Tree Detection

Data Science Institute
COLUMBIA UNIVERSITY

Juncheng Pan, Tianhao Wu, Yanbing Chen, Zehua Zeng, Zezhong Fan
Anjani Prasad Atluri, Harini Srinivasan
Aayush Kumar Verma, Sining Chen

Data Science Capstone Project with IBM

Introduction

Our project aims to detect trees from a geographical map. Based on tree detection result, further meaningful applications include tree health and species prediction, and carbon dioxide amount prediction. In this way, actions can be taken such as putting more taxation on companies that release higher amount of carbon dioxide, which protects the environment. Tree detection also helps to identify the percentage of trees cut down each year and necessary actions can also be taken from the result. Moreover, some utility companies use power lines which are close to trees. If trees are identified, companies will be able to plan ahead and avoid area with trees.

Methods and Challenges

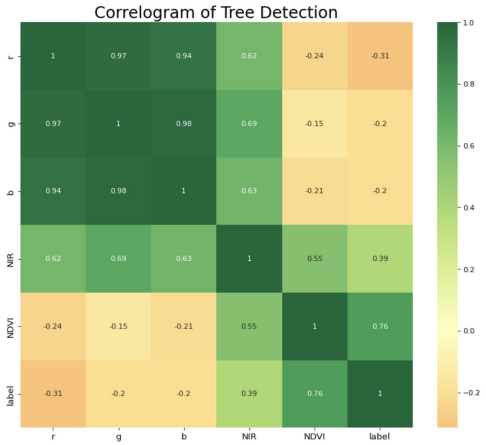

Figure 1. Correlations between features of input images. Color channels include red, green, blue, and Near-Infrared (NIR)

Models used:

- Random Forest
- ANN
- CNN

Challenges encountered:

- Huge Map Size
- Proper Metrics
- Confusion of Tree and Grass

Solutions:

- Picture Clipping
- Evaluate Model using F-1 Score and IoU Score
- Assign higher weights to grass samples

Results

The result of predicted tree label using CNN is shown in Figure 2. The training accuracy is 95% with 0.9532 F1 score. The test accuracy is 76%. The model has a good performance distinguishing between tree and grass. Compared to CNN, random forest has severe overfitting issues with a test accuracy of 70%.

The training accuracy for ANN achieves 94.33% and validation accuracy achieves 94.32%. The test accuracy achieves 80.2% on image of 1667x1668 pixels. ANN model can also differentiate grass and trees to a certain degree. However, ANN still have kind of overfitting problem and space to improve the performance.

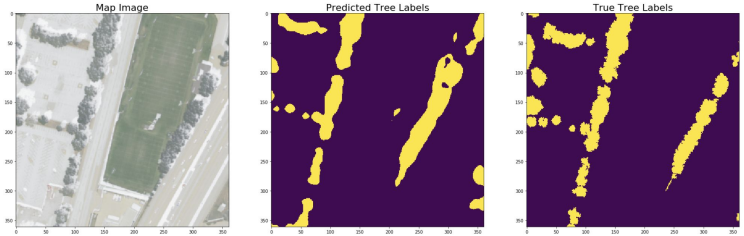

Figure 2. Predicted tree labels using Convolutional Neural Network (CNN). Yellow parts are tree labels.

Conclusions and Next Steps

Both random forest and ANN perform well in distinguishing between buildings and vegetations, but so far only CNN can distinguish between tree and grass by assigning higher sample weights to grass. The accuracy of CNN can be further improved if we can find a better threshold for separating tree and grass and adjust sample weight correspondingly.

Acknowledgments

We would like to thanks Data Science Institute and IBM company for the support of datasets and the project.

References

[1] GISGeography. (2022, May 30). What is NDVI (normalized difference vegetation index)? GIS Geography. Retrieved October 12, 2022, from https://gisgeography.com/ndvi-normalized-difference-vegetation-index.