Università degli studi di Milano-Bicocca

Signal and Imaging Acquisition and Modelling in Environment

Final Project

Forest segementation on Sentinel-2 images trained on an external dataset

Author:
Jan Fiszer - 897193 - j.fiszer@campus.unimib.it

June 3, 2023

Abstract

Remote sensing has witnessed significant advancements in the past decade thanks to the rapid development of machine and deep learning. Data captured by satellites tend to be enormous and manual analysis is very tedious and time-consuming. Therefore, AI use becomes beneficial. Once a satellite is launched, it provides valuable images without any financial or time contribution, so there is no concern about the amount of data to train a model. A popular use of earth images is observing environmental changes such as deforestation. While the widely recognized deforestation in the Amazon rainforests would be easily detectable, this study focuses on a more localized and patriotic perspective—the Puszcza Karpacka region in southern Poland. U-Net [1] model was trained on an external dataset with the further aim of observing forest changes between 2017 and 2022 through the application of this model on Sentinel-2 images. However, despite efforts, the expected results were not achieved. The challenges encountered and potential solutions are extensively discussed within the paper, offering valuable insights for future research and development in this domain. The code is available at: https://github.com/janfiszer/deforestation-in-Poland.

1 Introduction

Deforestation has emerged as a critical issue in the current era, contributing significantly to the climate crisis. While the clear-cutting of vast rainforest expanses garners attention, there exist covert methods that make deforestation less detectable, such as the substitution of old trees for young ones or the selective felling of individual trees, avoiding satellite surveillance. Although these practices have a comparatively lesser impact than large-scale deforestation, they still disrupt wildlife and can pose substantial challenges to local ecosystems if broadly exploited.

By training the U-Net model to perform binary image segmentation on a Kaggle dataset, and further use on Sentinel-2 images, it was aimed to verify if the changes are visible in the Puszcza Karpacka region. Then by comparing the difference in model outputs, it would be possible to observe the forests' changes.

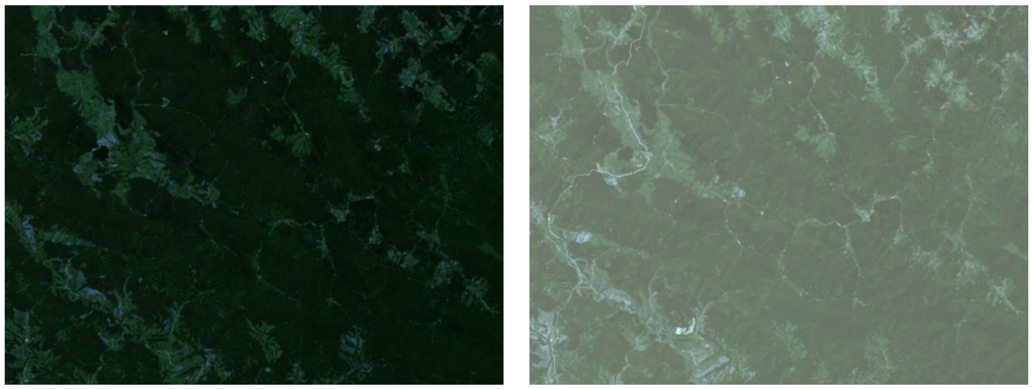
Figure 1: Two sample images from the training dataset with corresponding grand truth masks. In the second example, we can see that some of them were created inaccurately.

2 Datasets

2.1 Training dataset

The training dataset utilized in this study, available on Kaggle, is a subset of a larger dataset described in [2]. It comprises 5108 aerial images of dimensions 256x256 pixels, with a resolution of 0.5 meters. These images were captured in Thailand, Indonesia, and India. Each image in the dataset is accompanied by its corresponding ground truth masks, as depicted in Figure 1.

During exploratory data analysis, it was observed that many masks of the images are very non-accurate (see Figure 1 and 4).



(a) July of 2017

(b) June of 2022

Figure 2: Two Sentinel-2 images of the Puszcza Karpacka region from different years.

2.2 Sentinel-2 Images

A total of four Sentinel-2 images were downloaded for this study, with two images captured in July 2017 and two images in June 2022 (see Figure 2). The selected area of observation spans across two Sentinel-2 tiles, requiring two images for each period. To ensure consistency in vegetation stages and forest conditions, images from the same season were chosen. Additionally, both tiles were subject to a maximum cloud coverage limit of 3%, with no observed cloud coverage within the region of interest. All of them have 10m resolution, which occurred to be a significant issue (better described in the section 4). Since only four images were required, they were obtained by browser editor provided by Google Earth Engine.

3 U-Net training

3.1 Data preparation

To facilitate data handling and memory management, the DataGenerator class was implemented, using TensorFlow datasets objects to load the dataset in batches. In order to align the image resolution with the Sentinel-2 imagery (10 meters), a progressive resizing approach was employed. Initially, the images were resized to 128x128, followed by further downsampling to 64x64.

However, it was observed that the resolution of 32x32 was insufficient to accurately distinguish forest areas for the model. Therefore, the final image size was set at 64x64, resulting in a resolution of 2 meters.

The 5108 images were divided into 4086 training and 1022 validation images (80:20 split).

3.2 Model architecture

The model consists of an encoder-decoder architecture inspired by U-Net [1], which proved to perform very well in image segmentation tasks. The encoder consists of a series of downsampling blocks, each consisting of two convolutional layers followed by max pooling and dropout. Convolutional layers of the first downsampling block have 32 filters, and subsequently, each subsequent downsampling block doubles the number of filters compared to the previous block.

The decoder part of the architecture comprises upsampling blocks, where each block consists of a transposed convolutional layer for upsampling, concatenation with the corresponding feature map from the encoder(skip connections), dropout regularization, and two convolutional layers. It has a symmetrical schema as the encoder also ends with 32 filters, followed by a sigmoid layer with one filter and 1x1 kernels to perform binary classification.

It is worth mentioning that skip connections are a crucial innovation of the U-Net that aids gradient flow and prevents its vanishing (for more details [1]).

3.3 Training

3.3.1 Hyperparameters

As stated in the previous section (Section 3.1), the input images were resized to a size of 64x64 pixels. Due to the limited memory capacity of a laptop, a batch size of 4 was selected during the training phase. The model underwent training for a total of 24 epochs, but it was observed that overfitting occurred during this process, as shown in Figure 3. As a result, the final model was trained for only 12 epochs. For more detailed information about initial parameters, please refer to the config.py file available in the GitHub repository.

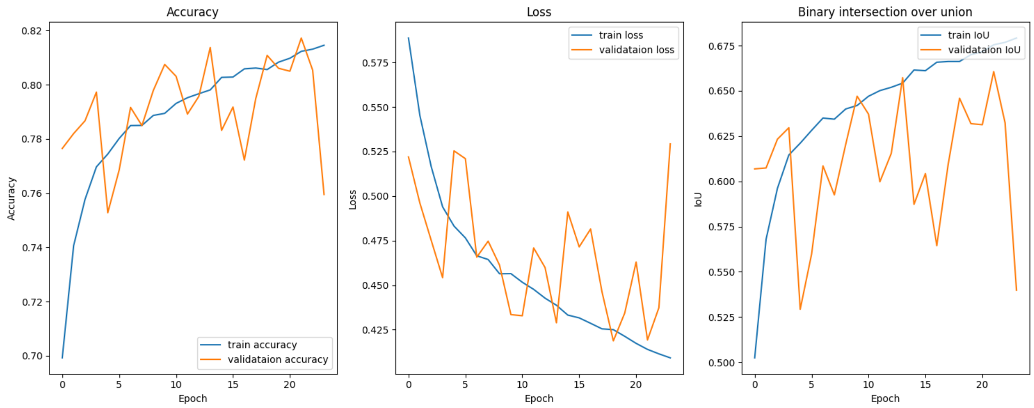


Figure 3: Learning curves of the model trained for 24 epochs, where the overfitting is visible. Thus, the model used for further predictions was trained just by 12 epochs.

3.3.2 Optimizer and loss funtion

The optimization algorithm utilized in this study was Adam [3]. It was employed with an initial learning rate of 0.001. For the task of binary image segmentation, the loss function employed was binary cross-entropy.

3.3.3 Metircs

In addition to the conventional metric of accuracy, the performance of the model was assessed using the binary intersection over union (IoU) metric, which is considered more appropriate for evaluating image segmentation tasks. IoU provides a measure of the overlap between the predicted and ground truth segmentation masks, allowing for a more comprehensive evaluation of the model's ability to accurately identify and delineate forest areas.

3.4 Results

After employing various approaches, the final model achieved the following results on the validation set:

```
accuracy=0.79,
binary_iou=0.63,
```

Although the obtained metrics values may not appear to be perfect, it was crucial to avoid overfitting the data, considering the imperfections in the ground truth masks. In fact, a slight underfitting was intentional, which indeed resulted in better outcomes compared to the original masks, as demonstrated in Figure 4. This highlights the model's ability to generalize well and improve upon the limitations of the provided ground truth masks.

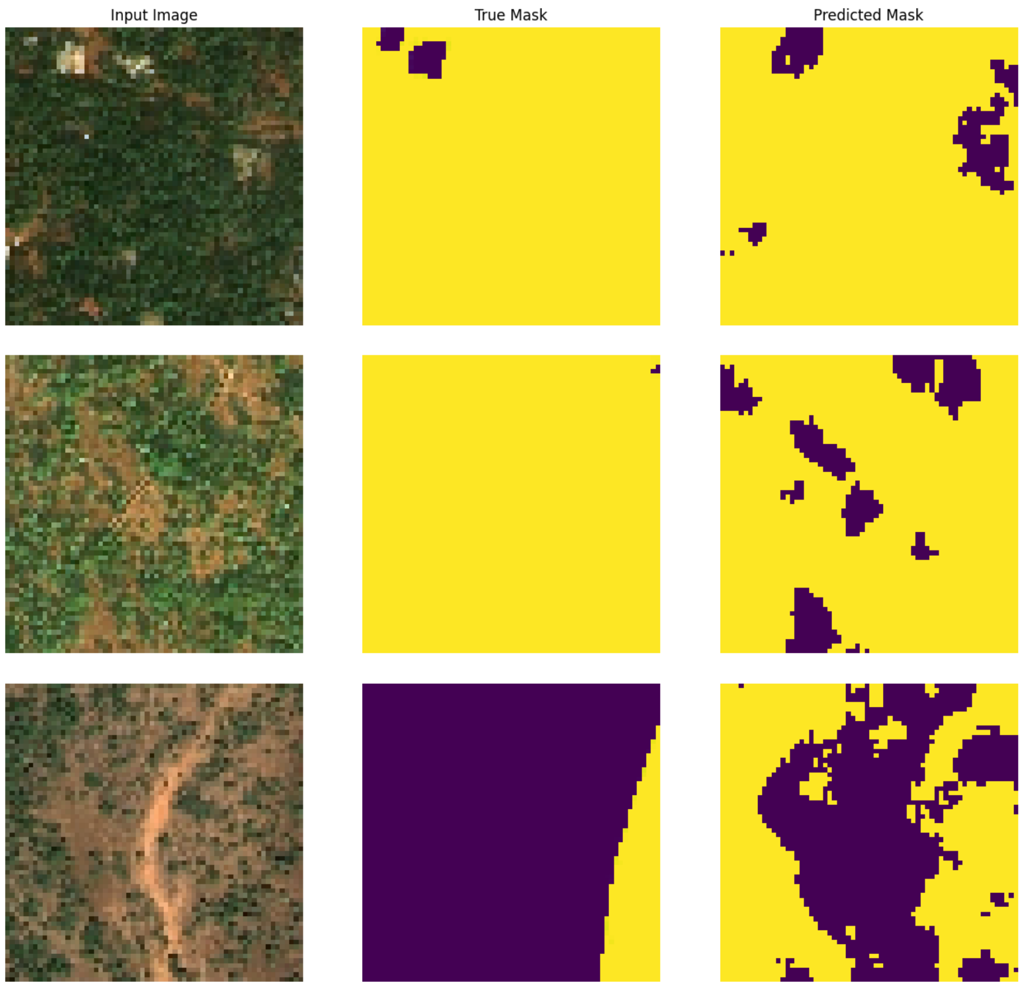


Figure 4: Prediction results with corresponding RGB images and grand truth masks. Proves the model's ability to deliver results better than the grand truth.

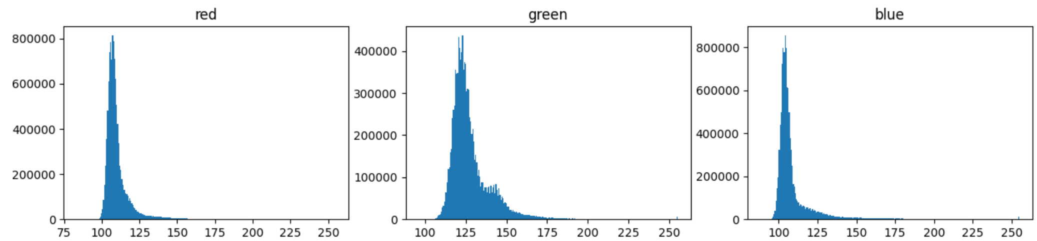


Figure 6: Histograms of the image from 2022 for each RGB channel before stretching.

4 Sentinel-2 images segmentation

The entire image was divided into patches of size 64x64 pixels to fit them in the trained U-Net and then subsequently merged to generate a complete segmented mask of the image.

Figure 2 illustrates that the forest area is visually distinguishable to the human eye, indicating that it should be feasible for a convolutional neural network (CNN) to detect and segment the forests. However, a noticeable challenge arose from the differences in color ranges between the two photos. The 2017 image appeared excessively dark, while the other image appeared overly white. This issue is evident from the histograms depicted in Figure 6 (2017 image considered), revealing that the full-color ranges were not adequately utilized. To address this, a histogram stretching technique was implemented, resulting in significantly improved images as shown in Figure 8, accompanied by their corresponding histograms (Figure 7).

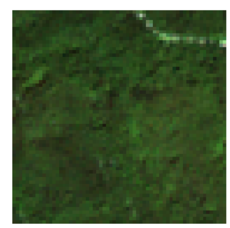


Figure 5: 64x64 image patch.

The prediction outcome depicted in Figure 9 was found to be unsatisfactory. Notably, the yellow areas in the predicted output, which were expected to have values close to one due to the use of the sigmoid activation function, instead had values of 0.001. Intriguingly, the predicted forest areas appeared to align with the streets in the photo, which is the opposite of the expected outcome.

Another potential factor that could have contributed to the unsatisfactory results was the variation in tree types, likely influenced by different climates. To investigate this possibility, the model was applied to a photo captured

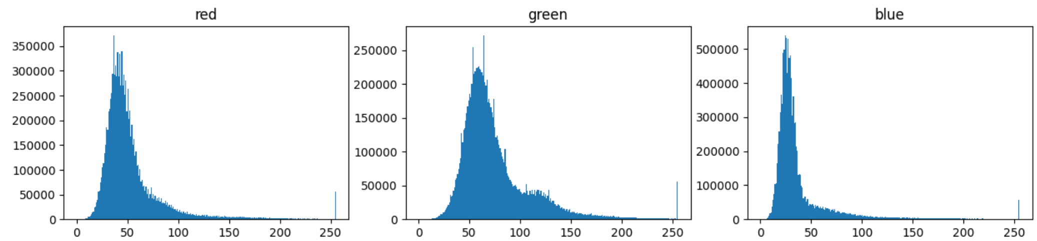
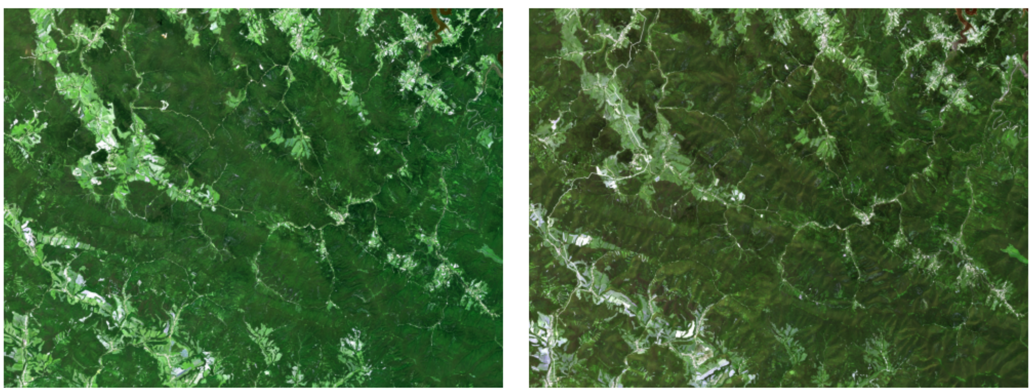


Figure 7: Histograms of the image from 2022 for each RGB channel after white stretching.



(a) July of 2017

(b) June of 2022

Figure 8: Two Sentinel-2 images after histogram stretching.

in the town of Jatam in northern India. This alternative try did not yield successful outcomes either, further emphasizing the issue of the limited resolution of Sentinel-2 images, which turned out to be the primary challenge. As observed in the sample image (Figure 5), the network faced difficulties in accurately segmenting the picture due to the small spatial context provided by the individual patches. The broader spatial context, which is accessible to human perception, plays a significant role in distinguishing the forests.

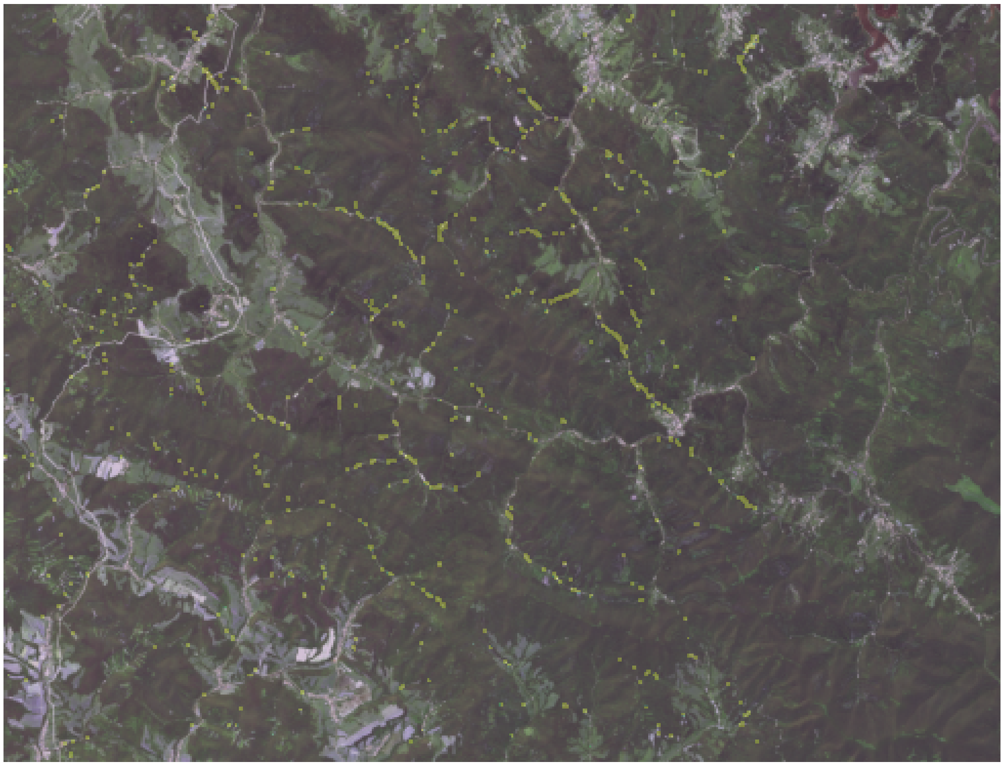


Figure 9: Prediction on the Sentinel-2 image after contrast stretching, with vmax=0.001 of matplotlib funtion imshow()

5 Possible solutions

This section presents potential solutions for further work and analysis, with the goal of segmenting Sentinel-2 images in mind.

5.1 More accurate training set

One possible solution is to enhance the model by providing more accurate input data. This can be achieved by training the segmentation model on higher-resolution images. Ideally, a resolution of 10m, similar to the Sentinel-2 images, would yield significantly better results. It is relevant to ensure that the training data is representative of the target climate zone and includes samples of similar forest types as the target one.

5.2 Unsupervised learning techniques

Another promising approach is to explore unsupervised learning techniques directly on Sentinel-2 images. This entails not only considering spectral properties but also incorporating spatial context, with a focus on capturing the homogeneous characteristics of the dense Puszcza Karpaca's forests. The utilization of contrast-stretched images in this process could potentially enhance the ability to capture the desired features.

5.3 Non-ML techniques

In addition, non-machine learning techniques offer potential solutions. For instance, the Normalized Difference Vegetation Index (NDVI) could provide valuable information for forest segmentation. However, distinguishing forests from other green areas, such as fields, may pose a challenge. Therefore, employing simple thresholding techniques, even within the visible range, could be effective. The distinct detectability of forests in the images presented in Figure 8 supports this idea.

6 Conclusion

Despite, a very well-performing segmentation model (Figure 4), the primary objective of the study wasn't obtained. By considering and exploring possible solutions from section 5, it is hoped that further advancements can be made in effectively segmenting Sentinel-2 images while addressing the specific challenges encountered in this study.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[2] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[3] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.