A DEEP LEARNING SYSTEM FOR LOCATING AND ANALYZING DEFORESTATION FROM SATELLITE IMAGERY

Robert Pendergrast

Columbia Engineering, Computer Science, Class of 2026

ABSTRACT

Motivated by the vastly increasing number of satellites in Low Earth Orbit (LEO), as well as the growing issue of deforestation, this report presents a method for detecting and locating areas of deforestation from satellite imagery. Original data sourced from the European Space Agency's Copernicus program depict areas of deforestation and untouched land in the Brazilian Amazon rain forest. These data are then labeled and curated into a dataset to train three different neural network architectures to recognize deforestation in a satellite image. The best model is chosen, and, using said model, a system is designed to map regions of deforestation in the Amazon rain forest.

1. INTRODUCTION

In the last few years alone, the number of man-made objects in Low Earth Orbit has skyrocketed at a near-exponential rate. The year 2023 saw the percentage of satellites on orbit increase by thirty-five percent, with the number reaching a total of 9,115 satellites (Statistica). This number is only expected to increase in the coming years. Here lies an incredible scientific opportunity. The influx of objects launching into orbit opens the door to vast amounts of satellite imagery that can be used for a variety of activities including climate monitoring, disaster management, economic forecasting, and more. This project looks at a climate-related use case for these images, namely, monitoring and analyzing deforestation in the Amazon rainforest.

Deforestation is an issue that continues to be at the forefront of the ever-growing climate crisis. The objective of this project is to develop a deep learning-based system that can not only classify images containing deforested regions, but also one that can, given a larger satellite image, highlight and map out deforested areas. Close attention was paid to the accuracy of the different models in their detection of deforested areas. Additionally, visual inspections of the region labeling will also be performed to ensure validity in the system's behavior.

2. METHODS

2.1. Data Acquisition and Processing

As with any deep learning project, the initial step in its completion was to locate potential data sources that would allow a model to identify deforestation in the Amazon. Initially, an attempt was made to use a well-known and existing dataset: the EuroSat dataset. The EuroSat dataset contains an amalgamation of 64 by 64 pixel satellite images from the European Space Agency's Sentinel-2 Satellite. The images are separated into 10 classes, which could potentially be used to classify deforested regions of the Amazon rainforest given a larger picture. However, after working with this dataset, a fine-tuned Resnet was ineffective in mapping scenes of deforestation because of discrepancies between the training data and the images being analyzed. As a result, it was necessary to curate an original dataset for this project. To do so, images from the European Space Agency's Copernicus Data Browser were sourced.

2.2. The Copernicus Browser

The Copernicus Data Browser is a free online tool used to display a variety of images and measurements taken from the European Space Agency's Sentinel-2 satellite. This is the same satellite that took the EuroSat images, indicating that it was a well-respected source of satellite imagery. The images captured cover a wide range of time periods, spanning roughly from 2015 to the present-day (The European Space Agency). Using the Copernicus Browser, four satellite image scenes were selected from the Amazon rainforest. Two scenes displayed a completely deforested area, and the other two were of untouched forest. Two of these images are shown in Figure 1.

2.3. Data Processing

To process each of these images into a viable dataset, a simple pipeline was followed. Each of the four primary images was segmented into a grid of 36 x 24 images, totaling 864 sample images per scene. With four scenes, there were 3456 images in total. These 'sub-images' were then used as the dataset that

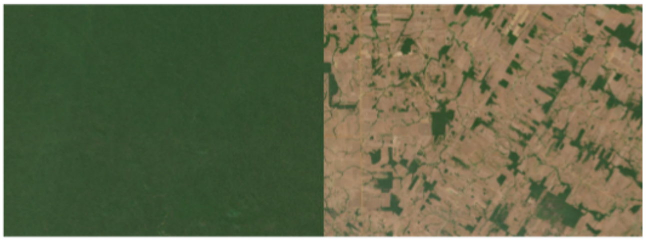

Fig. 1. Two of the selected scenes from the Sentinel-2 Data. The scene on the right displays untouched rainforest, whereas the scene on the left is deforested (European Space Agency).

the deep learning models would be fine-tuned on. The dimensionality of the training images (50 x 50 pixels) was chosen to maximize the number of training samples while also maintaining the integrity of the image data. A diagram illustrating the image processing procedure is given in Figure 2:

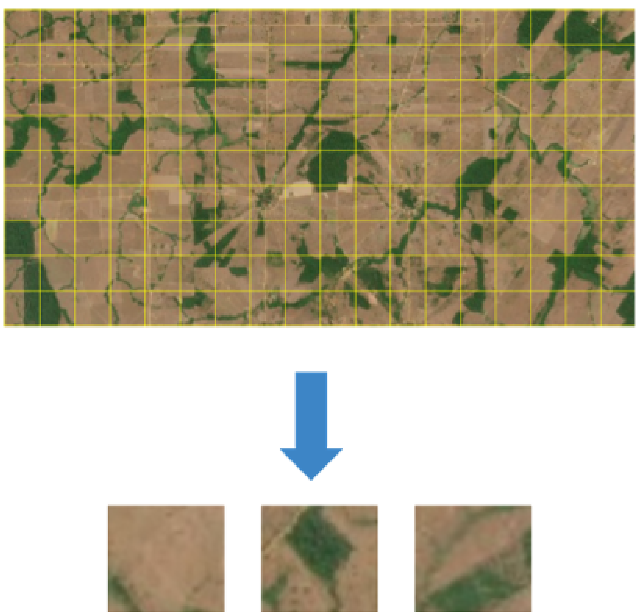

Fig. 2. An illustration of the image processing step. Each satellite scene is divided into sub-scenes, which are then fed into the deep learning models.

The dataset was then divided into training, validation, and testing sets, following a traditional 80-10-10 split using Pytorch's Datasets library. Then, using Pytorch's DataLoader Library, training, testing, and validation datasets were created using batch-sizes of 256, 1, and 1 respectively.

2.4. Model Training and Selection

With the data curated, computer vision models were subsequently trained. For this experiment, three separate models were trained and their respective accuracies and losses were compared. The model with the combination of the overall best accuracy, loss, and runtime was then chosen to analyze additional scenes of Amazonian deforestation. The three models chosen for fine-tuning were as follows: Resnet18, Resnet50, and the ViT-16 (Vision Transformer). Resnet18 was chosen since the Resnet architecture has a proven history of being able well-suited for fine tuning tasks, as discussed in class. Experimenting with Resnet50 was a logical next step to determine the effect of training a deeper neural network on the dataset. Finally, the Vision Transformer was selected to experiment with the potential advantages of using a transformer model, rather than the traditional convolutional architectures typically used for image processing. Brief explanations of each of these architectures are given in the following sections.

2.4.1. Resnet

The Resnet is a state-of-the-art convolutional neural network architecture that is well suited for classification tasks. Developed in 2015 by He et al, the Resnet proved to be a pivotal development in the field of deep learning. The Resnet architecture works by introducing skip-connections into the convolutional network structure. These skip connections (shortcut connections) work so address the issues of exploding and vanishing gradients (He et al). He et al's work proved that the addition of shortcut connections to a convolutional network significantly improved the networks performance. Both the Resnet18 and Resnet50 models used in this project were based on the work of He et al. All credit is attributed to them.

2.4.2. Vision Transformer

The Vision Transformer was ideated in 2020 by Dosovitskiy et al, leveraging existing transformer architectures used for natural language processing to tackle image tasks. As with traditional transformer models, the Vision Transformer processes positionally encoded image pieces. To do this, The Vision Transformer segments input images into flattened 'patches', which are then fed into the encoder architecture. The encoder then classifies the image (Dosovitskiy et al). The Vision Transformer used in this project was based on the work of Dosovitskiy et al. All credit is attributed to them.

Each of these models was trained on the previously described dataset, and their performances were evaluated. Then, the best model was chosen to work in combination with a Python script using the OpenCV library to localize and highlight areas of deforestation from larger satellite scenes. These new scenes were the same size as the initial four images used to create the dataset. However, the chosen images depict the same region, taken seven years apart. The chosen model was then used to visually gauge how much deforestation had occurred in that region over those seven years.

3. RESULTS AND ANALYSIS

3.1. Model Analysis and Comparison

Once each of the models was fine-tuned to the dataset, their performances were evaluated. Although each of the models had demonstrated to be successful in delineating between forested and deforested lands, they were compared against each other to determine which model was the best. A summary of this evaluation is provided in this section.

3.1.1. Resnet18 Performance

The fine-tuned Resnet18 performed very well with the training data. Over the course of the model's training epochs, the models performance converged to a final validation accuracy of 99.7 percent and a final validation loss of just 0.04. When tested on the test set, the model was able to successfully classify 100 percent of the images, indicating that a very strong model had been produced.

3.1.2. Resnet50 Performance

The fine-tuned Resnet50 performed very similarly to the Resnet18 model. With a final training accuracy and loss of 99.7 percent and 0.20 respectively, it is safe to say that the model performed very well. Just like the Resnet18 model, the fine-tuned Resnet50 scored 100 percent when labeling the testing data, further evidencing the strength of its performance. It is important to note, however, that the testing accuracy for both of these two models varied from between 0.98 to 1.0 depending on the data split. Regardless, these accuracies were very high, and a sign that the models were training well.

3.1.3. Vision Transformer Performance

Finally, the Vision Transformer's training demonstrated its ability to accurately label images of deforestation just like the previous two models. Interestingly, the Vision Transformer model had a constant 99.7 percent accuracy throughout each of its training epochs, not showing an initial low accuracy like the previous two models. Likewise, its validation loss remained relatively constant at 0.20. Regardless, the high accuracy and low loss suggest that it performs well. This was indeed the case, as it was able to classify the test set with 100 percent accuracy.

Graphs depicting the individual accuracies and losses of each of the three models are given in Figure 5, as well as their respective accuracies compared against each other.

3.2. Image Analysis

While all the models performed exceptionally well, the final model selected was the Resnet18, due to its perfect performance on the test set in combination with its faster runtime and lower loss. The fine-tuned Resnet18 model was used to create a Python script to localize deforested areas. This end result is visualized in Figure 6:

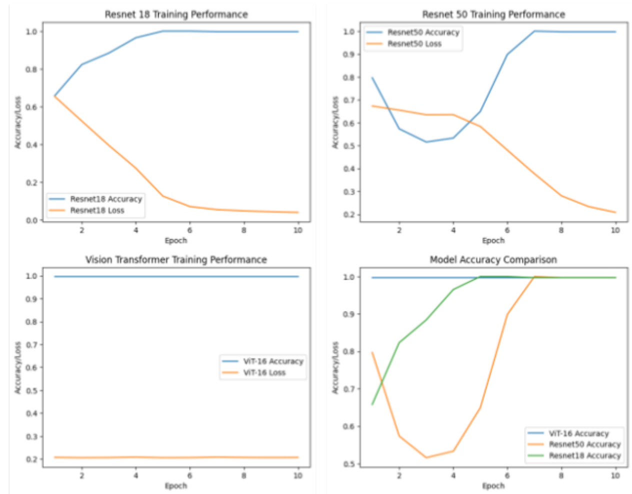

Fig. 3. Plots displaying the validation accuracy and loss for the fine-tuned Resnet18 (top left), Resnet50 (top right), Vision Transformer (bottom left), as well as the three models' validation accuracies compared to each other (bottom right).

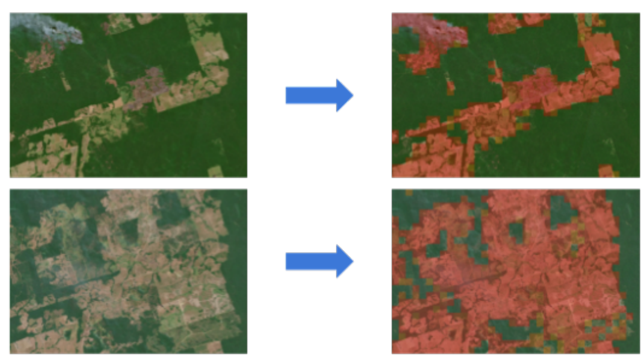

Fig. 4. Highlighted regions of deforestation in the same scene from 2017 (top) and 2024 (bottom). Deeper red highlights indicate that the model is confident in the presence of deforestation while orange and yellow highlights suggest that the model is less confident.

A brief visual inspection of these images suggests promising (and scary) results. The applied Resnet18 model was able to clearly identify most of the deforested regions of the given scenes. There are, however, a few interesting features to note. The model labeled the clouded portion in the upper left-hand corner of the 2017 image as deforested land. This is likely due to the fact that clouded images were intentionally left out of the dataset – lending itself to the downsides of using a contrived dataset like the one from this project. Additionally, there are some patches of deforested land that were labeled untouched. This can be attributed to inaccuracies in the model. Regardless, it is clear that the model was able to classify these regions very well. The highlighted regions showcase the frightening trend of environmental damage occurring in the Amazon rainforest.

4. CONCLUSION

In total, a comprehensive system was built to detect and localize areas of deforestation from satellite images. After processing and creating a unique dataset and fine-tuning three different deep learning models, a system capable of classifying scenes of deforestation was created. These models were then used in conjunction with a Python script to localize and highlight areas of deforestation from within a larger satellite image.

Although the system built in this project was successful at both classifying and locating areas of deforestation, there were still many challenges that were encountered during this projects' creation. As mentioned before, the initial ideation of this project involved using a widely used dataset (The EuroSat dataset) that included a variety of classes. The intent here was to develop a system that could not only identify areas of deforestation, but also that could label the cause of said deforestation. However, this approach was a dead end as the models' accuracy was not good enough to develop a reliable working system.

This project lays the foundation for future work. More Sentinel-2 data can be acquired from different regions and hand-labeled to classify why deforestation had occurred. Potential labels include deforestation caused by wildfire, logging, or farming. This system in turn could draw even greater insights into the deforestation occurring in both the Amazon and around the world. Regardless, the system developed in the project was able to highlight deforested areas and serves as a proof of concept for future systems to be developed.

5. CONTRIBUTION STATEMENT

As the sole member of this project, I personally took on every step of designing and creating the computer vision system. This includes project ideation, gathering the data, processing the data, training the three models, model analysis, system analysis, report writing, and presentation creation. The data is sourced from the European Space Agency's Copernicus Browser. Resnet network architectures based on the work of He et al. Vision Transformer architecture based on the work of Dosovitskiy et al.

6. REFERENCES

Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *arXiv*, 2021, Accessed 9 December, 2024.

European Space Agency, *Copernicus Browser*, https://browser.dataspace.copernicus.eu/.

European Space Agency, *Sentinel-2*, https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2

He et al. "Deep Residual Learning for Image Recognition." *arXiv*, 2015, Accessed 9 December, 2024.

JSR. "Number of Active Satellites from 1957 to 2023." Statista, Statista Inc., 15 Jan 2024, https://www.statista.com/statistics/897719/number-of-active-satellites-by-year/.