Comparative Analysis of Deep Learning Architectures for Deforestation Detection in the Amazon Rainforest Using Sentinel-2 Multispectral Imagery

Programming for Modern Machine Learning
Department of Informatics
Ruhr Universität Bochum

Piotr Oleksy
B.Sc. Geoinformatics

Julian Bröcker
B.Sc. Applied Computer Science

Rodrigo Lopez Murguia
B.Sc. Computer Science

Abstract

This project addresses the critical issue of illegal deforestation in the Amazon Rainforest, an ecosystem that has lost nearly 10% of its area in the last two decades. By utilizing multi-spectral satellite imagery and deep learning models, we conducted a comparative analysis of five convolutional neural networks architectures for land cover segmentation: U-Net, U-Net++, Pyramid Attention Network (PAN), Feature Pyramid Network (FPN), and DeepLabV3+. The models were trained with a ResNet34 backbone and trained using Dice Loss to optimize class overlap. The results indicate that U-Net and U-Net++ are the top performers, achieving higher Intersection over Union and Precision, as well as Recall. To ensure practical utility of the models for temporal change detection, a robust post-processing pipeline was developed to mitigate incorrect predictions due to atmospheric interference.

1 Introduction

The goal of the project was finding deforestation in the Amazon Rainforest using Computer Vision models and comparison of effectiveness of the models.

The Amazon Rainforest represents over a half of the rainforest area in the world. It is home to over 30 million people. Its 5.5 millions square kilometers account for almost a half of the world's rainforest area and its around 400 billion trees from 16,000 different species, 26,000 of different plant species and over 5,000 animal species create one of the most diverse ecosystems in the world. Many of the areas stretching in this area are yet to be explored, inaccessible to a modern human hiding secrets of uncontacted tribes and ancient history. It is a jewel of Earth's nature, sadly often exploited for unjustified human gain.

Deforestation has always been a problem in economically poor areas rich in forests. Clearing forests not only guarantees one-time gain through wood sales, but profit long-term, mostly as new farmland. From the year 2001 to 2020 the Amazon lost almost 550,000 km^2 of forest, almost 10% of its own area, and an area equivalent to metropolitan France. The forecasts show that even though new regulations and international outrage the phenomenon not only does not stop, but it might accelerate. Studies show that in the years 2021-2025 we might have lost another 240,000 km^2, which is an area equivalent to the entire United Kingdom. Of course, some clearances are legal with the purpose of building roads and infrastructure, which is a price we have to pay for the region's development. Some is the effect of wildfires, which might be even more dangerous, but sources say that around 90% of deforestation in the Amazon is caused by illegal actions.

In our project we are using Computer Vision models to segmentate biomes in satellite imagery, which later are used to analyse the change in time and find deforestation sites. We are using freely accessible Sentinel-2 data, European Space Agency's observation mission from the Copernicus programme. The mission consists of 3 satellites providing multi-spectral data with resolution spanning from 10 to 60 meters per pixel. For our project we use 4 out of 13 bands: Red, Green, Blue and Near Infrared. Those bands are the most useful for our purpose - discriminating lush green forest vegetation from different biomes. Naturally, using additional bands would provide further help, e.g. aerosol detection, cloud, especially cirrus, detection as well as additional predictors for the forest and non-forest areas. The reason behind using only those four bands was the found dataset, that contained both satellite images and their masks which massively reduced workload in data collection. We decided to stick with the dataset because of lack of availability of dataset applicable for our purposes and we were afraid that building a big enough dataset on our own would consume too much time. Processing and segmenting images manually would rather be a photogrammetry, and not a machine learning task.

2 Related Work

A number of articles have been published about handling similar issues, one of which is the source of our dataset (David and Ce, 2022). The authors focus on different U-Net architectures: Attention U-Net, U-Net, Residual U-Net as well as ResNet50-SeggNet and FCN32-VGG16 and compare their effectiveness. According to the authors their best performing model was Attention U-Net. Another article (Sartor et al., 2025) focused on detecting deforestation in Ivory Coast, where forests are cut down mainly for the purpose of creating farmland for cocoa production. Their best performing model turned out to be regular U-Net.

3 Approach

Our dataset consists of 499 training images with assigned masks saved as tif files. The training dataset is balanced (50.0%-50.0%) between forest and non-forest class. The validation had 100 images and masks consisting of 17.8% forest and 52.2% non-forest. Finally the testing dataset had 20 images: 49.8% forest and 50.2% non-forest. As mentioned in the introduction the images are four-band Sentinel-2 satellite imagery, all in the area of Amazon Rainforest. The bands used were:, Red - most useful for identifying vegetation types and soils, strongly reflected by dead foliage, Green - most useful for shadow areas and textures, Blue - most useful for soil and vegetation discrimination and forest type mapping and Near Infrared (NIR), best for biomass content, as well as detecting and analyzing vegetation. We trained five models: U-Net, U-Net++, Pyramid Attention Network (PAN), Feature Pyramid Network (FPN) and DeepLabV3+.

- U-Net is a convolutional neural network. Its contracting path uses small filters to scan images and identify features and aims to decrease the spatial dimensions of the image. It consists of repeated convolutions and applying activation functions. It is connected to decoder (expansive path) upsampling the images and creating a segmentation map. It uses skip connections to recall features that might have been lost in the downsampling contracting path.
- U-Net++ is an extension of the U-Net architecture. It uses a nested skip connection network improving accuracy. Also, it uses deep supervision providing more guidance, such as additional loss functions or predictions.
- Pyramid Attention Network (PAN) uses residual neural network architecture for the encoding which then is being handled by Feature Pyramid Attention (FPA). It combines features from three pyramid scales. The pyramid structure uses different scale information. The decoder - Global Attention Upsample (GAU) provides global pooling passing the data through 3x3 convolution on the low-level features and generating global context through 1x1 convolution.
- Feature Pyramid Network (FPN) creates a pyramid-like structure of an image at multiple scales. It starts with a bottom-up approach starting with extraction of low-level features, getting later to the high-level. Then it takes the top-bottom way where higher-level features are upsampled slowly getting layer-by-layer to the low-level ones. Both paths are connected with lateral connections transferring high-resolution details.
- DeepLabV3 is a Deep and Fully Convolutional Neural Network which uses Atrous Convolutions (convolutions with kernels with potential gaps) capturing information at different scales. It also uses Atrous Spatial Pyramid Pooling (ASPP), which applies sets Atrous Convolutions with different rates, then concatenates the output.

4 Experiment

To train and evaluate the models proposed in the approach section, we used the *segmentation_models_pytorch* library built on the PyTorch framework.

4.1 Architecture of the Model

Every model was trained with a *resnet34*, a 34-layer model pretrained with *imagenet*, this should be deep enough to understand the texture of the vegetation in the images without being very computationally heavy. The model receives 4 channels of input corresponding to the Red, Green, Blue and NIR (Near Infrared) given by the Sentinel 2 satellite. For the loss function, the Dice Loss is preferred for segmentation models, given that it maximises the intersection area over foreground while minimizing the union over foreground in a stable way during the training process. This is much better than using Cross Entropy Loss, which may lead to the model trying to guess that "everything is a forest" and getting a low error, but failing to ever converge to a desirable point. For the learning rate, we use Adam, which is an adaptive approach that checks the history of gradients to adjust the learning rate, leading to stable convergence.

4.2 Training of the model

For the training of the model back propagation was used with early stopping in order to prevent overfitting. During the training of the model metrics are retrieved (such as average loss, precision, Intersection over Union, and the confusion matrix values. These were used to generate a final graphic report of the training process. The best model found through the epochs is saved for later use.

4.3 Deforestation Detection

To evaluate the forest cover loss, multiple images of the same spot over time were evaluated To evaluate the models for forest monitoring, we chose critical "hotspots" of forest loss using "Global Forest Watch", which is an open-source, online platform managed by the World Resources Institute (WRI) that provides real-time data on satellite imagery. Once the area selection is done, the imagery was explored via the Copernicus Browser to visually inspect cloud cover and land features before extraction. The images were downloaded using the Sentinel Hub API in .tiff format. Later, the model performs segmentation on each of the satellite images, one by one, chronologically and saves the location of the forest areas to compare with the result of the following images. If an area of the forest from the past image turns to non-forest, deforestation is assumed. One of the biggest challenges faced was cloud handling.

4.4 Cloud Handling

To avoid false positives while detecting deforestation it was important to deal with clouds, which always get classified as non-forest by the models. Clouds' shadows were often also recognised as non-forest class. As a first line of defence we use simple thresholding technique, a basic photogrammetry technique, clouds usually are easily segmentable in that way, but since we lack some bands we found that this technique alone would not be sufficient. As an additional measure we use automatically generated masks from Copernicus. Even though the masks would not be useful for our main model training (classes: vegetation/not-vegetated - forests are in the same class as fields, meadows), they contain both cloud and cloud shadow classes. Both our own threshold-based mask as well as Copernicus' make up the final mask. If we find changes in the masked areas, the change is ignored.

As a final measure of avoiding false detection the code does not report deforestation immediately when unmasked (not in a cloudy/shadowy area) change occurs. To avoid detecting deforestation, when the mask accuracy is not perfect, or "random" classification changes, such as sudden change in classification of pixels near the cut-off point, e.g. because of the changes in lightning in the imagery, each point for which change from forest class into non-forest class waits until it stays the same for three next images in a row. If a change back into the forest class is observed, the counter is resetted.

5 Results

The next section contains the results of the performance of the top performer models during their training. The model training metrics for the other models trained are found in Figures 5, 6 and 7 in the Appendix.

Figure 1: The U-Net resulted in a precision around 95-98%, an IoU of 95% and had the best training/validation loss curves.

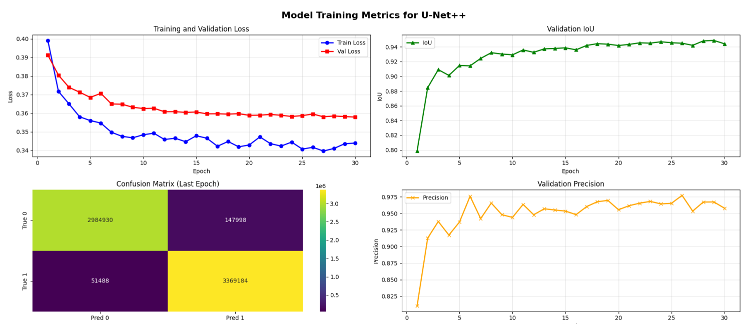

Figure 2: The U-Net++ had also really good results in the training phase

After the training, a general performance test was made on the test set, which gave us the following results:

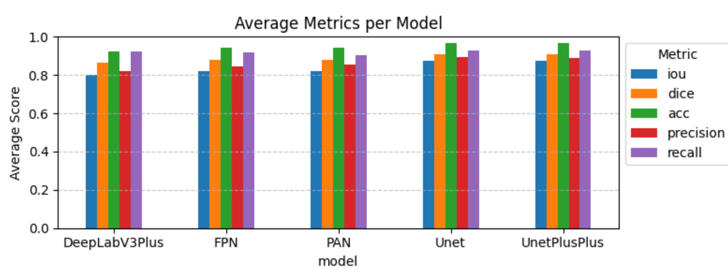

Figure 3: Average metrics of each models, including Intersection over Union, Dice, Accuracy, Precision and Recall

The distribution of TP, FP, TN and FN across test samples is illustrated in Figure 8 in the Appendix.

A sample run of deforestation detection performed with the UNet++ model shows how the deforestation is detected step-by-step. Naturally, the biggest progress occurs when the sky is clear. If the sky was fully clouded there would be no progress at all, though a full cloudiness comes with a high probability of rain, during which the deforestation would most likely not be progressing . A potential usage of Sentinel-1 radar imagery is a room for improvement in cloud handling.

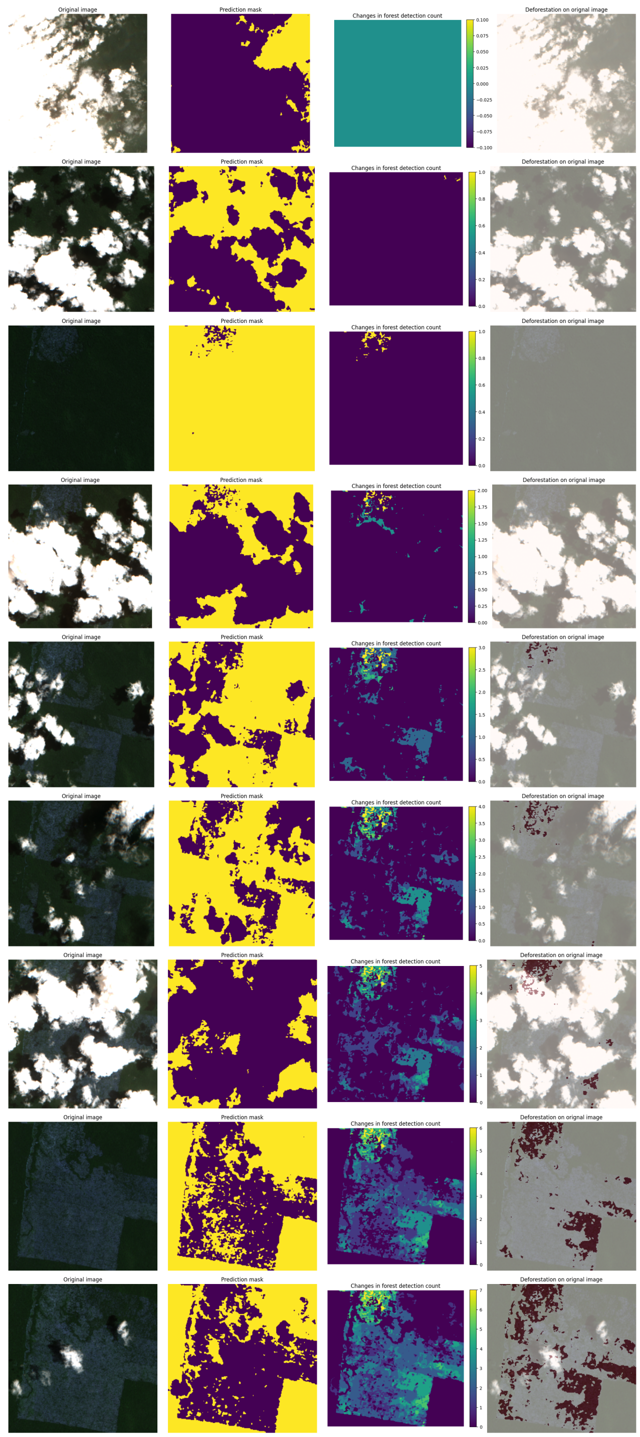

Figure 4: Visual evaluation of the model's performance in the presence of atmospheric interference. The columns demonstrate (from left to right) the original satellite imagery with varying degrees of cloud opacity, the generated prediction masks, the forest detection confidence levels, and the final segmentation overlay. This sequence highlights the architecture's ability to maintain land cover classification consistency despite significant visual obstructions.

6 Conclusion

The top performers for this task were shown to be the U-Net and Unet++, with higher precision and IoU, which suggests that the skip connections in the U-Net are particularly effective to detect the little details on the land. This successfully demonstrated the application of these deep learning architectures for land cover segmentation using satellite imagery and cloud handling. As this was our first collaborative project on machine learning, the initial phase was marked by a significant learning curve, as we lacked a clear roadmap to reach our objectives, on which we also had problems defining. The project also remains with areas of opportunity, for example, the data diversity of the dataset was not that good, as well as implementing hyperparameter tuning techniques could further refine the performance of the U-Net models.

References

Shanna Hanbury. 2025. 91% of brazilian amazon deforestation last year was illegal, report finds. Accessed: 2026-01-28.

InfoAmazonia. 2023. Deforestation in the amazon: past, present and future. Accessed: 2026-01-28.

David John and Ce Zhang. 2022. An attention-based U-Net for detecting deforestation within satellite sensor imagery. *International Journal of Applied Earth Observation and Geoinformation*, 107.

Gabriele Sartor, Matteo Salis, Stefano Pinardi, Özgür Saracik, and Rosa Meo. 2025. Deep learning tools to support deforestation monitoring in ivory coast using SAR and optical satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 144.

Sentinel Hub. Sentinel-2 bands. Accessed: 2026-01-28.

Wikipedia contributors. 2026. Amazon rainforest. Accessed: 2026-01-28.

A Training metrics of the models

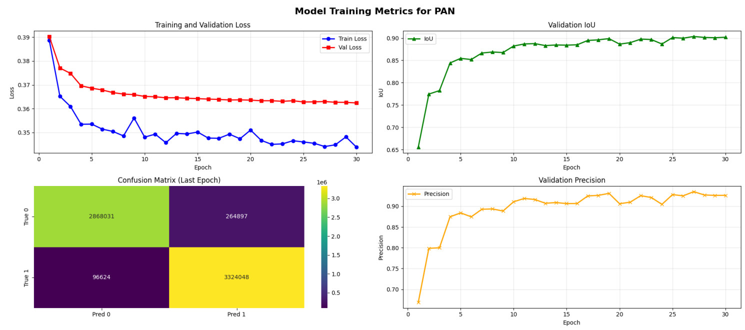

Figure 5: The PAN model reached a peak IoU of 91% and a precision of 94%

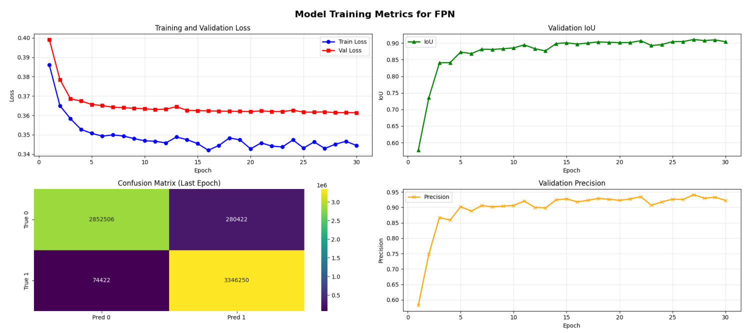

Figure 6: The FPN model reached a stable IoU of around 90% and precision of 93%

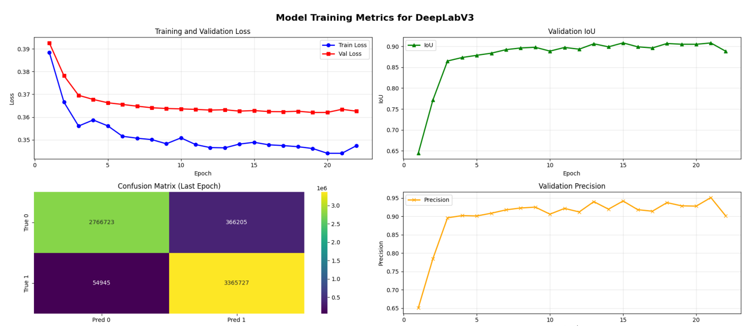

Figure 7: DeepLabV3+ IoU reached about 91% and the precision oscillated between 90-95%

B Confusion matrix components

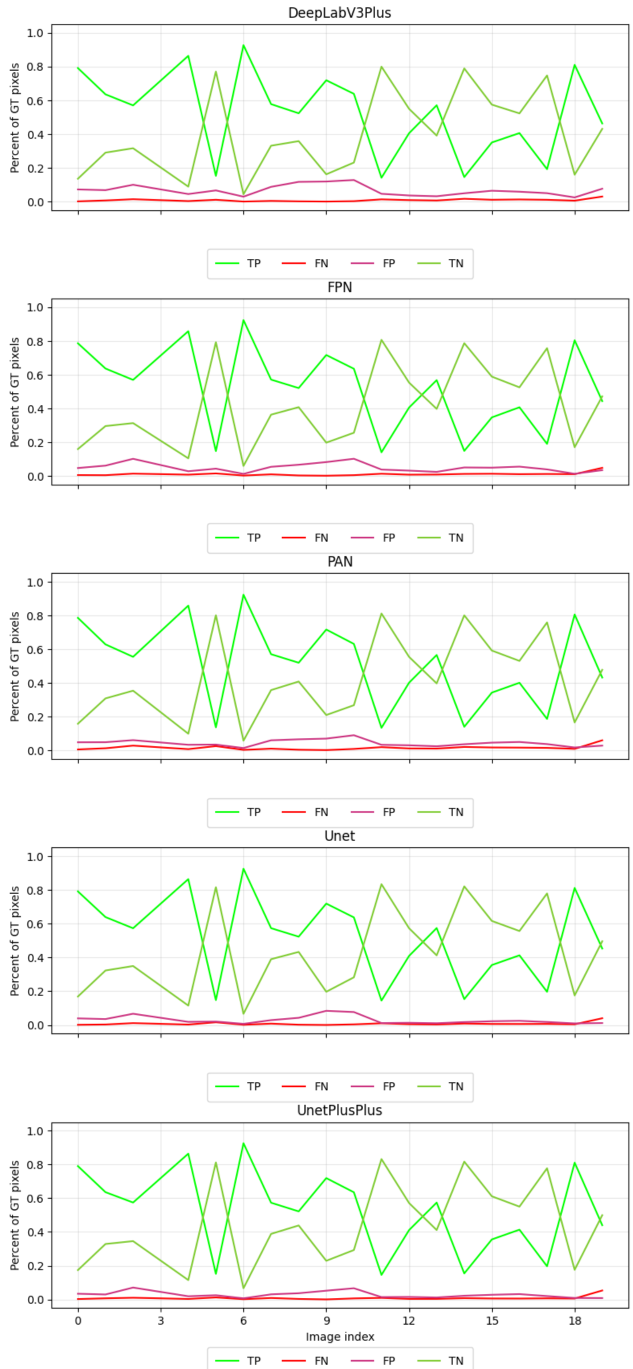

Figure 8: Comparative analysis of confusion matrix components (TP, FN, FP, TN) across 20 test samples for all evaluated architectures. The plots illustrate the pixel-wise classification performance consistency for DeepLabV3+, FPN, PAN, U-Net, and U-Net++.