RESIST: Resource Efficient Satellite Image Segmentation Tool for Curvilinear Structure Segmentation

Charuka Rathnayaka[a], Gowantha Silva[a], Kavinda Pathirana[a], Thanuja Ambegoda[a], Roodra Manogaran[b], Suniti Karunatillake[b]

[a]*Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka*
[b]*Department of Geology and Geophysics, Louisiana State University, Baton Rouge, United States of America*

Abstract

Segmentation of curvilinear structures in satellite images is a crucial task in various domains, including planetary science and agriculture, but is limited by computational resource constraints. To address this challenge, we propose a novel Resource Efficient Satellite Image Segmentation Tool (RESIST) that preserves both global information and pivotal image features. RESIST employs parallel segmentation models with different patch sizes to capture contextual information and fine details. Then it combines their predictions to minimize false detections. Furthermore, RESIST employs postprocessing layers that take advantage of the continuous nature of curvilinear structures to enhance segmentation accuracy. Experimental results on Martian satellite images demonstrate that RESIST outperforms state-of-the-art models in terms of accuracy and performance, making it accessible to researchers with limited computational resources and advancing image segmentation techniques for various applications.

Keywords: Deep Learning, Computer Vision, Satellite Image Segmentation, Semantic Segmentation, Martian Inverted Channel Segmentation, Curvilinear Structure Segmentation, Inverted Channels

1. Introduction

Satellite image segmentation is considered one of the key research areas in both foundational domains like planetary science and industrial domains like

agriculture. To obtain segmentation results with high accuracy, supervised deep neural network-based models have to be used. Satellite imagery data are often found in high resolution. Hence, the processing and training of these satellite images is restricted by computational resource constraints.

To overcome this challenge, downsampling of the images or patch extraction technique is commonly used. However, these approaches have their drawbacks. Downsampling an image results in losing pivotal image feature details, and patch extraction causes a loss of global information. Employing either of these strategies alone will yield low accuracy in the resulting segmentation.

As a solution, we propose a novel Resource Efficient Satellite Image Segmentation Tool (RESIST) that preserves both the global information and the pivotal image features. The complete end-to-end pipeline of this tool was developed for segmenting inverted channels in satellite images of Mars. Topographically inverted channels on Mars are continuous positive relief channels that form sinuous patterns and provide compelling evidence of early fluvial activity [1, 2]. However, manual annotation of these channels is laborious and has a high likelihood of inconsistencies (i.e., low replicability) across researchers. Automating this mapping process can support ongoing explorations in regions such as Aeolis Dorsa [2]. In this context, we presented an abstract at the Lunar and Planetary Science Conference (LPSC) 2023, organized by the Lunar and Planetary Institute and NASA, with a new automated approach for inverted channel segmentation using deep learning
[3].

RESIST employs two parallel segmentation models, using image patches of two different sizes as input. Larger patches capture contextual information, while smaller patches capture fine details. The predicted results from both models are combined to minimize false detections of inverted channels. These patches are then stitched together to reconstruct the original image layout. To validate the curvilinear structure, the bounding boxes are drawn around the identified inverted channels and stretched to detect over- lapping segments. Non-overlapping segments are discarded as they do not preserve the curvilinear structure of inverted channels. Discontinuities be- tween inverted channel segments are connected, and missed interior regions are corrected to obtain the final segmentation results, as shown in Figure 1c. The main contributions of the project are listed below.

1. A context-enhanced segmentation approach, which achieves higher ac-

curacy with minimal computational resource requirements.

2. An extended bounding box overlap approach, a gap-filling layer, and an interior region-filling layer as postprocessing layers, which significantly improve the segmentation accuracy in curvilinear structures.
3. A novel labeled dataset to train and test models for inverted channel segmentation of Martian satellite images.

The segmentation results of RESIST were compared with state-of-the-art segmentation models using three evaluation metrics: F1 score, Jaccard score, and AUC score. The comparison revealed that RESIST outperformed the state-of-the-art models, producing significantly better results. This suggests that RESIST is highly effective in accurately segmenting inverted channels in satellite images of Mars, surpassing existing approaches in terms of segmentation accuracy and performance.

The practical implications of RESIST include its accessibility to researchers with limited computational resources, enabling accurate segmentation results in fields such as planetary science and agriculture. The use of parallel segmentation models with different patch sizes and validation of curvilinear structures using stretched bounding boxes in RESIST has the potential to advance image segmentation techniques, opening doors for applications in remote sensing, geospatial analysis, and computer vision.

2. Related Work

The automation process of annotating inverted channels on Mars using satellite images has received limited exploration. Therefore, there is a lack of specific literature on the automatic segmentation of inverted channels in satellite images of Mars. However, many researchers have carried out work on the detection and segmentation of other geological features within the Martian landscape [4].

Y. Wang et al. have developed an automatic object detection model to identify dark slope streaks on Mars which take the shape of dark thin stripes [4] or fans. Using gradient and regional grayscale information, the regions of interest are identified. Local binary patterns are calculated for the extracted regions. Finally, they are fed into a DDS classifier implemented using the AdaBoost machine learning algorithm.

In the domain of satellite image classification and segmentation, many works have been carried out for the satellite imagery of Earth. Detecting

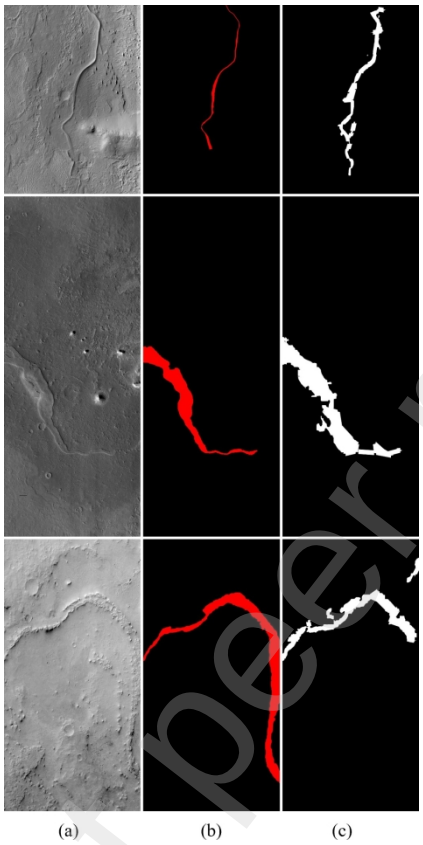


Figure 1: Segmentation results obtained by the model. (a) HiRISE image; (b) Ground truth mask; (c) Final Prediction

water bodies from satellite images is considered to be crucial in improving and managing urban water systems to solve environmental issues and allow timely flood protection planning [5]. Utilization of deep convolutional neural network models has been used for water body detection during the last decade [6]. A novel effective deep convolutional neural network model has been proposed by Kunhao Yuan et al. for water body segmentation using satellite images, which is developed to extract features from multispectral imagery in order to enhance the performance of the process [7].

L. Rubanenko et al. have developed a model using a Mask R-CNN that can automatically segment crescent-shaped sand dunes found on Earth and Mars called barchan dunes [8]. The input satellite images are first converted

into a feature map by feeding it through a backbone neural network. Next, the feature map with the highlighted objects is passed through a region proposal network. Finally, the regions of interest are passed through an ROI align algorithm, and the information is sent through several layers with varying loss functions to segment detected objects.

Since the morphology of inverted channels resembles curvilinear structures, our proposed model can be inspired by work on medical image seg- mentation, especially retinal vessel segmentation. V. Cherukuri et al. have proposed a regularized deep network for vessel segmentation of retinal im- ages [9]. This network domain consists of two parts. First, a representation network is trained to identify curvilinear features in retinal images. Then a task network uses the representation layer features to identify the features at a pixel level. Finally, the filters of the representation layer are optimized using the task network parameters. The authors also propose to implement a multi-scale version of the representation layer using filters that can handle varying thicknesses of retinal vessels.

Among the retinal vessel segmentation models, the U-Net model proposed by Ronneberger et al. introduced the feature map jump connection technique and was able to produce promising results [10]. This proposed U-net model is extensively used in the domain of medical image segmentation. In the advancement of automated medical image segmentation, a substantial number of optimized models based on U-net were later introduced [11]. The top per- forming two state-of-the-art models, basic U-Net [10] and Deform U-Net [12] which is an upgraded version of U-Net, exhibit some noticeable flaws when generating results. These models get confused about the vessel boundary or overlook the retinal vessels around the intersection points due to noise in the images [13]. Liangzhi Li et al have proposed a model named IterNet to overcome this issue, where a standard U-Net uses the raw input images to analyze and map them into a rough segmentation map. Then to optimize the already generated segmentation map, the authors integrate an iteration of mini U-Nets which use the output of the second last layer of its precedent model as its input [13]. The IterNet model has been able to generate better results compared to the previous state-of-the-art models for the commonly used datasets in the field of retinal vessel segmentation [13].

L. Mou et al. proposed one such novel segmentation network (CS2-Net) that consists of a self-attention mechanism in both the encoder and decoder to extract rich hierarchical representations of curvilinear structures [14]. Apart from encoder and decoder modules, the methodology uses a “Channel and

Spatial Attention Module (CSAM)". Features extracted from the input data by the encoder are fed to the CSAM. CSAM generates channel-spatial focus-aware expressive features. To capture more boundary information and segment curvilinear structures, the authors have used a 1 x 3 and 3 x 1 convolutional kernel. The decoder finally reconstructs the curvilinear features to produce segmentation results.

In [15], C. Guo et al. proposed a network named Spatial Attention U-Net. The SA-UNet is considered a lightweight network, and it doesn't require thousands of annotated training images. Data augmentation was done to efficiently use the available annotated images. A spatial attention module was introduced to the basic U-Net. The attention map was inferred across the spatial dimension. Then the attention map was multiplied by the feature map to refine adaptive features. Test results showed that the proposed model outperformed state-of-the-art models like DEU-Net, Vessel-Net, and AG-Net.

S.A Kamran et al. claim that these U-Net based segmentation methods do not perform well in extracting macrovascular structures as they lose resolution in the encoding process and the loss cannot be recovered in the decoding process [16]. Hence, they propose the generative adversarial network, named RV-GAN which uses a pair of generators and a pair of discriminators. Plus, for adversarial training of their model, they introduce a novel weighted feature matching loss with inner and outer weights to combine with reconstruction and hinge loss. The authors have tested the model against a few U-Net based models and a few GAN models using CHASE-DB1, DRIVE, and STARE datasets.

In the field of biomedical image analysis, Ambegoda et al. introduced a novel approach for segmenting neuron membranes in 2D electron microscopy images by incorporating local topological constraints [17]. The proposed method takes pixel-wise membrane probability maps as inputs and formulates the segmentation task as an edge labeling problem on a graph. The authors assert that their proposed method enhances the accuracy of neuron boundary segmentation compared to conventional segmentation approaches by effectively addressing gap completion and minimizing topological errors. When considering works done to handle high-resolution images in segmentation tasks, Y. Wang et al. propose a resource-efficient method for segmenting high-resolution volumetric microCT images [18]. To reduce memory requirements and processing time, the authors use a combination of 3D convolutional neural networks and a novel memory-efficient data sampling strategy. The proposed method involves training a small network on a downsampled version of the data and using it to perform initial segmentation. The segmentation results are then used to train a larger network on a higher- resolution version of the data, enabling the segmentation of the entire image while keeping memory usage low.

This approach proves to be highly beneficial when dealing with extensive image datasets that exceed the memory capacity of a typical computer system equipped with relatively limited GPU and RAM resources.

3. Dataset

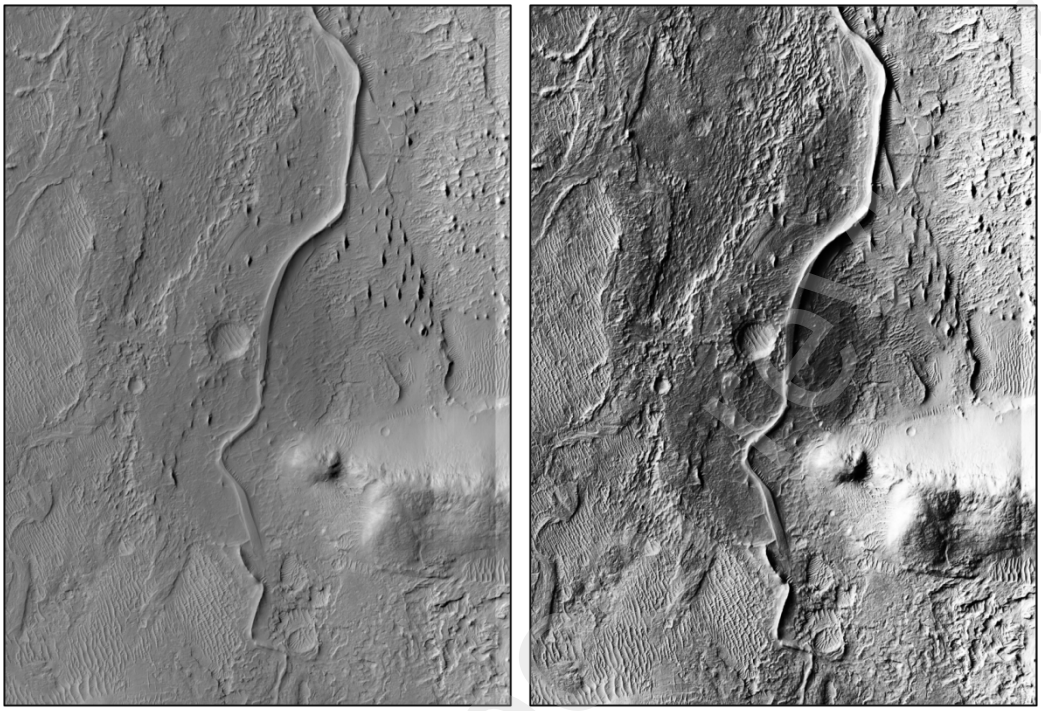
To train the segmentation models used in RESIST, a novel dataset was prepared due to the lack of a labeled dataset of inverted channels in satellite images of Mars. The dataset consists of 23 HiRISE images of the Martian terrain from two different regions; Aeolis Dorsa and Miyamoto Crater. The images are 2048 pixels in width and in the range of 2836-10624 pixels in height. The file size of an image is in the range of 2-10 MB. The dataset was annotated using the cloud-based image annotation platform Dataloop AI [19], creating masks for each image to distinguish the inverted channels as a separate class from the background.

Although the number of original images is small, they contain a wealth of information covering vast areas of Mars. To fully utilize this data, we divided the images into smaller sections for model training. Two subsets were created: one with 256x256 pixel images, resulting in 3664 samples, and another with 512x512 pixel images, producing 924 samples. Training the models on these subsets allowed us to extract valuable features from the terrain and improve the model's ability to generalize effectively. More details on how these subsets were created are provided in the preprocessing section under the methodology.

4. Methodology

4.1. Preprocessing

In this study, we aimed to build a segmentation model capable of accurately identifying the Inverted Channel class in satellite images. To enhance the effectiveness of model training, we applied various data preprocessing strategies and augmentation techniques, ensuring that the available data was used to its fullest potential.



(a) (b)

Figure 2: Effect of histogram normalization. (a) Grayscale HiRISE image; (b) Histogram normalized HiRISE image

First, we applied histogram normalization, which is a common technique used in image processing that involves adjusting the brightness and contrast of an image to improve its visual quality. This technique helps to remove variations in the illumination of the image, resulting in a more consistent and standardized image dataset. We applied histogram normalization to our grayscale images to reduce the impact of lighting variations and enhance the contrast of the images, as shown in (Figure 2).

Furthermore, the original images in the dataset were quite large, making them challenging to handle without access to high-performance computing resources. To address this issue, we divided each original image into smaller image patches of a manageable size. Specifically, we created two separate datasets, each containing image patches of different sizes. The first dataset comprised image patches with dimensions of 256x256, while the second dataset contained image patches with dimensions of 512x512.

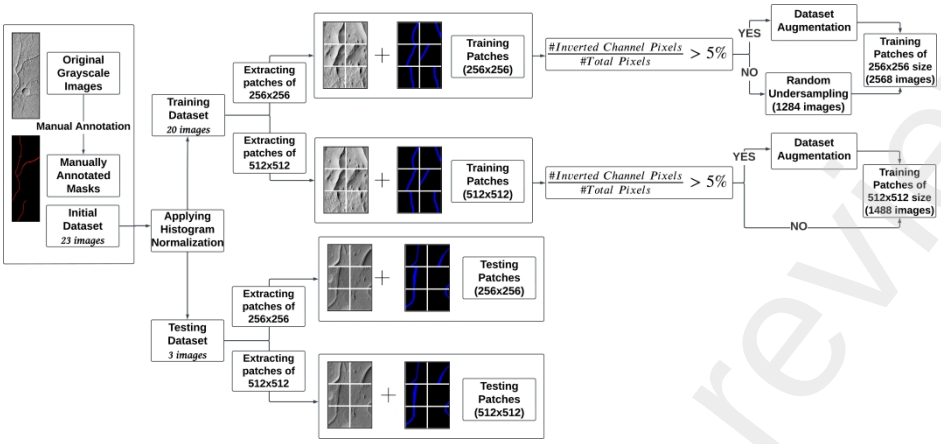


Figure 3: Architecture of the data preprocessing layer.

Our approach was based on context enhanced segmentation, which means that we aimed to incorporate additional information about the surrounding area of each pixel to improve the accuracy of our segmentation results. By generating two separate datasets with different image patch sizes, we could optimize our training process and effectively leverage the available data to achieve robust and accurate segmentation results.

Class imbalance is a common issue in machine learning, and our dataset was no exception. Specifically, there was a significant class imbalance between the Inverted Channel class and the Background class in each satellite image.

To address this issue, we employed a two-part approach to create a balanced dataset of 256x256 image patches. First, we divided the original dataset based on the percentage of Inverted Channel class presence in each image patch. Specifically, we extracted image patches with more than 5% Inverted Channel class presence and applied augmentation techniques such as random rotation and additive Gaussian noise to expand the dataset. Meanwhile, image patches with less than 5% Inverted Channel class presence were randomly undersampled to reduce the imbalance.

After this process, we combined both parts of the dataset to create a final dataset of 256x256 image patches. Notably, the count of image patches with more than 5% Inverted Channel class presence (1284 image patches) was

equal to the number of image patches with less than 5% Inverted Channel class presence (1284 image patches). As a result, the final dataset was not biased towards either class and was suitable for training a robust and accurate segmentation model.

Overall, our approach allowed us to maximize the amount of information available in the data and improve the effectiveness of our segmentation model by addressing the class imbalance issue in our dataset. By creating a balanced dataset, we could train a model that was equally effective at classifying both the Inverted Channel class and the Background class.

Our objective for the 512x512 patch dataset was to train a model that would perform well in real-world scenarios where the Inverted Channel class is significantly underrepresented compared to the Background class in satellite images. However, we also wanted our model to learn all the important features of the Inverted Channel class.

To achieve this objective, we followed a similar approach with the 256x256 dataset. We divided the dataset into two parts based on a 5% threshold for the presence of the Inverted Channel class in each image patch. For the image patches above the threshold, we used more data augmentation techniques such as random rotation, additive Gaussian noise, and flipping to create a larger and more diverse set of training examples. By applying these techniques, we aimed to improve the model's ability to recognize the important features of the Inverted Channel class.

Unlike our approach with the 256x256 dataset, we did not use undersampling in this case. We wanted our model to be trained on a dataset that accurately reflects the class imbalance in real-world scenarios, so we kept all the image patches with less than 5% Inverted Channel class presence. By doing so, we ensured that our model learned to identify the Inverted Channel class even when it was a minority class in the image.

Finally, we combined both parts of the dataset to create a final 512x512 image patch dataset that contained a total of 1488 images. By following this approach, we were able to train a segmentation model that could accurately classify both the Inverted Channel class and the Background class in real-world satellite images, even when the Inverted Channel class was under-represented (Figure 3).

4.2. Context Enhanced Segmentation

The two datasets of different patch sizes were employed to train two separate models while keeping the aforementioned objectives at the forefront

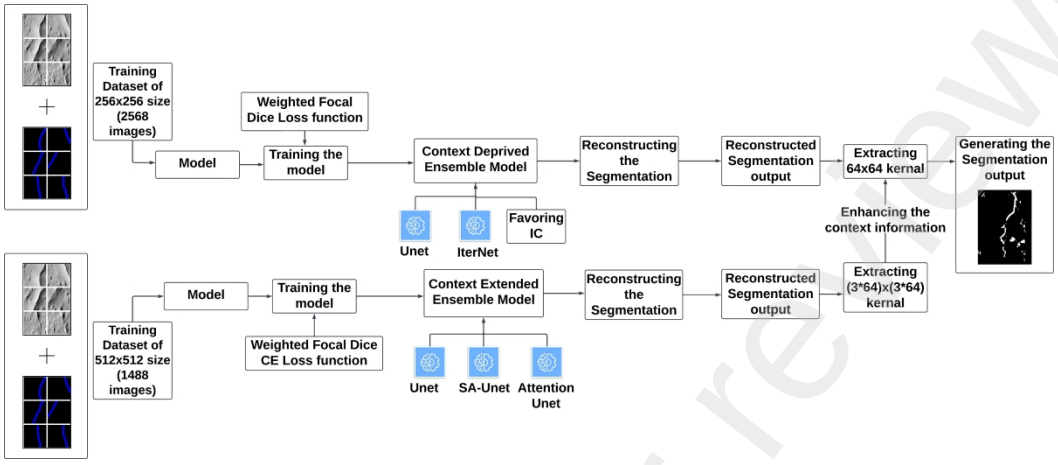


Figure 4: Architecture of the Context Enhanced Model.

of our approach. A *Context Deprived Model* was trained using the 256x256 patch dataset, and a *Context Extended Model* was trained using the 512x512 patch dataset. Figure 4 illustrates the high-level architecture of the *Context Enhanced Model*, offering a comprehensive overview of the system's design and the integration of its components.

4.2.1. Context Deprived Model

The dataset used for this model consists of images that are 256x256 in size. These images have a limited amount of contextual information about the overall satellite image. Additionally, the model was optimized to perform well on the Inverted Channel class.

To construct the model, we employed an ensemble approach by combining two distinct models: U-Net and IterNet. We experimented with several segmentation models and found that U-Net and IterNet consistently outperformed the others in accurately identifying curvilinear structures in Martian satellite images. The IterNet architecture is an extended variation of the U-Net architecture. In a standard U-Net, input images are processed to create an initial segmentation map, which is then refined to improve accuracy. In IterNet, multiple mini U-Nets are used iteratively, with each mini U-Net taking the output of the second-to-last layer of the previous mini U-Net as input and generating a more accurate segmentation map. The ability of these two models to capture fine-grained details and maintain spatial consistency made them the most suitable choices for our task.

For the model training, we used a custom loss function called "Weighted-FocalDiceCELoss" that combines both the Dice Loss and the Focal Loss to address class imbalance issues and improve segmentation performance. The function combines the two losses using a weighted sum, where the Dice Loss is weighted at 70% and the Focal Loss is weighted at 30%.

The Dice Loss[20] is a similarity-based loss function that measures the overlap between the predicted segmentation map and the ground truth segmentation map. It is an effective loss function for segmentation tasks as it is sensitive to both false positives and false negatives. However, it does not directly address the issue of class imbalance in the dataset.

In contrast, the Focal Loss[20] is a variant of the Binary Cross-Entropy Loss that incorporates increased weights for misclassified examples, thereby achieving a more balanced impact of each class on the overall loss function. This is especially important in the case of class-imbalanced datasets, where the model might have a tendency to overfit the majority class.

By combining the two losses, we get the advantages of both. The Dice Loss helps to ensure that the predicted segmentation map is accurate and similar to the ground truth segmentation map. The Focal Loss, on the other hand, helps to balance the influence of each class on the overall loss, making the model less biased towards the majority class. This combination helps to create a more effective training signal for the model and can lead to better segmentation performance, especially when dealing with class-imbalanced datasets.

Following the completion of training, we used a testing dataset of 256x256 patches to evaluate the efficacy of our model. The testing dataset was input into the model, generating predictions for each patch. We specifically sought to prioritize the identification of the Inverted Channel class, by taking the union of the two models: U-Net and IterNet. This ensured that the model would be biased towards predicting inverted channels even when the probability of this class was low.

After generating predictions for all patches, we then stitched them together to reconstruct the segmentation in the original size of the satellite image. The primary aim of this segmentation was to minimize the number of false negatives while simultaneously favoring positive results. By prioritizing the detection of inverted channels, we hoped to improve the accuracy of our model and reduce the risk of missing important features. The architecture diagram of the context deprived model is represented in Figure 5, which portrays the system's underlying components and their interactions,

elucidating the model's design and functionality.

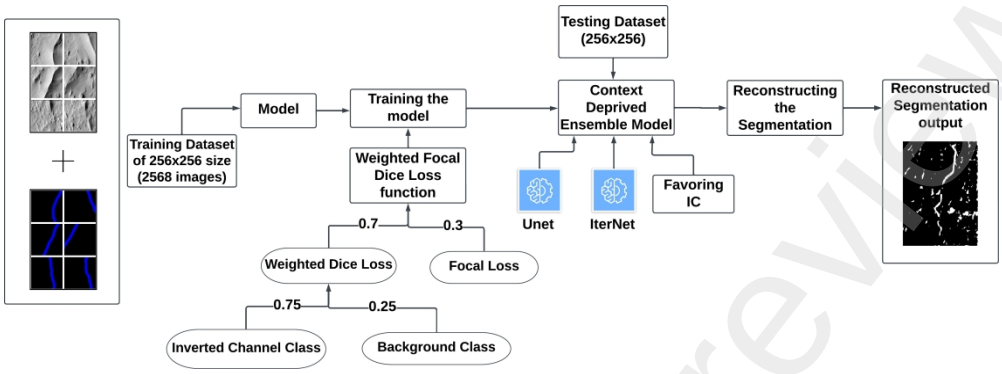


Figure 5: Architecture of the Context Deprived Model.

4.2.2. Context Extended Model

The dataset comprising 512x512 satellite images provided significantly more contextual information than the 256x256 image patches (4x times higher context). Our goal was to build a model that minimizes false positives while maintaining a high true positive rate. This approach enabled the model to accurately predict the fundamental outline of the inverted channels.

For the model, we employed a probability-voted ensemble approach by combining U-Net, SA-UNet, and Attention U-Net. We tested multiple models and selected these three based on their superior performance in segmenting curvilinear structures.

Attention U-Net is an enhanced version of U-Net that integrates attention mechanisms to improve segmentation accuracy. The attention mechanism allows the network to focus on the most relevant regions of the input image, making it particularly effective for segmentation tasks. The architecture consists of an encoder network that extracts features from the input image and a decoder network that reconstructs the segmentation mask. The attention mechanism is applied at multiple levels of the decoder, enabling the net- work to selectively emphasize important features and improve segmentation precision.

The SA-UNet extends the U-Net by adding a spatial attention module to the encoder network. The spatial attention module uses the feature maps from the encoder to learn a set of attention maps that highlight the relevant

regions of the image for segmentation. The attention maps are then used to weigh the feature maps before they are passed to the decoder network.

For the model, we used a custom loss function called "WeightedFocalDiceCELoss" that combines the Dice Loss, Focal Loss, and Cross-Entropy Loss. The Dice Loss component is used to measure the overlap between the predicted segmentation masks and the ground truth masks. The Weighted Dice Loss in this function uses a weighted average of the Dice Loss scores for the foreground (inverted channel) and background (non-inverted channel) classes. This loss helps to ensure that the model produces accurate and precise segmentation.

The Focal Loss component is designed to address class imbalance issues in the dataset. In many segmentation tasks, the foreground object (inverted channel) is much smaller in size compared to the background (non-inverted channel), which can lead to class imbalance. The Focal Loss applies a weighting factor to the loss function that increases the contribution of hard-to-classify examples, thereby improving the model's ability to handle imbalanced datasets.

The Cross-Entropy Loss[20] component is used to penalize misclassifications between the predicted and ground truth masks. This loss measures the difference between the predicted segmentation masks and the ground truth masks using the Cross-Entropy Loss function. This loss is embedded to penalize misclassifications.

The overall loss function is a combination of these three loss components, where the Weighted Dice Loss is given the highest weight (50%), followed by Focal Loss (30%), and Cross-Entropy Loss (20%). The loss function is designed to improve the model's ability to accurately segment the target object while minimizing false positives.

After the completion of the model training, we evaluated its effectiveness on a testing dataset comprising 512x512 image patches. We used the model to generate predictions for each patch, with the specific objective of reducing the false positive rate. After obtaining predictions from the U-Net, SA-UNet, and Attention U-Net models, we employed a probability voting ensemble approach to combine them and generate the final prediction. Subsequently, we stitched together all the predictions for the test patches to reconstruct the segmentation in the original image size. The primary aim of this model was to minimize the number of false positives while maintaining a high true positive rate. Consequently, the model was expected to provide a basic outline of the Inverted Channel class as the prediction. Figure 6 outlines the architectural

diagram of the context extended model.

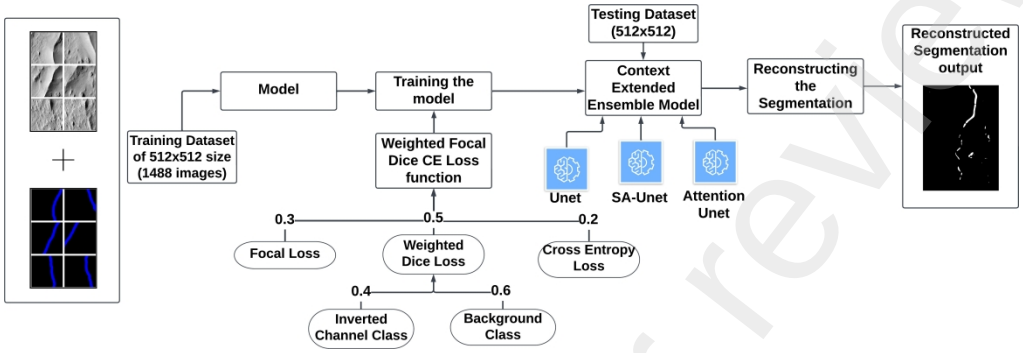


Figure 6: Architecture of the Context Extended Model.

4.2.3. Context Enhanced Model

In our efforts to optimize the performance of our image segmentation model for segmenting inverted channels, we recognized the trade-off between patch size and contextual information. Therefore, we designed two separate models, each trained on image patches of different sizes. The first model was trained on lower-sized patches, which provided less contextual information but proved effective at reducing the false negative rate. On the other hand, the second model was trained on higher-sized patches, providing 4 times the amount of contextual information compared to the first model, and was successful in reducing the false positive rate.

To achieve the best possible results, we employed a context enhanced segmentation technique that combined the strengths of each model while minimizing their individual weaknesses. This allowed us to achieve a higher level of accuracy in segmenting inverted channels while reducing both false positive and false negative rates.

In the context of enhanced segmentation, we aim to improve the accuracy of the segmentation prediction obtained from a satellite image by using both context deprived, and context extended models. The process involves using the prediction of the context deprived model as the base, and the prediction of the context extended model as a guide.

To achieve this, we iteratively select a 64x64 kernel from the prediction of the previously trained context deprived model (6.25% of the trained patch size). For each instance, we mapped it to the context extended model and

extracted an extended kernel ((3x64)x(3x64)) from the prediction of that model.

If the 64x64 kernel from the base has Inverted Channel class pixels inside the selected kernel; we check whether the Inverted Channel class is present in the extended kernel. If it is present in the extended kernel, we do not make any changes to the base prediction. However, in the event that the extended kernel does not include the Inverted Channel class, we proceed to remove the inverted channel-marked pixels within the kernel of the base prediction.

After iterating over the entire image, we can generate a segmentation with fewer false positives while maintaining the true positives in the base prediction. Our approach demonstrates the importance of balancing contextual information and patch size in image segmentation and highlights the benefits of combining multiple models to achieve optimal results.

4.3. Postprocessing

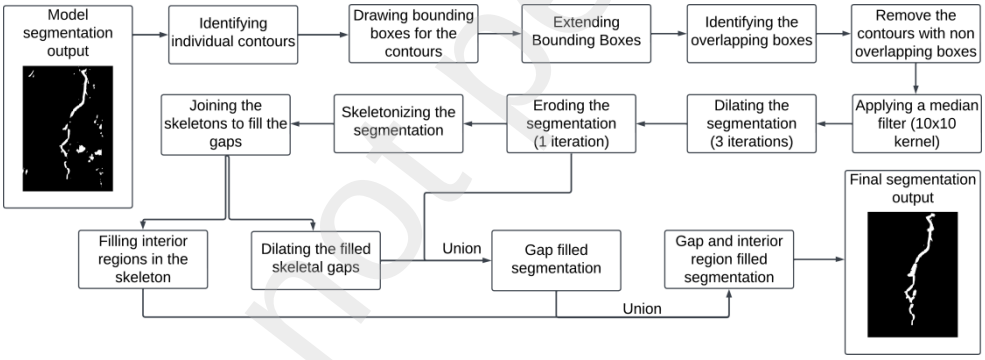


Figure 7: The flow of the postprocessing layer.

The restricted size of our dataset and the presence of ambiguities in identifying the inverted channels posed a significant challenge in our project. To address this issue, we incorporated a series of postprocessing layers to improve the accuracy of the segmentation process. The primary postprocessing layer that we employed was the extended bounding box overlap technique, which helped to refine the segmentation results by improving accuracy. In addition to this, we also used several other postprocessing layers to fine-tune the segmentation results. Figure 7 shows how the proposed postprocessing layer works, highlighting its components and how they work together to

achieve its overall function.

Overall, our approach was effective in enhancing the accuracy of the segmentation process and overcoming the challenges posed by the limited size of our dataset and the ambiguities in identifying inverted channels. This methodology can be extended and applied to other similar projects to overcome similar challenges.

4.3.1. Extended Bounding Box Overlap

In our project, we were tasked with segmenting inverted channels from satellite images. By capitalizing on the curvilinear and continuous nature of inverted channels, we were able to exploit their characteristics to successfully eliminate false positives in the segmentation process using the following approach.

Our approach involved initially identifying the contours predicted as inverted channels and marking their boundaries. Next, we created separate bounding boxes for each individual segment and extended them in the direction of their length. This extension was accomplished by multiplying the bounding box's width by a ratio of the image height to its width, enabling us to extend each bounding box based on its length.

We then selected the longest bounding box and marked it as an inverted channel segment. We repeated this process, identifying additional bounding boxes overlapping with the previously marked inverted channel segment and marking them as inverted channel segments as well. This iterative process continued until no additional bounding boxes were found as inverted channel segments. Finally, we eliminated all pixels within the bounding boxes that were not identified as the Inverted Channel class, while preserving the remaining inverted channel pixels. This technique significantly improved the accuracy of the segmentation by removing a higher number of false positives. By leveraging the unique curvilinear and continuous structure of inverted channels, we were able to develop an effective and efficient approach to accurately segment them from satellite images.

4.3.2. Morphological Operations

To address the issue of false negatives in our model, we integrated several morphological operations into our pipeline. First, we applied a median filter with a 10x10 kernel size immediately following the bounding box overlap layer. This helped to remove noise from the image and smooth out any rough regions, reducing the number of false positives.

Next, we applied a dilation operation with a 5x5 kernel for 3 iterations to expand the inverted channel structure. The dilation operation added pixels to the edges of the true positive regions, making it more likely that these regions would be included in the final output.

Finally, we applied an erode operation with a 5x5 kernel for 1 iteration to refine the boundaries of the predicted regions. The erode operation removed pixels from the edges of the predicted regions, making the boundaries sharper and more accurate.

The combination of these morphological operations was successful in improving the accuracy of our model by reducing the number of false negatives and improving the true positive rate (Table 7). By incorporating these operations into our pipeline, we were able to capture more of the true positive regions that were previously missed by the model, resulting in more reliable and accurate predictions.

4.3.3. Gap Filling Layer

The segmentation output still had discontinuities. To fill these gaps and connect discrete segments in the segmentation output, we used the following approach. First, we skeletonized the segmentation output. This basically thinned down the inverted channel predictions into skeletal segments that are 1 pixel in width. Next, we identified the coordinates of the end pixels of the skeletal segments and joined the discrete skeletal pieces that are less than 200 pixels apart. These newly introduced gap fills were then dilated to create a segmentation output that only consisted of the gaps to be filled. Finally, the union of the earlier segmentation output and the gap fills was taken, to produce an output with fewer discontinuities (Figure 8).

4.3.4. Interior Region Filling Layer

Finally, to remove interior regions of false negatives in the segmentation output, we applied morphological operations to fill in interior pixels of connected components in the gap filled skeleton. The final output was taken as the union of earlier created gap filled segmentation output and the interior region filled skeleton (Figure 9).

5. Results

5.1. Evaluation Metrics

In our project, we chose three metrics to evaluate the segmentation results: F1 score, Jaccard score, and AUC score.

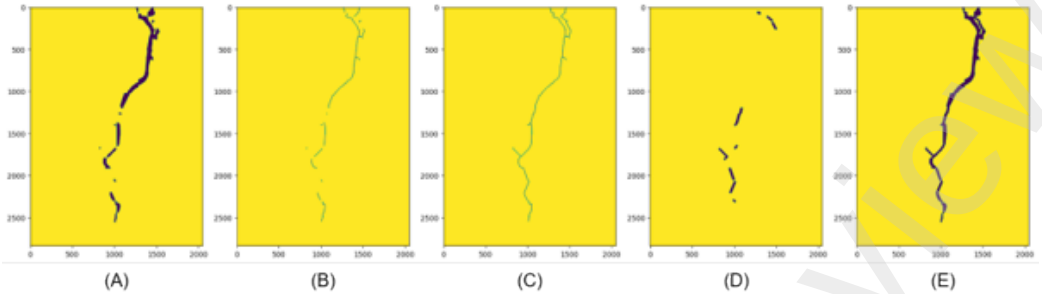


Figure 8: The gap filling layer. (A) Prediction; (B) Skeleton of the prediction; (C) Gap filled skeleton; (D) Dilated gap fillings; (E) Union of the prediction and the dilated gap fillings

F1 score is a metric that combines precision and recall into a single score, which makes it useful for evaluating binary classification tasks like segmentation. Jaccard score, also known as Intersection over Union (IoU), measures the overlap between the predicted segmentation mask and the ground truth mask. AUC score, short for area under the receiver operating characteristic curve, is a metric commonly used in binary classification tasks, that measures the trade-off between true positive rate and false positive rate.

By using multiple metrics to evaluate the segmentation results, we were able to gain a complete understanding of the model's performance. Each metric provided a different perspective on the model's accuracy and helped to identify areas where the model may be struggling.

5.2. Experiments, Results and Observations

This section provides a detailed description of the experimental design, the comparisons, and the results of our proposed computer vision pipeline. All the experiments were performed on a Tesla K80 GPU with 12GB of VRAM and an Intel Xeon E5-2680 v4 CPU with 15GB of RAM.

The first step in our experimental approach involved using state-of-the-art semantic segmentation models, which have demonstrated promising results in the medical domain. Specifically, we trained four models - U-Net, IterNet, SA-UNet, and Attention U-Net - using our initial dataset. We used a Dice Loss function to train our model for image segmentation. To account for class imbalance, we weighted the loss function with a weight of 0.8 for the Inverted Channel class and 0.2 for the Background class. This helps the model learn to better predict the rare, Inverted Channel class.

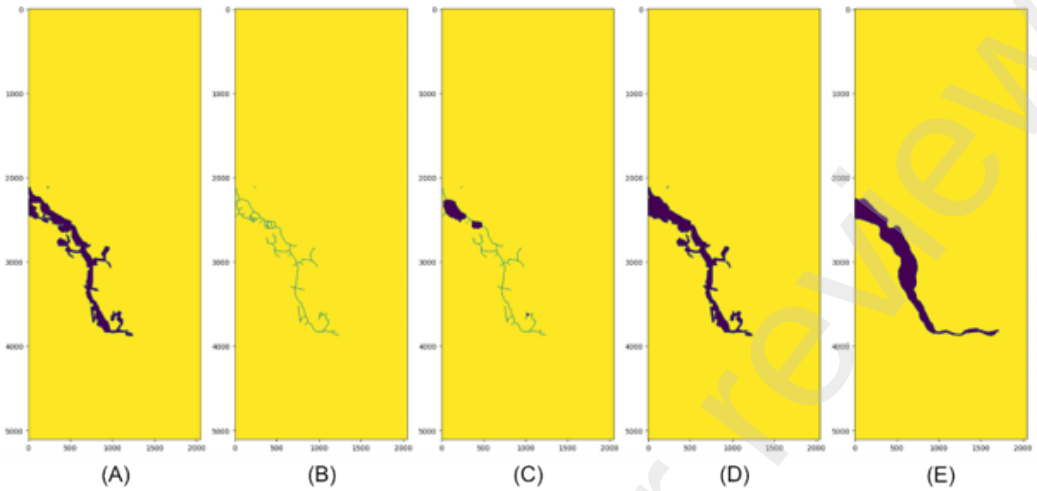


Figure 9: The interior region filling layer. (A) Prediction (contains empty interior regions); (B) Skeleton of the prediction; (C) Interior region filled skeleton; (D) Union of the prediction and the interior region filled skeleton; (E) Ground truth

To ensure that our work could be replicated without the need for high computational power, we fed the complete images after downsampling to train the models on the free tier of Google Colab. This allowed us to efficiently train the models and obtain initial results for further analysis. The results obtained on the downsampled images are presented in Table 1.

Model	Image Size	F1 Score	Jaccard Score	AUC score
U-Net	512x512	0.968023	0.938545	0.540723
U-Net	1024x1024	0.956656	0.917147	0.600646
IterNet	512x512	0.895795	0.811802	0.650264
IterNet	1024x1024	0.936776	0.881296	0.627325
SA-UNet	512x512	0.962754	0.929027	0.580050
SA-UNet	1024x1024	0.958053	0.919859	0.550386
Attention U-Net	512x512	0.962449	0.927915	0.593755
Attention U-Net	1024x1024	0.919556	0.853082	0.640787

Table 1: Performance comparison on downsampled images.

After careful observation of our initial results, we realized that additional preprocessing techniques could be implemented to improve the accuracy of our models. To address this, we converted the HiRISE images to grayscale, as they were originally captured in this format. This conversion helped us

to simplify the input data and reduce the dimensionality of the input space, which in turn reduced the complexity of our models. This approach is particularly useful when dealing with limited computational resources, as we were able to reduce the overall computational burden without sacrificing accuracy. Secondly, we applied histogram normalization as a data preprocessing step. Histogram normalization is a common technique used in image processing that involves adjusting the brightness and contrast of an image to improve its visual quality. This technique helps to remove variations in the illumination of the image, resulting in a more consistent and standardized image dataset. We applied histogram normalization to our grayscale images to reduce the impact of lighting variations and enhance the contrast of the images.

Table 2 demonstrates the results obtained for each of the models mentioned above with data preprocessing steps.

Model	Image Size	F1 Score	Jaccard Score	AUC score
U-Net	512x512	0.936672	0.882253	0.614845
U-Net	1024x1024	0.850239	0.745228	0.666388
IterNet	512x512	0.926448	0.863352	0.638022
IterNet	1024x1024	0.905472	0.881476	0.645879
SA-UNet	512x512	0.922798	0.858314	0.636789
SA-UNet	1024x1024	0.951467	0.907968	0.576847
Attention U-Net	512x512	0.927536	0.866751	0.643732
Attention U-Net	1024x1024	0.948797	0.903757	0.602081

Table 2: Performance comparison on downsampled images with data preprocessing steps.

During the initial experiments, we observed that downsampling the original satellite images into smaller sizes (512x512 or 1024x1024) resulted in the loss of valuable information for the segmentation task. Due to resource limitations, we were unable to train a model in the original size of the satellite images, which would've required significantly more computational resources. To address this issue, we adopted an image patching concept, in which the original satellite images were split into smaller patches that could be fed into the model for training. This approach allowed us to retain more of the original image information, while keeping the computational requirements within reasonable limits. In the testing phase, we similarly split the original test images into patches and used the model to generate segmentation predictions for each patch. These segmented patches were then reconstructed to create the final segmentation mask in the size of the original image.

We experimented with training models with patch sizes of 512x512, 256x256, and 128x128, using the weighted Dice Loss function. By training with different patch sizes, we were able to evaluate the trade-offs between computational efficiency and segmentation accuracy and identify the optimal patch size for our specific task and dataset.

To evaluate the performance of our image patching approach, we conducted experiments using different patch sizes (512x512 and 256x256) on four different models. The results for each combination of model and patch size are summarized in Table 3.

Model	Image Size	F1 Score	Jaccard Score	AUC score
U-Net	512x512	0.950834	0.907154	0.635816
U-Net	256x256	0.946436	0.898592	0.668614
IterNet	512x512	0.961085	0.925794	0.618314
IterNet	256x256	0.943973	0.894079	0.651937
SA-UNet	512x512	0.965424	0.933881	0.559545
SA-UNet	256x256	0.960936	0.9253724	0.601852
Attention U-Net	512x512	0.956512	0.917293	0.650440
Attention U-Net	256x256	0.942418	0.891560	0.634345

Table 3: Performance comparison of models on patched images.

From the above results, we observed that the models trained with smaller image patch sizes, such as 256x256, were able to achieve higher true positive rates compared to the models trained with larger patch sizes (512x512).

The reason for this is that the smaller patch sizes allowed the models to extract fine details from the satellite images, but at the same time, these models had little context information of the overall image, leading them to predict other ambiguous features like ridges and impact craters as inverted channels, resulting in a higher rate of false positives. In contrast, the models trained with larger patch sizes were able to capture more context information of the overall image, but at the cost of lower true positive rates, as they were not able to extract the fine detailed features necessary for accurately identifying inverted channels. However, these models had a relatively lower rate of false positives due to their higher contextual understanding.

To overcome this challenge, we developed a novel context enhanced segmentation approach that combines the strengths of both models. Our approach involved training two models, one with a small patch size and one with a large patch size. We then used the small patch size model to predict the inverted channels within the image patches and used the large patch size

model to provide contextual information to the small patch size model. By integrating the predictions of both models, we were able to achieve significantly higher true positive rates while maintaining a lower false positive rate compared to the models trained with either small or large patch sizes alone. We further optimized two models using two separate loss functions to achieve their respective objectives. The first model used the Weighted Dice Focal Loss function, while the second model used the Weighted Dice Focal CE Loss function.

To evaluate the performance of the two models, we conducted a comprehensive analysis and recorded the results for both context deprived model (Table 4, Figure 10) and context extended model (Table 5, Figure 11).

Model	Image Size	F1 Score	Jaccard Score	AUC score
U-Net	**256x256**	**0.939910**	**0.886814**	**0.730585**
IterNet	**256x256**	**0.916006**	**0.845446**	**0.736352**
SA-UNet	256x256	0.910169	0.835873	0.692126
Attention U-Net	256x256	0.951043	0.907555	0.649751

Table 4: Performance comparison for the context deprived model.

Model	Image Size	F1 Score	Jaccard Score	AUC score
U-Net	**512x512**	**0.958880**	**0.921661**	**0.650927**
IterNet	512x512	0.916566	0.903432	0.678986
SA-UNet	**512x512**	**0.963057**	**0.929398**	**0.607861**
Attention U-Net	**512x512**	**0.953748**	**0.912060**	**0.684585**

Table 5: Performance comparison for the context extended model.

We observed that using a single model is not sufficient to achieve high accuracy, especially for context deprived and context extended images. Therefore, we used an ensemble of models to generate better results. For the context deprived model, we used U-Net and IterNet as base models and manually favored the prediction for the Inverted Channel class. This helped us to generate more accurate results for images with low context (Figure 12). For the context extended model, we combined U-Net, SA-UNet, and Attention U-Net models into a probability voting ensemble model. This approach helped us to improve the accuracy of the model for images with high context (Figure 13). After creating the two ensemble models, we used the context enhanced approach to combine the strengths of each model to generate better

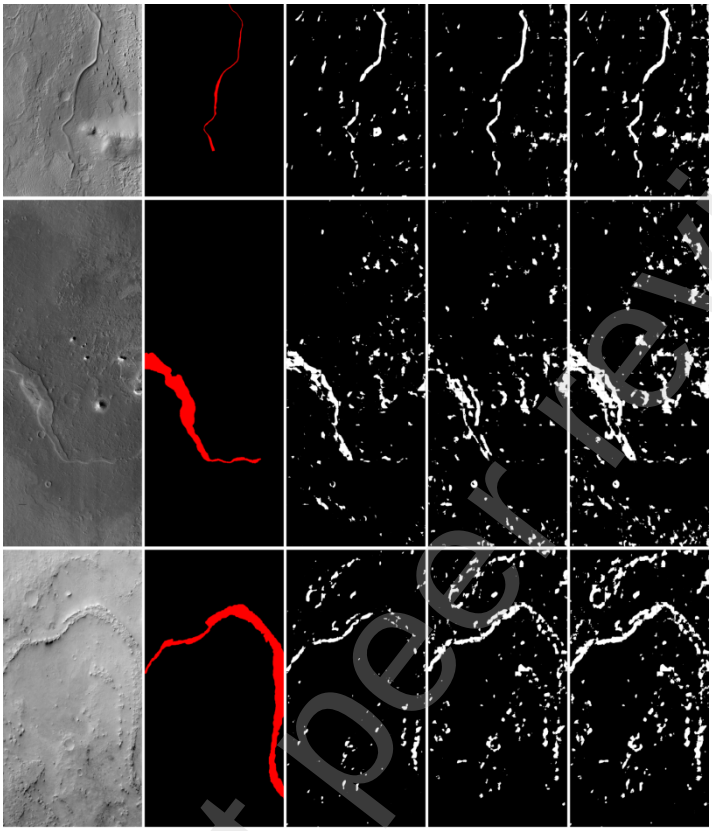


Figure 10: The predictions obtained for the context deprived model. (a) HiRISE image; (b) Ground truth mask; (c) U-Net model prediction; (d) IterNet model prediction; (e) Context deprived model prediction.

results. This approach allowed us to leverage the strengths of each model to overcome their respective weaknesses. Our ensemble approach improved the accuracy of our image segmentation models significantly as shown in Table 6.

Model	F1 Score	Jaccard Score	AUC score
Context Deprived Model	0.922642	0.856652	0.770441
Context Extended Model	0.961424	0.926323	0.628339
Context Enhanced Model	**0.942190**	**0.891185**	**0.767422**

Table 6: Performance comparison for the context enhanced model.

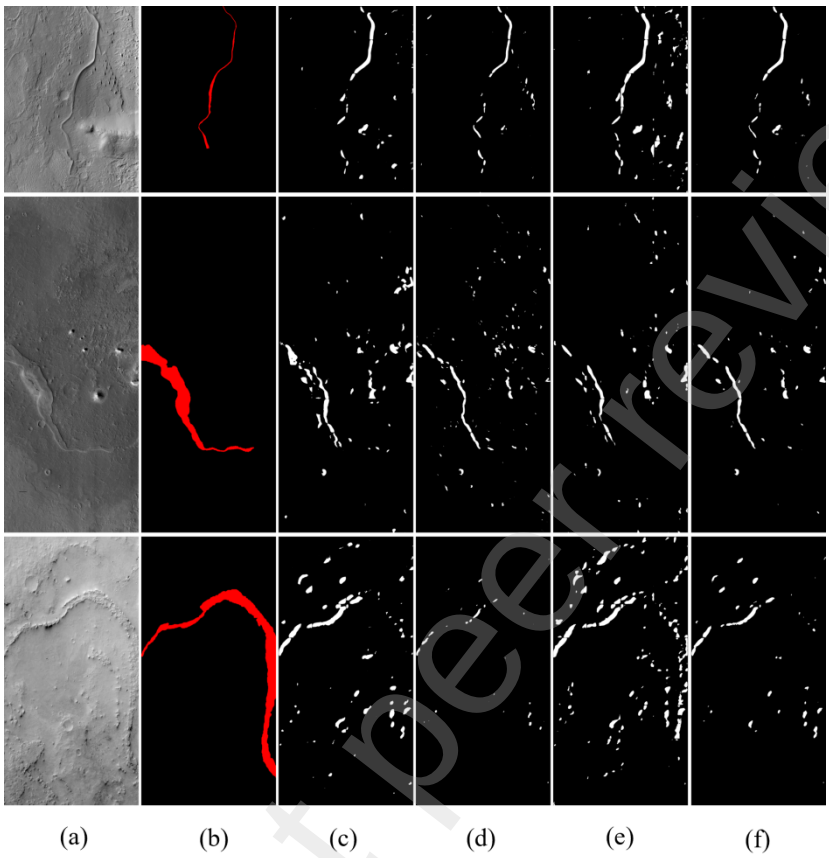


Figure 11: The predictions obtained for the context extended model. (a) HiRISE image; (b) Ground truth mask; (c) U-Net model prediction; (d) SA-UNet model prediction; (e) Attention U-Net model prediction; (f) Context extended model prediction.

After applying the context enhanced model, we noticed that there were still a significant number of closely grouped false positives in the segmentation output. To address this issue, we employed an extended bounding box overlap method. Specifically, we first identified each of the inverted channel segments in the image and drew a bounding box around them. We then extended these bounding boxes and accepted any overlapping boxes as inverted channels.

The extended bounding box overlap method proved highly effective in improving the segmentation results. By incorporating this approach, we were able to significantly reduce the number of false positives and achieve a more

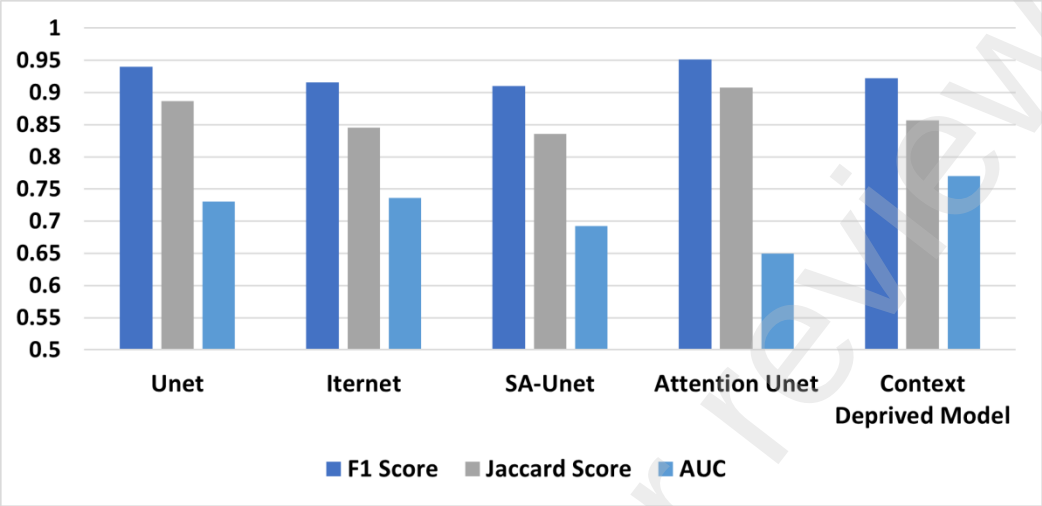


Figure 12: Visualization of performance comparison for the 256x256 dataset trained models

accurate segmentation output. Table 7 and Figure 14f demonstrate the final results generated by our pipeline after integrating the extended Bounding Box Overlap layer, Gap Fill layer and Interior Region Fill layers.

Model	F1 Score	Jaccard Score	AUC score
Context Enhanced Model	0.942190	0.891185	0.767422
Context Enhanced Model with Postprocessing Layers	**0.966907**	**0.944311**	**0.846299**

Table 7: Performance comparison for the final model with the postprocessing layer.

6. Discussion

The proposed novel computer vision pipeline is specifically designed to segment inverted channels from HiRISE images of Mars while minimizing computational resources. To achieve this, we explored two options for processing the satellite images: downsampling and patch-based training. Down- sampling the images resulted in poor segmentation results (Table 2), which prompted us to pursue the second option of dividing the images into smaller patches for training the segmentation models. With this approach, we separately generated predictions for each patch and then reconstructed the entire

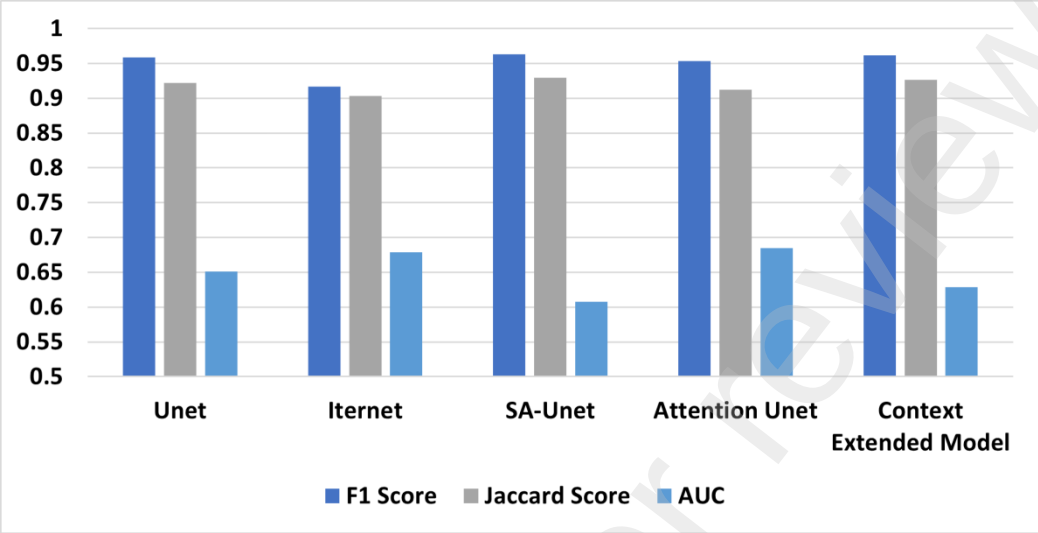


Figure 13: Visualization of performance comparison for the 512x512 dataset trained models

image using these predictions. Although this method required additional processing steps, the results demonstrated significant improvement in segmentation accuracy compared to downsampling (Table 3), thus validating our decision to use patch-based training for our resource-efficient satellite image segmentation tool.

But splitting the high-resolution image into small patches resulted in a loss of crucial contextual information. To address this issue, we propose a context enhanced model that combines two models: a context deprived model trained on small image patches to capture fine details and a context extended model that aims to capture the contextual information of the satellite image. The combined context enhanced model outperformed other state-of-the-art models in terms of accuracy while using the same computational resources (Table 6).

Although the context enhanced model generated satisfactory results, we observed that the segmentation prediction had discontinuities and false positives (Figure 14e). To address this challenge, we developed a novel post- processing layer specifically designed for curvilinear structure segmentation. This layer comprises three components: a bounding box overlap layer, a gap filling layer, and an interior region filling layer. The bounding box overlap layer takes the continuous and curvilinear nature of the structures into ac-

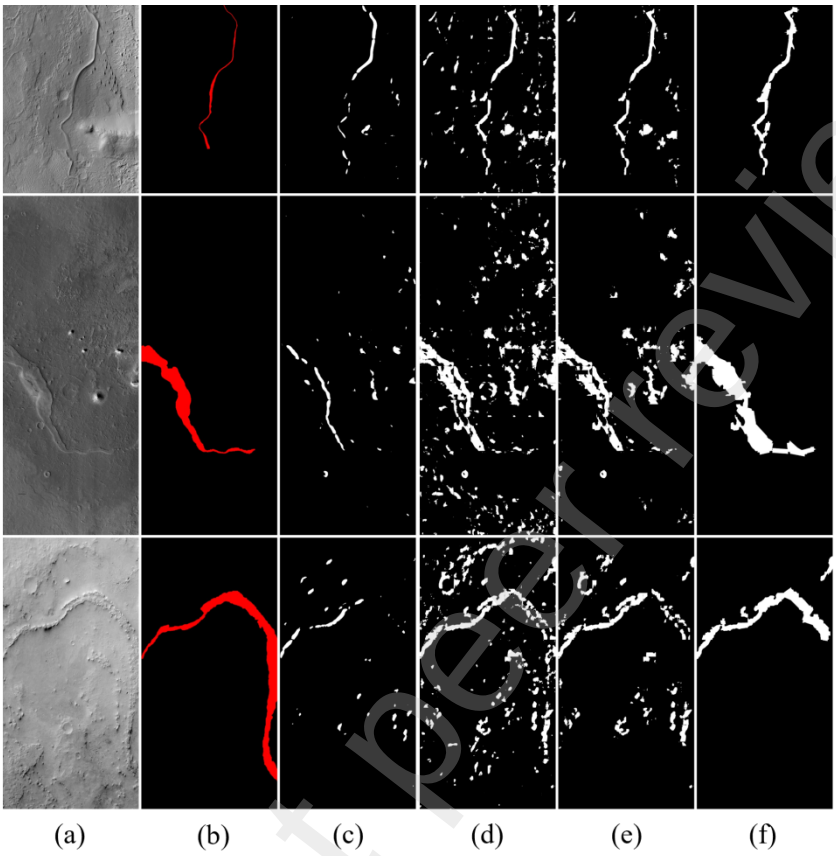


Figure 14: Visualization of the segmentation results obtained for each model. (a) HiRISE image; (b) Ground truth mask; (c) Context extended model prediction; (d) Context deprived model prediction; (e) Context enhanced model prediction; (f) Final prediction.

count and draws extended bounding boxes around the predicted inverted channel components. It iteratively selects the overlapping Inverted Channel class components and significantly reduces the false positives in the segmentation prediction. The gap filling layer and the interior region filling layer focus on the continuous nature of the inverted channel and remove false negatives to improve the segmentation accuracy.

Our proposed computer vision pipeline shows promising results for Mar- tian inverted channel segmentation and outperforms other state-of-the-art models (Figure 15). We believe that our approach can be applied to other curvilinear structure segmentation tasks and can contribute to the development of resource-efficient computer vision pipelines for remote sensing applications.

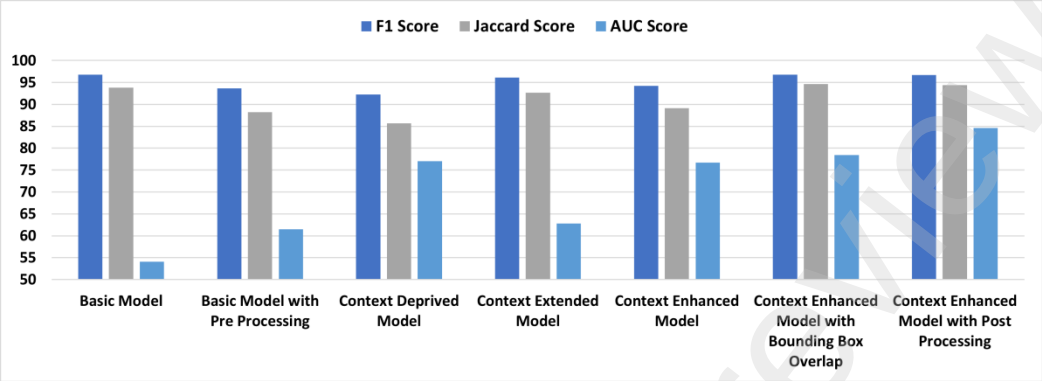


Figure 15: Comparison of Performance: A Visualization.

7. Conclusion

In this paper, we proposed a novel context enhanced computer vision pipeline for segmenting curvilinear structures from high resolution satellite images, specifically inverted channels on the Martian surface. By splitting the images into small patches and using a combination of context deprived and context extended models, our pipeline was able to outperform other state-of-the-art models in terms of segmentation accuracy while consuming fewer computational resources. The postprocessing layer implemented in our pipeline further improved segmentation accuracy by reducing noise and filling gaps in the segmented structures. The proposed computer vision pipeline was able to generate promising results by improving the accuracy and efficiency of curvilinear structure segmentation from high resolution satellite images and can also be used to segment similar curvilinear structures on Earth or other planets.

References

[1] J. Davis, M. Balme, P. Grindrod, R. Williams, and S. Gupta, "Extensive Noachian fluvial systems in Arabia Terra: Implications for early Martian climate," *Geology*, vol. 44, no. 10, pp. 847–850, 2016. https://doi.org/10.1130/78247.1.

[2] A. Lefort, D. M. Burr, R. A. Beyer, and A. D. Howard, "Inverted fluvial features in the Aeolis-Zephyria Plana, western Medusae Fossae Formation, Mars: Evidence for post-formation modification," *J. Geophys. Res. Planets*, vol. 117, no. E3, Mar. 2012. https://doi.org/10.1029/2011je004008.

[3] K.P.G. Pathirana, C.B. Rathnayaka, W.G.C. Silva, T.D. Ambegoda, R. Manogaran, and S. Karunatillake, "RESIST: Tool to Automatically Segment Martian Inverted Channels in HiRISE Images," *Lunar Planet. Sci. Conf.*, 2023. https://www.hou.usra.edu/meetings/lpsc2023/pdf/1821.pdf.

[4] Y. Wang, K. Di, X. Xin, and W. Wan, "Automatic detection of Martian dark slope streaks by machine learning using HiRISE images," *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 12–20, 2017. https://doi.org/10.1016/j.isprsjprs.2017.04.014.

[5] Z. Shao, H. Fu, D. Li, O. Altan, and T. Cheng, "Automatic Water-Body Segmentation From High-Resolution Satellite Images via Deep Networks," *Remote Sens. Environ.*, vol. 232, p. 111338, 2019. https://doi.org/10.1016/j.rse.2019.111338.

[6] Z. Miao, K. Fu, H. Sun, X. Sun, and M. Yan, "Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 602– 606, Apr. 2018. https://doi.org/10.1109/LGRS.2018.2794545.

[7] K. Yuan, X. Zhuang, G. Schaefer, J. Feng, L. Guan, and H. Fang, "Deep-Learning-Based Multispectral Satellite Image Segmentation for Water Body Detection," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 7422–7434, 2021. https://doi.org/10.1109/JSTARS.2021.3098678.

[8] L. Rubanenko, S. Pérez-López, J. Schull, and M. G. A. Lapôtre, "Automatic Detection and Segmentation of Barchan Dunes on Mars and Earth Using a Convolutional Neural Network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 9364–9371, 2021. https://doi.org/10.1109/JSTARS.2021.3109900.

[9] V. Cherukuri, V. K. B.G., R. Bala, and V. Monga, "Multi-Scale Regularized Deep Network for Retinal Vessel Segmentation," *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 824–828, 2019. https://doi.org/10. 1109/ICIP.2019.8803762.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lect. Notes Comput. Sci.*, pp. 234–241, 2015.

[11] X. Guo et al., "Retinal Vessel Segmentation Combined With Generative Adversarial Networks and Dense U-Net," *IEEE Access*, vol. 8, pp. 194551–194560, 2020. https://doi.org/10.1109/ACCESS.2020. 3033273.

[12] Q. Jin, Z. Meng, T. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, 2019. https://doi.org/10.1016/j.knosys. 2019.04.025.

[13] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal Image Segmentation Utilizing Structural Redundancy in Vessel Networks," *Proc. IEEE Winter Conf. Appl. Com- put. Vis. (WACV)*, pp. 3645–3654, 2020. https://doi.org/10.1109/ WACV45572.2020.9093621.

[14] L. Mou et al., "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging," *Med. Image Anal.*, vol. 67, 2021. https: //doi.org/10.1016/j.media.2020.101874.

[15] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation," *Proc. Int. Conf. Pattern Recognit. (ICPR)*, pp. 1236–1242, 2021. https:// doi.org/10.1109/ICPR48806.2021.9413346.

[16] S. Kamran, K. F. Hossain, A. Tavakkoli, S. Lee Zuckerbrod, K. Sanders, and S. Baker, "RV-GAN: Segmenting Retinal Vascular Structure in Fundus Photographs using a Novel Multi-scale Generative Adversarial Network," 2021. https://arxiv.org/pdf/2101.00535v2.pdf.

[17] T. D. Ambegoda and M. Cook, "Efficient 2D neuron boundary segmentation with local topological constraints," *arXiv preprint arXiv:2002.01036*, 2020.

[18] Y. Wang, L. Blackie, I. Miguel-Aliaga, and W. Bai, "Memory-efficient Segmentation of High-resolution Volumetric MicroCT Images," *Proc. Int. Conf. Med. Imaging with Deep Learning*, PMLR 172, pp. 1322–1335, 2022.

[19] Dataloop, "Dataloop," Online, 2021. https://dataloop.ai.

[20] S. Jadon, "A survey of loss functions for semantic segmentation," *IEEE Xplore*, Oct. 01, 2020.