Assignment 2 Image Segmentation

By: Benjamin Trenery

**Introduction**

Image segmentation is a method for partitioning an image into multiple segments. Each segment represents regions that are meaningful. It is a way for detecting features in an image. One model that is used for image segmentation is a u-net. U-nets are a type of convolutional model that is able to segment input images. It uses both encoders and decoders in order to accomplish this task. Encoders contract the series in order to extract the hierarchical features from the image. Then, the decoder expands the path by upsampling in order to gradually recover the spatial information of the features. It combines the features learned during the encoding step with the overall image features. Skip connections are used in order to connect layers in the encoder and decoder step. It allows the network to retain spatial information and mitigate the vanishing gradient problem. In the final layer of the model, it generally uses a 1 by 1 convolutional layer using a softmax activation function in order to create a pixel wise classification probabilities for each class in the problem that it is trying to identify. The inputted data are masks and images. Images are the original format and masks are outlines of the important parts of the image.

**Background**

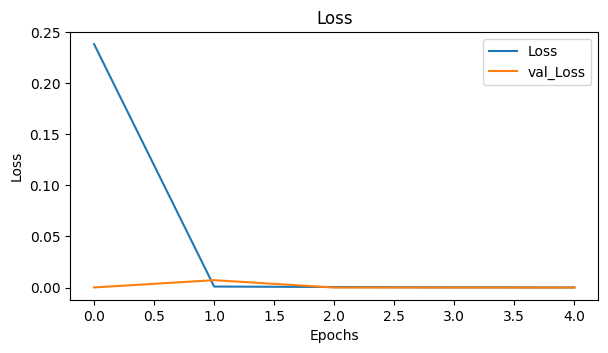
This assignment uses an image segmentation model. The goal of this model is to identify forests in an image. When developing new areas, it is important to identify forests before initiating construction in order to preserve the ecosystem and find out how much forest that there is in the area. This is considered a binary segmentation task. It uses a u-net architecture as the baseline of the model. The dataset came with a csv file, images, and masks. Due to the nature of the problem, a simpler model would not be able to identify the features as well as a convolutional neural network.

**Methods**

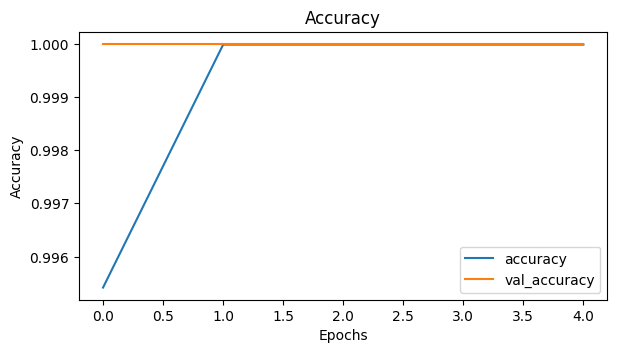
For this assignment, I could not access the GPU for a larger scale model. I also was not able to create the saliency map in order to identify what features the model was identifying. However, I did create a model that examined 5108 aerial images that were 256 by 256 pixels. They were environmental images containing forests. The dataset came from kaggle. Due to being only a binary segmentation task, there was only 1 class. It was detecting whether parts of an image were of a forest or not. The model was with a batch size of 60 for 5 epochs. It used an Adam optimizer with a 0.01 learning rate. Dropout and batch normalization were used for model generalization. The model was normalized to be values between 0 and 1. Each of the images and masks were converted to numpy arrays. Data augmentation was also utilized for image generalization by using vertical and horizontal flips, zoom and rotation. The images were colored, so it used 3 channels for the rgb scale.

**Results**

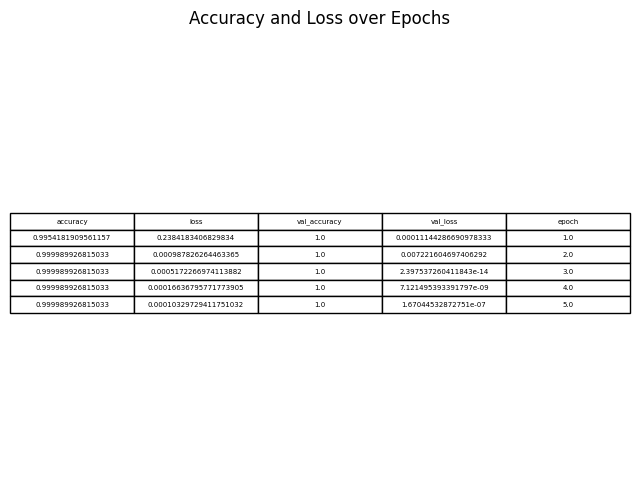
The model had a low loss and high accuracy for the training and testing data. The loss for the training data has a maximum of about 0.24 and a minimum of about 0 For the validation data, it has a maximum and minimum of about 0. The accuracy for the training and validation data has a minimum and a maximum of about 100 percent. Figure 1 shows the training data has a worse initial loss than the validation data. Figure 2 displays the initial accuracy of the training data was worse than the validation data.



**Figure 1 U-net model showing epochs on the x-axis and loss on the y-axis. The loss of the training data set is compared to the loss of the validation data set.**



**Figure 2 U-net model showing epochs on the x-axis and accuracy on the y-axis. The loss of the training data set is compared to the loss of the validation data set.**

****

**Figure 3 U-net model performance in terms of accuracy and loss for training and validation data over epochs**

**Discussion**

Overall the model was very accurate at classifying forests in environmental images. The model also had a low loss value. Higher accuracy values and lower loss values are indicators of better models. By being able to identify forests, it helps developers to better know where to do their construction and how to best preserve the environment. Due to the model already having a high accuracy rate, it is unlikely that other changes are needed to improve the model. However, due to not being able to see the saliency map, one can not determine if the model is examining the right features. There was an example from a study for identifying wolves, where the model identified wolves based on the snow on the background rather than on their features. This shows that while the model might be accurate, it is not learning the right features. In this case, since I could not make the saliency map, I can not determine whether the correct features were being learned. Also, potentially, it could struggle in places with different environments or lighting. None of the images were taken at night nor did they contain snow. Potentially, further improvement could be made to the model by providing more images of different environments.