



# Automated Multi-class Crater Segmentation in Mars Orbital Images

Shrey Malvi\*  
smalvi@asu.edu  
Arizona State University  
Tempe, Arizona, United States

Mirali Purohit  
mpurohi3@asu.edu  
Arizona State University  
Tempe, Arizona, United States

Hitansh Shah\*  
hshah79@asu.edu  
Arizona State University  
Tempe, Arizona, United States

Jacob Adler  
jbadler2@asu.edu  
Arizona State University  
Tempe, Arizona, United States

Niketan Chandarana  
nchandar@asu.edu  
Arizona State University  
Tempe, Arizona, United States

Hannah Kerner  
hkerner@asu.edu  
Arizona State University  
Tempe, Arizona, United States

## ABSTRACT

Crater mapping and counting are critical analyses in many planetary science investigations, as the size-frequency distribution of impact craters can be used to measure the age of a planet's surface and interpret its geologic history. Crater counting is extremely tedious — counting hundreds to thousands of small features in a small region could take days to months for a trained planetary scientist. Previous work has demonstrated the feasibility of using computer vision techniques to automatically map and count craters in Mars orbital images using semantic segmentation. We present an improved approach for binary and multi-class semantic segmentation of craters in THEMIS daytime thermal infrared images using U<sup>2</sup>-Net and U-NetFormer with template matching. Our approach is the first method to perform multi-class segmentation of craters using computer vision. Our binary segmentation approach outperforms previous approaches that used semantic segmentation and template matching. A new global high-resolution image mosaic dataset of Mars (CTX, 5 m/px) is now available, but to date, no studies have benchmarked automated crater counting methods for this improved dataset. Toward this goal, we applied the THEMIS-trained model to out-of-domain CTX datasets and evaluated results quantitatively using the DoMars16 benchmark dataset and qualitatively using global mosaic tiles. We show that the THEMIS-trained models effectively segment craters in CTX images without additional fine-tuning. The code can be found [here](#).

## CCS CONCEPTS

• **Computing methodologies** → **Image segmentation**; *Neural networks*.

\*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*GeoAI '23, November 13, 2023, Hamburg, Germany*

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0348-5/23/11...\$15.00

<https://doi.org/10.1145/3615886.3627748>

## KEYWORDS

Multi-class crater classification, semantic segmentation, out-of-distribution evaluation, U-Net, template matching, remote sensing

### ACM Reference Format:

Shrey Malvi, Hitansh Shah, Niketan Chandarana, Mirali Purohit, Jacob Adler, and Hannah Kerner. 2023. Automated Multi-class Crater Segmentation in Mars Orbital Images. In *6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI '23)*, November 13, 2023, Hamburg, Germany. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3615886.3627748>

## 1 INTRODUCTION

In planetary science, crater detection is widely explored and an essential task to understand the history and age of surfaces of other planets. It can help to determine the relative ages of geologic units (i.e., which surfaces are older than others) and can estimate the absolute ages (e.g., 3.8 Ga) of surfaces based on the assumed flux of impacts on each planet [1, 6, 13, 16, 27, 32]. Additionally, it also helps with safe path analysis and hazard detection for rover navigation and landing. Studies of Mars' surface particularly rely on crater counting, as regions can vary in age from > 4.0 Ga to present, and the age bears strongly on the interpreted climate history and formation process. In general, more craters and more *large* craters indicate an older surface. Also, assessments of the size distribution of craters and identifying types of craters can be quite informative. A number of martian image and topographic datasets exist with global coverage of the planet (from THEMIS - THERMAL EMISSION IMAGING SYSTEM [8], MOLA - Mars Orbiter Laser Altimeter [25], and CTX - ConTeXt camera [2], etc.) making crater detection possible at various resolutions over the globe.

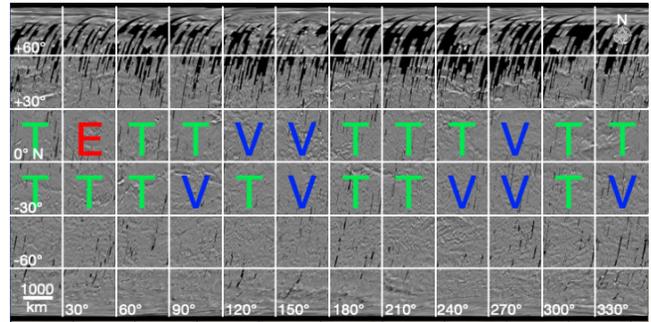
Despite the availability of large and diverse datasets, there are several challenges in crater detection. Manual detection of craters is extremely tedious and time-consuming because planetary scientists map craters by a visual inspection for most studies, looking through images rendered within geospatial software over a period of days to years. Hence, although there are terabytes of data available from various satellites, labeled data is scarce. Additionally, craters vary substantially in their appearance: craters can be very small or very large, are typically circular but might sometimes have a more elliptical shape due to the angle of impact, can be overlapping, and can be partially buried by younger lava flows or landslides. Furthermore, the choice of which craters to map based on the type of crater

observed may vary depending on the intended use case—for example, certain crater morphologies (i.e., buried or secondary craters) should not be included in crater counts used for age calculations [18]. For this reason, multi-class crater detection can help scientists categorize and filter by different types of craters that should be considered in a given study.

In the early stages of automating the task of crater detection, researchers employed classical methods like Support Vector Machines [33], Continuously Scalable Template Matching [30], Feed-forward Neural Networks, and Ensemble Methods [10]. These methods showed promising results, however, they had several drawbacks including complex frameworks and suffering from variations in crater appearances due to different illumination conditions [7]. In recent years, many deep learning (DL) based Crater Detection Algorithms (CDAs) have been proposed but have some important limitations. Previous methods are limited to binary classification or segmentation and do not perform multi-class segmentation. Previous studies have also primarily used the THEMIS global thermal image dataset [12, 28] for crater mapping instead of the newer, higher-resolution CTX (5 m/pixel resolution) dataset [11] that is now globally available and would provide more detailed crater maps and counts. To overcome these limitations, we created a pipeline (shown in Figure 2a) that uses newer Deep Learning-based models and template matching as a post-processing step. We performed both binary and multi-class crater segmentation on the THEMIS dataset. We compare our method with a recent study by DeLatte et al. [9], our method outperforms theirs in terms of recall and f1-score by about 10%. To the best of our knowledge, we are the first to propose an automated multi-class crater segmentation pipeline using deep learning models. Multi-class segmentation maps would allow planetary scientists to filter crater maps or counts according to the relevant morphologies for a given study, for example, to ignore partially or completely buried impact craters to enable a more precise estimation of the surface age. Toward the ultimate goal of automated crater mapping in the newer, higher-resolution CTX global image dataset, we applied the THEMIS-trained binary segmentation model to out-of-domain CTX images. We evaluated results quantitatively using the DoMars16 benchmark dataset [34]. These out-of-domain experiments show promising results for leveraging THEMIS-trained models for crater mapping in higher-resolution CTX images without additional training or fine-tuning.

## 2 RELATED WORK

Planetary scientists largely developed the field of crater counting in the Apollo era, as astronauts collected and returned samples of rocks from different regions of the Moon which were used to calibrate functions to estimate the absolute age of a surface based on the size-frequency distribution of craters on the surface. Historically, counting craters in a region was a manual task, with domain experts mapping the craters by eye and hand in images taken by Earth-based telescopes, flyby missions, orbiters, and lander cameras. The counted craters from a region were then plotted as a size-frequency distribution and compared to other regions on the Moon or other planets [16]. In the last decade, studies proposing automated pipelines for counting craters of different sizes in spacecraft images have been published. Early methods relied on continuously



**Figure 1: Train/val/test splits from THEMIS data. Green tiles (T) for training, blue tiles (V) for validation, and red tiles (E) for testing.**

scalable template matching [30], support vector machines [33], decision trees [26], feed-forward neural networks, or ensemble methods [10]. Some showed promising results for Mars and the Moon, however, their frameworks were complex, not scalable for global-scale processing, and not robust to the diverse surface conditions that influence crater appearance such as illumination angle, crater shape, or atmospheric perturbations. [20] and [24] were amongst the first studies to propose an approach based on Convolution Neural Networks (CNNs), and more specifically a U-Net-based approach. These initial studies focused on automated crater mapping in lunar orbital image datasets. Taking inspiration from the template matching algorithm in [24], [9] proposed a U-Net-based model using the global Mars THEMIS thermal image dataset, and modified the U-Net architecture with 7x downsampling for semantic segmentation and replaced max pooling with average pooling which led to a dramatic reduction in training time. [17] proposed a Nested Attention-Aware U-Net (NAU-Net) model based on the U-Net++ architecture for lunar crater detection, which has a faster convergence rate than U-Net. [14] introduced a new image cropping method, Pyramid Cropping, which uses multi-scale crater segmentation for detecting lunar craters, allowing craters as large as 200 km in diameter to be counted. Most recently, [7] implemented a Martian Crater U-Net (MC-UNet) approach for Martian craters using the THEMIS dataset, which has the fastest convergence rate and better precision score compared to prior methods. In addition, apart from U-Net-based methods, YOLO-based methods have been also utilized to build CDA [3, 35]. In terms of deep learning-based multi-class crater detection, there has been no prior research conducted. The most closely related work we found was presented by Lagain et al. [19] wherein they introduced a Crater Detection Algorithm with cluster analysis to identify secondary craters along with primary craters. With respect to CTX, Benedix et al. [3] claimed that they found 60k craters using their CDAs but there was no mention of a definite data processing pipeline and evaluation. We opted to continue the research with U-Net-based methodology as they have achieved state-of-the-art results for crater counting. Hence, we have used newer and more advanced architectures: U<sup>2</sup>-Net and U-NetFormer for binary as well as multi-class semantic segmentation and defined a pipeline for evaluation.

### 3 DATASETS

The primary dataset used in this research is the THEMIS global daytime infrared image map of Mars from 2006 [28]. In order to train a model for crater segmentation, we used annotated labels of Martian craters from the manually annotated catalog provided by Robbins and Hynes [23]. Class labels for multi-class segmentation were obtained from the catalog by Lagain et al. [18]. For out-of-domain evaluation using CTX images, we used the global image mosaic available at the Murray Lab site [5] and the DoMars16 benchmark dataset [34].

#### 3.1 THEMIS

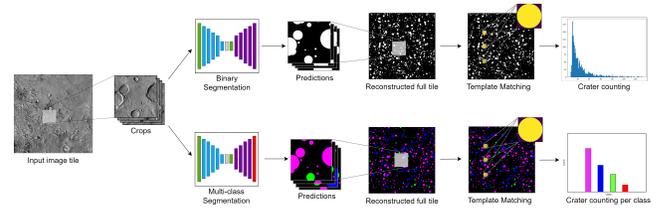
The THERmal EMission Imaging System (THEMIS) global dataset contains  $30^\circ$  by  $30^\circ$  tiles (7680 by 7680 pixels), ranging from  $-90^\circ$  to  $90^\circ$  latitude and  $0^\circ$  to  $360^\circ$  longitude. In total, the dataset is organized into 72 tiles covering the entire surface of Mars (Figure 1). We have used the tiles from the equatorial region of  $-30^\circ$  to  $30^\circ$  latitude and  $0^\circ$  to  $360^\circ$  longitude (24 tiles in total) for this study since this has been the region of interest of most previous studies including [9] and tiles in this region have the fewest missing pixels. The resolution of these images is 256 pixels per degree (ppd), or 231.55 meters per pixel (m/px). It must be noted that we have used the older version of the THEMIS dataset to be consistent with the work of [9] for fair comparison and since [23] had also provided the annotations on the same version. As a separate experiment, we trained and evaluated our models on the newer version of THEMIS [29] as well, however, the results were very similar.

**3.1.1 Binary Segmentation:** In 2012, Robbins and Hynes (RH2012) [23] provided a global catalog of martian craters which were annotated manually by experts on the THEMIS Daytime infrared (IR) global mosaic. This catalog contains 384,343 crater annotations, each having diameter  $\geq 1$  km, and other information if it could be determined including position, morphology, ejecta morphology, and modification state. In 2021, Lagain et al. [18] provided a modified version of the RH2012 crater database to correct for the populations of craters that were previously misclassified and identify those that should be excluded for crater counting (e.g., buried craters).

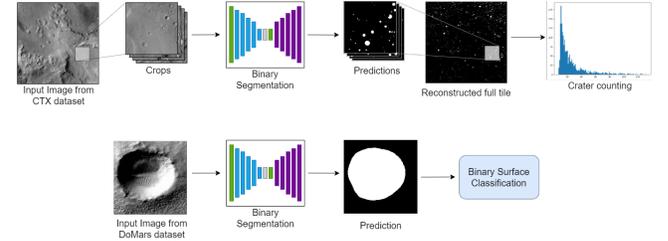
**3.1.2 Multi-class Segmentation:** Lagain et al. [18] classified craters into 4 categories: Other, Secondary, Layered, and Buried. An example of each category is shown in Figure 3. Layered craters are primary impact craters that have a continuous and circular or radial layered ejecta morphology. Secondary craters are craters that are formed as an after-effect from a larger primary impact crater, along a radial direction of the primary impact crater, often having a more elliptical or herringbone shape and being very small in size. Buried craters are impact craters that are partially or completely filled with sedimentary or volcanic materials. All other impact craters that cannot be classified into either of the three categories are classified as Other.

#### 3.2 CTX

The ConTeXt Camera (CTX) has been operational since 2006 on the Mars Reconnaissance Orbiter (MRO) and has collected high-resolution images covering the entirety of the planet ( $\geq 99.5\%$ ). The Murray Lab [5] Global CTX Mosaic dataset merges the best CTX



(a) Pipeline for binary segmentation and multi-class segmentation of THEMIS images followed by crater counting and per class crater counting.



(b) Illustration of out-of-domain dataset evaluation on CTX and DoMars16.

**Figure 2: Schematic representation of training and out-of-domain evaluation pipeline.**

images in each region to create a seamless global image mosaic at 5 meters per pixel [21]. Each image is a  $4^\circ$  by  $4^\circ$  tile, having a resolution of 47420 by 47420px.

#### 3.3 DoMars16

DoMars16 is a publicly available dataset for geologic landform classification based on CTX [34]. The dataset contains 16150 data samples from 5 different thematic group classes which have 15 different sub-classes including crater and crater field. We employed this dataset for binary classification, meaning that the model’s objective was to determine whether a given sample contained one or more craters by segmenting the craters within the sample. From all the classes of DoMars16, we designated two classes, “Crater” and “Crater Field”, as the positive class, while the remaining 13 classes were designated as the negative class for binary segmentation. Notably, some data samples in DoMars16 have multiple classes within one sample. For instance, a sample generated from the B20\_017281\_2002\_XN\_20N118W CTX image might encompass features like cliffs, craters, channels, and ridges simultaneously. To ensure a fair evaluation of the model’s performance on the negative class, we exclusively assessed samples that did not contain craters or crater fields. Number of samples per class is reported in Appendix Table 8. In other words, the model was expected to predict the negative class correctly for these samples because they lacked any craters.

### 4 METHOD

We have followed a 2-step approach: Segmentation of impact craters using a Deep Learning semantic segmentation algorithm, followed

by instance-level counting of impact craters using template matching. This approach is inspired by the work of DeLatte et al. [9], which used a U-Net-based approach for semantic segmentation. First, we improved the binary segmentation using recent and more advanced Deep Learning architectures. Then, we modified the binary segmentation approach to perform multi-class segmentation. Figure 2a represents an overview of our crater segmentation system. Lastly, we used the trained models to predict and count the craters from CTX images. Figure 2b shows the overall evaluation pipeline.

## 4.1 Semantic Segmentation

Semantic Segmentation is the task of assigning a class label to every single pixel of an input image, such as classifying a pixel as belonging to a particular object or background. In this study, we employ Semantic Segmentation methods based on U<sup>2</sup>-Net and U-NetFormer to determine whether a pixel belongs to a crater or not (binary segmentation) and to also determine the crater class of the pixel (multi-class segmentation).

**4.1.1 Binary Segmentation.** We used the U<sup>2</sup>-Net implementation for training the Semantic Segmentation model on the THEMIS dataset images. U<sup>2</sup>-Net, with a deep architecture, is known to produce sharper and more refined masks compared to U-Net without significantly increasing the computational cost [22]. We used the same data processing steps given by DeLatte et al. to have a fair and equal comparison of our methods with the previous ones. [9]. We have used two types of prediction masks: filled and edge masks containing craters from a range of 2 to 32 km in radius. The edge masks are created with a radius of 8px thickness. We split the dataset (24 tiles) as 15 tiles for training, 8 tiles for validation, and, 1 tile for testing shown in Figure 1. We divided each tile into 512 by 512px crops which resulted in 225 images/crops per tile and fed each crop as an input to the segmentation model. We obtained a 512 by 512px semantic map as a result of the segmentation model and then reconstructed the tile by stitching all the predicted semantic map crops together followed by template matching for crater counting.

**4.1.2 Multi-class Segmentation.** U<sup>2</sup>-Net, being a CNN-based method, possesses strong capability for capturing local information by adopting hierarchical representation. However, the convolution operation limits the model from capturing the long-range dependencies or global context. U-NetFormer, which uses attention blocks as backbone, has demonstrated the ability to extract global context and obtained state-of-the-art results for image classification tasks [31]. This global contextual information helps with a better or more accurate classification of each pixel, thus helping with assigning a proper class to the pixels and at large, to the craters. To validate the performance of U-NetFormer for crater detection, we trained this model for binary segmentation using the same data split mentioned in the section 4.1.1. Then, we trained for the multi-class data using the annotations provided by Lagain et al. [18]. Section 4.4 mentions the data processing and implementation details for training the multi-class segmentation model.

## 4.2 Template Matching and Crater Counting

Template Matching involves sliding a template (filled circles or circular rings in our case, depending on the annotations) over an entire

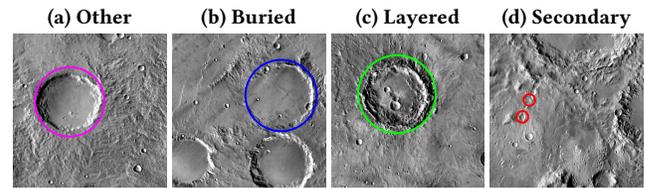


Figure 3: Class-wise samples from Lagain annotations [18].

larger image and evaluating the similarity between the template and each overlapping region of the image. We defined these templates for the range of 2-32 km, which corresponds to the range of 8.6px to 138.2px in the image. Then, we traverse the entire semantic map with these templates, one at a time, and filter out the circles with a cross-correlation coefficient value higher than 0.50 (template threshold) in the case of circular rings and 0.65 in the case of filled circles. We will discuss this disparity in the correlation coefficients and the need for it in the results section. For further refinement, i.e., to remove duplicate or overlapping craters, we set the *lat-lon-thresh*: minimum latitude and longitude distance between two candidate craters and *rad-thresh*: minimum radius delta between candidate craters to be considered as distinct craters. We have used the default parameters for *lat-lon-thresh* and *rad-thresh* as given by Silburt et al. [24]

## 4.3 Out-Of-Domain Evaluation Pipeline

**4.3.1 CTX.** The next big step in Martian studies, specifically crater detection and surface age estimation, is using higher resolution Mars orbital images, such as the ones taken through the Context Camera (CTX). However, unlike the THEMIS dataset, there is no global set of annotations available for the CTX images, which is why one cannot train a Semantic Segmentation model using these images at a similar scale as the THEMIS dataset, in terms of the geographical area covered. One can either create a set of annotations on a smaller geographical region and train a model on it, or use a model trained on similar data. To validate the effectiveness of our model, we performed a Binary Segmentation on one CTX tile containing the location of the MER Spirit rover mission in Gusev crater and surrounding region *MurrayLab\_GlobalCTXMosaic\_V01\_E172\_N-16*. For counting the craters, we used the HoughCircles method from the OpenCV library [4], and removed all the duplicate craters using Non-Max Suppression - this method is much faster as compared to Template Matching albeit not as accurate, which is why we only used it for this high-resolution CTX image

**4.3.2 DoMars16.** DoMars16 is evaluated using trained binary segmentation models, U<sup>2</sup>-Net and U-NetFormer. For DoMars16, we evaluate models based on their performance for binary classification, specifically, determining whether a given image contains a crater or not. For positive samples (crater and crater field), if the model successfully segments the crater area (indicating the detection of a crater), we classify that sample as a True Positive. Conversely, for negative samples (other landforms), where the model doesn't segment (predicting the absence of a crater), we categorize them as True Negatives. This approach helps in gauging the model's

effectiveness in correctly refraining from crater segmentation when there are no craters in the samples.

#### 4.4 Implementation Details

For binary segmentation, we trained three models U-Net, U<sup>2</sup>-Net and U-NetFormer. We trained the U-Net model using the same setting provided by DeLatte et al. [9] to use it as a baseline. In the case of U<sup>2</sup>-Net, we trained the model using Dice Loss [15] instead of Binary Cross Entropy which resulted in faster convergence. The input images were normalized in the range of -1 to 1 as a preprocessing step and the model was trained for 40 epochs with a batch size of 8 and the learning rate was set to 1e-3. Whereas for U-NetFormer, we opted for the same data preprocessing and training strategy as mentioned by [31] and used Dice Loss only on the output of the feature refinement head from the model. As this model is a lightweight implementation, we trained it for 100 epochs with a batch size of 16 and with an initial learning rate of 6e-4 and a cosine scheduler. All models were trained with Adam optimizer with default parameters.

In the case of multi-class, we trained U<sup>2</sup>-Net and U-NetFormer models. For both models, we replaced the last layer of the model with 'softmax' and defined the out channels = 5 (four for classes and one additional channel for background). Figure 3 shows the crater of each class represented by color mapping. We trained the model using a categorical-cross-entropy loss function. The rest of the parameters and settings have been kept the same as the binary segmentation model.

We used the following system for this study: Intel(R) Xeon(R) Gold 6148 CPU @ 2.40GHz and 16 GB Tesla V100-SXM2 GPU.

## 5 RESULTS AND ANALYSIS

### 5.1 Binary Segmentation

As mentioned in section 4.1.1, we trained binary segmentation models for two types of predictions: i) filled masks, ii) and edge masks. We used the dice coefficient as the metric for evaluating the model training. Previous studies have used accuracy as a measure for evaluating the models, however, it must be noted that this could lead to misleading results. The predicted masks and ground truth masks have a significantly larger number of black pixels (normal terrain) as compared to white (craters) pixels, which is why, the accuracy will always be high regardless of how good or bad the model performed. Table 1 shows the dice coefficient for all three models along with the number of epochs for which the models were trained. From these results, we can say that the visual predictions and the ultimate crater count for U<sup>2</sup>-Net and U-NetFormer will outperform those of U-Net. Figure 4 and 5 show the qualitative results, i.e., input crops of 512 by 512px, its ground truths, and the predictions of all three models for filled masks and edge masks respectively.

In Template Matching, we compare a crater prediction with circular templates of varying sizes as mentioned in 4.2, and determine if the cross-correlation coefficient is greater than a certain threshold to establish whether the prediction is actually a crater or not. Generally, this threshold is set at 0.50. Theoretically, in terms of circular rings, a partially detected crater might have just a semi-circular arc as the semantic segmentation prediction. However, in terms of a filled circle, a partially detected crater would be a little

Type	Model	Epochs	Train Dice Score	Validation Dice Score	Test Dice Score
Filled mask	U-Net	500	0.88	0.70	0.75
	U <sup>2</sup> -Net	40	0.82	0.76	0.81
	U-NetFormer	100	0.83	0.81	<b>0.84</b>
Edges	U-Net	500	0.59	0.59	0.60
	U <sup>2</sup> -Net	40	0.76	0.72	0.68
	U-NetFormer	40	0.73	0.74	<b>0.68</b>

**Table 1: Binary segmentation performance on training, validation, and testing for filled and edge prediction.**

Tile Id	Model	Crater Count	Precision	Recall	F1-Score
13 (Test)	U-Net	1364	0.86	0.60	0.71
	U <sup>2</sup> -Net	1644	0.92	0.72	0.81
	U-NetFormer	1728	0.92	0.76	<b>0.83</b>
16 (Validation)	U-Net	442	0.88	0.69	0.77
	U <sup>2</sup> -Net	538	0.92	0.84	<b>0.88</b>
	U-NetFormer	534	0.92	0.836	0.87
21 (Validation)	U-Net	476	0.86	0.78	0.82
	U <sup>2</sup> -Net	561	0.93	0.92	<b>0.93</b>
	U-NetFormer	539	0.96	0.88	0.92

**Table 2: Binary segmentation results for filled mask prediction at template threshold = 0.65.**

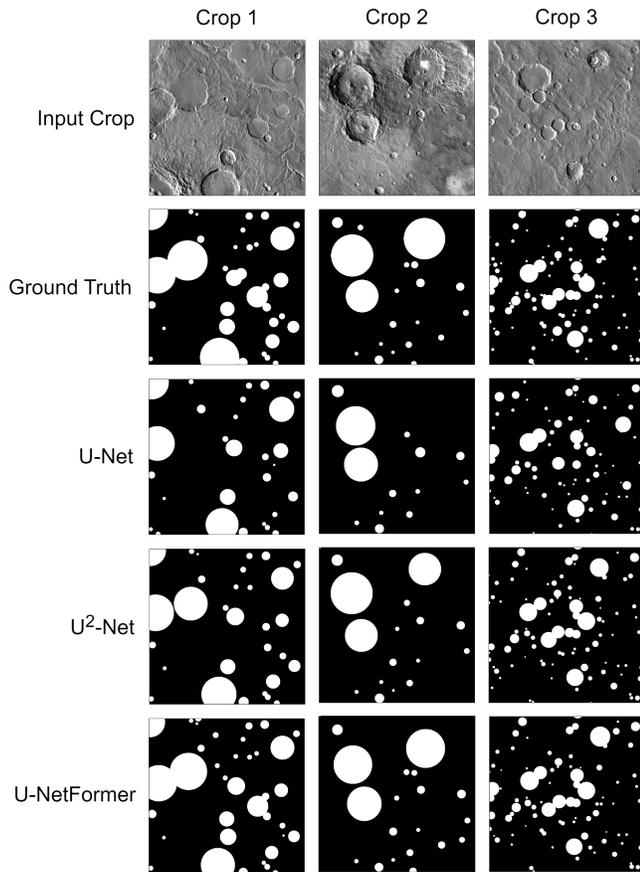
Tile	Model	Crater Count	Precision	Recall	F1-Score
13 (Test)	U-Net	1581	0.96	0.70	0.81
	U <sup>2</sup> -Net	1846	0.93	0.82	0.87
	U-NetFormer	1917	0.92	0.85	<b>0.88</b>
16 (Validation)	U-Net	523	0.96	0.82	0.88
	U <sup>2</sup> -Net	569	0.94	0.89	<b>0.91</b>
	U-NetFormer	565	0.93	0.88	0.90
21 (Validation)	U-Net	548	0.97	0.90	0.94
	U <sup>2</sup> -Net	566	0.96	0.93	<b>0.95</b>
	U-NetFormer	556	0.95	0.92	0.93

**Table 3: Binary segmentation results for edge mask prediction at template threshold = 0.50.**

bit more than a semi-circle, something like a gibbous moon. In that case, taking a cross-correlation coefficient value of 0.5 (template threshold) would lead to several duplicates, which might not get removed even after de-duplication. To verify this, we carried out a series of experiments, taking different cross-correlation coefficients in both cases. Through these experiments, we proved that our hypothesis was true in the case of filled masks, and the best result was obtained in the case of higher thresholds of 0.60 and 0.65. The detailed results from these experiments can be found in the Appendix. The best crater-counting results for filled masks and edge masks are presented in Table 3 and Table 2, respectively. Both U<sup>2</sup>-Net and U-NetFormer have shown about 10% improvement in terms of F1-Scores across all test and validation tiles. The detailed results of all the experiments have been provided in the appendix.

### 5.2 Multi-class Segmentation

Through the binary segmentation experiments, we found out that the edge masks perform better than the filled masks. So, for multi-class segmentation, we trained the model using the edge masks.

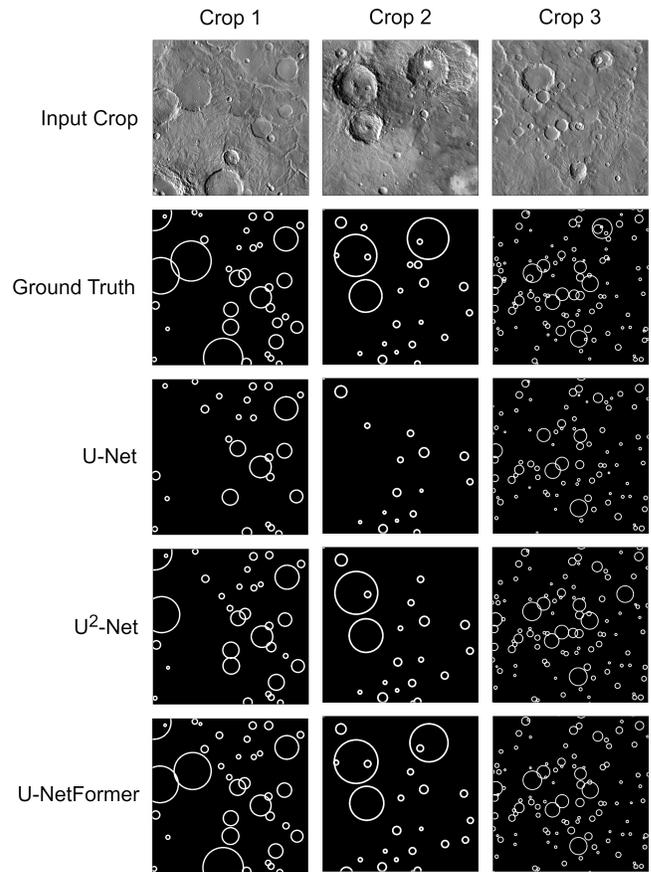


**Figure 4: Binary segmentation results for filled mask prediction on tile 13 (test).**

Table 4 shows that the training loss was almost the same for both models, but the difference starts to show in the validation loss. The test loss for the U-NetFormer model is quite low as compared to the U<sup>2</sup>-Net model. For crater counting, we modified the template matching algorithm. Firstly, we converted the predicted mask to single channel and performed template matching to find the craters from which we obtained x, y, and radius information. Then, to get class information of a particular crater, we took the maximum color pixels from the bounding box of that crater and assigned the class corresponding to that color. After performing the modified template matching, we got the crater counts of each of the 4 classes and evaluated the models' performance. Table 5 shows that for tile #13 (test), the f1-scores of all the 4 classes for the U-NetFormer model were better than the corresponding scores for the U<sup>2</sup>-Net model, thus proving the hypothesis we provided in 4.1.2. Figure 6 shows the qualitative comparison of both the models which also validates that U-NetFormer generates better masks reliably.

### 5.3 Out-Of-Domain Evaluation

**5.3.1 CTX.** We evaluated both, U<sup>2</sup>-Net and U-NetFormer-based models against the CTX tile as mentioned in section 4.3. The prediction masks can be seen in Figure 7 for the 2 models. A higher-resolution version of both predictions is available at [link](#). From a



**Figure 5: Binary segmentation results for edge mask prediction on tile 13 (test).**

Model	Train Loss	Validation Loss	Test Loss
U <sup>2</sup> -Net	0.04	0.08	0.22
U-NetFormer	0.04	0.05	<b>0.14</b>

**Table 4: Multi-class segmentation performance based on categorical cross-entropy loss on training, validation, and testing for edge prediction.**

visual inspection of the full-scale images, we can see that the U-NetFormer model has predicted more craters than the U<sup>2</sup>-Net-based model, however, the latter was able to predict craters of slightly larger size in terms of radius. U-NetFormer with 24669 predicted craters performs better as compared to U<sup>2</sup>-Net in terms of crater count which only predicted 20698 craters of range 70m to 800m.

**5.3.2 DoMars16.** We evaluated both the models on data samples belonging to the crater and non-crater classes as mentioned in 4.1.2. Both the models performed well in identifying craters and not identifying non-craters. The overall class-wise results have been mentioned in Table 6. The detailed results of each individual class and their performance have been reported in the Appendix Table 8.

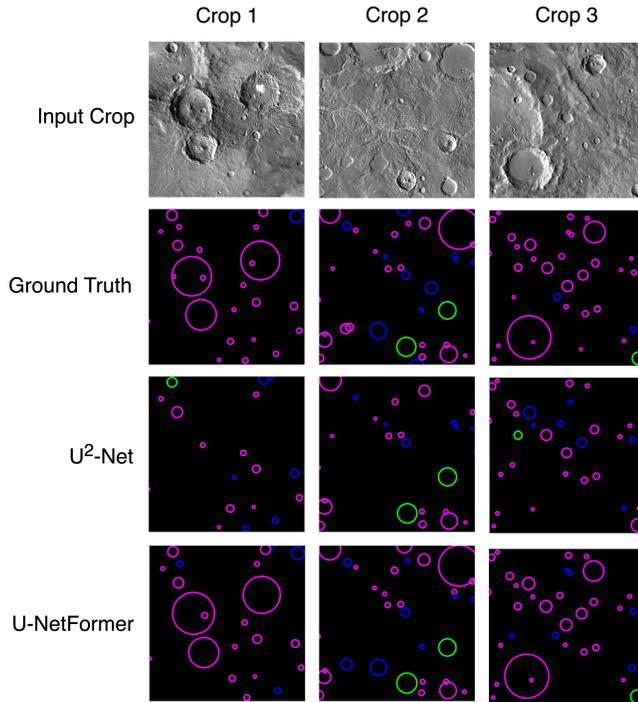


Figure 6: Multi-class segmentation results for edge prediction on tile 13 (test).

Model	Crater Category	Precision	Recall	F1-Score
U <sup>2</sup> -Net	Other	0.84	0.64	0.72
	Layered	0.67	0.31	0.42
	Buried	0.43	0.23	0.3
	Secondary	0	0	0
U-NetFormer	Other	0.83	0.69	<b>0.75</b>
	Layered	0.47	0.46	<b>0.47</b>
	Buried	0.59	0.46	<b>0.52</b>
	Secondary	0.14	0.13	<b>0.13</b>

Table 5: Multi-class segmentation results for edge mask prediction at template threshold = 0.5 for tile 13 (test).

Region	Accuracy		F1-score	
	U <sup>2</sup> -Net	U-NetFormer	U <sup>2</sup> -Net	U-NetFormer
Crater	92.5	<b>95.12</b>	96.1	<b>97.49</b>
Non-crater	<b>93.33</b>	91.99	<b>96.8</b>	95.69

Table 6: Average accuracy and average F1-Score on DoMars16 for Crater (positive) and Non-crater (negative) class.

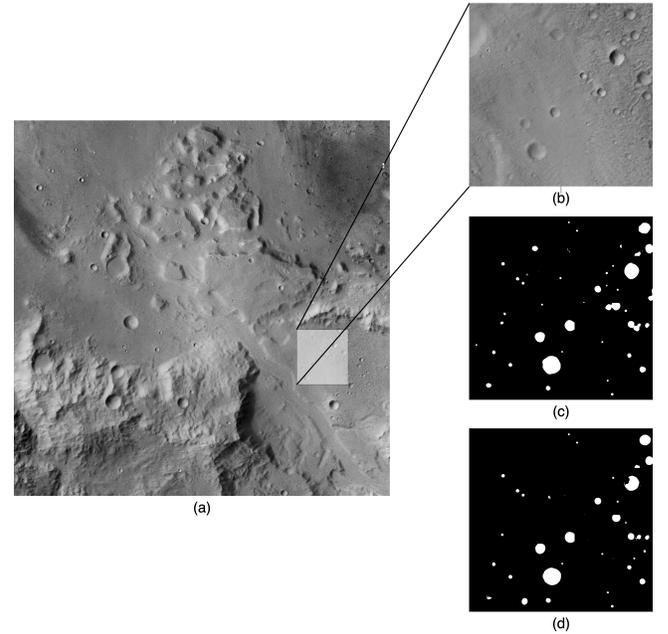


Figure 7: (a) Crop of  $1^\circ$  by  $1^\circ$  of the CTX test tile. (b), (c), and (d) represent the zoomed-in region of the crop and its corresponding predicted filled masks for U<sup>2</sup>-Net and U-NetFormer models respectively.

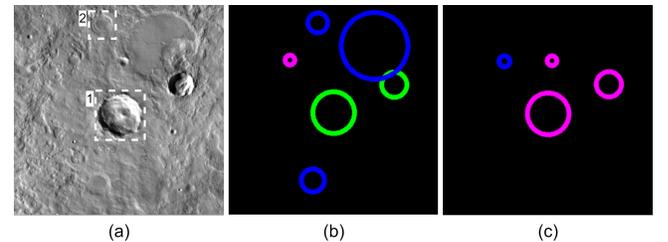
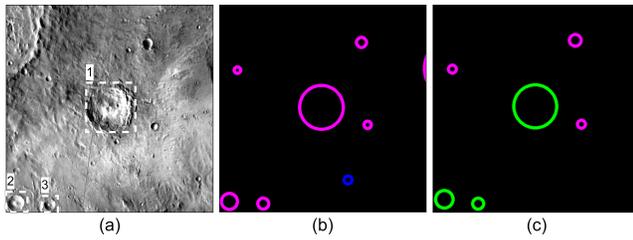


Figure 8: Importance of context in multi-class: (a) Crop of a small area from tile #13 (test) showing an actual image, (b) U-NetFormer predicted mask, (c) U<sup>2</sup>-Net predicted mask.

## 6 DISCUSSION

### 6.1 Importance of Context

In the process of semantic segmentation, the context of a pixel refers to the neighboring region of the pixel. As mentioned in Section 4.1.2, the surrounding region of a crater plays an important role in determining the class of the crater. Figure 8 shows an example from each class and it can be observed that there is an important context in classifying the type of crater. In Figure 8, (a) is a crop of a small area from tile #13 (test), and the 2 craters bounded by boxes 1 and 2 belong to layered and buried categories respectively. (b) and (c) are the predicted masks of the same crop from the U-NetFormer and U<sup>2</sup>-Net models, respectively. As stated previously, the U-NetFormer model takes into consideration the context of the neighboring pixels before assigning them a semantic value. The



**Figure 9: Crop of a small area from tile #13 (test) showing (a) the actual image, (b) ground truth annotations, (c) U-NetFormer predicted masks**

result of this is visible, as it was able to predict and categorize both the craters in question correctly, whereas the  $U^2$ -Net-based model categorized crater 1 as belonging to the “Other” class instead of the “Layered” class, and it could not predict either of the 3 buried craters, let alone categorizing them.

## 6.2 Inconsistencies in Data

This is one other critical factor that affects the results of the multi-class segmentation to a large extent. We have considered the catalog proposed by Lagain et al. [18] as the ground truth in the case of multi-class segmentation, but we did not take into consideration the possibility that there might be erroneous class labels in the catalog. One such example is shown in Figure 9, where (a) is a crop of a small area from tile #13, (b) shows the ground truth provided by Lagain et al. [18] in their catalog and (c) is the predicted mask of the same region using the U-NetFormer model. In (a), the craters bounded by boxes 1, 2, and 3 all match the definition of a layered crater, and yet, they have been marked as belonging to the “Other” crater class. On the other hand, our model predicted them to be of the layered class, which in a general sense should be the right prediction but is considered as an incorrect prediction as the catalog says otherwise. There were 3 misclassified craters in such a small region of just 1 tile — across the entire tile, there would be enough such instances to have a notable effect on the crater counts of the multi-class segmentation.

## 6.3 Limitations and Future Work

When we talk about tasks related to categorization or classification, any model tends to do better when it has near-equal data points of all the classes. However, that was not the case with this study. The 4 crater classes were severely unbalanced, with only 403 secondary craters and 2475 layered craters in the training tiles, as compared to 11362 craters of the “Other” class. This is one of the reasons why the model did not have the best multi-class segmentation results. To handle this issue, we can consider using a weighted categorical loss entropy function, assigning weights to each class inversely proportional to the number of instances of each class. We can also fine-tune the binary segmentation model for multi-class segmentation by initializing the model weights learned from the binary segmentation task.

The craters we identified in the CTX tile belong to a range of 70m to 800m radius, whereas in reality, there would be craters of all sizes. Since we have used the same model for this task, we had

to split the CTX tile into crops of size 512 by 512px and used a similar range of 14 to 160 pixels for the maximum radius parameter. We can generate crops of different sizes (1024 by 1024px, 2048 by 2048px, and so on) and scale them down to the 512 by 512px scale, adjust the pixel range so as to cover craters of all sizes, and predict the craters from each of these different scale crops. This approach might help identify craters of a larger range of sizes and could lead to an automated generation of ground truth crater annotations for CTX images.

As mentioned in Section 4.3.1, there is no high-resolution (based on CTX or HiRISE) annotated dataset available publicly for crater detection or crater segmentation. Also, domain experts will take a significant amount of time manually annotating the data, as the data is in terabytes. However, automated CDAs can be helpful in reducing human efforts and making it a less time-consuming task. Predictions or outputs from CDAs can be considered as a starting point and experts can use them to generate annotations by rectifying the incorrect or mismatched predictions, which is a less tedious task. Also, high-resolution satellite images are helpful in identifying small craters (the ones which are undetectable in THEMIS or low-resolution satellite images due to a lower meters/pixel resolution). As we proposed above, CTX data with different scales can be used to detect craters of all sizes. Thus, space scientists can use our model to generate a baseline set of annotations for high-resolution data, as our results prove that our model is capable of detecting craters on high-resolution data as well (shown in Figure 7).

## 7 CONCLUSION

Crater Counting is a critical part of larger studies related to surface age estimation, safe path analysis, rover landing, and other planetary research. Various Deep Learning-based models have been used in the past for this specific purpose, but there was never a definitive State-of-the-Art (SOTA) model. We achieved better performance than the previous state-of-the-art U-Net-based model by employing the  $U^2$ -Net model for binary segmentation. However, for multi-class segmentation, the U-NetFormer model surpassed the  $U^2$ -Net model. Both of these methods demonstrated approximately a 10% improvement in f1-score compared to the U-Net-based model for binary segmentation. Moreover, none of the previous methods addressed multi-class segmentation. Multi-class crater segmentation is critical as it helps get a more accurate estimate of the surface age. As per the authors’ knowledge, this is the first attempt to perform an automated multi-class crater segmentation and the results are promising given the challenges related to the class imbalance. We further proved that our hypothesis regarding the importance of context for multi-class crater segmentation was true, which is the reason why U-NetFormer performs better than  $U^2$ -Net. Detecting craters on higher-resolution CTX images is also something researchers have not explored in the past. We used the same model on this Out-of-distribution data which is of a completely different scale, and yet, both our models were able to detect craters pertaining to a similar range in terms of the pixel size. We also performed a similar evaluation on the CTX-based DoMars16 dataset and our model had exceptional results in identifying craters and not identifying non-craters, with an accuracy of 95.12%. We hope that our work helps space scientists with the various space

exploration tasks and that it inspires computer scientists to build more robust models for Binary and Multi-class Segmentation of not just craters, but other planetary surface features as well.

## REFERENCES

- [1] Nadine G Barlow. 1990. Constraints on early events in Martian history as derived from the cratering record. *Journal of Geophysical Research: Solid Earth* 95, B9 (1990), 14191–14201.
- [2] JF Bell III, MC Malin, MA Caplinger, J Fahle, MJ Wolff, BA Cantor, PB James, T Ghaemi, LV Posiolova, MA Ravine, et al. 2013. Calibration and performance of the Mars reconnaissance orbiter context camera (CTX). *International Journal of Mars Science and Exploration* 8 (2013), 1–14.
- [3] GK Benedix, Anthony Lagain, Kevin Chai, S Meka, S Anderson, C Norman, PA Bland, Jonathan Paxman, MC Towner, and Tele Tan. 2020. Deriving surface ages on Mars using automated crater counting. *Earth and Space Science* 7, 3 (2020), e2019EA001005.
- [4] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).
- [5] California Institute of Technology - Division of Geological and Planetary Sciences. [n. d.]. The Bruce Murray Laboratory for Planetary Visualization. <http://murraylab.caltech.edu/CTX/>.
- [6] Clark R Chapman and Kenneth L Jones. 1977. Cratering and obliteration history of Mars. *Annual Review of Earth and Planetary Sciences* 5, 1 (1977), 515–538.
- [7] Dong Chen, Fan Hu, P Takis Mathiopoulos, Zhenxin Zhang, and Jiju Peethambaran. 2023. MC-UNet: Martian Crater Segmentation at Semantic and Instance Levels Using U-Net-Based Convolutional Neural Network. *Remote Sensing* 15, 1 (2023), 266.
- [8] Philip R Christensen, Bruce M Jakosky, Hugh H Kieffer, Michael C Malin, Harry Y McSween, Kenneth Nealson, Greg L Mehall, Steven H Silverman, Steven Ferry, Michael Caplinger, et al. 2004. The thermal emission imaging system (THEMIS) for the Mars 2001 Odyssey Mission. *Space Science Reviews* 110 (2004), 85–130.
- [9] Danielle DeLatta, Sarah T. Crites, Nicholas Guttenberg, Elizabeth J. Tasker, and Takehisa Yairi. 2019. Segmentation Convolutional Neural Networks for Automatic Crater Detection on Mars. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (2019), 2944–2957.
- [10] Kaichang Di, Wei Li, Zongyu Yue, Yiwei Sun, and Yiliang Liu. 2014. A machine learning approach to crater detection from topographic data. *Advances in Space Research* 54, 11 (2014), 2419–2429.
- [11] JL Dickson, LA Kerber, CI Fassett, and BL Ehlmann. 2018. A global, blended CTX mosaic of Mars with vectorized seam mapping: A new mosaicking pipeline using principles of non-destructive image editing. In *Lunar and planetary science conference*, Vol. 49. Lunar and Planetary Institute The Woodlands TX, USA, 1–2.
- [12] CS Edwards, KJ Nowicki, PR Christensen, J Hill, N Gorelick, and K Murray. 2011. Mosaicking of global planetary image datasets: 1. Techniques and data processing for Thermal Emission Imaging System (THEMIS) multi-spectral data. *Journal of Geophysical Research: Planets* 116, E10 (2011).
- [13] William K Hartmann and Gerhard Neukum. 2001. Cratering chronology and the evolution of Mars. In *Chronology and evolution of Mars: Proceedings of an ISSI Workshop, 10–14 April 2000, Bern, Switzerland*. Springer, 165–194.
- [14] Zhonghua Hong, Ziyang Fan, Ruyan Zhou, Haiyan Pan, Yun Zhang, Yanling Han, Jing Wang, Shuhu Yang, and Yanmin Jin. 2022. Pyramidal Image Segmentation Based on U-Net for Automatic Multiscale Crater Extraction. *Sensors and Materials* 34, 1 (2022), 237–250.
- [15] Pavel Iakubovskii. 2019. Segmentation Models Pytorch. [https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch).
- [16] Boris A Ivanov. 2001. Mars/Moon cratering rate ratio estimates. *Space Science Reviews* 96, 1–4 (2001), 87–104.
- [17] Yutong Jia, Lei Liu, and Chenyang Zhang. 2021. Moon impact crater detection using nested attention mechanism based UNet++. *IEEE Access* 9 (2021), 44107–44116.
- [18] Anthony Lagain, Sylvain Bouley, David Baratoux, Chiara Marmo, François Costard, O Delaa, A Pio Rossi, M Minin, GK Benedix, M Ciocco, et al. 2021. Mars Crater Database: A participative project for the classification of the morphological characteristics of large Martian craters. (2021).
- [19] Anthony Lagain, Konstantinos Servis, GK Benedix, Christopher Norman, Seamus Anderson, and PA Bland. 2021. Model age derivation of large martian impact craters, using automatic crater counting methods. *Earth and Space Science* 8, 2 (2021), e2020EA001598.
- [20] Christopher Lee. 2019. Automated crater detection on Mars using deep learning. *Planetary and Space Science* (2019).
- [21] Michael C Malin, James F Bell III, Bruce A Cantor, Michael A Caplinger, Wendy M Calvin, R Todd Clancy, Kenneth S Edgett, Lawrence Edwards, Robert M Haberle, Philip B James, et al. 2007. Context camera investigation on board the Mars Reconnaissance Orbiter. *Journal of Geophysical Research: Planets* 112, E5 (2007).
- [22] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. 2020. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern recognition* 106 (2020), 107404.
- [23] Stuart J Robbins and Brian M Hynek. 2012. A new global database of Mars impact craters  $\geq 1$  km: 1. Database creation, properties, and parameters. *Journal of Geophysical Research: Planets* 117, E5 (2012).
- [24] Ari Silburt, Mohamad Ali-Dib, Chenchong Zhu, Alan P. Jackson, Diana Valencia, Yevgeni Kissin, Daniel Tamayo, and Kristen Menou. 2018. Lunar crater identification via deep learning. *Icarus* (2018).
- [25] David E Smith, Maria T Zuber, Herbert V Frey, James B Garvin, James W Head, Duane O Muhleman, Gordon H Pettengill, Roger J Phillips, Sean C Solomon, H Jay Zwally, et al. 2001. Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research: Planets* 106, E10 (2001), 23689–23722.
- [26] TF Stepinski, Wei Ding, and R Vilalta. 2012. Detecting impact craters in planetary images using machine learning. In *Intelligent data analysis for real-life applications: Theory and practice*. IGI Global, 146–159.
- [27] Robert G Strom, Steven K Croft, and Nadine G Barlow. 1992. The Martian impact cratering record. *Mars* (1992), 383–423.
- [28] THEMIS Science Team. 2006. THEMIS Day IR Global Mosaic. [https://www.mars.asu.edu/data/thm\\_dir/](https://www.mars.asu.edu/data/thm_dir/) (2006).
- [29] THEMIS Science Team. 2014. THEMIS Day IR Global Mosaic. [https://www.mars.asu.edu/data/thm\\_dir/](https://www.mars.asu.edu/data/thm_dir/) (2014).
- [30] T. A. Vinogradova, M. Burl, and Eric Mjolsness. 2002. Training of a crater detection algorithm for Mars crater imagery. *Proceedings, IEEE Aerospace Conference* 7 (2002), 7–7.
- [31] Libo Wang, Rui Li, Ce Zhang, Shenghui Fang, Chenxi Duan, Xiaoliang Meng, and Peter M Atkinson. 2022. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 190 (2022), 196–214.
- [32] SC Werner and KL Tanaka. 2011. Redefinition of the crater-density and absolute-age boundaries for the chronostratigraphic system of Mars. *Icarus* 215, 2 (2011), 603–607.
- [33] Philipp Georg Wetzler, Rie Honda, Brian L. Enke, William J. Merline, Clark R. Chapman, and Michael C. Burl. 2005. Learning to Detect Small Impact Craters. *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Volume 1* 1 (2005), 178–184.
- [34] Thorsten Wilhelm, Melina Geis, Jens Püttschneider, Timo Sievernich, Tobias Weber, Kay Wohlfarth, and Christian Wöhler. 2020. Domars16k: A diverse dataset for weakly supervised geomorphologic analysis on mars. *Remote Sensing* 12, 23 (2020), 3981.
- [35] Sudong Zang, Lingli Mu, Lina Xian, and Wei Zhang. 2021. Semi-supervised deep learning for lunar crater detection using ce-2 dom. *Remote Sensing* 13, 14 (2021), 2819.

## A SUPPLEMENTARY RESULTS

Tile #	Tile name	Model	Template Threshold	Matched Craters (TP)	Detected Craters (TP + FP)	Craters in range (TP + FN)	Precision	Recall	F1-Score	MaxR
13	thm_dir_N00_030_0_30_30_60_filled.png	U-Net	0.5	1568	4656	2256	0.34	0.70	0.45	138
			0.55	1525	2438	2256	0.63	0.68	0.65	134
			0.6	1465	1878	2256	0.78	0.65	0.71	130
			0.65	1364	1576	2256	0.87	0.60	0.71	128
13	thm_dir_N00_030_0_30_30_60_filled.png	U <sup>2</sup> -Net	0.5	1749	4019	2256	0.44	0.78	0.56	138
			0.55	1754	2102	2256	0.83	0.78	0.80	134
			0.6	1705	1897	2256	0.90	0.76	0.82	129
			0.65	1644	1774	2256	0.93	0.73	0.82	129
13	thm_dir_N00_030_0_30_30_60_filled.png	U-NetFormer	0.5	1825	3895	2256	0.47	0.81	0.59	138
			0.55	1832	2220	2256	0.83	0.81	0.82	137
			0.6	1774	1976	2256	0.90	0.79	0.84	130
			0.65	1728	1877	2256	0.92	0.77	0.84	133
16	thm_dir_N00_120_0_30_120_150_filled.png	U-Net	0.5	470	901	638	0.52	0.74	0.61	99
			0.55	462	618	638	0.75	0.72	0.74	100
			0.6	451	539	638	0.84	0.71	0.77	101
			0.65	442	499	638	0.89	0.69	0.78	102
16	thm_dir_N00_120_0_30_120_150_filled.png	U <sup>2</sup> -Net	0.5	549	937	638	0.59	0.86	0.70	135
			0.55	547	637	638	0.86	0.86	0.86	136
			0.6	546	605	638	0.90	0.85	0.88	103
			0.65	538	579	638	0.93	0.84	0.88	104
16	thm_dir_N00_120_0_30_120_150_filled.png	U-NetFormer	0.5	547	894	638	0.61	0.86	0.71	138
			0.55	546	628	638	0.87	0.85	0.86	102
			0.6	546	604	638	0.90	0.85	0.88	122
			0.65	534	578	638	0.92	0.84	0.88	123
21	thm_dir_N00_270_0_30_-90_-60_filled.png	U-Net	0.5	516	1114	606	0.46	0.85	0.6	138
			0.55	506	737	606	0.69	0.83	0.75	121
			0.6	496	622	606	0.80	0.82	0.81	103
			0.65	476	549	606	0.87	0.79	0.82	103
21	thm_dir_N00_270_0_30_-90_-60_filled.png	U <sup>2</sup> -Net	0.5	574	1014	606	0.57	0.95	0.71	135
			0.55	574	679	606	0.85	0.95	0.89	115
			0.6	569	632	606	0.90	0.94	0.92	110
			0.65	561	599	606	0.94	0.93	0.93	110
21	thm_dir_N00_270_0_30_-90_-60_filled.png	U-NetFormer	0.5	550	750	606	0.73	0.91	0.81	133
			0.55	548	598	606	0.92	0.90	0.91	112
			0.6	544	575	606	0.95	0.90	0.92	97
			0.65	539	562	606	0.96	0.89	0.92	98

**Table 7: Quantative comparison of three models for different template thresholds using tile 13 (test), tile 16(validation), and 21(validation) on binary segmentation. "Tile name" represents the corresponding name of the file given in the dataset by [9] and "MaxR" represents the maximum radius crater found (pixels).**

Object Class	Feature	Total # of samples	Accuracy		F1-Score	
			U <sup>2</sup> -Net	U-NetFormer	U <sup>2</sup> -Net	U-NetFormer
Crater	Crater	1164	95.27	97.16	0.97	0.98
	Crater Field	1342	89.71	93.07	0.94	0.96
Non-crater	Aeolian Curved	632	90.03	86.55	0.95	0.93
	Aeolian Straight	475	97.68	97.05	0.99	0.99
	Channel	112	91.96	88.39	0.96	0.94
	Cliff	266	95.86	80.45	0.98	0.89
	Gullies	493	96.96	98.58	0.98	0.99
	Mass Wasting	263	98.48	95.82	0.99	0.98
	Mixed Terrain	45	88.89	93.33	0.94	0.97
	Mounds	514	80.93	77.82	0.89	0.88
	Ridge	186	95.16	95.70	0.97	0.98
	Rough Terrain	33	93.94	90.91	0.97	0.95
	Slope Streaks	49	100	100	100	100
	Smooth Terrain	88	100	100	100	100
Textured Terrain	23	91.30	91.30	0.95	0.95	

**Table 8: Total number of samples for Crater (positive) and Non-crater (negative) classes and performance in terms of accuracy and F1-Score on DoMars16.**