Pradeep Kumar Baisakh Paper check

By Pradeep Kumar Baisakh

Automating The Process of Linear Feature Extraction of Europa using Machine Learning

[1]Pradeep Kumar Baisakh
School of Physical Sciences
National Institute of Science Education and Research, Bhubaneswar
pradeepkumar.baisakh@niser.ac.in
[1]Anshuman Panda
School of Mathematical Sciences
National Institute of Science Education and Research, Bhubaneswar
anshuman.panda@niser.ac.in
Dr.Subhankar Mishra, Dr.Guneshwar Thangjam

April 20, 2024

Abstract

Europa, Jupiter's icy moon, captivates with its unique geology, hosting a vast ocean beneath its icy exterior, sustained by Jupiter's tidal forces. The red scars, Lineas, add intrigue, their crimson hue sparking questions about Europa's geophysical processes. Discovering this ocean challenges traditional habitable zone concepts, expanding the search for life beyond star-centric boundaries. The project focuses on Lineas, aiming to make a linea detection model using machine learning. Here we tried building a CNN based segmentation model which resulted with training IoU 32%, testing IoU 29.6% and an f1 Score of 53.33%.

1 Introduction

Europa, Jupiter's intriguing icy moon, defies expectations with a hidden ocean beneath its thick icy crust, sustained by Jupiter's tidal forces. Despite its desolate appearance, Europa's potential for life challenges conventional habitable zone theories. The red scars, known as Lineas, add mystery to its surface, formed by a blend of tidal forces and asynchronous rotation. The project aims to explore Lineas, unravel their mysteries, and analyze salts, hoping to discover organic compounds. The detection of hydrated salts fuels anticipation for potential organic findings on Europa, offering a glimpse into the broader possibilities of habitable icy moons beyond our solar system.

Machine learning methods could be utilized in the scientific exploration of celestial bodies such as Europa, one of Jupiter's moons. These techniques aid in the automated detection of distinctive surface features, such as "lineas". By analyzing image data with

ML algorithms, researchers can uncover insights into Europa's geological characteristics and potential for sustaining life.

Objective: To build an image segmentation model that can automate the process of linea detection in satellite images.

Motivation : The ultimate goal of this project is to do the spectral analysis of the ocean of europa to find the presence of any prebiotic molecule or some useful salts thatcould lead to prebiotic compounds. An accurate detection model can be integrated with the available spectral data for analysis of only the linea regions. Since, only through the lineas or cracks in the ice we can observe the ocean beneath, which is our focus. This paper presents a CNN based architecture as a segmentation model for binary classification, where the classes are linea or no linea.

2   The Dataset:

The images were sourced from Galileo's Solid State Imaging instrument and are publicly accessible through NASA via PDS-atlas.Here is the link for the unlabelled dataset. The images were previously labeled by some researchers from the University of Bern[4]. There were 1,023 images of size 224×224 and masks(labeled images) were also provided by them. Some example photos and their masks are given below in fig.1[1].

All the codes were done in Python.
The images and masks were read using the cv2 package and stored as NumPy arrays. The data was then shuffled to avoid any inductive biases if present and make the data more randomized for training. Then the entire dataset was split into 75% and 25% for training and testing respectively.

3   The Model

We built a CNN-based model using Pytorch. The data was passed through six convolutional layers. Each convolutional layer is basically a filter matrix that convovles across the entire image and its output acts as input for the next layer and the features are extracted in this method.The convolutional layers are followed by batch normalization then followed by six fully connected layers, which were passed through ReLu activation functions, except for the last layer, which was passed through a softmax activation function. The softmax function gives the probability of the presence of linea at a data point.

3.1   Training and Evaluation

The entire process was performed using Nvidia RTX 2080Ti GPUs.
The model is trained using the extracted training data and PyTorch's DataLoader. The data was trained over 16 epochs, iterating through the training data in batches. The loss was calculated using Binary Cross Entropy loss[1]. BCELoss penalizes the model based on how much the probability is deviating from ground truth. For higher probability it

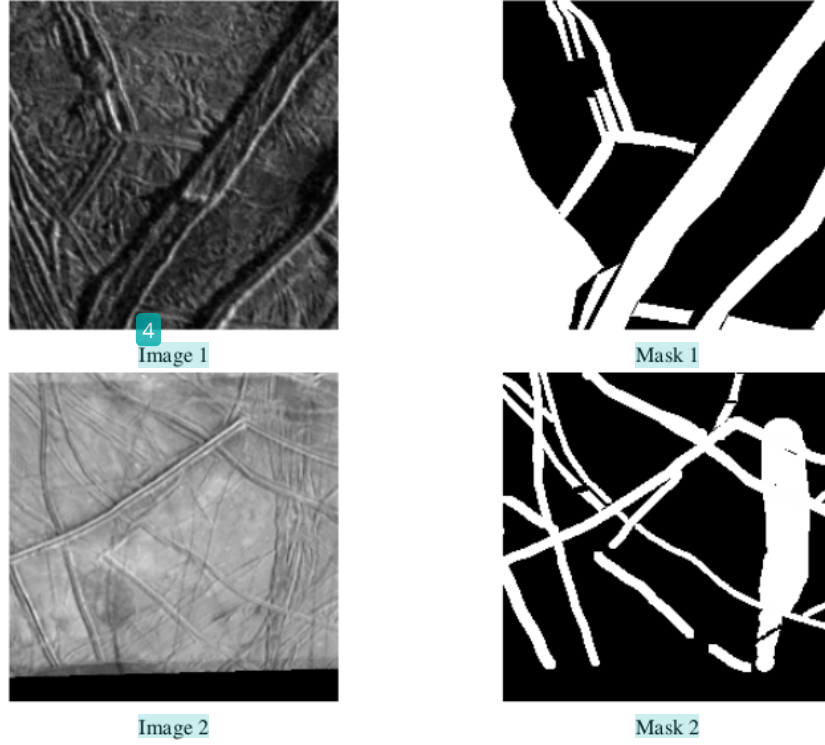|	
Image 1	Mask 1
Image 2	Mask 2

Figure 1: Example images and their masks from the dataset

penalizes less while for lower probablity it penalizes the model more. Then the model tries to minimize the loss while training.

$$BCELoss(p,y) = -\frac{2}{N}\sum_{i=1}^{N}[y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)] \qquad (1)$$

where,
y_i = actual label for the sample
p_i = predicted probability that sample belongs to linea class

The model was evaluated based on the Intersection over Union (IoU2) metric, which basically gives us the measure of spatial overlap of the predicted image and mask. It is defined as the intersection of the predicted image and the mask divided by the unionof the predicted image and the mask. The idea is to reward True Postiives, but punish both false positives and false negatives.

$$IoU = \frac{(Predicted \cap Mask)}{(Predicted \cup Mask)} \qquad (2)$$
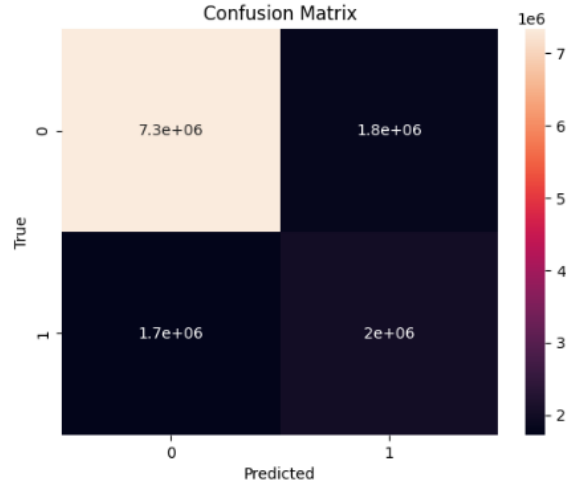
3

Figure 2: Confusion Matrix

Training IoU = 32%   **Testing IoU = 29.6%**
The confusion matrix was also calculated and given below in fig.2. From the confusion matrix,
Precision = 52.63%   **Recall = 54.05%**
f1 Score = 53.33%

3.2   Predicting a new image

An unseen image was loaded through the trained model for prediction. The output prediction is threshold to obtain binary mask predictions. An example prediction is provided in fig.3. You can observe many similarties in the prediction and ground truth image, but this is not good enough to be used for scientific purposes.

4   Conclusion

The results of this model are not good enough for it to be usable. The dataset is really old; the Galileo mission was launched in the year 1989, and the next mission by NASA is planned to be launched this year in October, which will reach Europe around 2030. So we will have to wait for the dataset then. Since the dataset was labeled manually, some form of active learning could also be used to label the dataset.
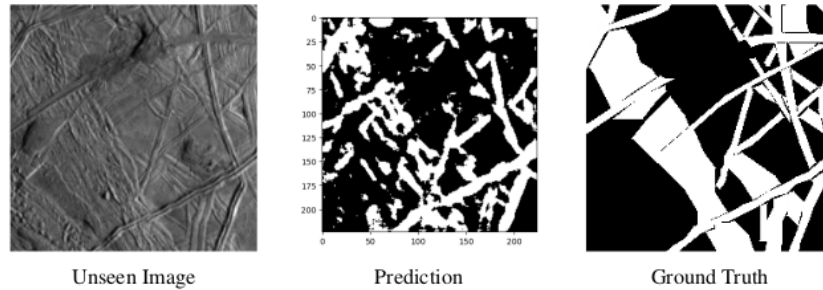
Unseen Image	Prediction	Ground Truth

Figure 3: Comparision between prediction of an unseen image and ground truth

References

[1]Haslebacher, C., Thomas, N., and Bickel, V. T., "LineaMapper: A deep learning-powered tool for mapping linear surface features on Europa",Icarus, vol. 410, 2024. doi:10.1016 j.icarus.2023.115722.

[2] Das, N. P., Nazareth, R. G., Mishra, S., and Thangjam, G. S., "Automated Detection of Lineaments on the Surface of Europa Using Machine Learning", in <i>53rd Lunar and Planetary Science Conference</i>, 2022, vol. 2678.

[3] Cao, G. , Wang, S. , Wei, B. , Yin, Y. , Yang, G.. (2013). A hybrid Cnn-Rf method for electron microscopy images segmentation. Journal of Biomimetics, Biomaterials, and Tissue Engineering. 18. 10.4172/1662-100X.1000114.

Pradeep Kumar Baisakh Paper check