CSCI 491/591 - Deep Learning - Final Report

Identifying Unmarked Trails Using GIS Data, Satellite Imagery, and Computer Vision:

Team Members:
Andrew Demaree

Team Number: Unknown

Abstract

This research project aims to identify unmarked mountain biking trails using GIS data, satellite imagery, and computer vision techniques. The analysis pipeline imploys image processing and deep neural network training and evaluation. Orthoimagery from the National Agriculture Imagery Program (NAIP) is processed to extract trail data, and high-resolution satellite imagery from the NAID is used to develop a custom implementation of DeepLabV3+ to detect and map potential unmarked trail systems. The model training process includes data preparation, model architecture selection, and hyperparameter tuning. Performance evaluation metrics such as Intersection over Union (IoU) and Frequency Weighted Intersection over Union (FWIOU) loss are used to evaluate the models performance. This work automates trail detection and contributes to sustainable trail management and access preservation. While specific results were not obtained due to computational constraints, the methods and pipeline components were individually validated and are ready for future implementation with more resources. The research offers a promising approach to address the challenges in accurate trail mapping and management, benefiting recreational activities and land managers.

Introduction

The purpose of this project is to develop an analysis pipeline to aggregate and analyze orthoimagery using machine learning, in order to identify potential unmarked or illegal trails used for mountain biking in the Pacific Northwest. The mountain biking community has spent the last two decades developing trails on both public and private land. Many of these trail systems have been left unmapped due to concerns about access, conflict with land managers, landowners, and public safety. However, in recent years there has been a push by local recreation managers to incorporate these trails into existing trail systems and to work with private landowners and local organizations to preserve access. The first step in this process is to identify and categorize trail systems that are not currently known to state and federal land managers. Traditional methods of mapping these trails, such as manual mapping or GPS, can be time-consuming, resource-intensive, and often yield inaccurate or unreliable results. This project aims to develop a machine learning model that leverages advanced computer vision techniques, such as object detection and semantic segmentation, to accurately identify and map unmapped mountain biking trails in the region. By utilizing high-resolution satellite imagery and comprehensive trail datasets, the proposed model will contribute to the ongoing efforts in sustainable trail management and preservation of access for recreational activities.

Background

Trail mapping and management have become essential components in preserving and maintaining natural landscapes while balancing recreational opportunities for a wide range of user groups, including mountain bikers. The accurate identification and mapping of existing and potential trail systems are crucial to land managers, local organizations, and private landowners in order to maintain sustainable access and minimize conflicts (Rails-to-Trails Conservancy, n.d.; U.S. Geological Survey, n.d.). Advancements in remote sensing and computer vision have enabled the development of new approaches to trail mapping using satellite imagery and machine learning techniques.

Mask R-CNN introduced by He et al. (2017), is a conceptually simple, flexible, and general framework for object instance segmentation. This model efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance. The method extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. Mask R-CNN has achieved good results when applied to instance segmentation, bounding-box object detection, and person keypoint detection, making it a promising tool for trail mapping applications.

In order to achieve pixel level resolution models such as DeepLabV3+, proposed by Chen et al. (2018) have been developed. It combines the advantages of spatial pyramid pooling module and encode-decoder structure for semantic segmentation tasks. By integrating multi-scale contextual information and gradually recovering spatial information, DeepLabV3+ captures sharper object boundaries and has demonstrated strong performance on the PASCAL VOC 2012 and Cityscapes datasets. This model can potentially be employed to enhance the accuracy of trail mapping and segmentation.

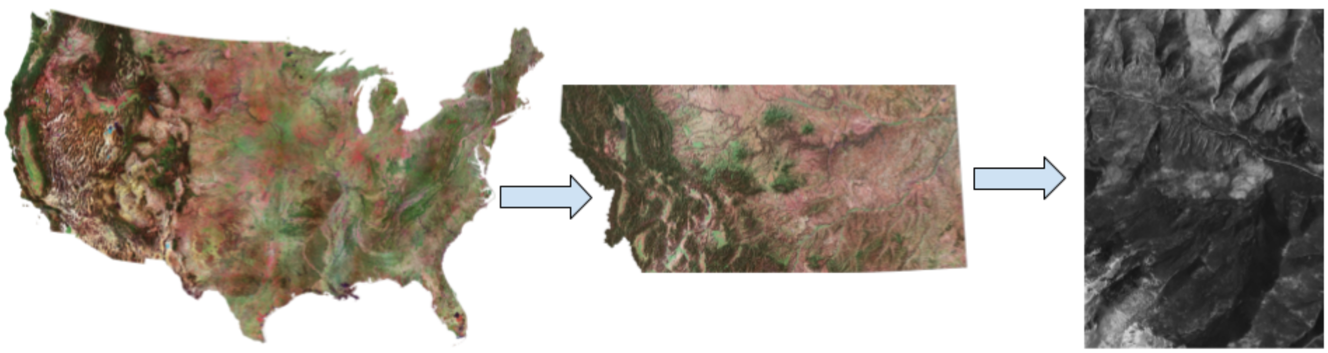
In the context of forest management, Jayathunga et al. (2019) demonstrated the potential of digital aerial photogrammetry (DAP) for assessing uneven-aged forest resources. They tested the performance of biomass estimation by varying the conditions of several factors, such as image downscaling, vegetation metric extraction, modeling method, and season. Their findings indicated that the performance of forest biomass estimation for uneven-aged forests varied with statistical representations and data sources, which suggests the importance of exploring different statistical approaches and data sources in trail mapping applications.

Finally, Stewart et al. (2020) has employed a deep learning model based on the U-Net architecture for road detection and monitoring in desert dune fields using Copernicus Sentinel-1 synthetic aperture radar (SAR) satellite data and Open Street Map (OSM) data for model training. Their methodology comprised two steps: processing time series of Sentinel-1 SAR acquisitions to produce multitemporal backscatter and coherence averages, and implementing the U-Net deep learning workflow. The approach achieved a performance evaluation of 84% to 89% in terms of Jaccard similarity coefficient across different dune fields in Africa and Asia, demonstrating its potential for similar applications in trail mapping.

This project aims to develop a machine learning model that leverages the capabilities of object detection and semantic segmentation algorithms, such as Mask R-CNN and DeepLabV3+, to identify potential mountain biking trails not currently mapped by state or federal entities in the US, using National Forest Service trail data and European Space Agency Sentinel-2 satellite imagery. By employing advanced computer vision techniques and referencing previous research findings, this project will contribute to the ongoing efforts in sustainable trail management, trail identification, and the preservation of access for recreational activities.

Data

For this project it was necessary to obtain high resolution orthoimagery for land areas across the entire United States. To achieve this goal tiled jpeg2000 images, with pixel resolutions between 1 and ½ meter, were obtained from the US National Map Website spanning the entire United States. Each tile is based on a 3.75 minute x 3.75 minute quarter quadrangle with a 300-pixel buffer on all four sides and is allowed to contain as much as 10% cloud cover. The tiles are geospatially referenced using the Universal Transverse Mercator (UTM) coordinate system based on the North American Datum of 1983 geodetic datum. In order to label the images with coordinates for existing trail systems, shapefiles for each of the US states were obtained from the USGS National Transportation Dataset. These shape files contain marked features for all transportation methods within each state. The trail features were extracted from these shapefiles and the statewide files were converted to individual shapefiles matching the dimensions of the orthoimagery tiles. Only orthoimages and their corresponding shape files were retained if the shapefiles for the geospatial region contained previously mapped trail features.



50 polygons (one for each state) x ~4,706 tiles per state = ~235,300 tiles total for the United States

Figure 1. A visualization of the approximate number of high resolution tiles contained in the dataset. It is important to note that this is a highly conservative estimate and the actual tile number is much larger because this estimate was calculated using the average state area and does not accurately account for extremely large states such as Texas and alaska.

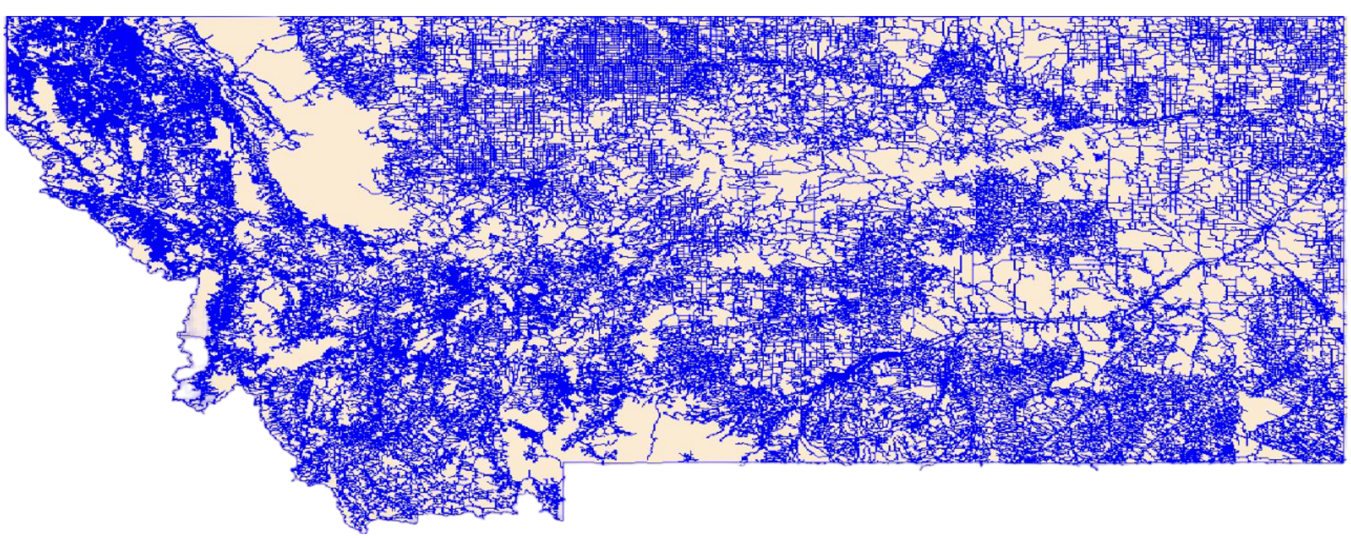


Figure 2. A visualization of the transportation features contained in the shapefile for the state of Montana.

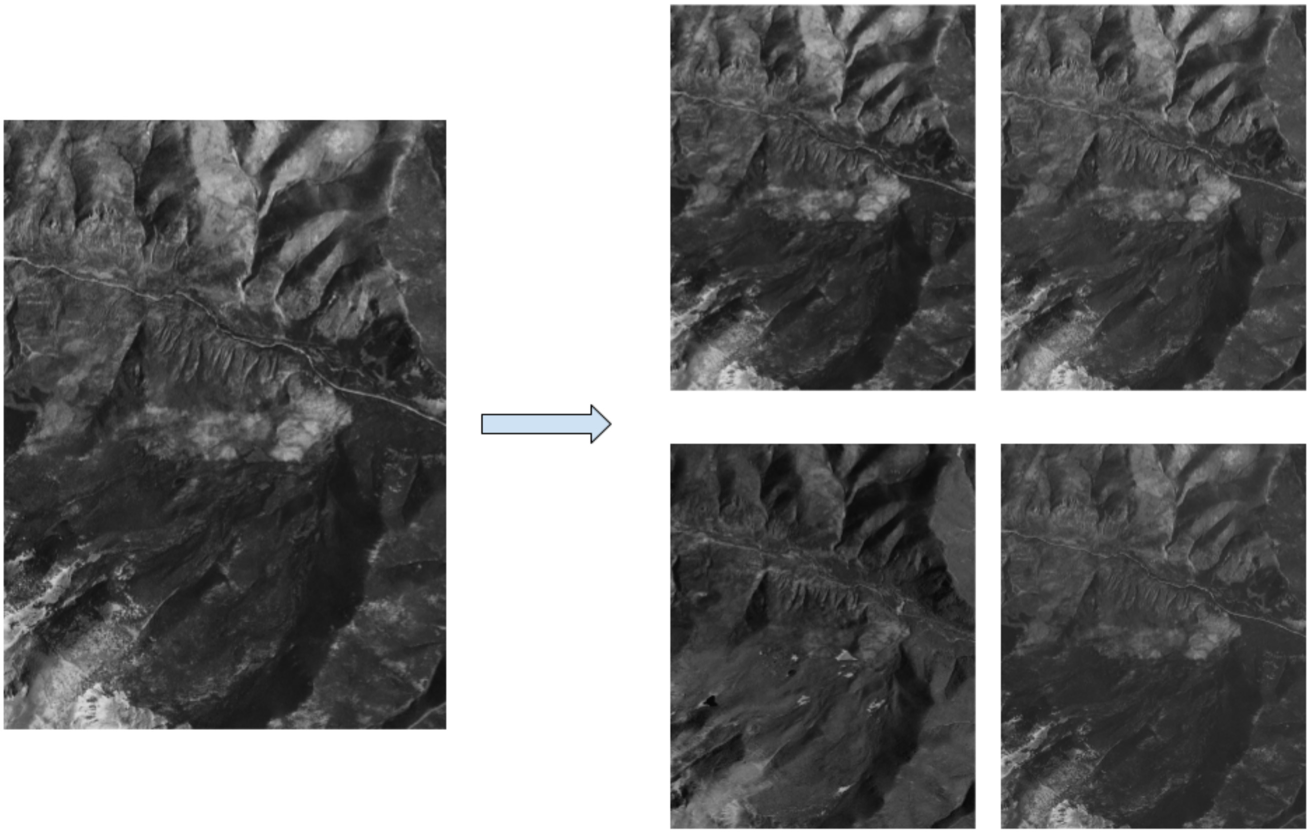


Fig 3. This figure represents the conversion from the original four channel image to four single channel mean normalized images of the same geospatial region.

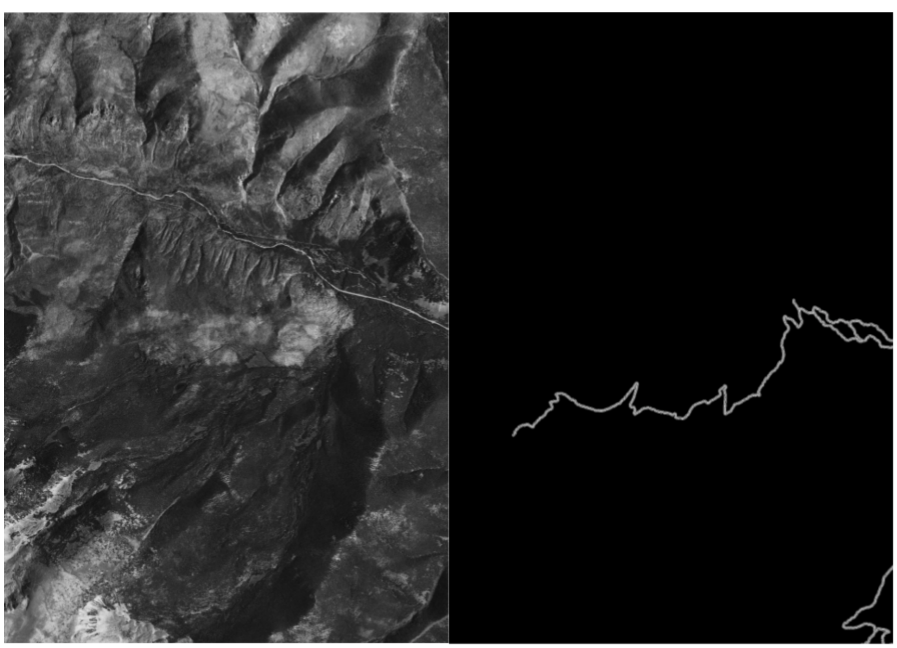


Fig 4. A side-by-side comparison of an original image tile and its corresponding binary mask showing the trail features in the image.

Experiments

The primary purpose of this project was to develop a pipeline for efficiently downloading orthoimages from the National Agriculture Imagery Program (NAIP), labeling the images using shape files for the trail networks mapped by the USGS across the United States, and training a Deep convolutional neural network designed for object detection and image segmentation to detect unmarked, mapped, or illegal trail systems. The experimental methods can be separated into two classes; image processing and model training and evaluation.

Image Processing:

The image processing required for this project focused on preparing orthoimages sourced from the NAIP using The National Map image downloader. The goal was to optimize the images for use with a deep learning model that incorporates pyramid pooling and atrous convolution techniques.

First, the orthoimage coordinates, which are geographical reference points for each image, were extracted from each image tile.The JSON formatted metadata for each tile is read in order to extract the bounding box metadata from them. The bounding box is used to determine the area of interest within the image, which is crucial for preparing data for the machine learning model. If no JSON files are found or the index is invalid then the tile was not used for further analysis. Next, The trail data is processed by reading shapefiles specific to each state and a subset of the GeoDataFrame (gdf) that falls within the previously obtained orthoimage coordinates is constructed. This subset represents the filtered trails within the area of interest identified by the bounding box coordinates. The resulting feature subset is analyzed to calculate the total trail length within the area of interest defined by the new shapefile. This calculation is performed by measuring the length of the geometry of each trail in the shapefile and summing them. Additionally, the total number of trails is computed. These calculations are then used to determine the orthoimagery tiles to be included in the final dataset.

To ensure optimal data preprocessing for the DeepV3+ deep neural network, the orthoimagery data is normalized. Normalization is crucial for standardizing the range of pixel intensity values. The original four channel JPEG2000 images are converted into tensors and normalized to have values between 0 and 1. The normalized data is then saved as individual single channel JPEG images and saved. The next step of the image processing pipeline involves creating binary masks from each of the shapefiles. Binary masks are created by rasterizing the shapefiles, resulting in an image where the trails are represented as 1s and the background as 0s. This mask serves as the ground truth for the model. The normalized single channel images and their corresponding binary masks are then aggregated, so they can be loaded into the model. This comprehensive image processing pipeline prepares the geospatial imagery data so that it is in a format that can be easily processed by the model. It involves extracting orthoimage coordinates, processing trail data, calculating trail lengths, normalizing orthoimagery data, and creating binary masks.

Model Training and Evaluation:

The model training and evaluation process consists of several experimental steps aimed at achieving effective performance and accuracy. In the first step, data preparation and preprocessing are conducted. A custom dataset class is developed to handle the preprocessed orthoimages and binary masks of trail systems. This class incorporates methods to process and transform the images, such as resizing and converting them into tensors, thus ensuring that the input data fits the format specifications for the models inputs. The architecture of the model is based on a modified DeepLabV3+ architecture, which is a convolutional neural network specifically designed for semantic image segmentation. To enhance its capabilities in detecting trails in diverse and potentially complex images, the architecture is enriched with atrous (dilated) convolutions and a pyramid pooling module. The atrous convolutions enable the model to operate at multiple scales, while the pyramid pooling module captures various levels of contextual information.

To train the model, a combination of Cross-Entropy loss and frequency weighted intersection over union (FWIOU) loss is employed. Cross-Entropy loss is commonly utilized in classification tasks, while FWIOU is a modification of the Intersection over Union (IoU) metric, tailored for evaluating segmentation tasks. Throughout the validation phase, the IoU metric is also utilized to assess the model's performance. To identify the optimal set of hyperparameters that maximize model performance, a grid search is performed over a specified range for each of the hyperparameters. These include parameters such as learning rate, batch size, number of epochs, network depth, number of levels in pyramid pooling, dilation rate, step size, gamma for learning rate scheduling, and kernel size. The grid search exhaustively explores these options to minimize validation loss and maximize the IoU metric, thereby optimizing the model's performance.

The model is trained using the Adam optimizer, with a learning rate scheduler employed to adapt the learning rate during the training process. To prevent overfitting, the model's performance is evaluated on a separate validation dataset. The training process involves forward and backward propagation, as well as parameter updates, to optimize the model's performance. After completing the grid search, the results are evaluated to select the set of hyperparameters that yield the best performance based on both the validation loss and IoU metric. The hyperparameters associated with the model achieving the lowest validation loss and the highest IoU are reported separately, providing insights into the trade-off between these two performance measures.

The significance and practical implications of this model and procedure lie in automating the process of trail detection in orthoimages. The aim is to optimize the model's ability to detect even unmarked trails, thereby enhancing the comprehensiveness and accuracy of the mapping output. Through preprocessing, modifying the model architecture, selection of appropriate loss functions and metrics, conducting hyperparameter tuning, training and validation, and evaluation the best-performing model can be selected.

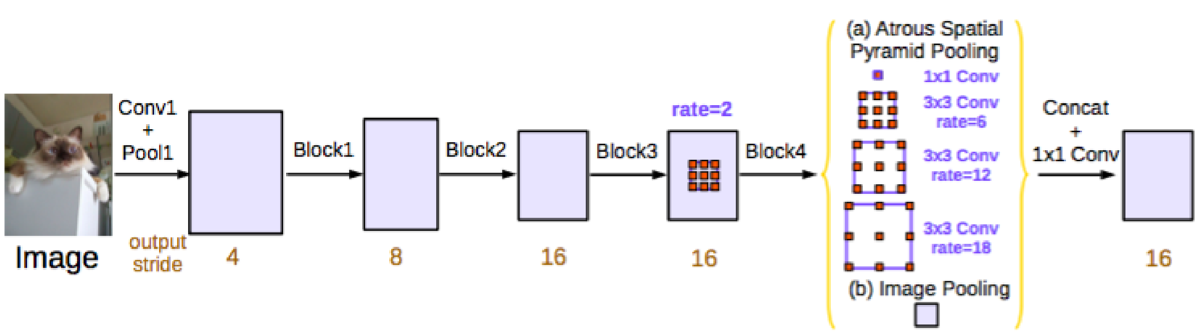


Figure 5. Basic principles of Atrous convolution and Pyramid Pooling. Adapted from "Rethinking Atrous Convolution for Semantic Image Segmentation," by L. Chen, G. Papandreou, Schroff, and H. Adam.

Results

Due to computational constraints no results were collected using this analysis pipeline. However the methods and classes developed in this pipeline were individually validated and will work for the intended dataset, given more time and computational resources. Efforts to increase the efficiency of the pipeline and acquire GPUs and increased storage space in order to perform a complete analysis are underway with the intention of obtaining and publishing the results. The Equations for the validation used to determine the optimal hyperparameters and the hyperparameters used in the gridsearch are listed below.

$$FWIOU\ LOSS = 1 - \sum(2 \times frequency \times intersection) / \sum(frequency \times union) + \varepsilon)$$

Frequency: The frequency or weight of each class in the dataset.
intersection: The pixel-wise intersection between the predicted and ground truth segmentation masks for each class.
Union: The pixel-wise union between the predicted and ground truth segmentation masks for each class.
ε : A small constant added for numerical stability to avoid division by zero errors.

$$Loss = \frac{1}{N} \times \sum(1 - (IA_i / UA_i)$$

N: The total number of samples or regions.
Intersection Area_i: The area of overlap between the predicted region and the ground truth region for the i-th sample.
Union Area_i: The combined area of the predicted region and the ground truth region for the i-th sample.

Hyperparameter	Range
learning_rate	[1e-5, 1e-4, 1e-3]
batch_size	[16, 32, 64, 128, 256]
epochs	[50, 100, 150, 200]
network_depth	[16, 34, 50, 101, 152]
num_levels	[2, 3, 4, 5]
dilation	[(1, 1), (2, 2), (4, 4)]

lr_step_size	[10, 20, 30]
lr_gamma	[0.1, 0.5, 0.9]
kernel_size	[2, 3, 4, 5]

Figure 6. The hyperparameters and their associated ranges to be used in a grid search to identify the optimal hyperparameters for the modified DeepV3+ Deep Convolutional Network.

Conclusions

This research project aimed to address the need for identifying unmarked mountain biking trails by leveraging GIS data, satellite imagery, and computer vision techniques. The proposed analysis pipeline outlined above, uses image processing in combination with deep neural network training to identify fine scale geospatial features. Despite not obtaining specific results due to computational limitations, the methods and classes developed within the pipeline were individually validated and demonstrated their suitability for the intended dataset. Future efforts will focus on increasing efficiency, acquiring additional computational resources, and completing a comprehensive analysis to obtain and publish the results.

This project highlighted the potential of using deep learning models, such as Mask R-CNN and DeepLabV3+, for trail detection and mapping. By leveraging advanced computer vision techniques, high-resolution satellite imagery, and comprehensive trail datasets, the proposed pipeline offers a promising approach to automate the trail identification process. The developed custom dataset class, model architecture modifications, and hyperparameter tuning demonstrate the potential to optimize model performance for this type of specific task. The findings of this project have practical implications for sustainable trail management and the preservation of access for recreational activities. Accurate identification and mapping of unmarked trail systems are essential for land managers, organizations, and private landowners to mitigate conflicts and incorporate these trails into existing trail networks. By automating the process and improving accuracy through computer vision, this research contributes to the ongoing efforts in trail management and enhances the comprehensiveness of trail mapping outputs.

Roles

All roles for this project were carried out by Andrew Demaree

References

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for Semantic Image segmentation. Computer Vision – ECCV 2018, 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

Conservancy, R.-to-T. (n.d.). Trail mapping and GIS: Rails-to-trails conservancy. Rails. Retrieved April 19, 2023, from https://www.railstotrails.org/our-work/trail-mapping-and-gis/

Finding yourself outdoors: U.S. geological survey. Finding Yourself Outdoors | U.S. Geological Survey. (n.d.). Retrieved April 19, 2023, from https://www.usgs.gov/news/national-news-release/finding-yourself-outdoors

Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2015.169
He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2017.322
Jayathunga, S., Owari, T., & Tsuyuki, S. (2019). Digital aerial photogrammetry for uneven-aged forest management: Assessing the potential to reconstruct canopy structure and estimate living biomass. Remote Sensing, 11(3), 338. https://doi.org/10.3390/rs11030338
PYTORCH documentation¶. PyTorch documentation - PyTorch 2.0 documentation. (n.d.). Retrieved April 19, 2023, from https://pytorch.org/docs/stable/index.html