Image segmentation of aerial photos

Andrea Paletto

University of Klagenfurt

July 5, 2023

Table of Contents

Introduction

In the context of transportation, pilots need to identify possible urgent landing areas in forests. Therefore, scientists have provided us with aerial photos of the forest and accompanying image masks that indicate the presence of birch trees in the images. These masks are binary, with a value of 0 representing areas without birch trees and a value of 1 indicating the presence of birch trees. Our objective is to utilize this dataset to train a model capable of performing semantic segmentation, accurately identifying and delineating birch trees in the images.

Data Description

The dataset consists of two main components:

- Images (X):
    - The "X" component of the dataset contains aerial photos of forests.
    - These images serve as the input data for the semantic segmentation model.
    - Each image represents a snapshot of the forest, captured from an aerial perspective.
- Masks (y):
    - The "y" component of the dataset contains image masks that accompany the forest photos.
    - The masks are binary, with a value of 0 representing areas without birch trees and a value of 1 indicating the presence of birch trees.
    - These masks provide ground truth labels for training the semantic segmentation model.
    - The goal is to train a model that can accurately identify and delineate the areas occupied by birch trees in the forest images.

Mathematical Description of the Proposed Model (pt 1)

The proposed model is based on the U-Net architecture, which consists of a contracting path and an expansive path. The architecture allows for effective feature extraction and precise localization.

- **Contracting Path:** The contracting path consists of multiple convolutional layers followed by max pooling operations. The purpose of the contracting path is to capture and compress the input information, gradually reducing the spatial dimensions while increasing the number of feature channels.
- **Expansive Path:** The expansive path performs upsampling followed by concatenation with features from the contracting path. It consists of multiple upsampling layers, each followed by a convolutional layer. The expansive path helps in recovering the spatial resolution while maintaining the learned features from the contracting path.

Mathematical Description of the Proposed Model (pt 2)

- **Parameters:** The proposed model offers various parameters that can be adjusted to optimize its performance:
  - **Number of Convolutional Filters:** The number of filters in each layer of the U-Net can be modified. Increasing the number of filters can enhance the model's capacity to capture complex features, but it also increases the computational complexity.
  - **Optional Dropout:** Dropout regularization can be applied to specific layers, such as conv2 and conv4, to mitigate overfitting. The dropout rate can be adjusted to control the extent of regularization.
  - **Tuning hyperparameters** like: LR, Dropout, filter and batch size

Algorithm (pseudo code)

My project is divided in the following part:

- First part: In the first part I tried with different combination of: learning rate, filter size and dropout. I took the best one.
- Second part: I used the best hyperparameters from the first part and I also used Data augmentation and binary focal loss technique to improve the result.
- Finally, in the third part I used Transfer learning technique, here the encoder part of the U-Net uses a pretrained VGG16.

System architecture (flow diagram)

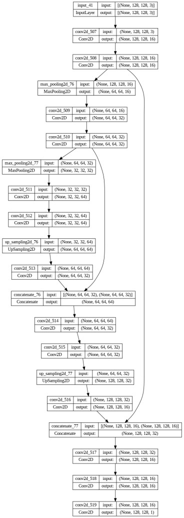

Figure: uNet architecture

Performance and model evaluation 1

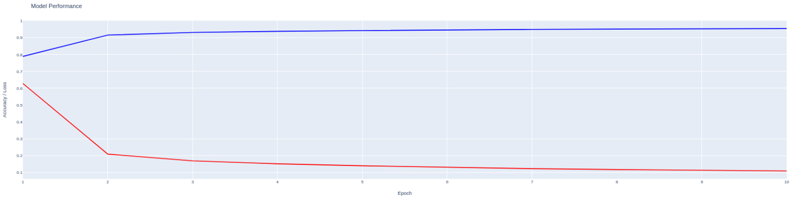

Figure: uNet with tuned hyperparameter (LR, Dropout, filter size)

Performance and model evaluation part 1.1

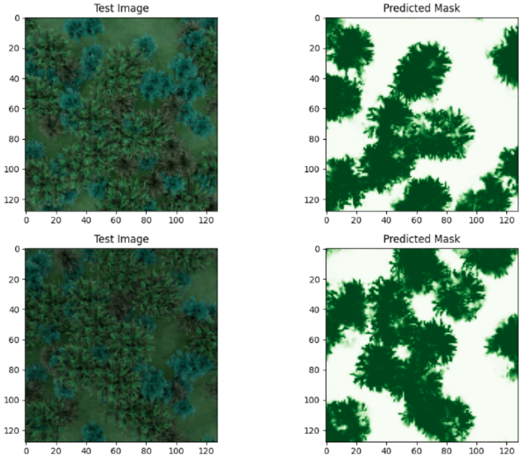

Figure: predicted mask with hyperparameter tuned

Performance and model evaluation 2

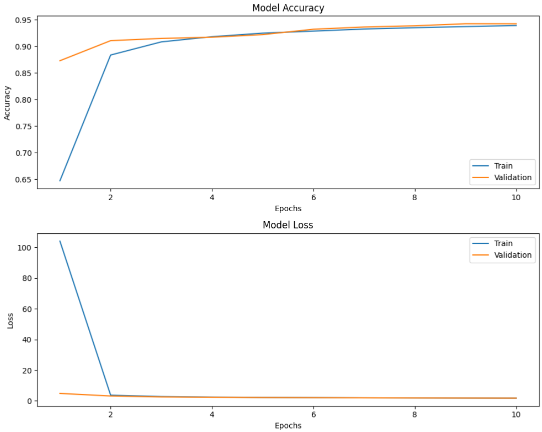

Figure: uNet accuracy with data augmentation without using dropout

Performance and model evaluation part 2.1

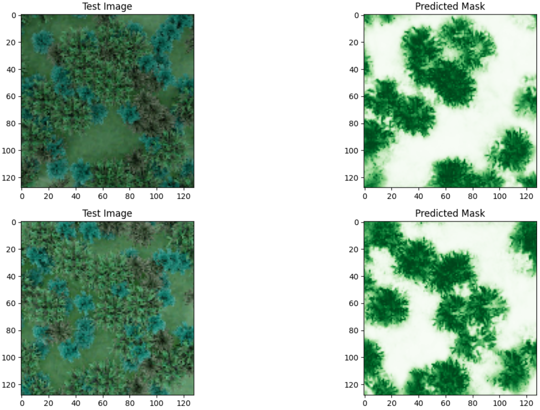

Figure: predicted mask with data augmentation without using dropout

Performance and model evaluation 2.3

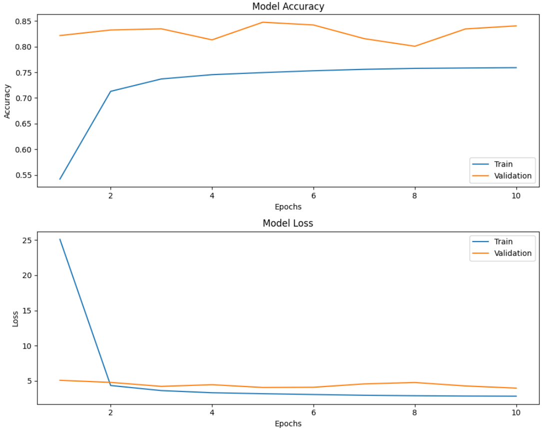

Figure: uNet accuracy with data augmentation using dropout

Performance and model evaluation part 2.4

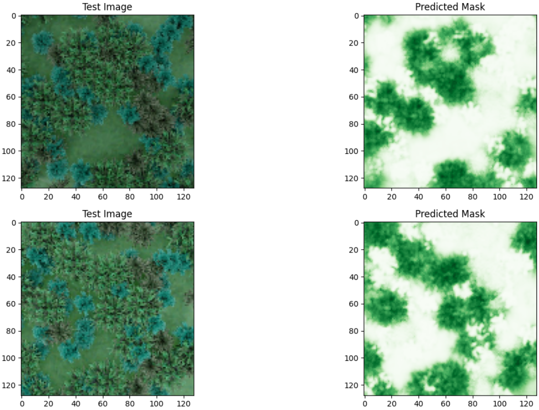

Figure: predicted mask with data augmentation using dropout

Performance and model evaluation 3

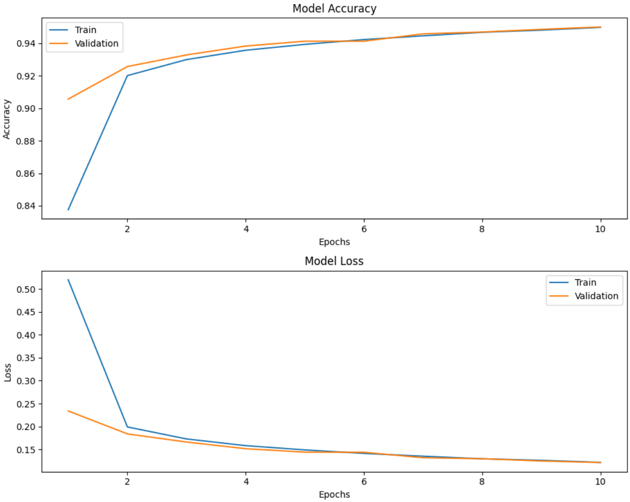

Figure: uNet and VGG16 accuracy and loss

Performance and model evaluation part 3.1

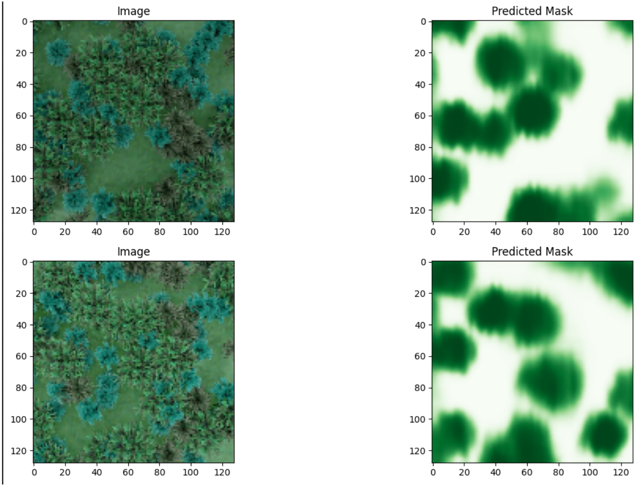

Figure: predicted mask uNet and VGG16

Performance and model evaluation final

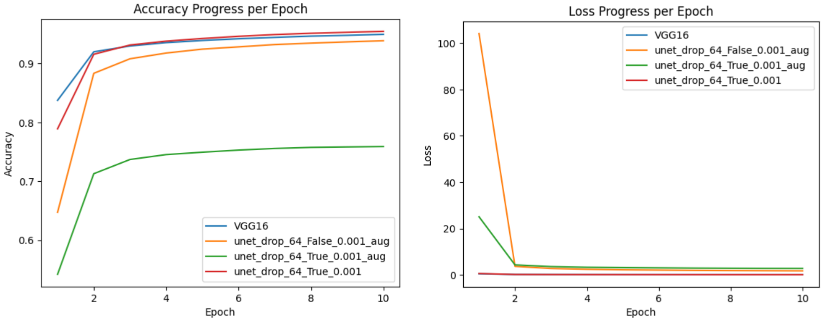

Figure: all result

Discussion regarding the model performance I

Approach 1: U-Net without Augmentation and binary focal loss

- Parameters: Number of convolutional filters = 64, Dropout = True, Learning rate = 0.001
- The model demonstrates an improvement in accuracy and a decrease in loss over the epochs.
- The accuracy starts at around 0.789198 and reaches 0.954768 after 10 epochs.
- The loss starts at a high value of 0.628280 and reduces to 0.108505 after 10 epochs.

Approach 2: Augmentation and binary focal loss with U-Net (Dropout)

- Parameters: Number of convolutional filters = 64, Dropout = True, Learning rate = 0.001

Discussion regarding the model performance II

- The model demonstrates an improvement in accuracy and a decrease in loss over the epochs.
- The accuracy starts at around 0.541945 and reaches 0.759083 after 10 epochs.
- The loss starts at a high value of 25.098499 and reduces to 2.824596 after 10 epochs.

Approach 2.1: U-Net with Augmentation and binary focal loss (No Dropout)

- Parameters: Number of convolutional filters = 64, Dropout = False, Learning rate = 0.001
- The model shows a significant improvement in accuracy and a substantial decrease in loss compared to the previous approach.
- The accuracy starts at around 0.647332 and reaches 0.938905 after 10 epochs.

Discussion regarding the model performance III

- The loss starts at a high value of 104.103943 and reduces to 1.769409 after 10 epochs.

Approach 3: Transfer learning with VGG16 encoder

- The model using VGG16 as the encoder exhibits the highest accuracy and lowest loss among the four approaches.
- The accuracy starts at 0.837588 and reaches 0.949746 after 10 epochs.
- The loss starts at 0.520351 and decreases to 0.121496 after 10 epochs.

Conclusion(Page 1)

Among the four approaches evaluated in this project, Approach 3, which utilized transfer learning with the VGG16 encoder, exhibited the highest accuracy and lowest loss.

- The accuracy of Approach 3 started at 0.837588 and reached 0.949746 after 10 epochs.
- The loss started at 0.520351 and decreased to 0.121496 after 10 epochs.

Approach 3 outperformed the other approaches in terms of both accuracy and loss. This is likely because:

- The VGG16 encoder, pre-trained on a large dataset, was able to extract more meaningful and representative features from the input images.
- This led to better classification performance.

Conclusion (Page 2)

However, it's worth noting that Approach 3 requires the use of transfer learning and a pre-trained model, which may not always be available or suitable for all scenarios. In cases where transfer learning is not feasible, Approach 2.1, which utilized augmentation and the U-Net architecture without dropout, showed a significant improvement in accuracy and a substantial decrease in loss compared to the other U-Net approaches.

- Approach 2.1 achieved an accuracy starting at 0.647332 and reaching 0.938905 after 10 epochs.
- The loss started at 104.103943 and reduced to 1.769409 after 10 epochs.

Data augmentation played a crucial role in improving model performance, and the U-Net architecture effectively learned and extracted features from the augmented data.

Conclusion (Page 3)

In conclusion, Approach 3 with the VGG16 encoder should be preferred if transfer learning is an option:

- Achieved the highest accuracy of 0.949746 and lowest loss of 0.121496 after 10 epochs.
- Benefited from the meaningful and representative features extracted by the pre-trained VGG16 encoder.

Approach 2.1 without dropout and with augmentation can be a viable alternative when transfer learning is not feasible or desirable:

- Demonstrated a significant improvement in accuracy, starting at 0.647332 and reaching 0.938905 after 10 epochs.
- Showed a substantial decrease in loss, starting at 104.103943 and reducing to 1.769409 after 10 epochs.

Both approaches provide valuable insights into achieving high accuracy and low loss in image classification, each with its own advantages and considerations.