MarsLS-Net: Martian Landslides Segmentation Network and Benchmark Dataset

Sidike Paheding[1], Abel A. Reyes[2], A. Rajaneesh[3], K.S. Sajinkumar[3], Thomas Oommen[4]
[1]Fairfield University [2]Michigan Technological University [3]University of Kerala
[4]University of Mississippi
spaheding@fairfield.edu, areyesan@mtu.edu, {rajaneesh, sajinks}@keralauniversity.ac.in,
toommen@olemiss.edu

Abstract

Martian landslide segmentation is a challenging task compared to the same task on Earth. One of the reasons is that vegetation is typically lost or significantly less compared to its surroundings in the regions of landslide on Earth. In contrast, Mars is a desert planet, and there is no vegetation to aid landslide detection and segmentation. Recent work has demonstrated the strength of vision transformer (ViT) based deep learning models for various computer vision tasks. Inspired by the multi-head attention mechanism in ViT, which can model the global long-range spatial correlation between local regions in the input image, we hypothesize self-attention mechanism can effectively capture pertinent contextual information for the Martian landslide segmentation task. Furthermore, considering parameter efficiency or model size is another important factor for deep learning algorithms, we construct a new feature representation block, namely Progressively Expanded Neuron Attention (PEN-Attention), to extract more relevant features with significantly fewer trainable parameters. Overall, we refer to our deep learning architecture as the Martian landslide segmentation network (MarsLS-Net). In addition to the new architecture, we introduce a new multi-modal Martian landslide segmentation dataset for the first time, which will be made publicly available at https://github.com/MAIN-Lab/Multimodal-Martian-Landslides-Dataset

1. Introduction

Landslides have an important role in shaping the landscapes [20] and thus, can fundamentally change the planet's morphological features. While terrestrial landslides are well-studied in detail, studies on the role of extraterrestrial landslides in modifying the terrain are modest in number. Mars, sobriquet the red planet, is being explored by several researchers to study early climate as the planet was warm and wet during the geological past [11]. Except for the evidence of life, early Mars mimicked the Earth, with imprints of natural hazards like landslides, floods, volcanism, dust devils, and meteorite impacts, which are omnipresent throughout the surface of Mars. Thus, Mars is an ideal candidate for studying climate change as all the ancient natural hazards are preserved in the most exquisite way [16, 19, 26]. Moreover, on a planet like Mars, which has made huge transitions in 'geological' scenarios from a water-bearing planet to a barren one, the study of landslides can help us discern the crucial climate change it witnessed.

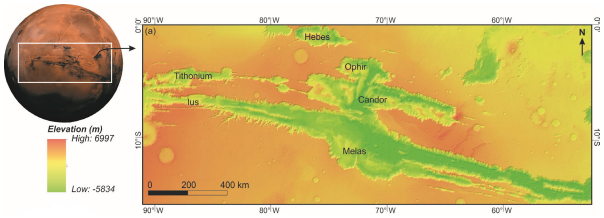

Figure 1. Location of Valles Marineris (VM) on Mars used for landslide segmentation in this study. The elevation map of VM draped over hillshade.

Identifying a potential area on Mars to study landslides left with a conspicuous choice of Valles Marineris (VM) shown in Fig. 1. This area is considered a museum of landslides [34], characterized by different types of landslides, such as debris flow, earth slide, rock avalanche as well as simple and complex landslides. Typically, landslide studies are conducted by means of visual interpretation of medium to high-resolution optical images, where geomorphologists are engaged to mapping terrestrial landslides [5, 34]. However, these methods are quite labor-intensive, subjective, and time-consuming. In addition, the study of Martian landslides has been hindered by the availability of adequate remote sensing data [6]. Therefore, there is a need to tackle

two main challenges for landslide studies: 1) build a standard and publicly available dataset for the research community, and 2) develop sophisticated computational methods to automate the landslide identification and mapping process, given the complexity of the morphological characteristics of the Martian terrain.

In order to accurately identify Martian landslides, feature extraction plays an important role. Morphological characteristics of the landslide have been found to be a key element for differentiating large Martian landslides [10,25,34]. Brunetti et al. [5] analyzed features such as shape, size, tone, mottling, and texture, to recognize and map landslides in the VM. Their study also suggested that an empirical relationship (i.e., a power law) exists between the volume and area of the Martian slides. Crosta et al. [10] mapped about 3,100 Martian landslides following standard geomorphological properties of VM landslides, such as the presence of scars, morphostructures, lateral levees, etc. A very recent study [34] found 15 critical features, such as length, width, relative relief, slope statistics, the slope of the scarp, and geology, that are useful for landslide categorization in the VM, Mars. Then they used traditional machine learning methods, such as logistic regression, which takes those extracted features to classify landslides into three categories, including debris flows, rock avalanches, and slumps. To sum up, all aforementioned studies have put much attention on geometric parameters to identify Martian landslides. In the work presented by [23], they adopted a Vision Transformer (ViT) [12] model to classify three types of landslides on Mars by dividing satellite images into patches, and demonstrated the strengths of deep learning models for Martian landslides classification. In this study, we introduce a new end-to-end Martian landslide segmentation model that fuses heterogeneous multi-modality imagery data in a contextually aware manner. The main contributions of this research are summarized as follows:

1. We introduce a Martian landslide segmentation dataset, which will be made publicly available to the research community through a repository link.

2. We propose an end-to-end deep learning pipeline for Martian landslide segmentation from multi-modal imagery data directly. Our analysis reveals the robustness of deep learning models to address the challenges in segmenting Martian landslides, based on the characteristics of terrestrial landslides.

3. A new deep learning-based segmentation model, namely the Martian landslide segmentation network (MarsLS-Net), is further proposed. The experimental results and ablation study show the effectiveness and superior performance of MarsLS-Net compared to state-of-the-art methods.

2. Related Work

In this section, we review the most relevant machine learning methods for landslide detection and segmentation.

2.1. Landslide Segmentation

Several approaches for landslide segmentation tasks reported the use of traditional machine learning and deep learning methods in combination with object-based image analysis (OBIA) techniques [1, 17, 24, 42]. Tavakkoli et al. [42] proposed the use of (OBIA) with a set of Machine Learning methods for landslide detection, in which the best performance was reported by using a stacked approach (ensemble methods). Furthermore, they commented the use of high-resolution satellite imagery does not guarantee a high performance of the model, suggesting that it is also required to identify parameters based on target object detection to achieve competitive performance. Keyport et al. [24] provided a comparative analysis for pixel-based landslide detection using very high-resolution (VHR) remotely sensed aerial images. They reported the analysis by pixel-based and object-oriented analysis (OOA) methods, concluding a better performance for landslide mapping obtained by the OOA method, with fewer false positives obtained in the predicted outcomes compared with the pixel-based approach. Achariyaviriya et al. [1], combined three different ResNet [21] models to reach generalization for landslide segmentation, in which each ResNet model was trained with a different set of layers, such as RGB, normalized difference vegetation index (NDVI) and slope factor (SP) as inputs, combining the generated features for each model to train a final decision tree classifier. They claimed an improvement in the performance by the use of additional layers rather than RGB images, in addition to a low computational cost by the use of a decision tree classifier.

2.2. U-Net based Segmentation Models

Recent works on landslide segmentation proposed to use of end-to-end convolutional networks, such as U-Net [37] to automatically segment landslides in several regions on earth [4, 30, 39]. Pedrosa et al. [39] explored deep learning-based architecture models to show a generalization capacity for automatic landslide segmentation in a mountainous region, located in the city of Nova Friburgo, Rio de Janeiro, Brazil. They implemented the vanilla version of U-Net architecture [37], reporting mixed performance of generalization for landslide segmentation on each of the test areas. In addition, they provided a discussion on the selection of image patching sizes, network layers, and post-processing procedures (e.g., morphological operations). Bragagnolo et al. [4] also described the performance of U-Net architecture for mapping landslide scars through the use of satellite imagery. They performed several experiments using scenes captured by the Landsat-8 satellite from a region with a

high prevalence of landslides in Nepal. The results of using U-Net architecture over this composed dataset indicate a performance improvement, compared to previous efforts in similar studies. Indeed, this study showed the capacity of convolutional neural networks (CNN) based architectures to detect landslides, however, there are still divergences between the edges when compared with the manually mapped landslide scars.

With the promising performance obtained by the implementation of end-to-end CNN models for automatic landslide detection, several modified versions of U-Net [2, 27, 29, 47] were also explored. Qi et al. [31] introduced ResU-Net for automatic mapping landslide, where they achieved a high accuracy using only spectral bands of satellite images. The architecture adopted a U-Net alike architecture with the addition of residual learning [21] in the convolutional blocks of the encoding path of the U-Net architecture. The results showed that their proposed architecture yields better performance than the baseline in terms of precision, recall, and F1-score. Furthermore, Ghorbanzadeh et al. [17] presented a comparative analysis of ResU-Net [47], rule-base object-based image analysis (OBIA), and the combination of both (ResU-Net-OBIA) for landslide detection. Their results showed that the combination of ResU-Net and OBIA offered promising prior knowledge added to the mapping landslide process, obtaining considerably better results in terms of precision, recall, and F1-score compared to each stand-alone method. It is worth mentioning that all aforementioned landslide detection algorithms are typically applied in a rural area, which helps to easily discriminate landslides from dense vegetation. However, on extraterrestrial surfaces such as Mars, the same contrast is not likely present, thus making the segmentation task considerably difficult.

2.3. Vision Transformer-based Models

Vision Transformers (ViT) [12] is the adaptation of Transformers [44] for the context of computer vision. Due to the outstanding performance achieved by the use of ViT for image classification, there have been several variations of ViT for segmentation tasks [7, 8, 40], and have been applied to conduct different tasks, including landslide segmentation. For instance, Tang et al. [41] introduced the use of SegFormer [46], a ViT-based model, for landslide segmentation tasks. They reported a detailed comparison of performance with several state-of-the-art deep learning (DL) architectures, including HRNet [45], DeepLabV3+ [9], Attention-UNet [27], U^{2}Net [32] and FastRCNN [35]. The proposed method was complemented with a post-processing procedure based on morphological operations, reporting outstanding performance in contrast with the other evaluated models. A recent work [36] introduced a contextual progressive layer expansion with self-attention, as the key component of a DL model for multi-class Martian landslide segmentation, which demonstrated the strength of the self-attention mechanism in landslide segmentation. In our proposed architecture, one of the major components of ViT, multi-head self-attention, is used to capture more relevant characteristics of landslides by leveraging the multi-modal imagery dataset for Martian landslide segmentation.

3. Methodology

3.1. Architecture

Overall architecture. The proposed Martian landslide segmentation network (MarsLS-Net) is illustrated in Fig. 2. Unlike the state-of-the-art deep learning-based segmentation models which are typically built upon encoder-decoder architecture, the proposed MarsLS-Net is composed by a series of repeated operational blocks, namely Progressively Expanded Neuron Attention (PEN-Attention), which does not follow traditional encoder-decoder structure. PEN-Attention blocks exploit the potential of deep feature representation only using Convolutional PE layers and attention mechanism, yielding a less complex but effective model that is easy to train and highly parameter efficient. The model has shown a relatively faster convergence during training with competitive performance when compared with well-established segmentation models (i.e. U-Net, Attention U-Net, ResU-Net). Details of the proposed architecture are provided in the following section.

Convolutional Progressive Expansion (ConvPE). As shown in Fig. 2 (c), Convolutional Progressive Expansion (ConvPE) layer is composed of four consecutive operations: a 2D convolution, batch normalization, ReLU activation, and followed by a progressive expansion (PE). The concept of PE is adapted from the work presented in [28, 29, 38], where each neuron from its input or feature maps is firstly progressively expanded using Maclaurin series expansion of a nonlinear function, and then added, being expressed as

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

where x stands for the unit to be expanded, c and p are the coefficients and powers in the Maclaurin series expansion. u is a pre-defined hyperparameter that controls the number of terms used in a series expansion.

Multi-head self-attention (MHSA). The MHSA layer is incorporated into our proposed PEN-Attention block to attend to important features from the previous PE layers by aggregating the knowledge explored by multiple heads. As described by Dosovitskiy et al. [12], self-attention is an attention mechanism characterized by the use of scaled dot-product for similarity estimation, which is mathematically

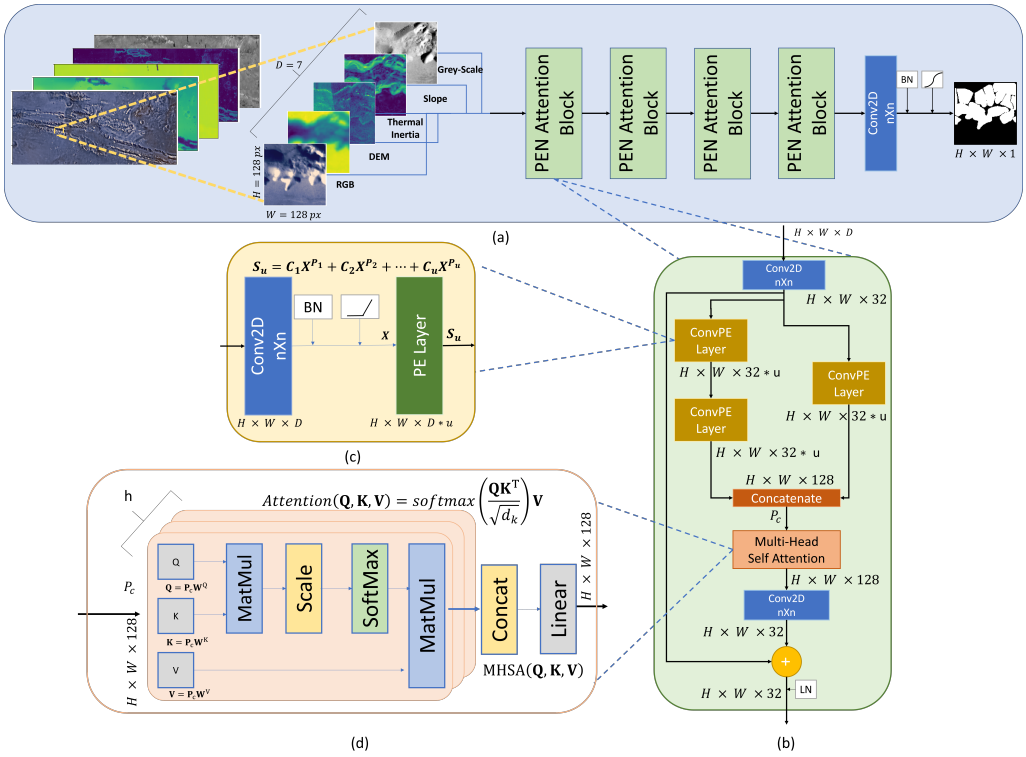

Figure 2. The proposed MarsLS-Net architecture. (a) the overall framework of MarsLS-Net model that is composed of a series of stacked PEN attention blocks, (b) represents the illustration of the multi-head self-attention mechanism, (c) provides details of the ConvPE Layer, and (d) illustrates the structure of the PEN attention block.

described as:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}})\mathbf{V} \quad (2)$$

where **Q**, **K**, and **V** are matrices representing queries, keys, and values, respectively. The dot-product of **Q** and the transpose of **K** acts as an attention filter. The $\sqrt{d_k}$ scales the dot-product, while the *softmax* function returns the weights which are then multiplied by **V**. In our proposed network, the output of **Q**, **K**, and **V** can be calculated as

$$\mathbf{Q} = \mathbf{P}_c\mathbf{W}^Q, \mathbf{K} = \mathbf{P}_c\mathbf{W}^K, \mathbf{V} = \mathbf{P}_c\mathbf{W}^V \quad (3)$$

where $\mathbf{P}_c$ represents the output from the concatenation of the two-branch ConvPE layer as shown in Fig. 2 (a).

The multi-head self-attention can be defined as the aforementioned attention mechanism repeated in parallel h times, where h represent the number of heads, with different learned projections $\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V$ for **Q**, **K**, and **V**, respectively. From that, the outputs are concatenated and linearly projected with $\mathbf{W}^O$ as

$$Multihead(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Concat(h_1, .., h_i, .., h_h)\mathbf{W}^O \quad (4)$$

where h_i can be defined as

$$h_i = Attention(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (5)$$

The MHSA layer is illustrated in Fig. 2 (d).

PEN-Attention Block. The PEN-Attention block is used to focus on regions that are relevant in terms of the global context of the concatenated PE layers, and thus generate a more accurate prediction in the output of the model. This block is composed by four components: 2D convolution, a two-branch based ConvPE layers, an MHSA layer, and a layer normalization [3], as shown in Fig. 2 (b). The two-branch ConvPE layers extract fine-grained features which are then concatenated, and subsequently fed to a MHSA layer. In addition, we added a skip-connection that bypasses ConvPE and MHSA layers to ease gradient flow in the network.

3.2. Implementation Details.

The proposed architecture is designed to be adjustable according to the need of tasks. In other words, different configurations can be easily obtained by setting hyperparameters in MarsLS-Net. For instance, we can select different numbers of terms (u in Eq. 1) in series expansion in the PE layer to produce various numbers of feature maps. This value was fixed to be 2 in our experiments to reduce

model complexity. The number of PEN-Attention blocks is denoted as the levels in the architecture, and this was set to be 4 in our case as shown in Fig. 2 (a). However, it can be increased to achieve deeper networks, and possibly, better accuracy. The number of filters in 2D convolutions on each ConvPE layer and on each PEN-Attention block is fixed to 48 with a kernel size of 5. In the ConvPE layer, we use dilated convolutions, instead of standard convolutions, with a dilatation rate of 2 and L_2 regularizer. The number of heads in the MHSA is set to be 2 considering both segmentation accuracy and memory efficiency. Experimental results on different kernel sizes and feature maps are provided in the Sec. 5.2. The use of stacked PEN Attention blocks are to encode the multi-model representation within a latent space and capture richer features.

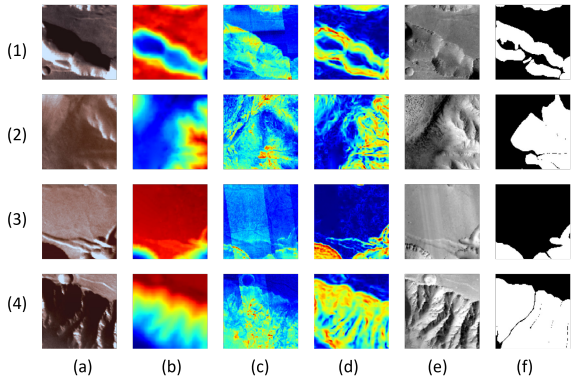

Figure 3. Sample of landslides regions (128×128) in the MMLS dataset with their respective segmentation mask and modalities: (a) RGB, (b) DEM, (c) thermal inertia, (d) slope, (e) gray-scale, and (f) ground truth.

Table 1. A quantitative summary of data distribution in the MMLS dataset.

Training Dataset	Validation Dataset	Testing Dataset	Total Patches
275	31	256	562

4. Dataset

To address the need for labeled data for the Martian landslide segmentation task, we created a landslide inventory for VM with a total of 2169 incidents. This inventory includes an existing landslide inventory of 682 incidents [34], and another newly mapped 1487 landslides. Landslides were vectorized from the Context Camera (CTX) imageries of the Mars Reconnaissance Orbiter (MRO) by NASA, which has a resolution of 5 m per pixel. The CTX data was obtained from the planetary data system (PDS) by NASA[1]. The digital elevation model (DEM) of the Valles Marineris is obtained from U.S. Geological Survey [15]. This elevation data is a combined product of NASA's Mars Orbiter Laser Altimeter (MOLA) of the Mars Global Surveyor spacecraft (MGS) and the European Space Agency's High-Resolution Stereo Camera (HRSC) of the Mars Express (MEX) spacecraft. It has a ground resolution of 200 meters per pixel. Imagery and elevation data for the Valles Marineris are brought into the ESRI ArcGIS software package and landslides were digitized based on morphological parameters described by Quantin et al. [33] and Crosta et al. [10]. RGB image data was obtained from USGS Astrogeology Science Center[2]. The image data has a resolution of 232 m per pixel. The RGB data was developed by NASA AMES from NASA's Viking Mission. Thermal inertia (TI) data of Valles Marineris region was obtained from USGS Astrogeology Science Center[3] [14], which have a spatial resolution of 100 m per pixel. These data were developed from Thermal Emission Imaging System (THEMIS) images of the 2001 Mars Odyssey orbiter mission. In addition, slope values of the study area are derived from the DEM using the Slope tool in the ArcMap. The RGB data, CTX data, DEM, slope, and TI were melded together as a single TIFF image using the Composite bands' tool in ArcMap. Consequently, we named our dataset as Multi-modal Martian LandSlide (MMLS) dataset.

Landslides Mapping. Landslides were vectorized manually in a GIS environment and were identified by scarp and surface morphology of the scarp and surface. The morphological properties include: 1) Scarps which are semi-circular steep, slope or escarpment, often perpendicular to flow direction [34, 43]; 2) The toes, as well as the depleted area, are characterized by heavily fractured surface [18]; 3) Martian landslides have irregular and jumbled morphology, and hence, exhibits hummocky structure (i.e. rounded hills or mounds) and lateral levees [10, 13, 22]. The landslides were mapped as polygons that included the zone of depletion and the entire run-out. Fig. 4 shows the visual distribution of Martian landslides in the Valles Marineris region, while Table 1 summarizes the quantitative distribution of the training, validation, and testing sets used in this study. All the modalities are downsampled to match the spatial resolution of the RGB modality, thus the resulting images can be fed to DL models for training. In Table 1, the size of each patch is fixed to 128×128 pixels, and we maintain the spatial resolution of each modality to 232 m per pixel.

5. Results and Discussion

Experimental setup. We empirically demonstrate the effectiveness of the proposed MarsLS-Net using the MMLS

[1]https://ode.rsl.wustl.edu/

[2]https://planetarymaps.usgs.gov/mosaic/Mars_Viking_MDIM21_ClrMosaic_global_232m.tif

[3]https://astrogeology.usgs.gov/

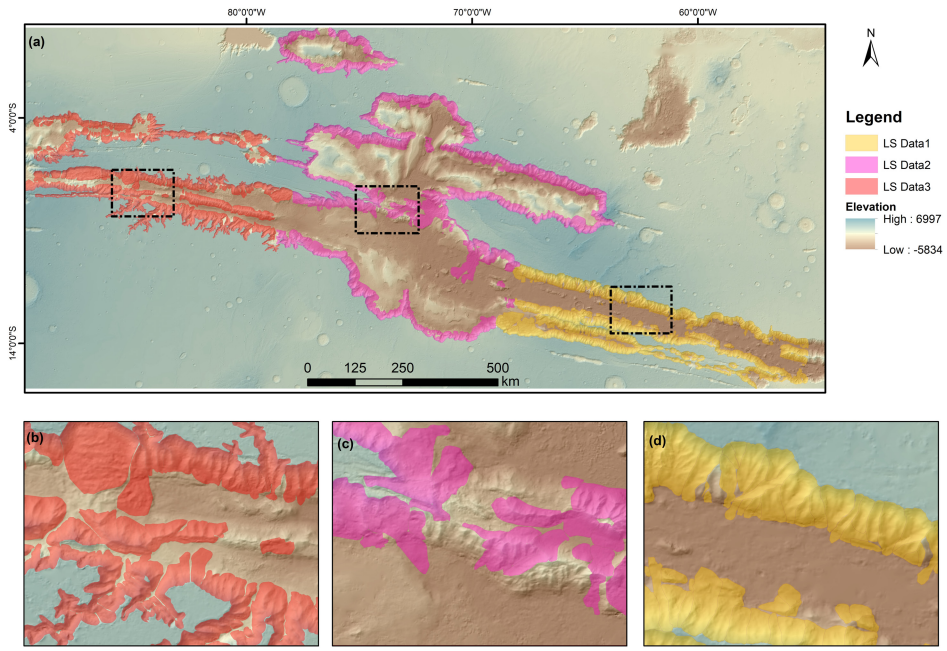

Figure 4. Distribution of Martian landslides in the Valles Marineris region. Region (b) and (d) are used to train and validate deep learning models respectively, while region (c) is used for performance testing.

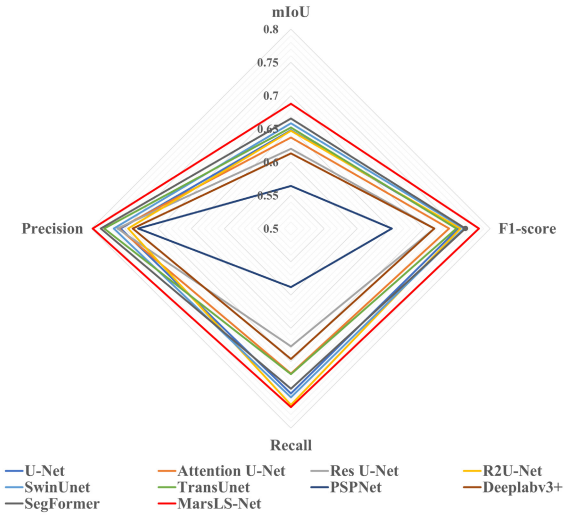

Figure 5. Spider plots representing the performance of the deep learning models that are experimented with in this study. The precision and F1-score are plotted on the X-axis (left and right angles), and the mIoU and recall scores are plotted on the Y-axis (top and bottom angles).

dataset and compare its performance with state-of-the-art deep network architectures. The performance of DL models is evaluated in terms of 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. Since most of the tested models converge within the range of 60-80 epochs, all the experiments, including competing models, run for a total of 100 epochs with a batch size of 16 samples per iteration. Each training process is monitored, and the best performance (mIoU) is saved. The training process was optimized with the Adam algorithm using a scheduled learning rate, initially set to 0.001 and reduced by a factor of 0.01 in every epoch, and binary cross-entropy is used as the loss function.

Table 2. Quantitative evaluation on the test set.

Method	mIoU	F1-score	Recall	Precision
U-Net [37]	0.6479	0.7519	0.7479	0.7558
Attention U-Net [27]	0.6370	0.7380	0.7182	0.7590
ResU-Net [47]	0.6196	0.7163	0.6769	0.7607
R2U-Net [2]	0.6474	0.7550	0.7651	0.7452
SwinUnet [7]	0.6583	0.7602	0.7539	0.7666
TransUnet [8]	0.6521	0.7488	0.7192	0.7810
SegFormer [46]	0.6656	0.7626	0.7409	0.7857
PSP Net [48]	0.5641	0.6516	0.5881	0.7304
Deeplabv3+ [9]	0.6131	0.7163	0.6961	0.7376
MarsLS-Net	**0.6878**	**0.7831**	**0.7684**	**0.7983**

5.1. Comparison with State-of-the-art Methods

The performance of our MarsLS-Net is compared with several the state-of-the-art segmentation models, including U-Net [37], Attention U-Net [27], ResU-Net [47], R2U-Net

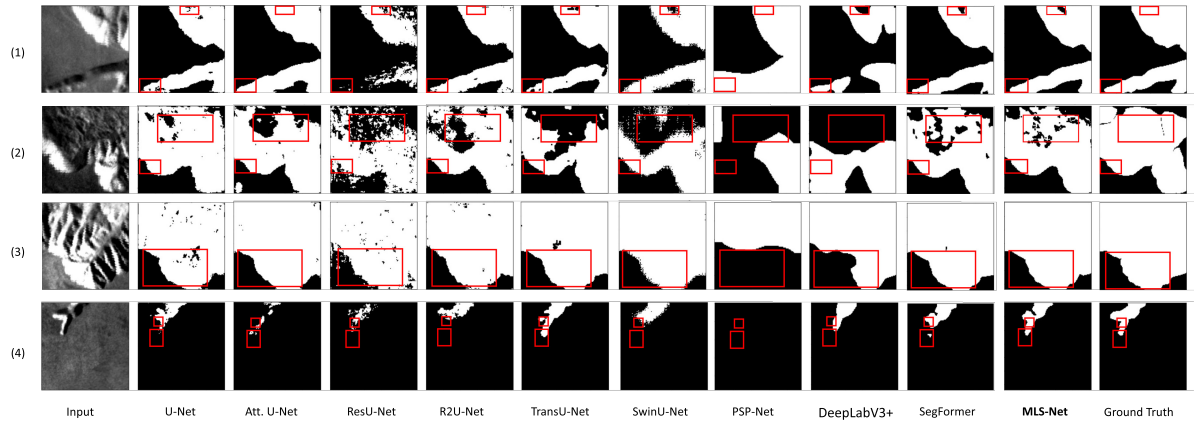

Figure 6. The illustration presents a visual comparison of Martian landslide segmentation results achieved through various deep learning approaches. The highlighted red boxes depict the areas where our approach adeptly predicts landslide-prone regions in accordance with the ground truth. These visual cues underscore the discernible differences in performance compared to other competing methods.

[2], SwinUnet [7], TransUnet [8], SegFormer [46], PSP-Net [48], and DeepLabV3+ [9]. Quantitative evaluation of our MarsLS-Net performance and the aforementioned models is provided in Table 2. It can be observed that the proposed deep network architecture outperforms all competing methods in terms of mIoU, recall, precision, and F1 score. The mIoU is a reliable metric for the segmentation task since this takes into account the relation between the predicted and ground truth annotation. From Table 2, it can be seen that the highest mIoU of 0.6878 is achieved by our MarsLS-Net, a considerable improvement over the second best performer (i.e., SwinUnet with a mIoU of 0.6583). Fig. 5 explains how our model performs when tested on four aforementioned evaluation metrics, while Fig. 6 shows a visual comparison of some sample results obtained by deep models experimented in this study. Specifically, a total of six random samples from the testing dataset are shown with the respective input image, segmentation mask, or ground truth, and the prediction of each model is used for comparison. As it is observed, our MarsLS-Net shows a more accurate segmentation for most of the cases.

5.2. Ablation Study

We conduct a set of experiments for ablation studies of our model from two different aspects: 1) the relative importance of different components in MarsLS-Net architecture, and 2) the sensitivity of various configurations of MarsLS-Net.

Importance of model components. In Table 4, we evaluate the performance of the proposed model by dropping one or more of the core components, including ConvPE layer, MHSA layer, and PE layers. As demonstrated in the table, the performance of MarsLS-Net decreases in the absence of any of the aforementioned components, in terms of mIoU, F1-score, and recall. For instance, mIoU is dropped

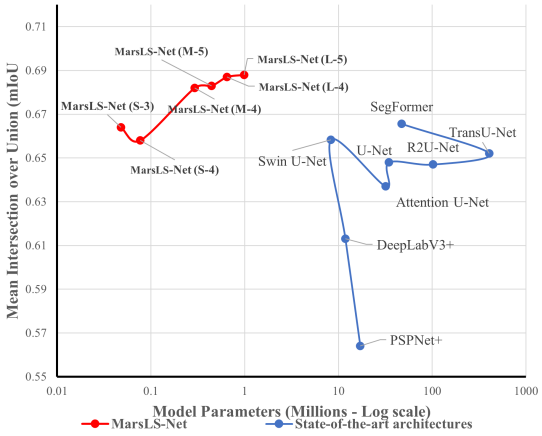

Figure 7. Model size vs. mIoU of different segmentation models experimented in this study.

by around 6.3% if PE and MHSA layers are excluded. This further confirms that the promising performance achieved by our proposed model is not a direct result of a particular component. Instead, its learning procedure can be seen as a form of episodic learning where every component has its own functionality.

Sensitivity of model configuration. Table 3 shows the performance of the proposed architecture by using different configurations in terms of the number of feature maps and the 2D convolution kernel size (i.e., the total number of PEN-Attention blocks). In each configuration, a name was denoted by the size of the architecture. For instance, in MarsLS-Net (S-3), S stands for a small model, a 3×3 kernel size used for each 2D convolution, along with 16 initial feature maps. We herein report experimental results by varying the initial feature maps to 16, 32, and 48, and the

Table 3. Comparison of different configurations of the proposed MarsLS-Net architecture.

Model Name	Initial feat. map	Kernel size	mIoU	F1-score	Recall	Precision	Model Parameters (Million)
MarsLS-Net (S-3)	16	n=3	**0.6635**	0.7703	0.7881	**0.7533**	0.0476
MarsLS-Net (S-4)		n=4	0.6581	**0.7774**	0.8447	0.7200	0.0770
MarsLS-Net (S-5)		n=5	0.6060	0.7542	**0.8875**	0.6557	0.1149
MarsLS-Net (M-3)	32	n=3	0.6635	0.7787	**0.8335**	0.7307	0.1770
MarsLS-Net (M-4)		n=4	0.6816	**0.7855**	0.8062	0.7659	0.2932
MarsLS-Net (M-5)		n=5	**0.6831**	0.7846	0.7954	**0.7742**	0.4427
MarsLS-Net (L-3)	48	n=3	0.6510	0.7739	**0.8496**	0.7105	0.3883
MarsLS-Net (L-4)		n=4	0.6866	**0.7847**	0.7822	0.7872	0.6487
MarsLS-Net (L-5)		n=5	**0.6878**	0.7831	0.7684	**0.7983**	0.9835

Table 4. Ablation study on the contribution of different components in the MarsLS-Net.

Method	mIoU	F1-score	Recall	Precision
w/o ConvPE	0.6442	0.7449	0.7267	0.7641
w/o MHSA	0.6450	0.7380	0.6921	0.7905
w/o PE & MHSA	0.6443	0.7372	0.6904	0.7908
Full version	**0.6878**	**0.7831**	**0.7684**	**0.7983**

2D convolution kernel size to 3, 4, and 5. For further reference, M stands for Medium, and L stands for Large. It is noticeable that the performance of the architecture tends to improve, in terms of mIoU, as the kernel size and initial feature map increase when the architecture is either medium or large. However, for the smaller version of the architecture, the trend goes in the opposite direction, reaching the best performance when the kernel size is 3 and the lowest performance with a kernel size of 5. It is important to point out that there is an increment of the computational cost as the feature map and the kernel size increase, as shown in the last column of the table. From this table, it can be observed that the best mIoU score (the most wildly metric used in the segmentation task) is achieved by setting a feature map of 48 and a kernel size of 5 (i.e., MarsLS-Net L-5) which is used in our proposed MarsLS-Net.

Figure 7 depicts a comparison of the parameter efficiency and segmentation accuracy, in terms of mIoU, of all variants of MarsLS-Net and the other competing models. As it is shown that the proposed MarsLS-Net with all its variants is consistently more parameter efficient and achieves a better mIoU score compared to the other state-of-the-art segmentation methods, which demonstrates the efficiency and effectiveness of the proposed model.

6. Conclusion

Detection of Landslides comprise a great importance for hazard assessment and environmental management. It is noteworthy to mention that while the developed model was specifically designed for Martian landscapes, its applicability extends to landslides on Earth as well. By adapting our methodology, we can potentially assist in automating the identification of landslides in terrestrial environments. In addition, regions like the Sahara, where vegetation cover is sparse, can benefit from our approach, as it does not rely on vegetation for landslide detection, similar to the Martian landscape.

Our proposed model, namely the Martian landslide segmentation network (MarsLS-Net), is a light-weighted deep neural network architecture. MarsLS-Net is composed by a series of progressively expanded neuron attention (PEN-Attention) blocks, where each PEN-Attention block consists of two core components: convolutional progressive expansion layer and multi-head self-attention layer. These two components were designed to obtain rich but also more relevant feature representation for segmenting Martial landslides. To test the effectiveness of our proposed architecture, we introduced a multi-modal Martian landslide segmentation (MMLS) dataset. This dataset will soon be made publicly available on a GitHub repository for the research community. We believe our study will establish a solid foundation for future research.

In addition, several interesting observations are drawn. (i) We find that it is possible to learn strong representations with a series of PEN-Attention blocks without introducing other complex deep network architectures. (ii) We empirically showed that both convolutional progressive expansion and multi-head self-attention mechanisms are key learners to extract pertinent information on landslides from multi-modal Mars imagery. (iii) We achieved superior results in segmenting Martian landslides compared to the state-of-the-art deep learning models, including the popular U-Net and its variants that have been explored for earth-based landslide segmentation tasks. In future work, we anticipate analyzing the contributions of each modality in the dataset for the Martian landslide segmentation problem and also explore innovative data fusion methods that can facilitate prediction outcomes.

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, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018. 3, 6, 7

[3] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016. 4

[4] L Bragagnolo, LR Rezende, RV da Silva, and JMV Grzybowski. Convolutional neural networks applied to semantic segmentation of landslide scars. *Catena*, 201:105189, 2021. 2

[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, 2

[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. *arXiv preprint arXiv:2105.05537*, 2021. 3, 6, 7

[8] 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. 3, 6, 7

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 3, 6, 7

[10] 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. 2, 5

[11] 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

[12] 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, 3

[13] A Dufresne and TR Davies. Longitudinal ridges in mass movement deposits. *Geomorphology*, 105(3-4):171–181, 2009. 5

[14] R. Fergason, 2014. `File:THEMIS_TI_Mosaic_Qual_30N240E_100mpp.cub`, USGS Astrogeology Science Center. 5

[15] RL Fergason, TM Hare, and J Laura. Hrsc and mola blended digital elevation model at 200m v2. *Astrogeology PDS Annex, US Geological Survey*, 2018. 5

[16] Emilie Gardin, Pascal Allemand, Cathy Quantin, Simone Silvestro, and Christophe Delacourt. Dune fields on mars: Recorders of a climate change? *Planetary and Space Science*, 60(1):314–321, 2012. 1

[17] Omid Ghorbanzadeh, Hejar Shahabi, Alessandro Crivellari, Saeid Homayouni, Thomas Blaschke, and Pedram Ghamisi. Landslide detection using deep learning and object-based image analysis. *Landslides*, 19(4):929–939, 2022. 2, 3

[18] 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. 5

[19] Matthew P Golombek, John A Grant, Larry S Crumpler, Ronald Greeley, Raymond E Arvidson, James F Bell III, Catherine M Weitz, R Sullivan, Philip R Christensen, LA Soderblom, et al. Erosion rates at the mars exploration rover landing sites and long-term climate change on mars. *Journal of Geophysical Research: Planets*, 111(E12), 2006. 1

[20] 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

[21] 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, 3

[22] JR Jenson. Introductory digital image processing: a remote sensing perspective.,(pearson prentice hall: Upper saddle river, nj). 2005. 5

[23] Sanjay Katta, Sidike Paheding, Thomas Oommen, Rajaneesh A, and Sajinkumar K. S. Categorization of Martian Landslides from Satellite Imagery Using Vision Transformer. AGU, Dec. 2021. 2

[24] Ren N. Keyport, Thomas Oommen, Tapas Ranjan Martha, Kochappi Sathyan Sajinkumar, and John S. Gierke. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. *Int. J. Appl. Earth Obs. Geoinformation*, 64:1–11, 2018. 2

[25] Baerbel K Lucchitta. Landslides in valles marineris, mars. *Journal of Geophysical Research: Solid Earth*, 84(B14):8097–8113, 1979. 2

[26] John F Mustard, Christopher D Cooper, and Moses K Rifkin. Evidence for recent climate change on mars from the identification of youthful near-surface ground ice. *Nature*, 412(6845):411–414, 2001. 1

[27] 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. 3, 6

[28] Sidike Paheding. *Progressively expanded neural network for automatic material identification in hyperspectral imagery*. PhD thesis, University of Dayton, 2016. 3

[29] 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

[30] Nikhil Prakash, Andrea Manconi, and Simon Loew. A new strategy to map landslides with a generalized convolutional neural network. *Scientific reports*, 11(1):1–15, 2021. 2

[31] Wenwen Qi, Mengfei Wei, Wentao Yang, Chong Xu, and Chao Ma. Automatic mapping of landslides by the resu-net. *Remote Sensing*, 12(15):2487, 2020. 3

[32] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern recognition*, 106:107404, 2020. 3

[33] Cathy Quantin, Pascal Allemand, and Christophe Delacourt. Morphology and geometry of valles marineris landslides. *Planetary and Space Science*, 52(11):1011–1022, 2004. 5

[34] 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, 2, 5

[35] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 3

[36] Abel A Reyes, Sidike Paheding, A Rajaneesh, KS Sajinkumar, and Thomas Oommen. C-ples: Contextual progressive layer expansion with self-attention for multi-class landslide segmentation on mars using multimodal satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 354–364, 2023. 3

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2, 6

[38] 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

[39] Lucas P Soares, Helen C Dias, and Carlos H Grohmann. Landslide segmentation with u-net: Evaluating different sampling methods and patch sizes. *arXiv preprint arXiv:2007.06672*, 2020. 2

[40] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7262–7272, 2021. 3

[41] Xiaochuan Tang, Zihan Tu, Yu Wang, Mingzhe Liu, Dongfen Li, and Xuanmei Fan. Automatic detection of coseismic landslides using a new transformer method. *Remote Sensing*, 14(12):2884, 2022. 3

[42] 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

[43] 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. 5

[44] 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. 3

[45] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3349–3364, 2020. 3

[46] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. 3, 6, 7

[47] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018. 3, 6

[48] 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. 6, 7