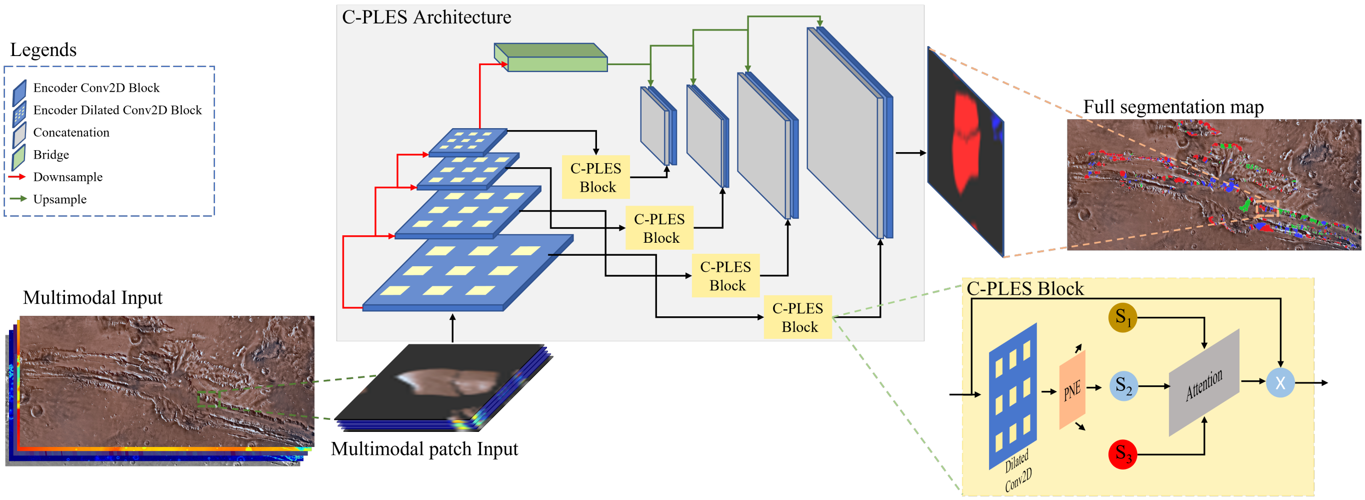


Figure 2. Detailed illustration of the proposed C-PLES architecture

ping larger landslides. However, all three approaches were unsuccessful in detecting minor landslides, and they had difficulty distinguishing individual landslides when they were close together. In summary, there is evidence that supports the use of multimodal inputs to improve the performance of landslide segmentation on Earth. However, it is worth mentioning that the contrast of the terrain and vegetation provides visual characteristics that eventually help to detect a landslide. In addition, another limitation lies in the presence of minor landslides, in which the models tend to fail to separate among individual landslides.

3. Methodology

The proposed C-PLES architecture for multi-class Martian landslides segmentation is illustrated in Figure 2. This architecture follows the pattern of an encoder-decoder structure, in which several components are incorporated to effectively extract pertinent features for multi-class landslide segmentation. Detailed explanations of different components in the C-PLES are described as follows.

3.1. Encoder-decoder network

An encoder-decoder architecture is a common approach used in semantic segmentation tasks [4,8,50]. This structure exploits multi-scale visual features extracted in the encoder side and recovers the spatial resolution in the decoder side [10]. However, this arrangement by itself tends to lose a certain range of contextual information. The encoder compresses the input into a lower-dimensional representation, which discards some of the spatial information of the original input [29], thus producing segmentation outputs with a lack of details.

3.2. Progressive Neuron Expansion

The concept of Progressive Neuron Expansion (PNE) is adopted from the work presented in [36, 42], in which each neuron from the input is progressively expanded following a Maclaurin series expansion of a nonlinear function. Every node, denoted as S_u, produced by the expansion is the result of the addition of the u subsequent component in the expansion. u represents a hyperparameter that controls the length (i.e., the number of terms in the Maclaurin series) of the expansion. The series expansion is mathematically expressed as follows:

$$S_u = \sum_{n=1}^{u} c_n x^{p_n} = c_1 x^{p_1} + c_2 x^{p_2} + ... + c_u x^{p_u} \tag{1}$$

where x represents the input neuron to be expanded, $c \in \{1, 1/2, 1/3\}$ and $p \in \{1, 2, 3\}$ are the non-trainable coefficients and powers in the Maclaurin series expansion of a nonlinear function, respectively.

3.3. Attention mechanisms

Attention mechanism in DL refers to a family of techniques that allow a neural network to focus on relevant parts of the input data, which is achieved by enabling the model to selectively attend to certain regions of an image, sequence, or another type of input while ignoring the remaining [20, 24, 34]. In this research, we explore the benefit of two particular attention mechanisms: dot-product attention and multi-head self-attention.

Dot-product attention, also known as Luong-style attention [32], is utilized, in this study, to leverage the production of enriched features from the PNE layer. This attention mechanism is typically used for neural machine translation [23,41,48]. The Luong-style attention involves a global attention model that considers all the sources in the input when generating the alignment scores and the context vector. To compute the alignment scores, an alignment model $a(.)$ is used, expressed as:

$$a_t(s) = align(\mathbf{h}_s, \mathbf{S}_i) \tag{2}$$

where $\mathbf{h}_s$ represents the current state to compute the alignment scores, and it is replaced by the $\mathbf{S}_1$ term from the PNE expansion. Meanwhile, the $\mathbf{S}_i$ term, $i \in [2 : u]$, represents the remaining terms on the expansion to be aligned with the source. The dot-product-based alignment model is mathematically expressed as

$$a_t(h_s, S_i) = \mathbf{h}_s^T \mathbf{W}_a \mathbf{S}_i \tag{3}$$

where $\mathbf{W}_a$ is a trainable weight matrix. The computed alignment annotations are used to generate a context vector, expressed as

$$c_t = \sum a_t \mathbf{S}_i \tag{4}$$

And finally, the attention is computed based on a weighted concatenation as follows

$$\tilde{S}_i = tanh(\mathbf{W}_c[\mathbf{c}_t; \mathbf{S}_i]) \tag{5}$$

where $\mathbf{W}_c$ is the trainable concatenation weight matrix.

On the other hand, the multi-head self-attention [46] is characterized by using the scaled dot-product attention mechanism for similarity estimations between the Key, Query, and Value matrices. In our proposed model, they are replaced by the expansion terms $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$, respectively, as described below

$$Attention(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3) = softmax(\frac{\mathbf{S}_1 \mathbf{S}_2^T}{\sqrt{d_{\mathbf{S}_2}}})\mathbf{S}_3 \tag{6}$$

where d is the dimension of $\mathbf{S}_2$ used to properly scale the dot product between $\mathbf{S}_1$ and $\mathbf{S}_2$ In addition, the multi-head self-attention can be summarized as a generalized version of the aforementioned self-attention mechanism repeated in parallel h times, in which h represent the number of heads with different learned projections. This is mathematically formulated as follows:

$$MultiHead(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3) = Concat(head_1, ..., head_h)\mathbf{W}^O \tag{7}$$

where $\mathbf{W}^O$ is the learnable weight matrix for the output, and

$$head_i = Attention(\mathbf{S}_1 \mathbf{W}_i^{S_1}, \mathbf{S}_2 \mathbf{W}_i^{S_2}, \mathbf{S}_3 \mathbf{W}_i^{S_3}) \tag{8}$$

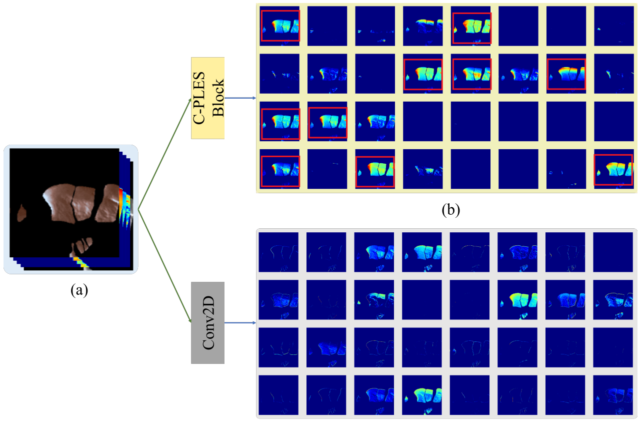
where $\mathbf{W}_i^{S_1}$, $\mathbf{W}_i^{S_2}$, $\mathbf{W}_i^{S_3}$ represent the learned projections for $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$, respectively.

Table 1. Detailed structure of the proposed C-PLES architecture

	Unit Level	Conv Layer	Filter	Dilation rate	Output size
Input					H x W x 7
Encoding	Level1	Conv 1	3 × 3, 64	2	H x W x 64
		Conv 2	3 × 3, 64	2	H x W x 64
	Level2	Conv 3	3 × 3, 128	2	H/2 x W/2 x 128
		Conv 4	3 × 3, 128	2	H/2 x W/2 x 128
	Level3	Conv 5	3 × 3, 256	2	H/4 x W/4 x 256
		Conv 6	3 × 3, 256	2	H/4 x W/4 x 256
	Level4	Conv 7	3 × 3, 512	2	H/8 x W/8 x 512
		Conv 8	3 × 3, 512	2	H/8 x W/8 x 512
Bridge	Level5	Conv 9	3 × 3, 1024	2	H/16 x W/16 x 1024
		Conv 10	3 × 3, 1024	2	H/16 x W/16 x 1024
C-PLES Block	Level1	Conv 11	1 × 1, 64	2	H x W x 64
	Level2	Conv 12	1 × 1, 128	2	H/2 x W/2 x 128
	Level3	Conv 13	1 × 1, 256	2	H/4 x W/4 x 256
	Level4	Conv 14	1 × 1, 512	2	H/8 x W/8 x 512
Decoding	Level4	Conv 15	3 × 3, 512	(-)	H/8 x W/8 x 512
		Conv 16	3 × 3, 512	(-)	H/8 x W/8 x 512
	Level3	Conv 17	3 × 3, 256	(-)	H/4 x W/4 x 256
		Conv 18	3 × 3, 256	(-)	H/4 x W/4 x 256
	Level2	Conv 19	3 × 3, 128	(-)	H/2 x W/2 x 128
		Conv 20	3 × 3, 128	(-)	H/2 x W/2 x 128
	Level1	Conv 21	3 × 3, 64	(-)	H x W x 64
		Conv 22	3 × 3, 64	(-)	H x W x 64
Output		Conv 23	1 × 1		H x W x 4

3.4. C-PLES: The Contextual Progressive Layer Expansion

The proposed DL architecture for multi-class landslide segmentation introduces a context enrichment Block, named C-PLES, that leverages the merits of the extracted features through the use of the PNE layers and the combination of global attention and local context, given by the dot attention mechanism and the dilated convolution, respectively. By the use of the C-PLES, we aimed to address the issue of context loss inherent in the encoder-decoder architectures. The C-PLES block is detailed and illustrated in Figure 2, and it is composed of an initial 2D dilated convolution. A dilated convolution increases the receptive field of the visual feature extractor and provides more contextual information from the encoder to map it with the decoder [47]. Then, a PNE layer takes the output of the dilated convolution to produce the expansion of each hidden neuron (i.e., $\mathbf{S}_1, \mathbf{S}_2$, and $\mathbf{S}_3$ as we used three terms in Maclaurin series expansion). This expansion enriches the visual feature representation, and the different numerical characteristics comprised by each term from the expansion generate a suitable scenario for the usage of attention mechanisms. Accordingly, the three nodes from the expansion are fed as the inputs for the attention layer (either dot-product attention or MHA). This attention layer helps the architecture to focus on a relevant part of the input. Considering computational efficiency, we select dot-product attention in our C-PLES block. In the end, the attention layer produced an attention mask that is multiplied by the original input of the C-PLES block. The output is then sent to the respective decoder. Mathematically, the C-PLES block can be represented as



(a)

(b)

(c)

Figure 3. Illustration of the feature maps generated by the use of the proposed C-PLES over one of the samples in the Mars multi-class landslide dataset. The sample (a), the feature maps (b) from the proposed C-PLES, and feature maps (c) from a regular 2D convolution.

follow:

$$C\text{-}PLES(\mathbf{X}) = attention(PNE(\hat{\mathbf{X}}))\mathbf{X} \tag{9}$$

where $\hat{\mathbf{X}}$ is a 2D dilated convolution of the input $\mathbf{X}$, this is used to generate a triple from the PNE (i.e. $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$) to be used as the key, query and value matrices in the attention function. Figure 3 illustrates visual evidence of the capabilities of the C-PLES to extract more relevant features compared to regular convolution operations.

3.5. Implementation details

The C-PLES architecture is depicted in Figure 2. Details of the composite architecture are provided in Table 1. The C-PLES architecture combines the merits of an encoder-decoder architecture with the addition of a C-PLES block which serves to join each level of the encoder side with the corresponding level at the decoder side. The architecture takes a $\mathbf{H} \times \mathbf{W} \times \mathbf{C}$ input, in which $\mathbf{H}$ and $\mathbf{W}$ could be either 64, 128, or 256 (The patch sizes), and C is the total number of bands (7 by default). This input data is processed to produce $\mathbf{H} \times \mathbf{W} \times 1$ segmentation mask. Each side of the architecture contains 4 levels in total. On each level, a set of two dilated 2D convolutions with a fixed kernel size (3×3) is performed with a down-sample operation at the end of the level on the encoder side. A set of two regular 2D convolutions with scaled-down feature maps and the same fixed kernel size with an up-sample operation at the end of each level on the decoder side. A bridge is connecting the encoder and decoder sides, transfers the enriched feature extracted in the encoder side, and starts up-sample the image until reaches the original input size. The C-PLES blocks are placed in the middle of each level as a link between the encoder and the decoder side, sharing the same feature maps as the corresponding level in the dilated 2D convolution operation which is followed by the PNE layer and the attention module to enhance the feature representation and the global context sent to the decoder side. It is worth mentioning a $ReLU$ activation function is added after each convolution operation. In addition, a dropout layer is also added to each set of convolutions to alleviate overfitting issues and stabilize the learning process during training.

In order to maximize the performance of the proposed C-PLES architecture, extensive experiments are performed to determine the best suitable loss function for the multi-class segmentation task. Since the dataset used in this study contains a significantly larger ratio of pixels belonging to the background, which makes the dataset heavily unbalanced, and thus we propose to fuse the focal loss and dice coefficient loss to effectively address this issue.

The focal loss [30] helps the model to effectively penalize easy and hard examples from imbalanced classes during the training process in a task such as image segmentation. Focal loss applies a modulating term ($(1 - p_t)^\gamma$) to the regular cross-entropy loss. This modulating factor reduces the loss contribution from easy examples and focuses the learning on hard-miss classified examples. Mathematically, the focal loss function is expressed as follows

$$\mathcal{L}_{FocalLoss} = -\sum_{i=1}^{i=n}(1 - p_i)^\gamma log_b(p_i) \tag{10}$$

where n is the number of samples and p is predicted probability. In order to address the class imbalance problem of positive and negative examples, a weighted parameter (α) is added as the inverse class frequency, extending the mathematical representation of the focal loss as:

$$\mathcal{L}_{FocalLoss} = -\sum_{i=1}^{i=n}\alpha_i(1 - p_i)^\gamma log_b(p_i) \tag{11}$$

In addition, dice coefficient loss (DCL) [26, 43] was also explored, considering that the dice coefficient is a popular metric to calculate similarities between images. The DCL is mathematically expressed as

$$\mathcal{L}_{DCL} = 1 - \frac{2\sum_i^N p_i g_i + 1}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + 1} \tag{12}$$

where 1 is added to the numerator and denominator to avoid undefined cases for scenarios in which $p = g = 0$.

We finally fuse both loss functions, denoted as $\mathcal{L}_{Total}$, in which a weight coefficient factor β is added to the focal loss to force the model to adjust the penalization for class imbalance cases. The $\mathcal{L}_{Total}$ can be formuated as

$$\mathcal{L}_{Total} = (1 - \beta)\mathcal{L}_{DCL} + \beta\mathcal{L}_{FocalLoss} \tag{13}$$

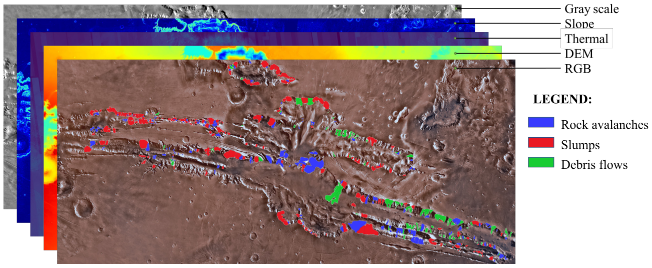


Figure 4. Images show the Valles Marineris regions on Mars. All the different modalities included in the dataset are stacked and aligned with the annotated landslides within the Valles Marineris region.

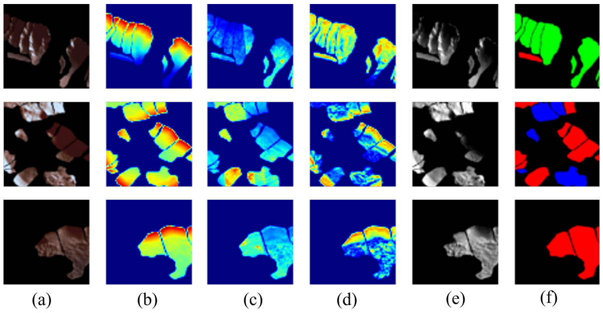


Figure 5. Samples from the Mars multi-class landslide dataset. This figure illustrates the different modalities: RGB (a), DEM (b), Thermal (c), Slope (d), GS (e), and the respective annotated segmentation outputs (f).

4. Dataset

The multi-class Mars landslide segmentation dataset is an imagery inventory for the Valles Marineris region, a vast system canyon that runs along the Martian equator surface. The dataset comprises a total of 682 manually annotated landslides, distributed as follows: 133 debris flow (DF), 275 rock avalanches (RA), and 274 slumps (S).

The image data used are the RGB, greyscale, DEM, slope, and thermal data. RGB data used is the Mars Viking Colorized Global Mosaic, which has a spatial resolution of approximately 232 meters. It is a derivative of NASA's Viking Mission to Mars [1]. The grey scale data is Context Camera (CTX) imagery of the Mars Reconnaissance Orbiter (MRO) mission, having a spatial resolution of 5 m and bandwidth of 600-800 nm. The elevation data of the study area, which is the DEM, is Mars MGS MOLA - MEX HRSC Blended DEM with a spatial resolution of 200 m [16]. The slope information is derived from the DEM using ESRI ArcGIS software and has a spatial resolution of 200 m. Thermal data is the thermal inertia of the Martian surface, generated from the Thermal Emission Imaging System (THEMIS) with a spatial resolution of 100 m. Figure 4 shows the composition of the multi-modal Martian imagery and the distribution of the multi-class landslides in the Valles Marineris region.

Landslides annotation.

The landslides were classified based on visual interpretation using satellite imagery in a Geographic Information System (GIS) environment, which was based on typical morphological features like flow direction, the shape of scarp [38, 45], area of depletion, and accumulation [19], irregular and jumbled terrain, hummocks, and lateral levees. These landslides were digitized as polygons.

Patch Sample landslides. The multi-modal composited image is available as a TIF file, with a dimension of 10,245 by 4,098 pixels with a total of 7 channels. In addition, the dataset is also available as a set of patches in several resolutions (without overlapping), distributed as follows: 122, 240, and 955 patches of 256 × 256, 128 × 128, and 64 × 64 respectively with the corresponding segmentation mask. Figure 5 illustrates a set of 256 × 256 samples from the proposed multi-class Martian landslide multi-modal dataset with the respective manually annotated segmentation mask.

5. Results and discussion

5.1. Experimental setup

We empirically demonstrate the effectiveness of the proposed C-PLES architecture using the multimodal Martian multi-class landslide segmentation dataset and compare its performance with several state-of-the-art deep network architectures, including U-Net [40], Attention U-Net [35], TransUNet [9], R2UNet [2], UNet3+ [25], UNet++ [51], and SwinUNet [7]. The performance of all models is evaluated in terms of the four standard assessment metrics: a) mean intersection over union (mIoU), b) precision, c) recall, and d) F1-score, all of which are typically used for landslide detection studies. All the experiments, including competing models, are run for a total of 150 epochs, with a batch size of 4, 8, and 16 samples for an input patch size of 256×256, 128×128, and 64×64, respectively. The $\mathcal{L}_{total}$ is used as a loss function in all the experiments as well, with β set as 0.7 after a grid search. The training process is optimized with the Adam algorithm using a scheduled learning rate, which is initially set to 0.001 and reduced by a factor of 0.01 in every epoch. As part of the pipeline in the training process, a data augmentation stage is added with a rotation range of 90 degrees, horizontal and vertical flip activated with zoom and reflect fill mode. The number of steps per batch size is computed as the ratio between the number of samples by the corresponding batch size. In addition, the validation loss was monitored to save checkpoints of the models with the best performance during the training process. In our exper-

[1] https://nssdc.gsfc.nasa.gov/planetary/viking.html

iments, the dataset was randomly split in a ratio of 8:2 for training and testing purposes, respectively. All models are trained using an NVIDIA A100 GPU.

5.2. Segmentation output performance comparison

For each image in the testing dataset, our model produced a segmentation mask that delineates the regions of interest of each type of landslide. The quantitative result to evaluate the performance of our C-PLES architecture and compare it with the aforementioned stare-of-the-art is shown in Table 2. As can be observed, the proposed C-PLES outperforms all the competing models in terms of mIoU and F1 scores for all three different patch sizes. The best mIoU (**0.5764**) is achieved by our proposed C-PLES method. In addition, the highest F1 score is also achieved by the C-PLES (**0.6948**). This is a promising indicator that our C-PLES model has the ability to effectively balance precision and recall across all classes, thereby achieving an overall superior landslide segmentation performance, identifying pixels that belong to a particular landslide class while minimizing the presence of misclassification.

Although our proposed C-PLES achieved competitive results in terms of recall among all the models evaluated, it is important to note that the precision metric is slightly lower compared to some state-of-the-art. This indicates a higher rate of false positives in the segmentation, as the model trade-off the performance of recall over precision to reduce missing important features in landslide segmentation. Future work may focus on improving the precision of C-PLES while maintaining a high recall rate, to achieve more accurate and complete segmentation of landslide areas. Figure 6 shows a visual comparison of the segmentation outputs of the proposed C-PLES and competing methods.

Figure 7 provides the segmentation output of C-PLES for the Martian Valles Marineris region. Four random samples are zoomed in and compared to the predicted segmentation with the corresponding ground truth. In the comparison, it is noticeable that the highest IoU (each class IoU) is for the Slump landslide class, while the lowest one is for the Rock Avalanches landslide class. Another visible aspect in Figure 7, is the lack of confidence in the model to identify the boundaries of the landslide, especially when different classes of landslides are very close among them. In some cases, the model tends to generalize the prediction of all the landslides within a patch input towards the class of the biggest landslide. This limitation may be linked to the quantitative results in terms of precision.

6. Ablation study

Contribution of image modality. Several image modalities were merged in the Martian multi-class landslide segmentation dataset. As mentioned before, the dataset contains a total of seven bands: RGB (3 bands), Gray-scale imagery, DEM, THEMIS, and Slope. In order to analyze the contribution of each modality to the performance of the proposed C-PLES, extensive experiments are conducted. The experimental results are detailed in Table 3, which is the result of testing on unseen 256 $\times$ 256 patches data samples from the dataset. The training hyperparameters are set as similar to the ones reported in Section 5, however, only mIoU and IoU (IoU_{DF} : Debris Flow, IoU_S : Slumps, and IoU_{RA} : Rock Avalanches) are considered as evaluation

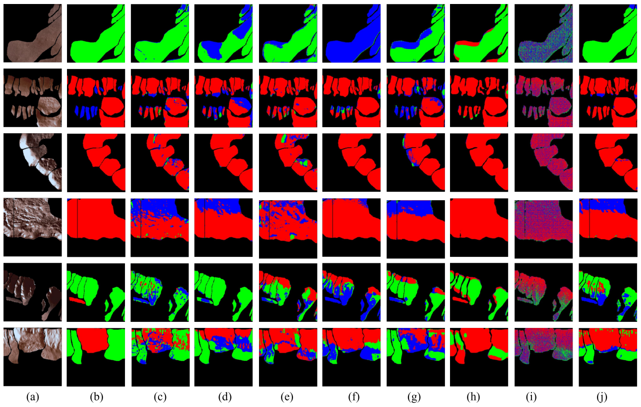


Figure 6. Visual comparison of multi-class Martian landslide segmentation outputs. (a) Input RGB image, (b) the segmentation mask (ground truth), (c) - (i) the segmentation outputs of experimented DL architectures: U-Net [40], Attention U-Net [35], TransUNet [9], R2UNet [2], UNet 3+ [50], UNet++ [51], and Swin-Unet [7], (j) the proposed C-PLES, respectively.

Table 2. Performance comparison of the proposed C-PLES against the state-of-the-art architecture for segmentation over the Martian multi-class landslide multimodal dataset. key: [**Best**, **Second Best**]

	Method	mIoU	F1 score	Recall	Precision
256 x 256	U-Net [40]	0.5111	0.6322	0.6447	0.6332
	Att. U-Net [35]	0.5387	0.6511	0.6514	0.6817
	TransUNet [9]	**0.5605**	**0.6835**	0.6517	**0.7502**
	R2UNet [2]	0.5449	0.5521	0.5514	0.6954
	UNet 3+ [25]	0.5491	0.6594	**0.6992**	0.6657
	UNet ++ [51]	0.4977	0.5963	0.6099	0.6005
	Swin-Unet [7]	0.4219	0.5122	0.5258	0.5288
	C-PLES (ours)	**0.5764**	**0.6948**	**0.6881**	**0.7165**
128 x 128	U-Net [40]	**0.5128**	**0.6341**	0.6312	0.6342
	Att. U-Net [35]	0.5089	0.6233	0.6190	0.6435
	TransUNet [9]	0.4986	0.6104	0.6154	0.6325
	R2UNet [2]	0.5098	0.6326	0.6406	0.6412
	UNet 3+ [25]	0.5121	0.6317	**0.6484**	**0.6508**
	UNet ++ [51]	0.4511	0.5460	0.5524	0.6192
	Swin-Unet [7]	0.4631	0.5703	0.5747	0.5761
	C-PLES (ours)	**0.5166**	**0.6471**	**0.6485**	**0.6454**
64 x 64	U-Net [40]	0.5384	0.6552	0.6503	0.6693
	Att. U-Net [35]	0.5550	**0.6702**	0.6672	**0.7039**
	TransUNet [9]	0.5354	0.6547	0.6448	0.6842
	R2UNet [2]	0.5243	0.6387	0.6476	0.6408
	UNet 3+ [25]	**0.5596**	0.6691	**0.6911**	**0.7038**
	UNet ++ [51]	0.4813	0.5713	0.6309	0.6095
	Swin-Unet [7]	0.4524	0.5455	0.5590	0.6245
	C-PLES (ours)	**0.5597**	**0.6894**	**0.6754**	0.7025

Table 3. Importance of different image modalities to the prediction accuracy.

Image modality	**mIoU**	**IoU$_{DF}$**	**IoU$_{S}$**	**IoU$_{RA}$**
RGB	0.3938	0.0001	0.5779	0.0001
RGB + DEM	0.5429	0.4127	0.5973	0.1627
RGB + THEMIS	0.4043	0.0091	0.5206	0.0917
RGB + Slope	0.4322	0.0099	0.5616	0.1587
RGB + GS	0.4089	0.0023	0.5619	0.0743
RGB + DEM + THEMIS	0.5215	0.3454	0.5854	0.1557
RGB + DEM + Slope	0.5355	0.4419	0.4827	0.2202
RGB + DEM + GS	0.4233	0.0064	0.6143	0.0730
RGB + DEM + THEMIS + Slope	0.5698	0.4840	0.5485	**0.2485**
RGB + DEM + THEMIS + GS	0.3937	0.0025	0.5773	0.0002
RGB + DEM +THEMIS + Slope + GS	**0.5764**	**0.4953**	**0.6153**	0.2052

Table 4. Model performance by varying components of the proposed C-PLES architecture. key: [**Best**, **Second Best**]

Model	**mIoU**	**F1 score**	**Recall**	**Precision**
C-PLES w/ dot self-attention	**0.5597**	**0.6894**	**0.6754**	**0.7025**
C-PLES w/ MHA	**0.5951**	**0.7150**	**0.7153**	**0.7175**
C-PLES w/o Attention	0.5479	0.6670	0.6590	0.6859
C-PLES w/o PNE	0.5513	0.6703	0.6657	0.6822

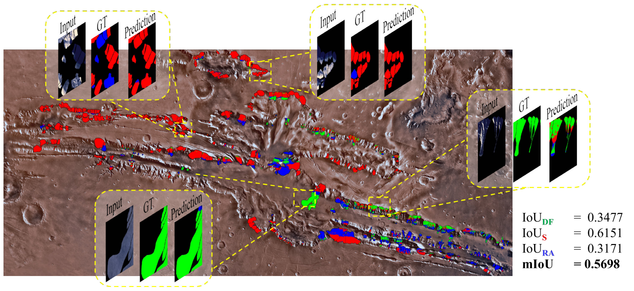


Figure 7. Segmentation map of the landslides in the Martian - Valles Marineris region. Four random regions are zoomed in to show in detail a comparison between the ground truth and the predicted segmentation. The IoU scores for each class are shown on the bottom right side of the figure.

metrics. On every round of training, a different set of modalities is used with the RGB modality as a baseline. According to the quantitative results, the best performance is reached by the full set of modalities. However, the best performance to segment the Rock Avalanches (IoU_{RA} of 0.2485) is given under the absence of the gray-scale image (RGB+DEM+THEMIS+Slope). For segmenting Debris Flow landslides, the absence of slopes reduces the ability of the network to discriminate this landslide from the others considerably. It is worth mentioning that, on the other hand, the presence of the gray-scale imagery improves the performance to identify Slumps landslides. Overall, our results suggest that the combination of the full set of modalities is beneficial to achieve better performance in general.

Importance of model components. To understand the inner working of the proposed C-PLES architecture, we evaluate the performance of the architecture with a different set of components that are present or absent. For instance, a major emphasis is on the overall impact of the inclusion or absence of the C-PLES block within the architecture. Table 4 summarizes the results obtained in this ablation study. Due to computational constraints, this experiment is performed with a 64 × 64 patch size dataset, allowing us to efficiently compare the performance of the proposed architecture but varying the attention mechanism. The best performance is reached with C-PLES architecture using the multi-head self-attention (MHA) mechanism. In this approach, the C-PLES block internally takes the output from the PNE layer as the query, key, and value matrices for the multi-head self-attention. The computation of multiple heads allows the model to capture a larger range of trained weights (a set of weights by each head) and thus capture different types of relationships between the different parts of the input sequence in comparison to a regular attention mechanism. In addition, as we hypothesized, the C-PLES architecture using the dot product attention mechanism, which is the version used in the majority of the experiments reported in this paper, provides better performance when it is compared with the version of the C-PLES without attention mechanism and without the PNE. Additionally, the C-PLES block under the absence of the PNE performed slightly better than its counterpart without an attention mechanism. This indicates the importance of the attention mechanism to selectively focus on relevant parts of the image and capture spatial relationships between objects and their context. Finally, the combination of attention with enriched feature representation by PNE provides a harmonically improved performance for Martian landslide segmentation.

7. Conclusions

In this research work, we proposed the C-PLES architecture for Martian multi-class landslide segmentation. The presented research opens the door not only to the analysis of the Martian landslides but also to the same task on Earth. The C-PLES block is introduced as a plug-and-play module that leverages the merits of two main components: the progressive expansion neurons and the attention mechanism. In addition, a Martian multimodal multi-class landslide dataset is introduced and used to evaluate the performance of the proposed method. Our experiments showed that the proposed C-PLES achieved state-of-the-art results for the Martian landslide segmentation task, which proves the importance of the C-PLES block that consists of progressive neuron expansion and attention mechanism to capture pertinent visual features from the different image modalities. The aforementioned dataset will be available to the research community.

References

[1] Witthawin Achariyaviriya, Toshiaki Kondo, Jessada Karnjana, and Takayuki Nishio. Landslide semantic segmentation using satellite imagery. In *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–4. IEEE, 2022. 2

[2] Md Zahangir Alom, Chris Yakopcic, Mahmudul Hasan, Tarek M Taha, and Vijayan K Asari. Recurrent residual u-net for medical image segmentation. *Journal of Medical Imaging*, 6(1):014006–014006, 2019. 6, 7

[3] Guendalina ANTONINI, Francesca ARDIZZONE, Mauro CARDINALI, Mirco GALLI, Fausto GUZZETTI, and Paola REICHENBACH. Surface deposits and landslide inventory map of the area affected by the 1997 umbria-marche earthquakes. *Bollettino della Società geologica italiana*, 121(1):843–853, 2002. 1

[4] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 3

[5] M Brunetti, M Cardinali, F Fiorucci, F Guzzetti, M Santangelo, P Mancinelli, G Komatsu, K Goto, and H Saito. Mapping, classification, and statistics of mass movements in valles marineris, mars. In *AGU Fall Meeting Abstracts*, volume 2011, pages EP43A–0662, 2011. 1

[6] Maria Teresa Brunetti, Fausto Guzzetti, Mauro Cardinali, Federica Fiorucci, Michele Santangelo, Paolo Mancinelli, Goro Komatsu, and Lorenzo Borselli. Analysis of a new geomorphological inventory of landslides in valles marineris, mars. *Earth and Planetary Science Letters*, 405:156–168, 2014. 1

[7] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, pages 205–218. Springer, 2023. 6, 7

[8] Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)*, pages 1–4. IEEE, 2017. 3

[9] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 6, 7

[10] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 3

[11] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1290–1299, 2022. 2

[12] GB Crosta, P Frattini, E Valbuzzi, and FV De Blasio. Introducing a new inventory of large martian landslides. *Earth and Space Science*, 5(4):89–119, 2018. 1

[13] Fabio Vittorio De Blasio. Landslides in valles marineris (mars): A possible role of basal lubrication by sub-surface ice. *Planetary and Space Science*, 59(13):1384–1392, 2011. 1

[14] Mingyu Ding, Bin Xiao, Noel Codella, Ping Luo, Jingdong Wang, and Lu Yuan. Davit: Dual attention vision transformers. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV*, pages 74–92. Springer, 2022. 2

[15] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[16] RL Fergason, TM Hare, and J Laura. Hrsc and mola blended digital elevation model at 200m v2. *Astrogeology PDS Annex, US Geological Survey*, 2018. 6

[17] Stefano Luigi Gariano and Fausto Guzzetti. Landslides in a changing climate. *Earth-Science Reviews*, 162:227–252, 2016. 1

[18] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2918–2928, 2021. 2

[19] Nancy F Glenn, David R Streutker, D John Chadwick, Glenn D Thackray, and Stephen J Dorsch. Analysis of lidar-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology*, 73(1-2):131–148, 2006. 6

[20] Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R Martin, Ming-Ming Cheng, and Shi-Min Hu. Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 8(3):331–368, 2022. 3

[21] Fausto Guzzetti, Alessandro Cesare Mondini, Mauro Cardinali, Federica Fiorucci, Michele Santangelo, and Kang-Tsung Chang. Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112(1-2):42–66, 2012. 1

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[23] Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*, 2019. 4

[24] Dichao Hu. An introductory survey on attention mechanisms in nlp problems. In *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2*, pages 432–448. Springer, 2020. 3

[25] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei

Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020. 6, 7

[26] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, pages 1–7. IEEE, 2020. 5

[27] Sanjay Katta, Sidike Paheding, Thomas Oommen, A Rajaneesh, and KS Sajinkumar. Categorization of martian landslides from satellite imagery using vision transformer. In *AGU Fall Meeting Abstracts*, volume 2021, pages P11B–04, 2021. 2

[28] Ren N Keyport, Thomas Oommen, Tapas R Martha, KS Sajinkumar, and John S Gierke. A comparative analysis of pixel-and object-based detection of landslides from very high-resolution images. *International journal of applied earth observation and geoinformation*, 64:1–11, 2018. 2

[29] Tao Lei, Risheng Wang, Yuxiao Zhang, Yong Wan, Chang Liu, and Asoke K Nandi. Defed-net: Deformable encoder-decoder network for liver and liver tumor segmentation. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 6(1):68–78, 2021. 3

[30] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 5

[31] BK Lucchitta. A late climatic change on mars. In *Lunar and Planetary Science Conference*, volume 15, pages 493–494, 1984. 1

[32] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015. 3

[33] Muhammad Maaz, Hanoona Rasheed, Salman Khan, Fahad Shahbaz Khan, Rao Muhammad Anwer, and Ming-Hsuan Yang. Class-agnostic object detection with multi-modal transformer. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part X*, pages 512–531. Springer, 2022. 2

[34] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62, 2021. 3

[35] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 6, 7

[36] Sidike Paheding, Abel A Reyes, Mohammad Alam, and Vijayan K Asari. Medical image segmentation using u-net and progressive neuron expansion. In *Pattern Recognition and Tracking XXXIII*, volume 12101, page 1210102. SPIE, 2022. 3

[37] Nikhil Prakash, Andrea Manconi, and Simon Loew. Mapping landslides on eo data: Performance of deep learning models vs. traditional machine learning models. *Remote Sensing*, 12(3):346, 2020. 2

[38] A Rajaneesh, CL Vishnu, T Oommen, VJ Rajesh, and KS Sajinkumar. Machine learning as a tool to classify extraterrestrial landslides: A dossier from valles marineris, mars. *Icarus*, page 114886, 2022. 1, 6

[39] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12179–12188, 2021. 2

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 2, 6, 7

[41] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*, 2015. 4

[42] Paheding Sidike, Vasit Sagan, Maitiniyazi Maimaitijiang, Matthew Maimaitiyiming, Nadia Shakoor, Joel Burken, Todd Mockler, and Felix B Fritschi. dpen: Deep progressively expanded network for mapping heterogeneous agricultural landscape using worldview-3 satellite imagery. *Remote sensing of environment*, 221:756–772, 2019. 3

[43] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017. 5

[44] Sepideh Tavakkoli Piralilou, Hejar Shahabi, Ben Jarihani, Omid Ghorbanzadeh, Thomas Blaschke, Khalil Gholamnia, Sansar Raj Meena, and Jagannath Aryal. Landslide detection using multi-scale image segmentation and different machine learning models in the higher himalayas. *Remote Sensing*, 11(21):2575, 2019. 2

[45] Miet Van Den Eeckhaut, Jean Poesen, Gert Verstraeten, Veerle Vanacker, Jan Moeyersons, Jan Nyssen, and LPH Van Beek. The effectiveness of hillshade maps and expert knowledge in mapping old deep-seated landslides. *Geomorphology*, 67(3-4):351–363, 2005. 6

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4

[47] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Dilated convolutional neural networks for cardiovascular mr segmentation in congenital heart disease. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers*, pages 95–102. Springer, 2017. 4

[48] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun,

Yuan Cao, Qin Gao, Klaus Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016. 4

[49] Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12104–12113, 2022. 2

[50] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 3, 7

[51] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018. 6, 7