Identification of Tree Canopy in Different Challenging Backgrounds from UAV Imagery Using Deep Learning

Tapos Biswas
Class Roll: JN-092-004
Registration No: 2018-326-855
Session: 2018–2019

Ipshita Ahmed Moon
Class Roll: SK-092-017
Registration No: 2018-926-868
Session: 2018–2019

A Project submitted for the degree of
B.Sc. Engineering in Robotics and Mechatronics Engineering

Department of Robotics and Mechatronics Engineering
University of Dhaka, Dhaka-1000, Bangladesh

December 2023

Supervisor:

Md. Shifat-E-Arman Bhuiyan
Lecturer
Department of Robotics & Mechatronics Engineering
University of Dhaka

Contents

List of Figures

List of Tables

Abstract

Accurate identification of tree species is crucial for effective forest operations and management. Unmanned aerial vehicles (UAVs) equipped with high-resolution imagery have emerged as indispensable tools for foresters. Deep learning techniques, notably Convolutional Neural Networks (CNNs), are widely adopted for feature extraction and classification. However, challenges persist, particularly concerning the complexity of data and software performance.

Recent advancements in deep learning methods, especially in object detection and segmentation, have sparked interest in their application for identifying tree canopies. In this study, we propose and evaluate the integration of CNN-based algorithms, specifically Faster R-CNN, Mask R-CNN, and YOLOv5, with high-resolution RGB data acquired from UAVs. Furthermore, we explore the potential of DeepLabV3+ and U-net, demonstrating their superior performance in landcover segmentation. Additionally, we incorporate the innovative Segment Anything Model (SAM) for enhanced landcover exploration.

To comprehensively assess our approach, we leverage four publicly available datasets: Tree Species Detection, NeonTreeEvaluation Benchmark, and Forest Damages - Larch Casebearer and DeepGlobe Land Cover. The experimental analysis yielded promising results, with mean average precision (calculated at IoU threshold 0.5) reaching approximately 79%, 58.4%, and 76.5% for the first three dataset and 99.76%, 99.96% accuracy on DeepGlobe land cover dataset. The associated processing times were remarkably efficient, ranging from approximately 0.5 to 1.6 milliseconds.

In summary, this study presents a holistic exploration of employing deep learning algorithms in conjunction with high-resolution UAV imagery for tree canopy and landcover identification. We demonstrate the effectiveness of Faster R-CNN, Mask R-CNN, and YOLOv5 algorithms across various datasets, shedding light on advancing forest monitoring and management practices. Furthermore, the incorporation of the Segment Anything Model (SAM) enriches our understanding of landcover dynamics, augmented by the superior performance of DeepLabV3+ and U-net in landcover segmentation, enhancing the potential for informed decision-making and sustainable resource management.

Chapter 1

Introduction

Tree canopy identification is a task that involves extracting the canopy of trees from remote sensing data such as aerial imagery, satellite imagery, or LiDAR data. The objective is to identify and delineate the boundary of individual trees, which is important for estimating their biomass, monitoring forest health, and assessing the effects of climate change on vegetation.

In recent years, tree canopy identification has gained significant attention due to its potential applications in agriculture, ecological monitoring, carbon sequestration, and forest management. Accurate and efficient methods for tree canopy identification can provide valuable insights into the forest structure and function, which are essential for understanding the ecological processes that govern forest ecosystems.

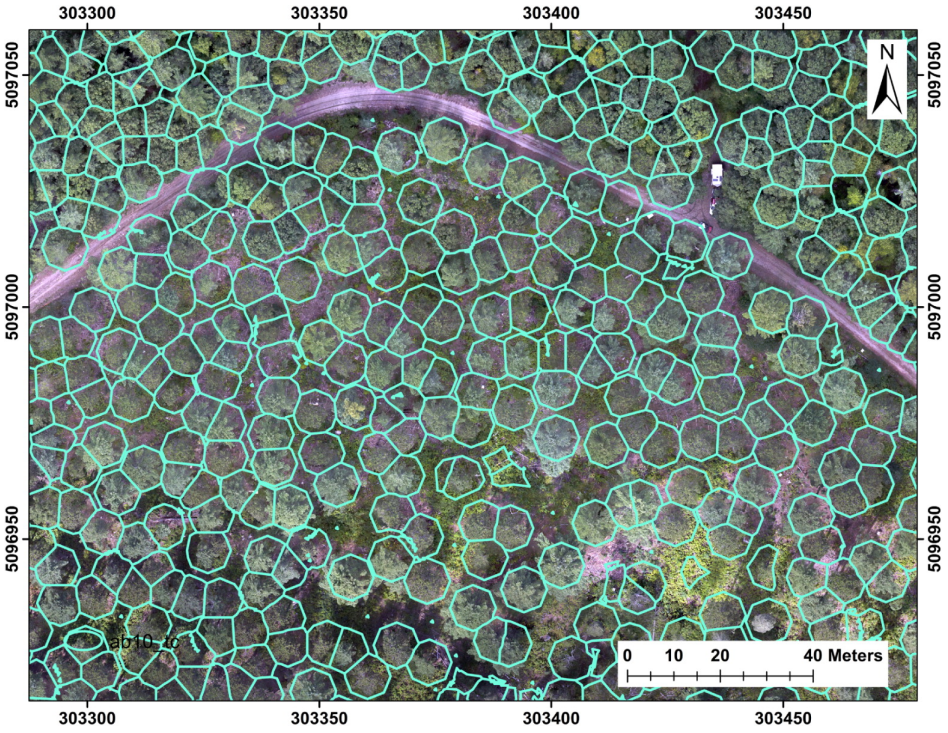

Figure 1.1: Tree Canopy Identification [22]

There are several challenges associated with tree canopy identification, including variations in tree species, leaf density, and background clutter. Additionally, the quality of the remote sensing data, such as resolution, noise, and atmospheric interference, can affect the accuracy of the identification results. Therefore, developing a robust and accurate method for tree canopy identification requires careful consideration of these factors.

In recent years, various approaches have been proposed for tree canopy identification, including machine learning-based methods or deep learning-based methods, rule-based methods, and hybrid methods. Deep learning-based methods, such as convolutional neural networks (CNNs) and random forests, have gained popularity due to their ability to learn complex features and patterns from data. Rule-based methods, such as mathematical morphology and region growing, rely on predefined rules to identify the canopy. Hybrid methods combine the advantages of both approaches and have shown promising results in tree canopy identification.

Figure 1.2: Unmanned Aerial Vehicle [23]

Unmanned aerial vehicles or drones can be used to capture the images. In recent years, UAVs have gained popularity in precision agriculture due to their ability to provide valuable insights and data. UAVs can help us make better decisions, improve productivity, and reduce costs. As the technology continues to improve and become more affordable, we can expect to see even more applications for drones in agriculture in the future.

1.1 Problem Definition

The main challenge addressed in this study is the accurate identification of tree canopies in various challenging backgrounds using UAV imagery. Traditional remote sensing techniques often fail to distinguish tree canopies from complex backgrounds, leading to inaccurate assessments. This research aims to develop a deep learning-based approach to overcome these limitations and improve the accuracy and efficiency of tree canopy identification in various challenging backgrounds.

1.2 Motivation

The significance of this research lies in its potential to advance our understanding of tree canopy identification and provide more accurate and efficient tools for environmental monitoring and management. Accurate tree canopy identification is crucial for monitoring forest health, understanding habitat distribution, and tracking changes in land cover. Furthermore, it supports sustainable forest management practices, contributing to climate change mitigation and preserving biodiversity. By leveraging deep learning techniques and UAV imagery, this research seeks to overcome the limitations of traditional remote sensing methods and enhance the accuracy and effectiveness of tree canopy identification in various challenging backgrounds.

1.3 Objective

The objectives of this project are as follows:

- To create a deep learning-based technique that can precisely locate tree canopies in multi-spectral UAV data.
- to improve its adaptability to various situations.
- To provide insights into the difficulties and potential of applying deep learning methods for identifying tree canopies in various difficult contexts.
- To contribute to the advancement of state-of-the-art remote sensing and deep learning techniques for tree canopy identification.

Chapter 2

Related Work

This literature review focuses on the identification and segmentation of tree canopies in difficult backgrounds using deep learning algorithms. The findings of this study will help enhance tree canopy mapping for environmental monitoring and decision-making.

2.1 Comparison of Unsupervised Algorithms for Vineyard Canopy Segmentation from UAV Multispectral Images [1]

2.1.1 Description

The focus of this research paper is the development and comparison of unsupervised algorithms for canopy segmentation in UAV-acquired imagery for precision agriculture applications, particularly in vineyards. This study develops and compares unsupervised canopy segmentation algorithms for UAV-collected vineyard precision agriculture data. HSV-based, DEM-based, and K-means unsupervised algorithms are recommended for visible spectrum or multispectral sensor data. RGB (Red-Green-Blue) and NRG (Near Infrared-Red-Green) photos from three different circumstances from two vineyards across two years (2017 and 2018) were utilized to assess the algorithms. Overestimation and underestimation indices

assessed the algorithms' vine identification accuracy. HSV-based methods overestimate RGB vegetation compared to DEM and K-means algorithms. HSV is more consistent for NRG pictures than DEM, which slightly overestimates plant coverage. K-means computational load increases as DEM quality falls, unlike DEM and HSV-based methods. Overestimation and underestimation indices were developed to quantify the algorithms' capacity for accurately identifying vines. HSV-based algorithms consistently overestimate vegetation in RGB imagery, whereas DEM and K-means algorithms tend to underestimate them. The HSV algorithm is more stable for NRG imagery than the DEM model, which slightly overestimates the amount of vegetation. the HSV-based algorithm and DEM algorithm have comparable computational times, whereas the K-means algorithm's computational demand increases as DEM quality diminishes.

2.1.2 Strengths

- Proposed unsupervised algorithms for canopy segmentation that can be utilized in precision viticulture applications for efficient crop management.
- Compared the performances of various algorithms for identifying vine canopies from multispectral UAV imagery.
- Introduced the concept of over-estimation and under-estimation indices to quantify the algorithms' ability to accurately identify vines under varying acquisition conditions.

2.1.3 Weaknesses

- The paper does not provide a comprehensive analysis of the limitations of the proposed algorithms in real-world scenarios or on a large scale.
- The paper acknowledges that the computational requirements of the k-means algorithm limit its utility for real-time identification in autonomous UAV monitoring. However, no specific solution or the alternative method is proposed to address this limitation.

- The DEM generation is a computational bottleneck, and its quality influences the identification of vines. It does not, however, provide a comprehensive analysis of the errors that may occur in the DEM model and their impact on the algorithm's accuracy.

2.2 Semantic Segmentation of Tree-Canopy in Urban Environment with Pixel-Wise Deep Learning [2]

2.2.1 Description

Using high-resolution aerial images, the study employs deep learning techniques to identify tree canopies in urban environments. The proposed method classifies tree-canopy and non-tree pixels pixel-by-pixel using a convolutional neural network (CNN). The authors apply their method to a dataset of aerial images from Vancouver, Canada, achieving high accuracy and demonstrating the potential of deep learning for segmenting tree canopies in urban environments.

2.2.2 Strengths

- Achieved high accuracy in semantic tree-canopy segmentation in urban contexts.
- The use of pixel-wise deep learning techniques facilitates the classification of individual pixels at a finer level, allowing for more precise identification of the tree canopy.
- The proposed method is scalable and can be applied to massive aerial image datasets.
- Addressed a significant environmental issue that may have applications in urban planning and management.

2.2.3 Weaknesses

- The proposed method requires the use of high-resolution aerial pictures, which may not be readily available in all circumstances.
- The study does not compare the proposed method to other state-of-the-art techniques for semantic segmentation of tree canopy.
- The study is limited to a specific geographical location and may not apply to other urban environments.

2.3 Canopy Segmentation and Wire Reconstruction for Kiwifruit Robotic Harvesting [3]

2.3.1 Description

The paper describes a method for segmenting kiwifruit canopies and cables to enhance robotic harvesting. The authors segmented fruit calyxes, branches, and wires in 327 images using DeepLabV3+. They devised a method for reconstructing discrete wire pixels utilizing the Progressive probabilistic Hough transform (PPHT) and removed noise by filtering lines that did not meet constraints. The training set was expanded to 1,566 images, and the authors compared the efficacy of uniform weights versus median frequency weights on the imbalanced dataset. Figure 2.1 demonstrate the block diagram of DeepLabV3+ for kiwifruit canopy image segmentation.

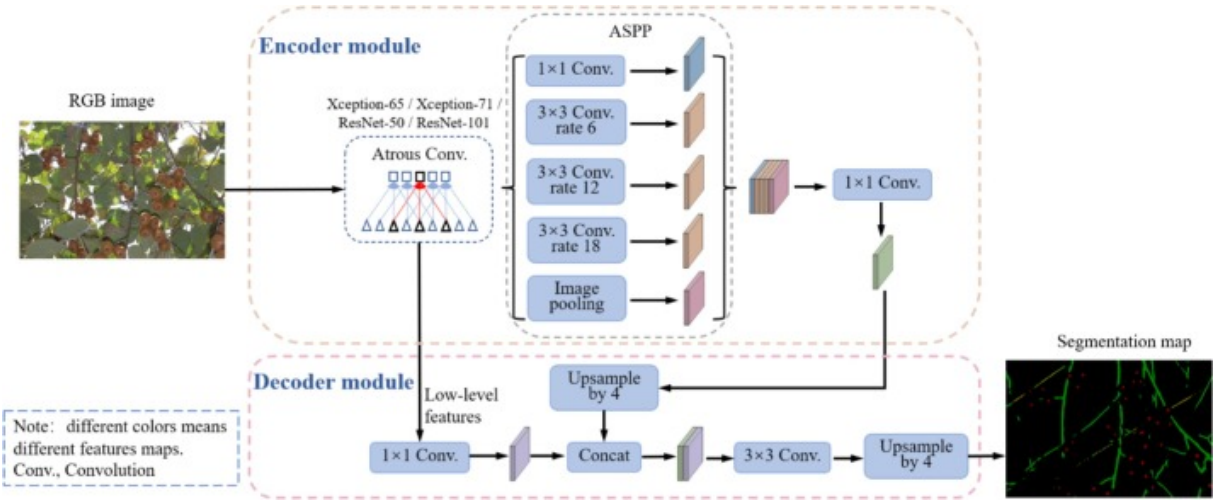

Figure 2.1: DeepLabV3+'s block diagram for segmenting kiwifruit canopies [3]

2.3.2 Strengths

- The use of a large and diverse dataset for training and testing, including imbalanced and occluded images, increased the applicability of the results to real-world scenarios.
- The use of DeepLabV3+ as a semantic segmentation network and ResNet-101 as the best backbone resulted in a high degree of accuracy when segmenting calyx, branch, and wire.
- The line-detection method based on PPHT was effective for reconstructing wires on segmentation maps with a high correct detection rate and rapid processing speed.

2.3.3 Weaknesses

- The article did not provide a comprehensive comparison with other cutting-edge semantic segmentation techniques for kiwifruit plantations.
- The paper lacked a comprehensive analysis of the effect of variable factors, such as illumination conditions and image resolution, on segmentation performance.

- The paper didn't talk about the suggested method's flaws or the problems that might come up when it's used in the real world.

2.4 Canopy Segmentation using ResNet for Mechanical Harvesting of Apples [4]

2.4.1 Description

This research paper concentrates on improving the efficiency of mechanical apple harvesting by analyzing tree canopies using semantic segmentation based on convolutional neural networks (CNN). The study utilized a pre-trained and modified ResNet-18 implementation of CNN and a Kinect V2 camera to acquire 253 images in a commercial "Fuji" apple orchard in Washington state. The images were divided into three classes of pixels, including trunk/branch, apples, and foliage, and the model's performance was assessed using three evaluation metrics. The results demonstrated high per-class accuracy for apples and leaves, but poor accuracy for trunk/branch due to a reduced proportion of pixels in each image. The study suggests that automating the location and swaying of trunks/branches under a dense canopy of leaves during apple harvesting seasons could increase the efficiency of mechanical apple harvesting techniques.

2.4.2 Strengths

- The proposed method is based on ResNet-18, a state-of-the-art deep learning architecture that has demonstrated superior performance in various computer vision tasks.
- The proposed method's accuracy and applicability were assessed using data from a real-world dataset gathered from a commercial apple orchard.
- The evaluation metrics utilized in this study are widely accepted metrics for semantic segmentation tasks, allowing for an objective and meaningful evaluation of the proposed method.

2.4.3 Weaknesses

- The proposed method was evaluated on a relatively small dataset of 253 images, which may not adequately represent the variations in the apple tree canopy under varying environmental conditions and geographical locations.
- The proposed method didn't work as well for the "trunk/branch" class as it did for the other two classes which may limit the applicability of the proposed method in practice.
- For the "trunk/branch" class, the suggested method may need to be improved and optimized more before it can be used to segment trees more accurately and quickly.

2.5 Deep Learning-Based Instance Segmentation Method of Litchi Canopy from UAV-Acquired Images [5]

2.5.1 Description

Using images taken by unmanned aerial vehicles, the research study suggests a deep learning-based instance segmentation strategy for litchi trees. (UAVs). In comparison to other approaches, this one has a more straightforward structure and fewer restrictions for the data form that must be used. While a partition-based method is utilized for segmenting high-resolution digital orthophoto maps (DOMs), a semi-automatic image annotation technique is developed to improve the efficiency of data pre-processing. This helps address the high computing needs. The lack of diversity that was present in the initial litchi dataset can be remedied by including data on citrus fruits in the training set.

2.5.2 Strengths

- The semi-automatic annotation method considerably increased the efficiency of data pre-processing and decreases labor requirements.

- The partition-based method permitted the processing of high-resolution DOMs and the incorporation of patch inference resulted in a unified segmentation result.
- The addition of citrus data to the training set improved dataset diversity and model performance.
- On the test set, the model obtained high precision, indicating effective instance segmentation.

2.5.3 Weaknesses

- The partition-based strategy raises computational costs, which may be difficult for some users.
- In certain instances, the labor-saving semi-automatic annotation method may still require substantial manual labor.
- The performance of the method under varying environmental conditions or in the presence of occlusions, which may impact real-world applications, is not addressed in the paper.

2.6 Image Segmentation of UAV Fruit Tree Canopy in a Natural Illumination Environment [6]

2.6.1 Description

The authors of this study propose an unsupervised image segmentation algorithm for the canopy of fruit trees under natural illumination conditions to improve pesticide application in plantations. Combining the shadow region luminance compensation method (SRLCM) with ensemble clustering yields high-quality color information for image segmentation using the suggested method. According to experimental data, the proposed technique outperforms widely used unsupervised methods such as K-means and GMM in terms of accuracy rate, recall rate, and

F1 score. The proposed technique can effectively minimize the disparity between shaded and unshaded canopies, resulting in a more comprehensive and accurate segmentation of the canopy as a whole. The research demonstrates the potential of computer vision technology in agriculture, and the proposed solution reacts faster than the GMM algorithm.

2.6.2 Strengths

- Combined SRLCM and ensemble clustering to obtain high-quality color features for image segmentation, which can enhance segmentation accuracy in environments with natural illumination.
- In terms of precision rate, recall rate, and F1-score, the suggested method outperformed commonly used unsupervised methods such as K-means and GMM.
- The proposed method takes slightly longer to respond than the K-means algorithm.
- The sample size used in the experiments is relatively small, which may affect the generalization of the proposed method to other orchards.

2.6.3 Weaknesses

- The tiny amount of the dataset utilized to test the suggested strategy might have constrained the study. A larger and more varied dataset might offer more details on the algorithm's performance in various settings.
- The paper compares the proposed unsupervised method solely with other unsupervised methods and not with supervised methods. A comparison with supervised methods could provide additional insight into the efficacy of the proposed method.

2.7 Object-based classification of urban plant species from very high-resolution satellite imagery [7]

2.7.1 Description

Utilizing extremely high-resolution satellite data, this study identifies, delineates, and classes urban plant species. The investigation will precisely characterize the location, species, and structure of the trees. Field operations are costly and data on residential gardens is scarce, so systematic data collection may be difficult. This study developed an object-based classification technique that employs new pertinent spectral and texture-based properties for each plant species to surmount the challenges posed by urban environments, species diversity, and tree proximity. For object-based classification of WorldView-2 satellite data, four spectral bands (blue, green, yellow, and red) and four textural qualities were most useful. In two study locations, a Random Forest classifier and ground validation identified 22 and 20 plant species, respectively. The research demonstrates that extremely high-resolution satellite imagery and object-based classification can accurately identify, demarcate, and classify urban plant species for urban planning and management.

2.7.2 Strengths

- Identified the most effective spectral and texture-based features for object-based classification using an object-based classification strategy.
- The possibility of satellite imagery of very high resolution for accurately defining urban vegetation is identified.

2.7.3 Weaknesses

- Possibility of misclassification due to the urban environment's complexity and the variety of plant species.
- The study only covers two different study regions, hence the results may not apply to other urban areas.

2.8 Computer Vision System for Detecting Orchard Trees from UAV Images [8]

2.8.1 Description

The authors of this paper investigate the use of deep-learning models for detecting apple trees in orchards using images captured by unmanned aerial vehicles (UAVs). The paper presents a novel approach to orchard tree inventory that aims to address the limitations of traditional manual methods. The paper highlights the potential of using UAVs and deep-learning models to automate orchard tree inventory and improve the accuracy and efficiency of crop monitoring. The paper provides valuable insights into the performance of two state-of-the-art tree detection methods, YOLO-V5 and DeepForest, through both qualitative and quantitative assessments. The authors show that DeepForest outperforms all of the tested models, achieving a high level of accuracy in detecting orchard trees from UAV images. In summary, the paper presents a significant contribution to the field of precision agriculture and computer vision. The research demonstrates the potential of using deep-learning models for detecting orchard trees from UAV images and provides insights into the strengths and weaknesses of the tested models. The paper also highlights avenues for future research to address the limitations and challenges of the proposed system, including scalability and dataset construction

2.8.2 Strengths

- To identify apple trees in an orchard using UAV RGB data, we examined two state-of-the-art techniques: YOLO and the DeepForest model.
- Presents a new dataset that can be used for further research in the field of orchard tree detection.

2.8.3 Weaknesses

- Experiment conducted on a limited dataset and the performance of the models on different types of orchards or crops is not investigated.

- Manual labeling of images used to construct the dataset can be time-consuming and costly, limiting the scalability of the proposed system.
- No comparison with other state-of-the-art methods for orchard tree detection, limiting the generalizability of the findings.

2.9 Identifying Streetscape Features Using VHR Imagery and Deep Learning Applications [9]

2.9.1 Description

This study looks at how deep learning and Earth observation datasets may be used to find fine-grained urban components that characterize streetscapes. The study focuses on recognizing eleven streetscape characteristics. The approach makes limited use of manual annotation development for model training and instead relies on publicly available data sources. The training dataset is manually produced, and the models are evaluated using a test dataset taken from the study area, as well as a model-specific evaluation of the test set of data. The results of these models are combined to form a geographical dataset and 3D views and street cross-sections for the city. The findings of the study are particularly relevant to the assessment of the potential for extracting street-level characteristics from open data sources and popular deep-learning architectures, which can assist in the investigation of urban data and associated research. The proposed models' application in other cities is uncertain because the streetscape components studied are not the only ones. Furthermore, due to data resolution constraints, the output is unsuitable for scientific measurements and computations. Overall, the study provides a practical method for classifying streetscape features and creating a GIS database. Figure 2.2 shows the flowchart of the overall methodology.

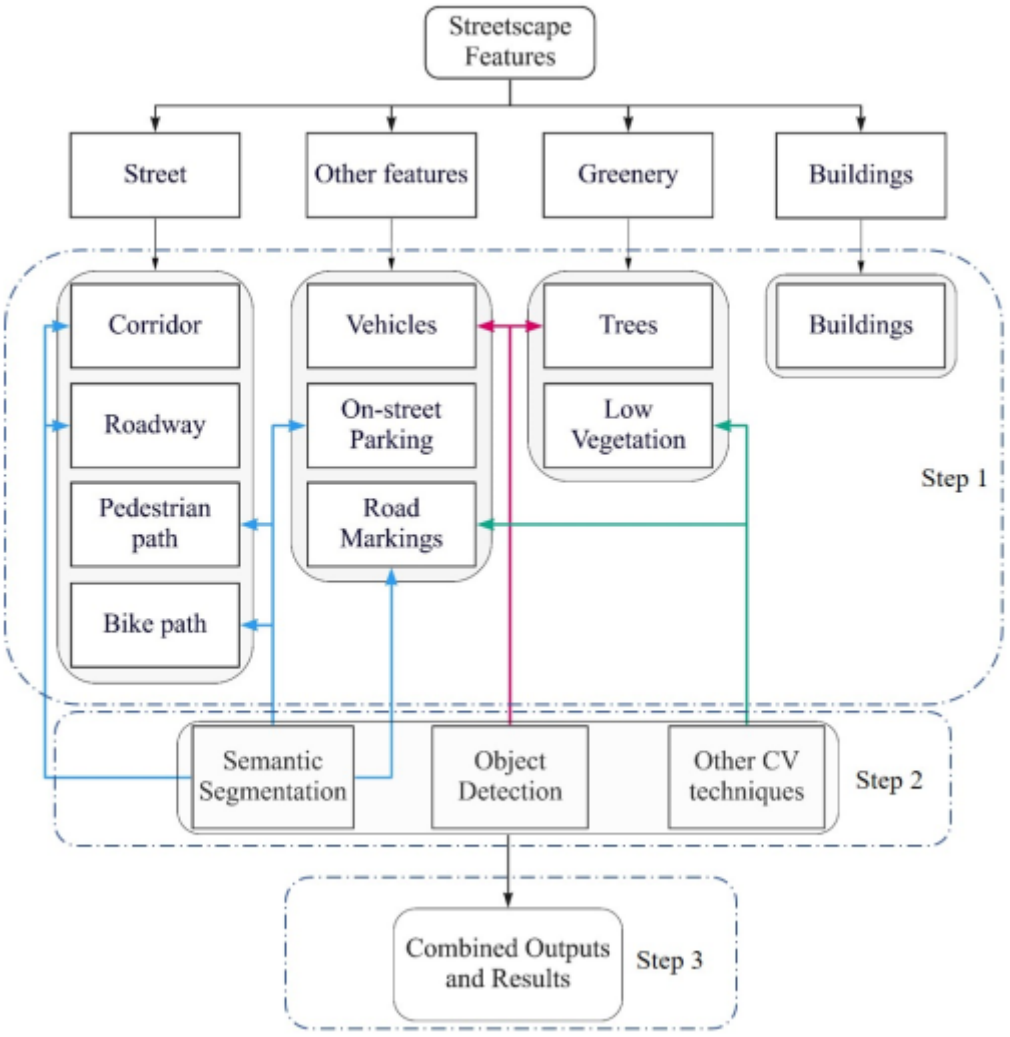

Figure 2.2: Flowchart displaying the study's general methodology [9]

2.9.2 Strengths

- This work makes it easier for other researchers to implement and reproduce the methodology by utilizing publicly available data and reducing reliance on human annotations during model training.
- The study evaluates the performance of all models using both model-specific and study-specific test datasets, ensuring better classification output.
- The study integrates the results to create a geospatial dataset and generates 3D views and street cross-sections for the city, providing more detailed visualizations of the identified features.

2.9.3 Weaknesses

- As this paper is based on urban regions, it is difficult to say whether it will be useful in rural or other areas.
- Due to the limited scope of this study's dataset, it is possible that it does not account for utility poles and other street furniture elements.

2.10 DeepForest: A Python package for RGB deep learning tree crown delineation [10]

2.10.1 Description

The paper proposes a deep-learning framework for automated tree crown delineation from high-resolution RGB images. The proposed method, called DeepForest, is based on the Faster R-CNN object detection architecture. The authors train the model on benchmark datasets, namely, the National Ecological Observatory Network (NEON). The experiments demonstrate that DeepForest outperforms existing state-of-the-art methods in terms of accuracy, efficiency, and generalizability. The authors provide an open-source software package that implements the DeepForest method in Python, making it easy to use and apply to a wide range of RGB imagery for tree crown delineation. While the proposed method shows significant promise for automated tree crown delineation, the authors note that several challenges remain, such as the robustness of the method to varying illumination and occlusions, the scalability of the method to large-scale datasets, and the generalizability of the method to other modalities such as multispectral or hyperspectral imagery.

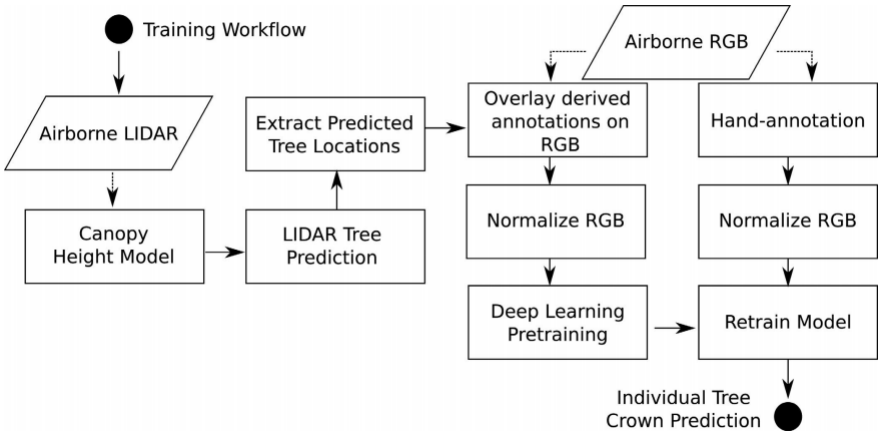

Figure 2.3: Prebuilt model training workflow. [10]

2.10.2 Strengths

- The proposed approach provides state-of-the-art performance on multiple benchmark datasets, indicating its effectiveness and superiority over existing methods.
- The authors provide a thorough evaluation of the method and conduct ablation studies to investigate the contribution of various components of the framework.
- The software package is open-source, easy to use, and can be applied to a wide range of RGB imagery for tree crown delineation.

2.10.3 Weaknesses

- The paper lacks a detailed discussion on the limitations and potential drawbacks of the proposed method. For instance, it is unclear how the method would perform under challenging conditions such as dense foliage, varying illumination, and occlusions.
- The authors do not provide any analysis of the computational complexity and memory requirements of the method, which could be critical for its practical deployment.

- The experiments are conducted only on RGB imagery, and it remains unclear whether the method would generalize to other modalities such as multispectral or hyperspectral imagery.

2.11 Comparing U-Net Convolutional Network with Mask R-CNN in the Performances of Pomegranate Tree Canopy Segmentation [11]

2.11.1 Description

The research investigates the application of deep learning methods for image segmentation in precision agriculture, especially for extracting ROIs from high-resolution UAV-based images of pomegranate tree canopies. Using aerial images of pomegranate trees, the authors evaluate the performance of U-Net and Mask R-CNN, the two convolutional network-based approaches. Mask R-CNN surpassed U-Net in terms of mAP, mAR, and numerous other measures, obtaining much greater performance on this application. Furthermore, utilizing pre-trained weights to initialize the Mask R-CNN model cut training time in half. The tests show the efficiency of deep learning approaches, especially Mask R-CNN, for pomegranate tree canopy segmentation, which may be employed in future precision agricultural research.

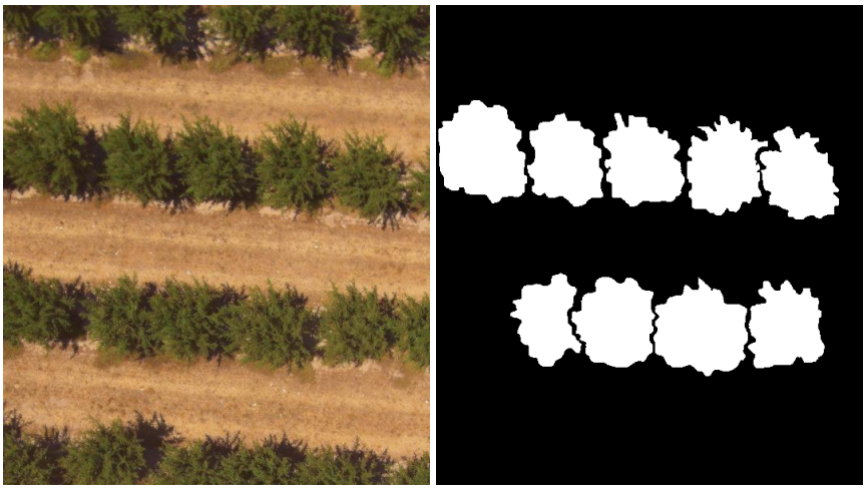

Figure 2.4: Aerial image on the left, ground truth label on the right (dataset samples, size: 534*600)) [11]

2.11.2 Strengths

- The authors give a useful dataset of high-resolution aerial photos of pomegranate trees taken during the growth season.
- A comparison of two CNN networks, U-Net and Mask R-CNN, reveals the relative merits and shortcomings of both approaches for pomegranate tree canopy segmentation.

2.11.3 Weaknesses

- The research is confined to the segmentation of pomegranate tree canopies, and the findings may not apply to other crops or applications.
- The authors point out several drawbacks of the approaches, such as the difficulty in recognizing some unlabeled occurrences and concerns with segmentation accuracy and completeness.
- The report makes no mention of the possible ethical concerns of deploying unmanned aerial vehicles and deep learning algorithms in precision agriculture.

2.12 Canopy Recognition of Cherry Fruit Tree Based on SegNet Network Model [12]

2.12.1 Description

The paper investigates the use of deep learning methods for image segmentation in precision agriculture, specifically for extracting ROIs from high-resolution UAV-based images of pomegranate tree canopies. Using aerial photos of pomegranate trees, the authors assess the performance of two convolutional network-based approaches, U-Net and Mask R-CNN. Mask R-CNN surpassed U-Net in terms of mAP, mAR, and numerous other measures, obtaining much greater performance on this application. Furthermore, utilizing pre-trained weights to initialize the Mask R-CNN model cut training time in half. The trials show the efficacy of deep learning approaches, especially Mask R-CNN. The network structure of SegNet is demonstrated in figure 2.5.

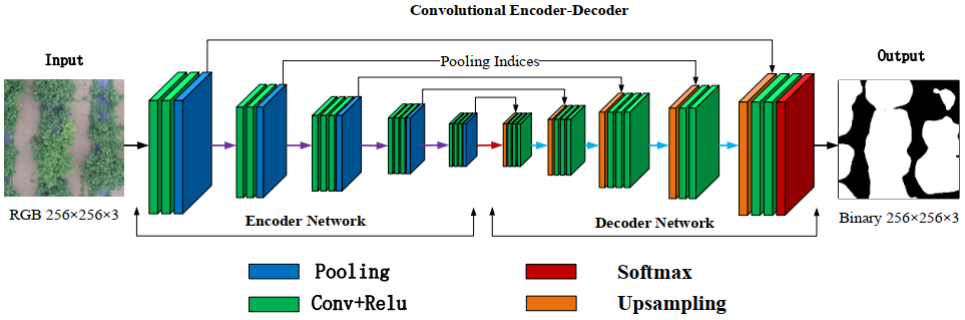

Figure 2.5: SegNet network structure diagram [12]

2.12.2 Strengths

- The study presents a method for reliably retrieving fruit tree canopy information from drone aerial photos, which is critical for precise variable spraying in current orchards.
- The study employs deep learning models and optimization strategies to increase identification accuracy, speed, noise management, and edge contour

smoothing, resulting in superior performance when compared to previous models.

- The study can be used to construct plant protection drones to prevent fruit tree diseases and insect pests.

2.12.3 Weaknesses

- The study solely looks at cherry trees in modern orchards and may not apply to other types of fruit trees or surroundings.
- The study does not compare the suggested strategy to existing ways for obtaining fruit tree canopy information in terms of cost-benefit analysis.

2.13 Segmentation of Green Vegetation of Crop Canopy Images Based on Mean Shift and Fisher Linear Discriminant [13]

2.13.1 Description

The research offers a hybrid segmentation technique called MS-FLD that combines mean shift (MS) and Fisher linear discriminant to increase the performance of crop picture segmentation. (FLD). For training data with long and narrow distributions, the MS-FLD technique employs a point-line-distance-based weighting mechanism. Across 50 soybean images and 20 weed photos utilized in the assessment, the proposed strategy functioned effectively and consistently. The MS-FLD outperforms other color-index-based approaches and corrects the problem of earlier systems' low segmentation rates in green regions with shadows. The proposed weighting method is suitable for any application where training data from the same class is distributed roughly along a line, and it may be extended to multi-class workloads. The major stages of the proposed method are shown in figure 2.6.

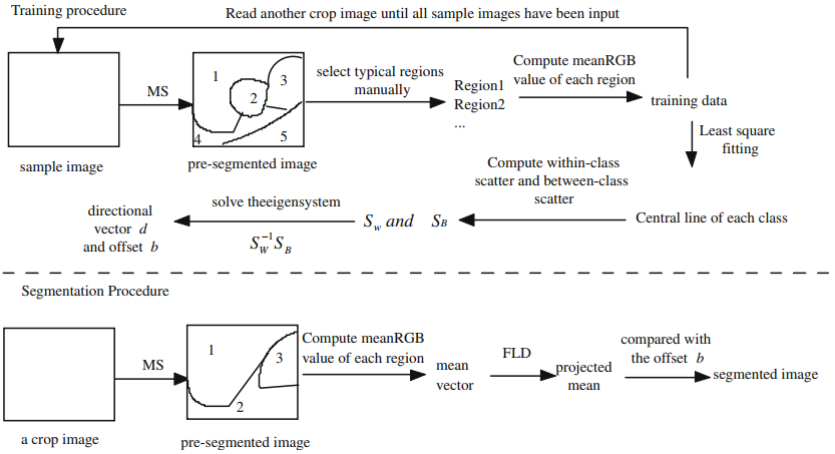

Figure 2.6: Significant phases of the MS-FLD [13]

2.13.2 Strengths

- The MS-FLD method outperforms prior color-index-based approaches in green regions with shadows and overcomes the issue of low segmentation rate, which is a significant improvement.
- The recommended weighting strategy is useful when training data for the same class are dispersed roughly along a line.
- The technique yields high-quality and dependable results in a wide range of practical applications.

2.13.3 Weaknesses

- Because of the time-consuming MS methodology, the MS-FLD technique is slower than other color-index-based techniques, and it may not be suitable for real-time applications that need fast processing.
- The MS-FLD method's performance in the study is only evaluated for two-class segmentation tasks; the expansion to multi-class issues is not extensively investigated.

- It is difficult to assess the overall performance of the MS-FLD approach because the study does not compare it to other segmentation algorithms that have been published in the literature.

2.14 Quantifying Urban Canopy Cover with Deep Convolutional Neural Networks [14]

2.14.1 Description

To effectively mitigate the effects of climate change, the urban canopy cover must be precisely and efficiently quantified. To get beyond the shortcomings of currently used manual and conventional computer vision techniques, the authors suggest using deep convolutional neural networks (DCNNs). They use datasets from several places (Oslo (Norway), Cambridge (USA), Johannesburg (South Africa), Sao Paulo (Brazil), and Singapore (Singapore)) to train and test the DCNN models, proving the superiority of their method in terms of scalability, accuracy, and computing economy. The use of the new technique in the Treepedia project, which has already affected public greening policy and stimulated interest in the subject, is demonstrated in the study to underline the relevance of their findings. Some examples are shown in figure 2.7.

Figure 2.7: Vertical vegetation classification using the DCNN semantic segmentation model. [14]

2.14.2 Strengths

- The use of the updated approach in the Treepedia project, which measured the amount of urban canopy cover in 22 cities across the world, illustrates the usefulness and effect of the research on areas of study and public greening strategies.
- The DCNN models outperformed the "threshold and cluster" method, lowering the mean absolute error in predicting the Green View Index (GVI) from 10.1
- The DCNN algorithm effectively processes a large number of Google Street View images, enabling the scalable study of urban greenery. According to the article, evaluating canopy cover in a metropolis like London may be done in an hour on a single machine, as opposed to the many days needed by conventional techniques.

2.14.3 Weaknesses

- 500 Google Street View photos from five different locations make up the very modest dataset used in this study for training and testing. The dataset's generalizability and robustness might be improved by increasing the number of cities it contains as well as the total amount of photos.
- Although the work shows that DCNN models are good at quantifying urban canopy cover, it is possible that more research and validation are needed to show that the method applies to a variety of geographical and environmental scenarios.
- Without detailed comparisons with other cutting-edge methods for assessing urban greenery, the article primarily contrasts the DCNN models with the threshold and cluster methods. A deeper comprehension of the benefits and drawbacks of the suggested technique would result from the inclusion of more comparison research.

2.15 Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on High-Resolution UAV Optical Imagery [15]

2.15.1 Description

By employing high-resolution RGB images from unmanned aerial vehicles (UAVs), the authors of "cite"urban2" give a detailed examination of five state-of-the-art fully convolutional networks (FCNs) a susceptible tree species' semantic segmentation. DeepLabv3+ Xception, FC-DenseNet, U-Net, SegNet and DeepLabv3+ MobileNetV2 are among the FCN concepts that have been studied. The work demonstrates that FCNs are capable of precisely segmenting trees and picking up on the distinguishing traits of the intended tree species. As the most accurate design, FC-DenseNet exceeds the other networks tested in terms of F1-score, intersection over union (IoU), and overall accuracy. SegNet performs admirably but is less reliable than U-Net, DeepLabv3+ MobileNetV2. A larger training dataset is necessary to properly leverage the DeepLabv3+ Xception architecture, which is more complex yet has the lowest accuracy scores. FC-DenseNet outperforms the other designs in terms of inference speed, although DeepLabv3+ Xception takes longer due to its greater complexity. when a fully connected conditional random field (CRF) is used to post-process network outputs, the segmentation quality is only marginally enhanced, but the processing time is significantly increased. Although DeepLabv3+ Xception takes longer because to its higher complexity, FC-DenseNet performs better in terms of inference speed than the other designs. Although the segmentation quality is only slightly improved following the post-processing of network outputs using a fully connected conditional random field (CRF), the processing time is substantially lengthened. The research offers insightful information on FCNs' effectiveness and computational effectiveness for the semantic segmentation of tree species. The findings aid in the identification of tree species and provide useful advice for choosing suitable FCN designs for UAV-based image processing.

2.15.2 Strengths

- The study provides a thorough knowledge of the merits and drawbacks of each evaluated architecture by a rigorous examination of the performance indicators, such as F1-score, overall accuracy, and intersection over union (IoU). Making well-informed decisions when choosing a suitable network for related activities is made possible by this research.

- The study focuses on the use of tree species delineation in practical applications, addressing the need for precise and effective segmentation techniques in the context of environmental monitoring and conservation. The utilization of high-resolution RGB photos captured by UAVs increases the research's relevance and application.

- The study evaluates the networks' computational effectiveness in terms of training and inference times. The trade-off between model complexity and computing demands may be better understood by the comparison of these measures, which is important for real-time and resource-constrained applications.

2.15.3 Weaknesses

- The performance of the evaluated architectures, notably DeepLabv3+ Xception, may have been impacted by the evaluation's data scarcity limits, which are acknowledged in the study. A more comprehensive and varied dataset would offer more.

- Although the research outlines the advantages of utilizing a fully connected conditional random field (CRF) for post-processing the network outputs, the enhancement in general accuracy metrics is minimal. The limited effect of post-processing on these measures raises concerns regarding the usefulness and efficacy of utilizing CRF, particularly in light of the large increase in processing time.

- The work emphasizes the need for more varied annotated databases to evaluate how well the evaluated designs generalize across various sensors, tree

species, and climatic factors. The assessment dataset's narrow focus could limit the findings' generalizability.

2.16 Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks[16]

2.16.1 Description

The pipeline for detecting trees in real environments using RGB data is presented in [16] and is based on deep learning. The authors employ an unsupervised LIDAR tree identification approach to retrain the neural network to overcome the problem of having little labeled training data. The model can recognize individual tree crowns with a recall of 69% and a precision of 60% when using this self-supervised technique in combination with a modest number of manually annotated images. The research delivers useful information and allows for cost-effective scaling of tree identification, with an emphasis on the potential of deep learning models in the field of ecology, forest management, and land utilization. In figure 2.8, an unsupervised light detection and ranging (LIDAR) detection creates initial training data for a deep learning model that is self-supervised. To generate the whole model, the model is then retrained using a limited number of hand-annotated trees.

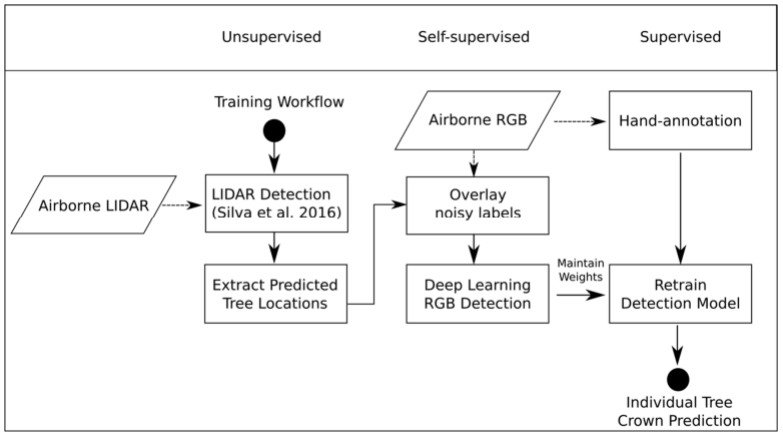

Figure 2.8: A block diagram of the proposed semi-supervised approach. [16]

2.16.2 Strengths

- The model works well when used to simulate a wide geographic region, demonstrating its capacity to represent tree traits in a variety of habitats and landscapes.
- The suggested pipeline shows potential for several uses, such as carbon dynamics analysis, ecology assessment, natural catastrophe recovery, and forest inventory.
- Custom regional tree detection models may be created affordably by combining pretraining with widely accessible high-resolution RGB images and unsupervised LIDAR data.

2.16.3 Weaknesses

- Unsupervised LIDAR pretraining, manual annotations, and RGB predictions all disagree on what defines a tree, which causes false positives and compromises the precision of crown area estimations.
- The model's effectiveness is evaluated using a small set of manually annotated images, which may not adequately reflect the complexity and unpredictability of different canopy conditions and ecosystems.
- The unsupervised LIDAR tree identification algorithm's performance greatly influences the self-supervised model's performance, which in turn influences the pipeline's total accuracy.

2.17 Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs [17]

2.17.1 Description

[17] provides a method for detecting distinct tree species using CNN and high-resolution images taken by RGB cameras mounted on unmanned aerial vehicles (UAVs). The authors analyze the performance of three state-of-the-art CNN-based algorithms on a dataset of 392 images: YOLOv3, RetinaNet, and Faster R-CNN. The dataset and a tight assessment system are utilized in the study to demonstrate how effectively RGB cameras and CNN-based algorithms execute tree detection tasks. According to the data, RetinaNet is the most precise. approach, with better accuracy in classifying tree species. The average precision of RetinaNet is very outstanding, displaying the precision with which it can detect trees. Furthermore, the study evaluates the computational efficacy of the methodologies and discovers that YOLOv3 is the fastest of the three, making it suitable for real-time applications. Figure 2.9 depicts the overall processing chain. The object identification technique in this figure is RetinaNet, however, alternative approaches (such as YOLOv3 and Faster-RCNN) can be used.

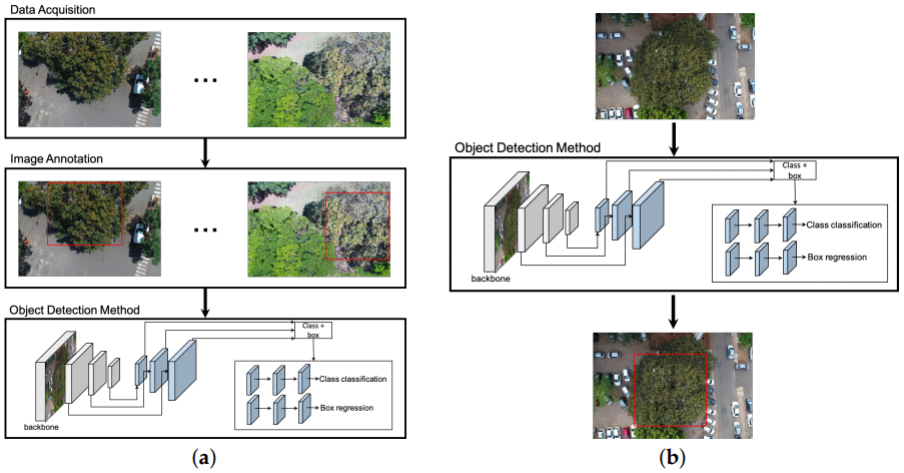

Figure 2.9: (a) A collection of annotated UAV images for training. (b) After training, the detection network was used to identify cumbaru trees in test images. [17]

2.17.2 Strengths

- The study compares three popular CNN-based detection methods, providing insights into their performance and allowing researchers and practitioners to make informed choices.
- RetinaNet attains a remarkable average precision of 92.64
- In addition to emphasizing their potential for population estimates and demography monitoring, the research underlines the viability of employing RGB cameras and UAV platforms for tree detection.

2.17.3 Weaknesses

- The dataset used in this study, which consists of 392 photos, does not adequately represent the variety and variability of tree species and environmental factors, which could limit the generalizability of the findings.
- The most computationally intensive approach is Faster R-CNN, according to the article, although it does not go into great detail on the resource needs and scalability problems, which could have an impact on actual implementation.

- The sorts of mistakes or misclassifications produced by the detection algorithms are not thoroughly covered in the study, even though doing so may provide light on their shortcomings and suggest areas for further development.

2.18 Tree Crown Detection and Delineation in a Temperate Deciduous Forest from UAV RGB Imagery Using Deep Learning Approaches: Effects of Spatial Resolution and Species Characteristics [18]

2.18.1 Description

[18] offers a thorough evaluation of two deep-learning-based techniques, Detectree2 and DeepForest to detect tree crown using UAV data in an alpine, temperate deciduous forest. In such forests, the transfer-trained Detectree2 approach is more suited and powerful for autonomously distinguishing individual tree canopies. The study emphasizes the significance of high-resolution imaging in improving detection accuracy. It also studies the impact of species diversity and geography on model performance. When high-resolution UAV images are available, the research suggests that deep-learning-based algorithms, when trained on adequate data, can be an effective tool for automated forest monitoring. The assessment flowchart for individual tree crown recognition and depiction from UAV RGB images in the deciduous forest, along with the essential processes and analysis is summarized in 2.10.

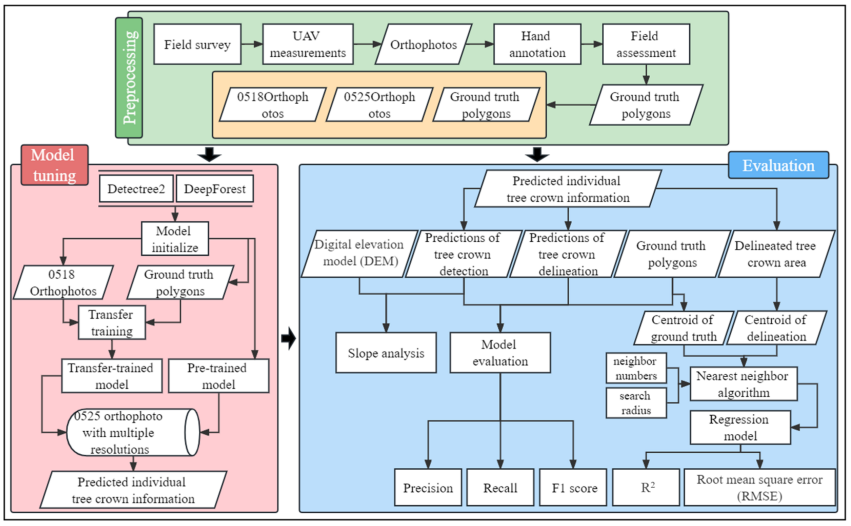

Figure 2.10: A graphical illustration of the basic procedures involved in the proposed method for the recognition and delineation of tree crowns. [18]

2.18.2 Strengths

- By instructing the deep learning models using UAV RGB images relevant to the target forest setting, the study evaluates the transferability of the learned patterns. This method shows how the models may be adjusted to various ecosystems and increase their usefulness in real-world applications.

- The effectiveness of deep learning models for detecting tree crowns under various topographical situations and tree species is examined in this research. This investigation contributes to a better understanding of the models' capabilities and limits by offering insightful information about the impact of topography and species diversity on the accuracy of the models.

- The accuracy of tree crown recognition is rigorously examined about picture spatial resolution. The report provides useful advice for data gathering and analysis in UAV-based forest monitoring by emphasizing the significance of greater resolutions for better outcomes.

2.18.3 Weaknesses

- While additional deep-learning models or other strategies for tree crown recognition are not studied, the article concentrates on contrasting DeepForest with Detectree2 techniques. A more thorough evaluation of the current methodologies would be possible by using a wider variety of comparison techniques.
- The use of reference techniques or ground truth data for validation is not covered in great detail in the study. The dependability and trustworthiness of the provided results would be increased by including a rigorous validation approach utilizing ground truth data.
- Although the possible use of phenological data to increase tree crown detection accuracy is briefly discussed in the study, there is no in-depth analysis or experimental testing. Expanding on this element might offer insightful information on utilizing phenological variability in future studies.

2.19 Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review[19]

2.19.1 Description

This paper [19] offers a comprehensive review of various machine learning classifiers used for land-use and land-cover classification based on satellite observations. This review delves into the efficacy and efficiency of these classifiers in interpreting and categorizing satellite-derived land-use patterns. The study underscores the importance of high-quality satellite data in ensuring classification accuracy. It also examines the influence of diverse terrains and spatial resolutions on the performance of these classifiers. When high-quality satellite observations are accessible, the research posits that machine learning classifiers, provided they are trained on ample data, can be instrumental for systematic land-use and land-cover analysis.

The review structure for land-use land-cover classification based on satellite observations, along with the principal methodologies and analysis, is summarized in ??.

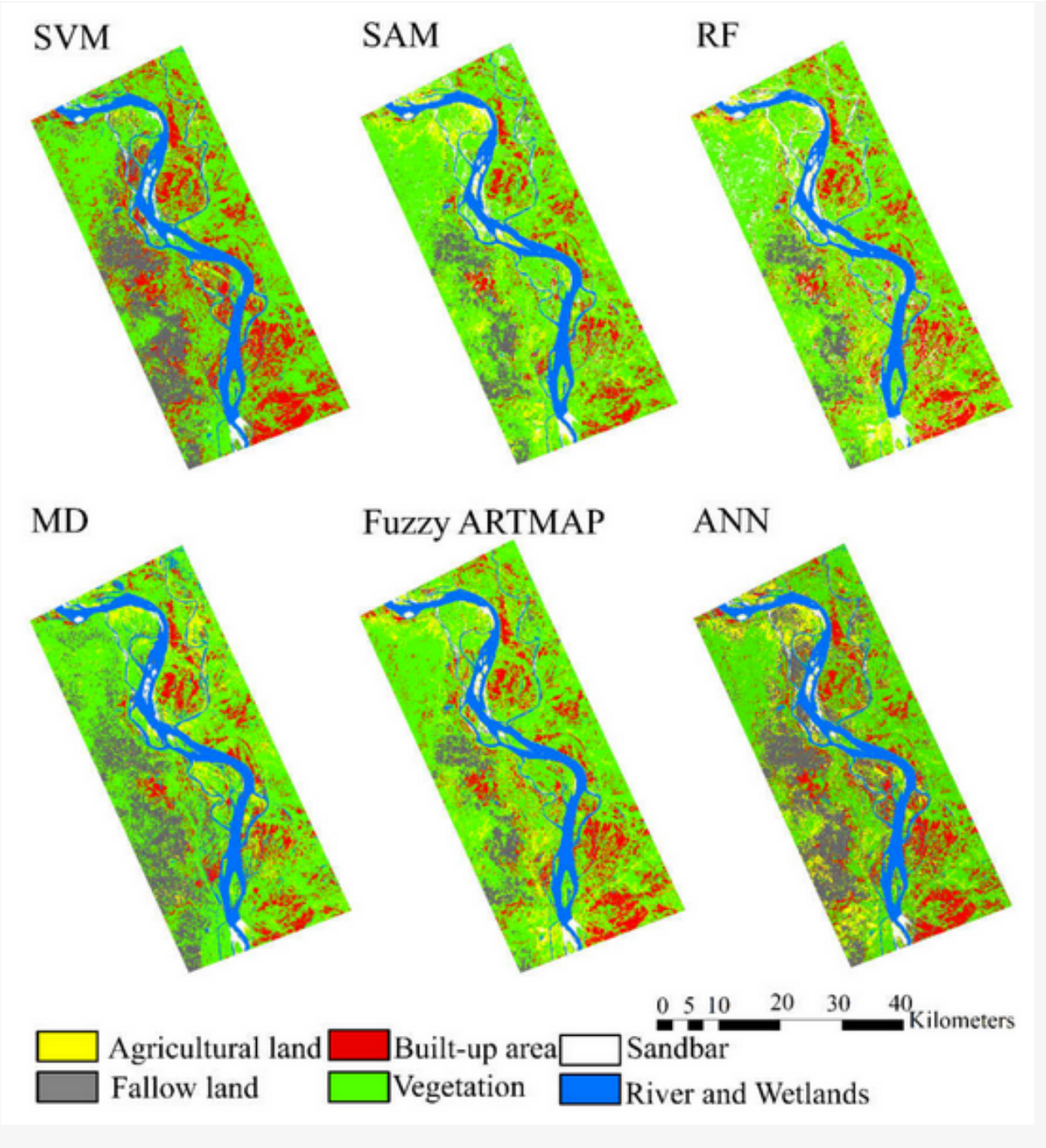

Figure 2.11: LULC map with different machine-learning technique [19]

2.20 Segmentation of satellite imagery using u-net models for land cover classification[20]

2.20.1 Description

The study focuses on utilizing satellite images to build accurate land cover classification maps using a modified U-Net structure within a neural machine learning model. By training and testing convolutional models for autonomous land cover mapping, the researchers hope to improve land cover mapping accuracy and change detection. The scientists used the BigEarthNet satellite image repository as well as an original dataset of Sentinel-2 photos and an Estonian land cover map from CORINE. The constructed classification model acquired a high F1 score of 0.749 for multiclass land cover classification, detecting errors in the BigEarthNet dataset. The segmentation models produce successful automatic land cover mappings, with a high IoU score for various land cover classes.

2.20.2 Strength

- The research provides a unique method for classifying land cover using a modified U-Net structure.
- The proposed method is tested on a big and difficult dataset, BigEarthNet.
- The suggested approach achieves good accuracy on both multiclass land cover classification and semantic segmentation tasks, as demonstrated in the study.

2.20.3 Weaknesses

- The research does not compare the suggested strategy to existing cutting-edge land cover classification algorithms.
- The report does not include a full analysis of the suggested approach's computing complexity and scalability.

2.21 Interactive segmentation in aerial images: a new benchmark and an open access web-based tool[21]

2.21.1 Description

Deep learning has made progress in the field of remote sensing, especially in satellite image segmentation. However, its use for interactive segmentation in land classification is limited. This study evaluates five main interactive segmentation methods on aerial images. It introduces a Cascading Forward Refinement approach, enhancing segmentation without the need for multiple models. While the Segment Everything Method (SAM) underperformed, the SimpleClick model outperformed. Based on this information, an online tool, RSISeg, was developed, providing improved interoperability and adaptability with remote sensing data.

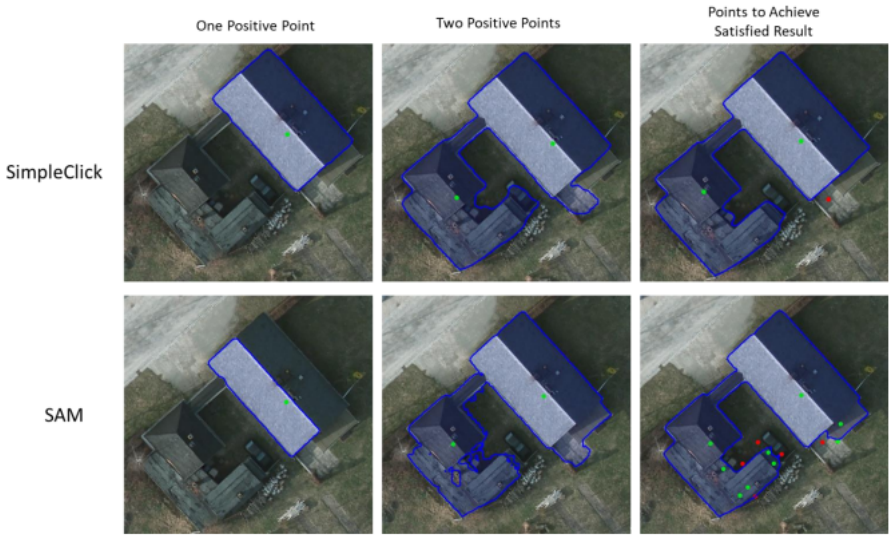

Figure 2.12: Comparison between SimpleClick and SAM through the process visualization of building segmentation.[21]

Evaluation study of interactive segmentation models for remote sensing. SimpleClick outperformed other competitors, especially the ICL model, demonstrating its adaptability in remote sensing missions. In contrast, SAM, despite having a lot of training data, did not perform well, showing that domain knowledge is as important as data volume. Specifically, SimpleClick segments buildings effectively

with a minimal number of mouse clicks, while SAM requires more input but still lacks precision.

2.21.2 Strengths

- Full analysis: The paper performs an in-depth comparative analysis of the different models, providing clear insights into their performance in remote sensing, which can guide future research and practical applications.models for detecting tree crowns under various topographical situations and tree species is examined in this research.
- Practical illustration: The study not only covers the theoretical performance of the models, but also provides concrete examples (such as building segmentation comparisons) to demonstrate the capabilities of different models in different scenarios. real situation.

2.21.3 Weaknesses

- Limited range: The article focuses primarily on comparing SimpleClick and SAM, perhaps ignoring the nuances and potential benefits of other models that may also be valuable in specific contexts.
- Possibility of overgeneralization: Although the article establishes the superiority of SimpleClick, it is possible that the model's performance is largely based on the specific characteristics of the data set. Wider application to a data set may yield different results.

2.22 Observations from Related Works

Some noticeable intuition can be made from reviewing these publications on tree canopy segmentation using deep learning. To begin with, deep learning is a very useful method for detecting tree canopies from remote-sensing images. A lot of research has shown that convolutional neural networks (CNNs) are exceptionally

good at this task. Transfer learning and Data Augmentation are also good approaches to this task. And, though CNN-based techniques are very excellent in terms of accuracy and other measures, they require a lot of training data. This can be problematic sometimes as annotated picture data is difficult to get. And, there's a trade-off between accuracy and computation complexity, more sophisticated models perform better than the simple model. So, advanced hardware is necessary to perform complex deep learning models.

Overall, the research shows that deep learning has a lot of potential for effectively segmenting tree canopies from remote sensing images, although data availability and processing requirements are still issues. Future research should focus on overcoming these challenges and exploring fresh applications of deep learning in forestry and environmental management.

Several significant discoveries can be made after reviewing 18 publications on tree canopy segmentation using deep learning. To begin, deep learning approaches have demonstrated great potential in effectively segmenting tree canopies from remote-sensing photos. Much research has shown that convolutional neural networks (CNNs) are exceptionally good at this task, especially when combined with other methods like transfer learning and data augmentation.

In addition to tree canopy segmentation, current research shows that the Segment Anything Model[33] (SAM) is extremely effective for correctly segmenting varied land cover elements within satellite data. SAM, a powerful object segmentation tool, accurately delineates essential features from high-resolution satellite pictures such as tree canopies, rivers, wetlands, and built-up areas. However, it is important to highlight that the success of SAM is strongly dependent on the availability of sufficient labeled training data and suitable computational resources, posing a significant difficulty in its application. Addressing these data limits and improving SAM's computing efficiency are critical steps toward realizing its full promise in advanced land cover research and effective management of the environment.

Chapter 3

Methodology

The methodology includes several stages, as shown in 3.1. These stages give a defined framework for carrying out the research and are intended to assist a thorough and methodical approach to the study working together to build a unified technique for attaining the study objectives. These stages contribute to the clarity, structure, and efficient implementation of the research approach.

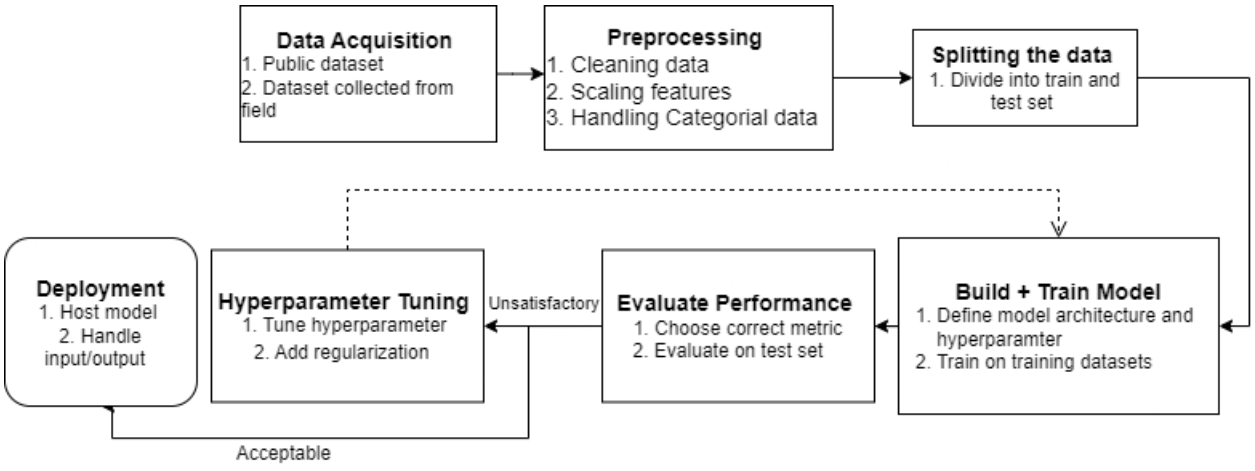

Figure 3.1: Methodology

The first stage is to collect high-resolution UAV images of a variety of conditions and difficult backgrounds. The dataset should include a range of situations.

subsequently, it is crucial to go on with accurate annotation of the obtained images. To do this, the areas of the imaging that correlate to the forest canopy must be properly labeled. The sections of the tree canopy must be precisely identified and marked throughout the annotating process. The dataset is enhanced with useful information that helps the deep learning model learn and differentiate tree canopy from other components found in the images by undertaking comprehensive and accurate annotation. The annotations are essential for training and assessing the model, ensuring that it can recognize and distinguish tree canopy in a variety of difficult backgrounds.

To increase the UAV imagery's quality and applicability for deep learning analysis, preprocessing is required after acquisition. Using methods like image enhancement, denoising, and geometric correction, includes eliminating sensor noise, artifacts, and distortions.

The annotated dataset must be appropriately subdivided to train and test the deep learning model. To create training, validation, and testing sets, divide the dataset. To reduce bias and help the model generalize, the splitting procedure should provide a broad distribution of difficult backgrounds in each subgroup.

Here, a suitable deep learning architecture for identifying tree canopies is selected, such as convolutional neural networks (CNNs). The number of layers, activation functions, and optimization techniques are taken into account when designing the network architecture. To implement the selected architecture, deep learning frameworks like TensorFlow or PyTorch are used. The model is then trained using the training set to optimize the model's parameters using backpropagation and gradient descent.

The trained model's performance is evaluated using the validation set. Recall, precision, F1-score, and Accuracy, are some of the measures used to gauge the model's proficiency in recognizing tree canopies in difficult backgrounds. To assess the model's efficacy, predictions are made and compared to actual data or expert annotations.

The performance of the model is optimized through hyperparameter tuning. Several hyperparameter values, including learning rate, batch size, and regularization

methods must be investigated. The hyperparameter space can be methodically searched using methods like grid search or random search. To determine the ideal setting, the model's performance is assessed with several hyperparameter values.

3.1 Dataset Description

3.1.1 Tree Species Detection Dataset

The first dataset used in this study named The Tree Species Detection [17] consists of 392 high-resolution RGB images captured by a 20-megapixel camera on a Phantom 4 UAV. Six missions were carried out during a seven-month period, from August 2018 to February 2019. The UAV flew at different heights ranging from 20 to 40 meters above the ground to capture the images.

Data was collected in three study zones in the urban section of Campo Grande municipality in the Brazilian state of Mato Grosso do Sul. These research areas have a combined area of around 150,000 square meters. The images depict the variation in tree species, appearance, scale, and lighting conditions seen in the research regions.

During the missions, a specialist manually annotated the dataset using LabelMe software. For each image, bounding boxes were placed around the cumbaru tree samples present in the scene. The bounding boxes were defined by specifying the upper-left and lower-right corner coordinates, accurately delineating the tree's location within the image.

Figure 3.2: Sample image of Tree Species Detection Dataset

3.1.2 NeonTreeEvaluation Benchmark Dataset

The second dataset, an extensive open-source Data for the NeonTreeEvaluation Benchmark [24] dataset was created to detect individual trees from airborne images. The dataset includes information from a variety of sources, such as RGB images, LiDAR tiles, hyperspectral files, and maps of canopy height at one meter.

For our research purposes, we focused exclusively on the RGB images available in the dataset. This dataset's subset has 214 RGB images in total. To facilitate analysis and evaluation, each RGB image in the dataset is paired with an annotation file in XML format, providing detailed annotations for individual trees present in the images. These annotations are useful for training and verifying tree identification algorithms since they provide crucial ground truth data., allowing for precise comparison and assessment of different approaches.

By leveraging these RGB images and their associated XML annotation files, we were able to conduct a thorough investigation into individual tree detection methods within the scope of our study.

Figure 3.3: Sample image of NeonTreeEvaluation Benchmark Dataset [24]

3.1.3 Forest Damages – Larch Casebearer Dataset

The third dataset named Forest Damages – Larch Casebearer [25] is an outcome of a collaborative project between the Swedish Forest Agency and Microsoft's AI for Earth program, aimed at aiding forest caretakers in quickly identifying and responding to threats caused by the Larch casebearer moth in Västergötland, Sweden.

The dataset consists of 1,543 high-resolution images captured during drone surveys conducted in five affected areas: Bebehojd, Ekbacka, Jallasvag, Kampe, and Nordkap, and their corresponding annotations. The annotations are provided in the Pascal VOC XML format, widely used for object detection tasks.

The collection contains exact bounding box annotations for trees classified as Larch or Other. In total, there are 101,878 annotated trees across the dataset, ensuring a diverse and extensive training sample.

Furthermore, some of them include annotations that describe tree damage in categories like Healthy (H), Light Damage (LD), High Damage (HD), and Other. These annotations enable the creation of damage categorization models. There are 44,522 annotated larch trees with damage-level descriptors in these batches.

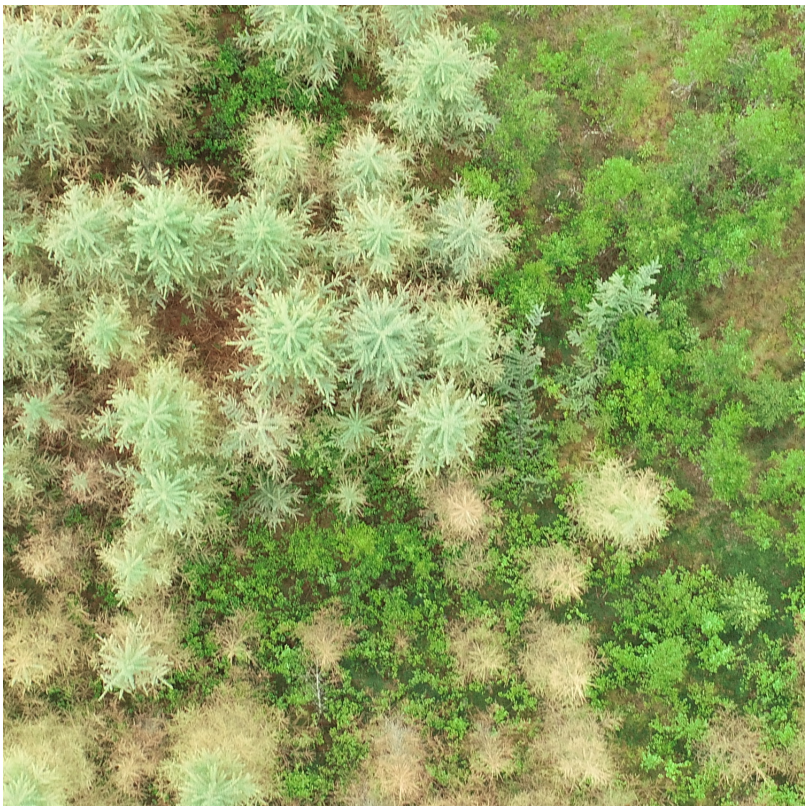

Figure 3.4: Sample image of Forest Damages – Larch Casebearer Dataset [25]

3.1.4 DeepGlobe Land Cover Classification Dataset

The DeepGlobe Land Cover Classification Challenge [26] intends to use satellite imagery to automate the segmentation and classification of distinct land covers, which is an important step toward sustainable development, autonomous agriculture, and urban planning. Participants are charged with identifying seven distinct land cover types, designated by specific RGB color codes, using 803 high-resolution (50cm pixel) training photos taken from DigitalGlobe and sized 2448x2448. These kinds span from metropolitan areas, denoted by a 0,255,255 RGB coding, to agricultural regions, woods, rangelands, water bodies, barren lands, and places deemed unknown, which frequently represent clouds or unidentified regions. While the

training set provides annotated masks to aid with classification, the challenge also includes 171 validation and 172 test photos without corresponding masks, encouraging real classification.

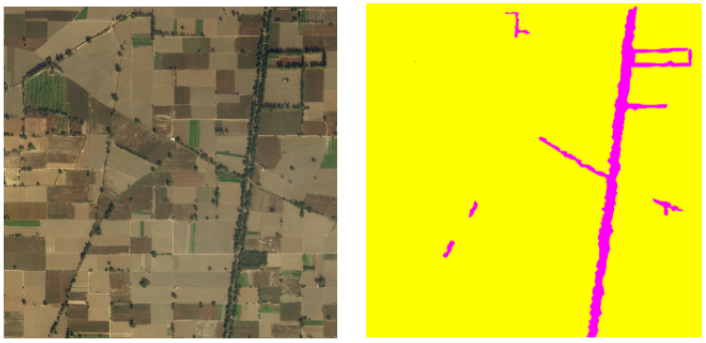

Figure 3.5: Sample image and ground truth image of DeepGlobe Land Cover Dataset [26]

3.2 Object Detection Models

3.2.1 Faster R-CNN

The faster R-CNN [34] architecture combines a region proposal network (RPN) with a shared convolutional neural network (CNN). By moving a small network across the CNN feature map, the RPN constructs possible bounding box proposals. For a set of predetermined anchor boxes, the RPN predicts the coordinates of the object's bounding box as well as the probability of objectness. These proposals are improved further utilizing fully connected layers after being classified and refined using a region of interest (RoI) pooling operation.

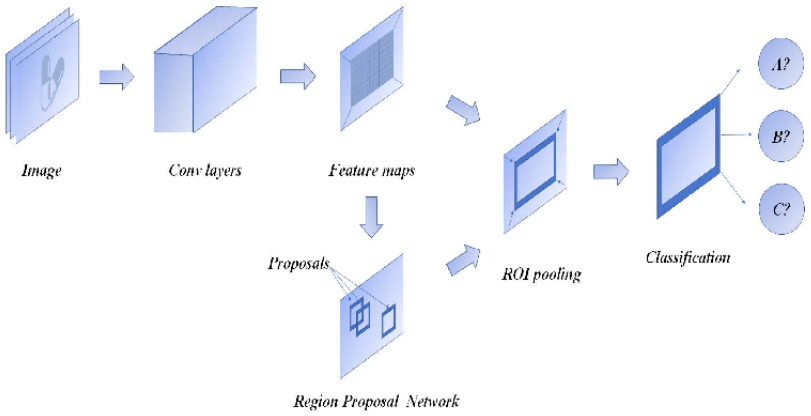

Figure 3.6: Faster R-CNN Architecture [27]

3.2.2 Mask R-CNN

Mask R-CNN [35] expands Faster R-CNN by including an extra branch for instance segmentation. It extends the previous architecture by adding a parallel branch that creates pixel-level masks for each proposed item. This branch extracts features from the suggested areas using RoI align and then applies a small FCN to forecast a binary mask for each region. The generated masks delineate the contour of the item with pixel-level precision, allowing for precise segmentation of various objects within an image.

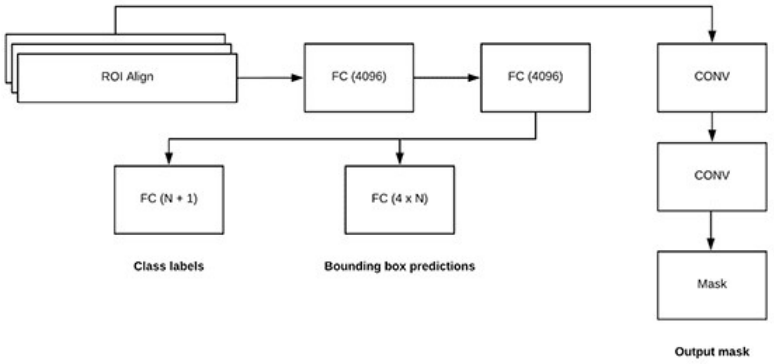

Figure 3.7: Mask R-CNN Architecture [28]

3.2.3 YOLOv5

YOLOv5 [36] is a highly sophisticated object identification model noted for its fast and precise performance. It uses a one-stage method, predicting class probabilities and bounding boxes in a single run. YOLOv5 is distinguished by its lightweight design, which includes anchor-based recommendations, focus loss for improved accuracy, and innovative data augmentation algorithms. The model has produced outstanding results in a variety of fields, gaining popularity for real-time object identification. Its combination of speed and accuracy makes it a popular choice in computer vision.

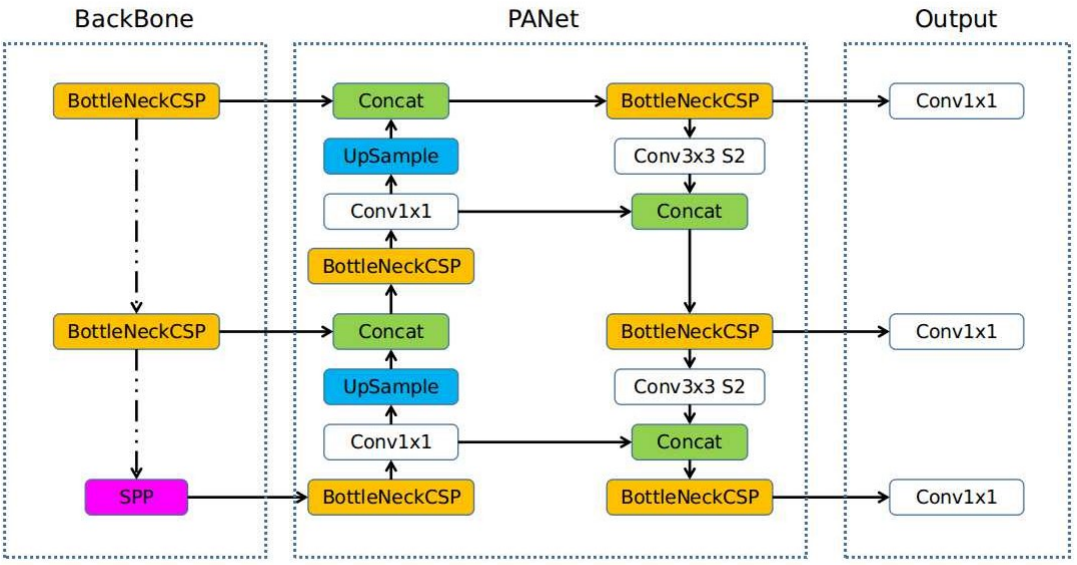

Figure 3.8: YOLOv5 Architecture [29]

3.3 Segmentation model

3.3.1 U-net

U-Net [30]is an innovative convolutional neural network model designed specially for biomedical image segmentation uses. Its symmetrical architecture is marked by

a narrowing path that captures context and a wider path that refines localization. Skip connections between mirrored layers in the encoder and decoder segments are used in the model to provide perfect localisation in the segmented outputs. U-Net's efficiency and accuracy in tasks, particularly when training data is scarce, have established it as a benchmark in medical imaging. Its capacity to produce high-quality segmentations with little input data has cemented its place as a top choice in the computer vision world.

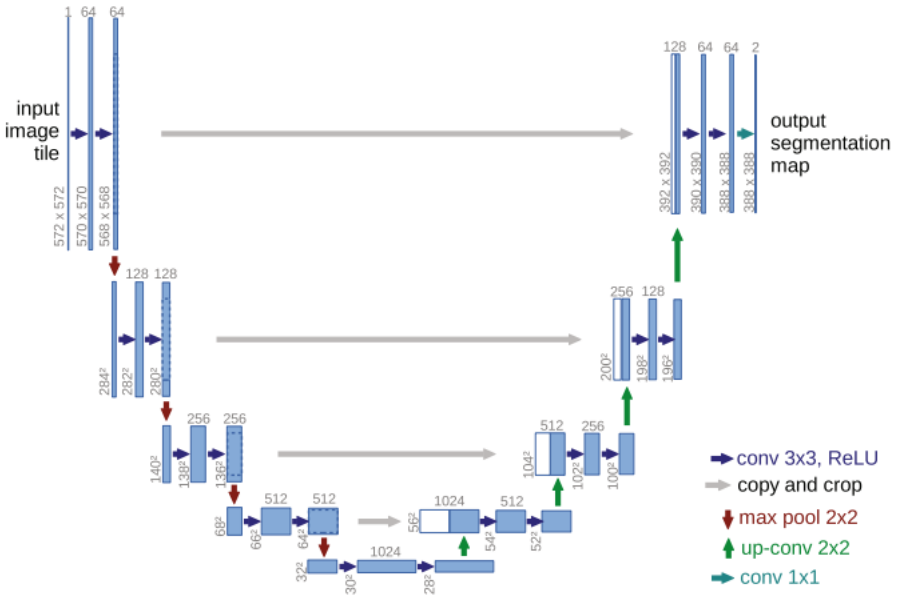

Figure 3.9: U-net Architecture [30]

3.3.2 DeepLabV3+

DeepLabv3+[31] is a modern convolutional neural network architecture designed for semantic segmentation. It is known for its unique technique, which includes atrous convolutions and spatial pyramid pooling modules, allowing the model to comprehend multi-scale contextual information without losing detail. This architecture can be enhanced further by adding image-level features and refining segmentations, particularly at object borders. DeepLabv3+ defines itself by producing dense and broad segmentations in complex scenarios, showing its advantage in a variety of benchmark datasets. DeepLabv3+ is widely known and used in the computer vision world because of its good efficiency and precision.

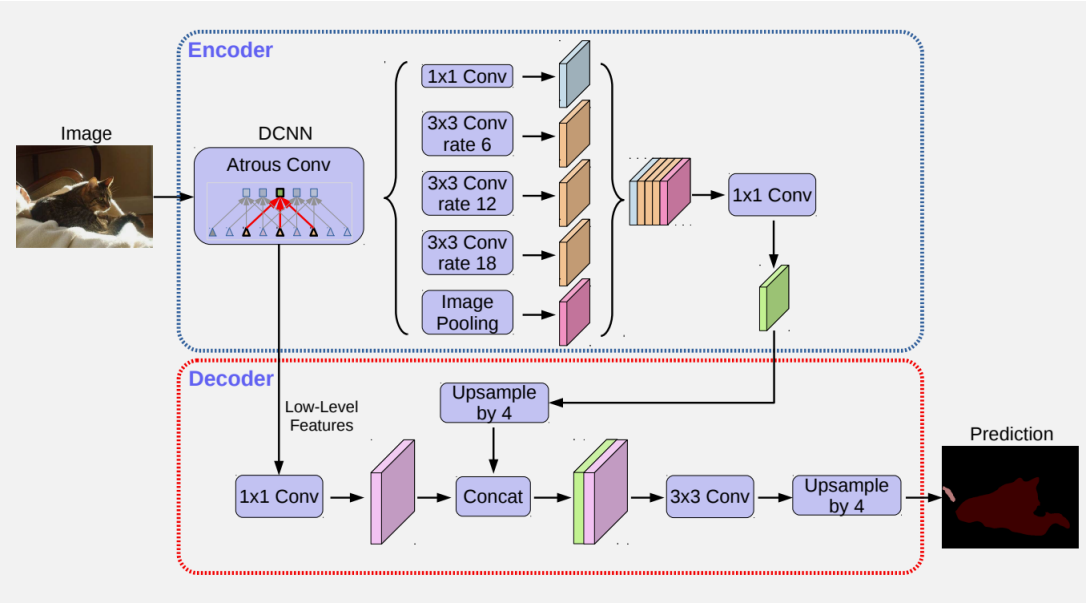

Figure 3.10: DeepLabV3+ Architecture [31]

3.3.3 Segment Anything Model

SAM (Segment Anything Model) [33] is a refined promptable segmentation model that can identify and segment any object in an image, even if it has never been seen before. SAM is trained on a huge dataset of over one billion masks, helping it to develop a broad grasp of object form and appearance. SAM's special architecture enables it to be both flexible and efficient. It can be given a range of inputs, such as text descriptions, bounding boxes, point clouds, and even other segmentation masks. SAM is capable of creating segmentation masks in real time, making it helpful in applications such as augmented reality and interactive image editing. SAM is a fascinating fresh idea that has the potential to change how we interact with images and the environment around us.

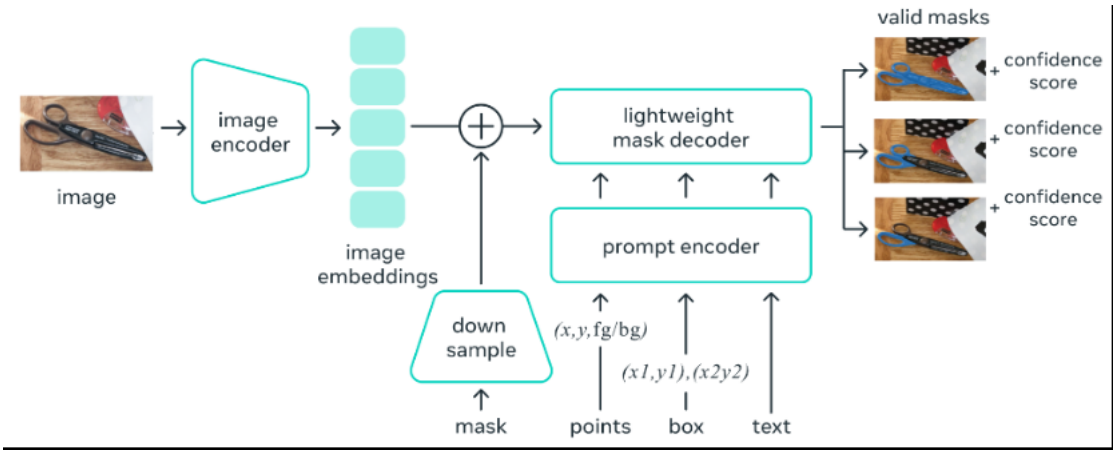

Figure 3.11: SAM Architecture [32]

SAM (Segment Anything Model) uses its extensive training data to learn the visual properties of objects and their borders, as well as the context of human instructions such as bounding boxes or pointers, to properly segment the picture. SAM develops high-resolution segmentation masks that properly outline the items of interest using this information.

3.4 Performance Metrics

Performance metrics are essential tools used to evaluate and quantify the effectiveness and accuracy of various models, algorithms, and systems. These metrics provide objective measures that help researchers and practitioners assess the quality of their work, make informed decisions, and compare different approaches.

Precision (P) is the percentage of accurately detected positive cases out of all positive instances predicted. It represents the capacity of the model to accurately identify positive cases. A high precision value suggests a low false positive rate, indicating a reliable model.

$$P = \frac{TP}{(TP + FP)} \tag{3.1}$$

The proportion of accurately detected positive cases out of all real positive instances is measured by recall (R), which is also referred to as sensitivity or true positive rate. It denotes the model's capacity to properly detect all positive instances while reducing false negatives. A high recall value indicates a low false negative rate, highlighting a comprehensive model.

$$R = \frac{TP}{TP + FN} \tag{3.2}$$

Mean Average Precision at 50 (mAP50) is a popular performance metric in object detection. It calculates the average accuracy of the top 50 recommendations or search results. mAP50 accounts for both precision and ranking quality, providing a measure of overall performance for systems that generate a ranked list of recommendations or search results.

The F1 score merges recall and precision into a single metric to assess the overall performance of the model. It is the harmonic mean of recall and precision, and it provides a balanced assessment of the two. Because it incorporates both false negatives and false positives. The F1 score is especially useful when there is an unequal distribution of the positive and negative classes in the dataset.

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{3.3}$$

3.5 Work already done

Our project involved training the YOLOv5 model on three available datasets. Additionally, we trained the preexisting models, Faster R-CNN and Mask R-CNN, using the Detectron2 library on the first two datasets [17] and [24]. To ensure effective training, we split the datasets into training, test, and validation subsets. This division allowed us to train the models on representative data while having separate datasets for performance evaluation.

For implementing Faster R-CNN and Mask R-CNN, we relied on the powerful Detectron2 library. However, we faced a challenge regarding the compatibility of the

library with the Pascal VOC dataset format. The annotation files provided with the dataset were in .xml format, so we had to convert them to the required JSON format compatible with Detectron2. This involved transforming the annotations into the COCO format, which Detectron2 supports.

During the conversion process, we encountered some corrupted files within the dataset. To ensure the integrity of the training process and avoid any biases or inaccuracies, we took the time to correct these files. This step was crucial to maintain the reliability and quality of the training data.

Unfortunately, our attempts to train and evaluate Faster R-CNN and Mask R-CNN on the third dataset [25] were unsuccessful. The recall and precision metrics were significantly low, indicating inadequate performance. We are exploring alternative strategies, fine-tuning parameters, and making necessary adjustments to get a satisfactory outcome from these models on the dataset.

On the other hand, YOLOv5, a different model architecture, presented its challenges. The annotation files provided with the dataset were not in the YOLO format, rendering them incompatible with direct training. Therefore, we had to recreate the annotation files in .txt format, adhering to the requirements of YOLOv5. This involved carefully aligning the annotations with the corresponding image files, enabling YOLOv5 to associate the objects correctly within the dataset.

Once the necessary data preparations were complete, we proceeded to train the models using the training and validation data from each dataset. This training phase involved optimizing the model parameters, adjusting the hyperparameters, and iteratively improving the models' performance. Training each model on its respective dataset allowed us to capture the unique characteristics and variations present in the data.

And then, we decided to train the land cover dataset using both U-Net and DeepLabv3+ models. Initially, our intention was to find a specific forest cover dataset, but due to difficulties in obtaining one, we selected the land cover dataset because both both have notable similarities. U-Net, with its excellent reputation in image segmentation and well-structured design, is our favorite starting point. After U-Net, we ventured into DeepLabv3+, an architecture praised for its

comprehensive image analysis and ability to detail complex segmentations. After testing both models, the results yielded significant insights. Notably, the land cover dataset serves as an indicator of forest cover, thus validating our approach. This effort emphasizes the importance of adaptability and innovation. When obstacles arise or resources are unavailable, it is essential to reassess and strategize. This project in particular broadened our horizons and introduced us to fascinating perspectives in the field of satellite photography.

We failed to train the land cover dataset on SAM before the deadline due to technical difficulties. We are currently investigating the issue and plan to address it in the future. We believe that SAM has the potential to be a viable tool for land cover segmentation, and we are committed to making it work. We understand that this is a setback, but we are confident that we can overcome it. We appreciate your patience and understanding as we work to make SAM a valuable tool for the land cover community.

Finally, to assess the effectiveness of our trained models, we evaluated their performance on the test set. This evaluation phase involved measuring various metrics, including recall, precision, and F1 score, to comprehensively understand the models' capabilities. The performance evaluation was crucial in determining the effectiveness and generalization ability of each model, ensuring that our trained models could perform well on unseen data.

A critical aspect of our project involves analyzing the outcomes of the various models we've trained. This comparison involves evaluating factors such as speed, memory utilization, and detection accuracy.

Through these extensive training and evaluation processes, we aimed to utilize the strengths of the three models and leverage them to accurately detect and classify objects within the datasets.

3.6 Work expected to be done

In the remaining stages of our project, there are several crucial tasks that we need to accomplish. These tasks are integral to our project's progression, and they will

significantly contribute to the quality and effectiveness of our work.

Firstly, we plan to focus on fine-tuning and optimizing our trained models. This entails a meticulous process of adjusting various hyperparameters, experimenting with different network architectures, and iteratively training and validating the models. The objective is to achieve the best possible performance and accuracy by striking a balance between precision and recall rates, reducing false positives, and improving object detection capabilities across our diverse datasets.

To evaluate the efficacy of our models, we will utilize a range of evaluation metrics. They will provide us with quantitative insights into how well our models perform on different datasets. By studying these metrics, we will obtain a better knowledge of each model's strengths and limitations, identify areas for development, and make better decisions to make better the performance of our models.

To enhance the robustness and generalization capabilities of our models, we plan to expand our datasets. This expansion involves collecting additional annotated images, diversifying the object instances, and introducing data augmentation techniques. Data augmentation includes applying transformations by rotating, scaling, or filtering to generate augmented samples. By increasing the size of our dataset and incorporating data augmentation, we aim to expose our models to a wider range of scenarios and improve their ability to accurately detect objects in real-world settings.

And also, we will explore the use of SAM for satellite image segmentation and plan to fine-tune the model to our specific problem. SAM is a promptable segmentation model that can segment any object in an image, making it promising for satellite image segmentation, where labeled data is scarce. We will fine-tune SAM on a dataset of labeled satellite images relevant to our problem, allowing it to learn the specific features of the objects we want to segment. Once fine-tuned, we can use SAM to segment satellite images of our target area, extracting valuable information such as land cover distribution or the presence of specific objects. We believe SAM has the potential to be a powerful tool for satellite image analysis, and we are excited to explore its use and see how it can solve real-world problems.

Furthermore, we will compare our results to state-of-the-art algorithms reported in the literature. This comparative analysis will provide us with valuable insights into the competitiveness and efficacy of our trained models, allowing us to gauge their performance against established benchmarks.

Chapter 4

Results

Our results are presented using a variety of datasets. Our research includes a thorough examination of these datasets, with an emphasis on determining what is driving the trends and patterns. We summarize our significant findings and observations in this section. In this section we've used box loss, class loss and objectness loss. Box loss measures the accuracy of predicted object positions and sizes compared to ground-truth bounding boxes. Class loss assesses the correctness of object classification predictions by evaluating class probabilities against actual labels. And objectness loss evaluates the model's ability to discern objects from the background within bounding boxes and assesses confidence in object presence.

4.1 Results on Tree Species Detection

This study used 392 high-res images from a Phantom 4 UAV, collected during six missions between August 2018 and February 2019 in Campo Grande, Brazil. The dataset shows tree variety. A specialist annotated the dataset with bounding boxes around cumbaru trees using LabelMe software. In our dataset examination, we checked the capabilities of Faster R-CNN, Mask R-CNN, and YOLOv5 frameworks, tapping into the strengths of these diverse deep learning architectures. In Figure 4.1, we can observe the training outcomes of Faster R-CNN on [17], including accuracy, box loss, and class loss. Looking at the first graph, we witness a remarkable improvement in accuracy until it reaches a certain point. As for the

second graph, which represents the box loss curve, it exhibits an initial increase, but after reaching a certain point, it begins to decline. The third graph displaying the class loss curve demonstrates a consistent decrease throughout the training process.

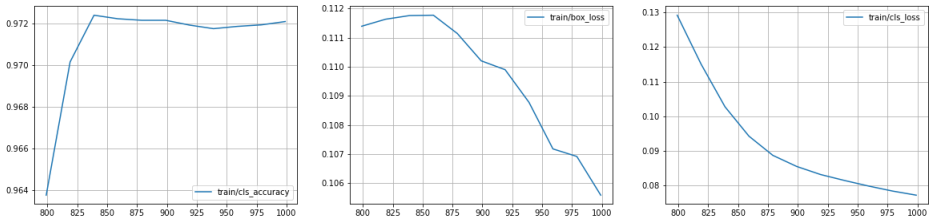

Figure 4.1: Accuracy and Loss Curves of Faster R-CNN on the Tree Species Detection with respect to Iterations

Figure 4.2 showcases the training results of Mask R-CNN on [17], highlighting the metrics of accuracy, box loss, and class loss.

Analyzing the initial graph, we can clearly observe a significant enhancement in accuracy until a certain threshold is reached. In regard to the second graph, which depicts the box loss curve, it initially shows an upward trend, but later experiences a decline after reaching a specific point. Moving on to the third graph portrays the class loss curve, which demonstrates a decrease with a steady period in the training process.

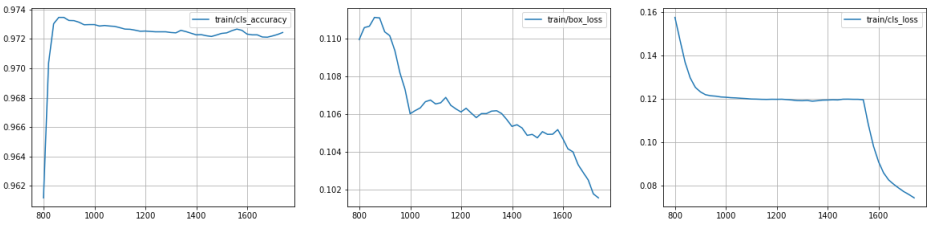

Figure 4.2: Accuracy and Loss Curves of Mask R-CNN on the Tree Species Detection with respect to Iterations

Figure 4.3 allows us to witness the training and validation results of yolov5 on the [17], specifically focusing on the metrics of box loss and object loss. Across all

the presented graphs, it is evident that there is a remarkable reduction in all the curves representing loss.

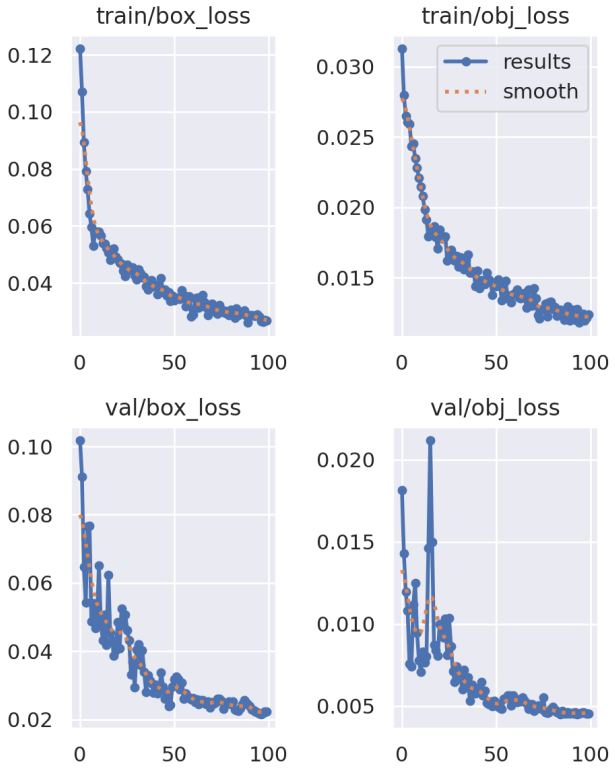

Figure 4.3: Loss Curves of YOLOv5 on the Tree Species Detection with respect to Epochs

The evaluation metrics of these methods are displayed in Table 4.1. The table provides a comprehensive overview and comparison of the performance of different methods for tree canopy identification dataset. The table includes various evaluation metrics and results obtained from each method, allowing for a quantitative assessment of their effectiveness.

Variant	P	R	mAP50	F1
Faster R-CNN	34.54%	49.4%	58.0%	40.6%
Mask R-CNN	41.1%	46.8%	67.3%	43.8%
YOLOv5	81.2%	77.2%	79.0%	79.2%

Table 4.1: The precision, recall, mAP50, and F1 score of different models on the Tree Species Detection Dataset

A comparison is made between the training time, inference time, and model size of these methods in Table 4.2. The complexity table provides a comparative

analysis of the computational complexity and resource requirements of different methods used on the dataset. It highlights various aspects of complexity, allowing researchers to assess the feasibility and practicality of each method.

Method	Training Time	Inference Time	Model Size
Faster R-CNN	2329 s	16ms	314 MB
Mask R-CNN	2228 s	16.1ms	334 MB
YOLOv5	576 s	1.6 ms	14.4 MB

Table 4.2: Complexity Comparison among different models on the Tree Species Detection Dataset

4.2 Results on Data for the NeonTreeEvaluation Benchmark

This result was done on the NeonTreeEvaluation[37] dataset. It has RGB, Lidar and Hyperspectral images , though we only did our training on just RGB images due to some problem with Lidar and hyperspectral images. The task was to detect the tree from the images with bouding box. And, Mask-RCNN , Faster R-CNN and YOLOv5 architecture were sued to train the model. The Figure 4.4 illustrates the training outcomes of Faster R-CNN on [16], focusing on the metrics of accuracy, box loss, and class loss. Upon examining the first graph, it becomes evident that there is a notable improvement in accuracy until a certain threshold is attained. As we shift our attention to the second graph, which represents the box loss curve, we observe a downward trend. Finally, in the third graph, the class loss curve is depicted, demonstrating a consistent decrease throughout the training process.

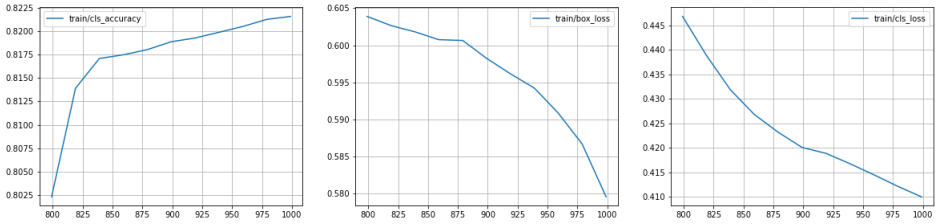

Figure 4.4: Accuracy and Loss Curves of Faster R-CNN on Data for the NeonTreeEvaluation Benchmark with respect to Iterations

Figure 4.5 provides a visual representation of the training results obtained from Mask R-CNN applied to [16]. It highlights key metrics such as accuracy, box loss, and class loss. Analyzing the initial graph, it becomes apparent that there is a significant development in accuracy up to a certain threshold. Shifting our focus to the second graph, which illustrates the box loss curve, we notice an initial upward trend, followed by a subsequent decline after reaching a specific point. Finally, the third graph showcases the class loss curve, which consistently decreases throughout the training process.

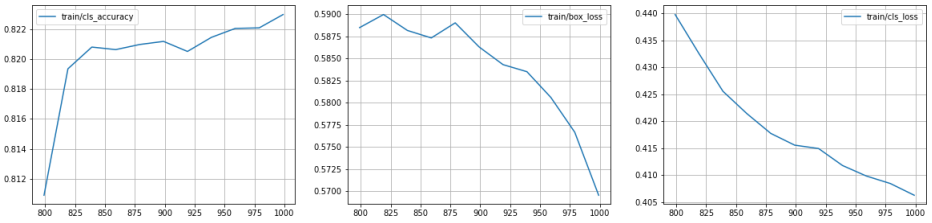

Figure 4.5: Accuracy and Loss Curves of Mask R-CNN on Data for the Neon-TreeEvaluation Benchmark with respect to Iterations

Figure 4.6 allows us to witness the training and validation results of yolov5 on the [16], specifically focusing on the metrics of box loss and object loss. Across presented the box loss graphs, it is evident that there is a remarkable reduction in all the curves representing loss. But object-loss curves display an upward trajectory indicating a decrease in accuracy. However, as the epoch progresses, the curve transitions to a downward trend, suggesting positive progress in learning.

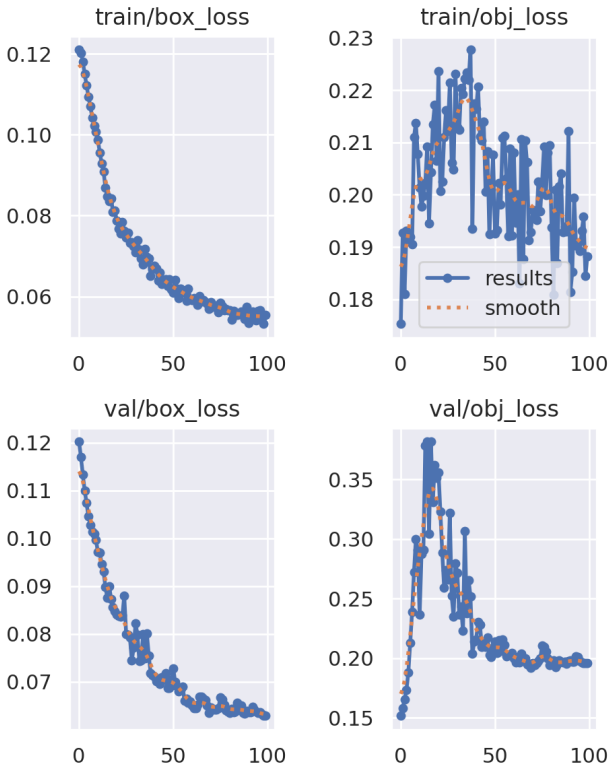

Figure 4.6: Loss Curves of YOLOv5 on Data for the NeonTreeEvaluation Benchmark with respect to Epochs

The evaluation metrics of these methods are displayed in Table 4.3. The table provides a comprehensive overview and comparison of the performance of different methods for tree canopy identification the dataset. The table includes various evaluation metrics and results obtained from each method, allowing for a quantitative assessment of their effectiveness.

Variant	P	R	mAP50	F1
Faster R-CNN	25.5%	36.5%	42.9%	30.0%
Mask R-CNN	30.17%	34.4%	49.53%	32.1%
YOLOv5	66.4%	56.7%	58.4%	61.2%

Table 4.3: The precision, recall, mAP50, and F1 score of different models on Data for the NeonTreeEvaluation Benchmark Dataset

A comparison is made between the training time, inference time, and model size of these methods in Table 4.4. The complexity table provides a comparative analysis of the computational complexity and resource requirements of different

methods used on the dataset. It highlights various aspects of complexity, allowing researchers to assess the feasibility and practicality of each method.

Variant	Training Time	Inference Time	Model Size
Faster R-CNN	374 s	8.5 ms	314 MB
Mask R-CNN	518 s	7.4 ms	334 MB
YOLOv5	460.8 s	1.0 ms	14.3

Table 4.4: Complexity Comparison among different models on Data for the NeonTreeEvaluation Benchmark Dataset

4.3 Results on Forest Damages – Larch Casebearer Dataset

The task of this dataset[38] is to detect the tree from a forest and also classify according to condition of the tree, like if it is damages or not. So far, we have analyzed the outcomes of YOLOv5 on dataset-3. The results have been depicted using graphs and tables to provide a comprehensive overview.

Figure 4.7 displays the confusion metrics. We can get the TP, FP, TN and FN from this figure. For Larch tree TP=0.75, FP=0.43, and FN=0.25 are respectively. For other tree TP=0.60, FP=061, and FN=0.4 are respectively. For background TP=0, FP=0.61,and FN=1 are respectively.

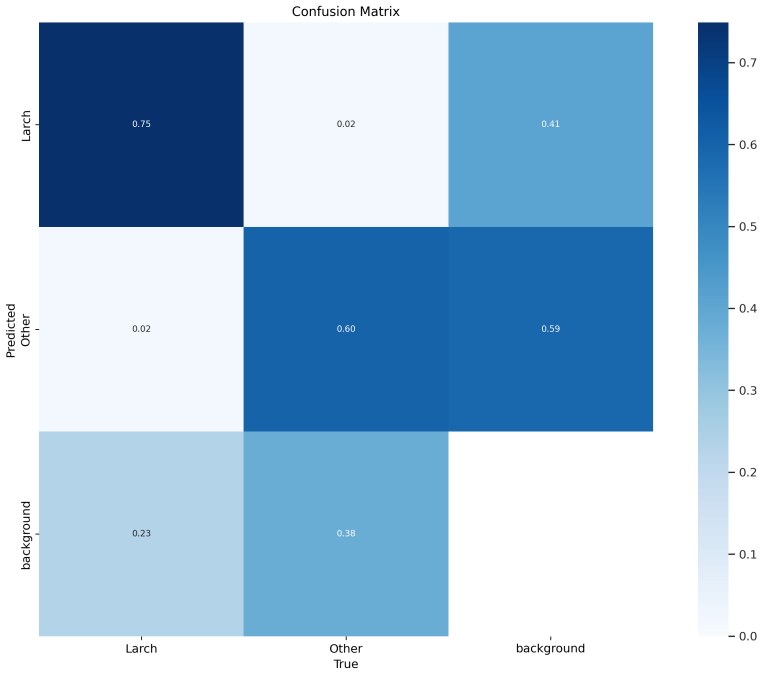

Figure 4.7: Confision matrix of YOLOv5 on Forest Damages – Larch Casebearer Dataset

The results are provided in Figure 4.8. The graphical representations in the graphs reveal notable fluctuations characterized by upward and downward movements across all the curves. These variations suggest the possibility of encountering overfitting challenges in the classification model. Specifically, as we observe the precision and recall values, they start to exhibit a decreasing trend beyond a certain threshold. This indicates that the model's performance becomes less accurate and less effective in correctly identifying and classifying instances as we move past that threshold. The declining precision and recall values imply that the model may be capturing noise or irrelevant patterns in the data, leading to reduced generalization capability and potentially compromising its overall predictive power. Therefore, it is crucial to address the overfitting concern and fine-tune the model to ensure better performance and reliability in real-world applications.

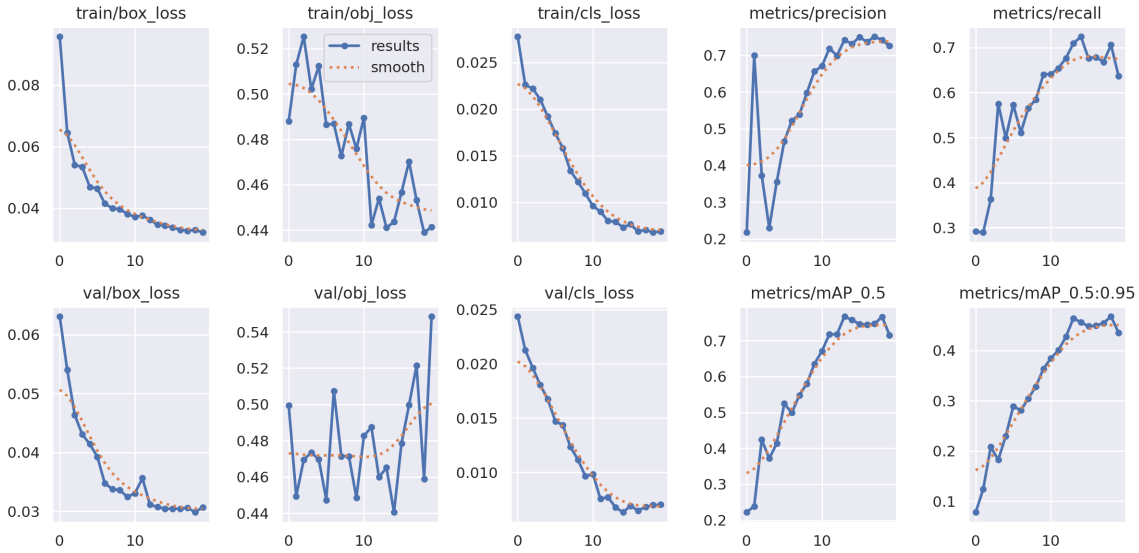

Figure 4.8: Results of YOLOv5 on Forest Damages – Larch Casebearer Dataset

Figure 4.9 illustrates the precision-recall curve. At the beginning of the curve, we observed relatively high precision values. This indicates that when the model classified instances as a tree, Larch or other, it was accurate in its predictions. As we moved along the curve, the precision values gradually decreased while the recall values increased. This indicates that the model started classifying more instances as tree, including some false positives, leading to a slight decrease in precision. The curve showcased the inherent trade-off between precision and recall. As we aimed for higher recall, the precision decreased, meaning that the model was classifying more instances as tree but with a higher likelihood of false positives. We observed a significant bend or inflection point in the curve. This bend indicated a critical threshold where the model transitioned from higher precision to higher recall. It represented the point where the model achieved the best balance between precision and recall.

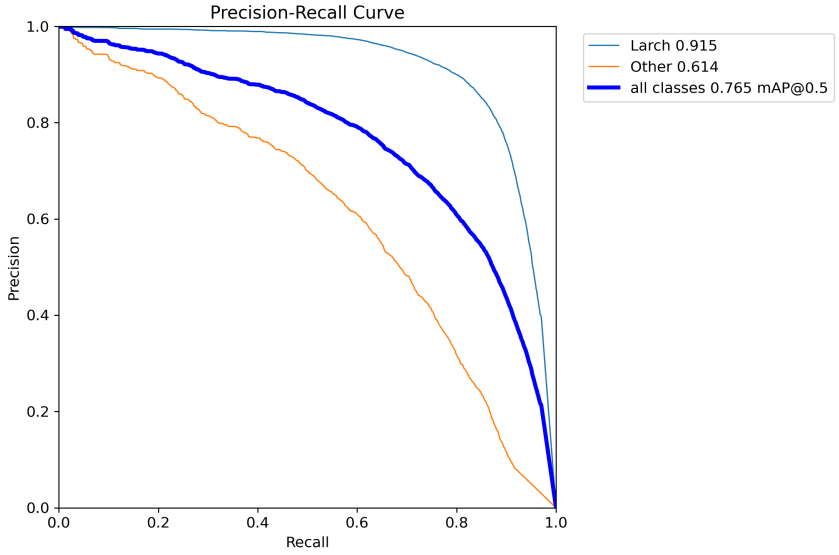

Figure 4.9: Precision-Recall graph

The evaluation metrics of YOLOv5 are presented in Table 4.5. The table provides the result or performance of the method for tree canopy identification on the dataset datasets. The table includes various evaluation metrics and results obtained from each method, allowing for a quantitative assessment of their effectiveness.

Variant	P	R	mAP50	F1
YOLOv5	74.2%	70.4%	76.5%	72.3%

Table 4.5: The precision, recall, mAP50, and F1 Score of YOLOv5 on Forest Damages – Larch Casebearer Dataset

Table 4.6 shows the training time, inference time, and model size of YOLOv5. The complexity table provides a comparative analysis of the computational complexity and resource requirements of different methods used on the dataset. It highlights various aspects of complexity, allowing researchers to assess the feasibility and practicality of each method.

Method	Training Time	Inference Time	Model Size
YOLOv5	4917.6 s	0.5 ms	15.4 MB

Table 4.6: Complexity of YOLOv5 on Forest Damages – Larch Casebearer Dataset

4.4 Results on DeepGlobe Dataset

The DeepGlobe Land Cover Classification Challenge [26] wants computers to automatically sort different types of land like cities, farms, forests, and water areas. This helps in better city planning, farming, and looking after the environment. So far, we trained this dataset in DeepLabV3+[31] and U-net[30] model. U-Net is a popular deep learning architecture commonly used for image segmentation tasks, known for its U-shaped structure that captures features at various scales. DeepLabV3+ is an advanced model for semantic image segmentation that combines spatial pyramid pooling and encoder-decoder structures to capture object details and boundaries. Both are employed to segment objects and areas within images with high accuracy.

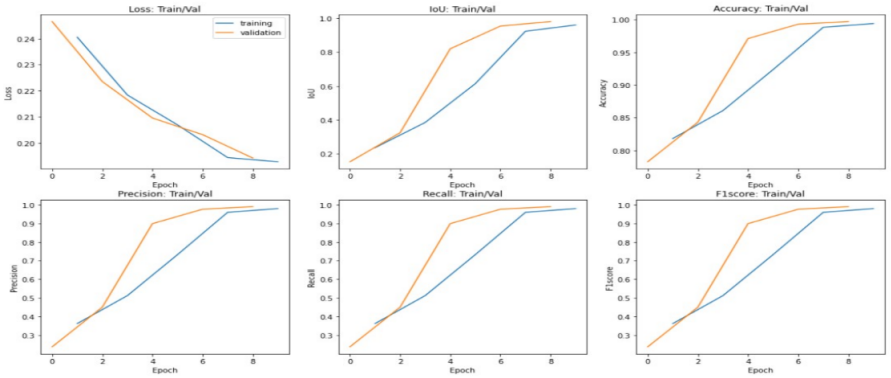

Figure 4.10: Loss, IoU, Accuracy, Precision, Recall and F1 Score Curves of U-Net with respect to Epochs

This graphical depiction depicts the performance of the U-net model on the DeepGlove Landcover dataset, displaying several performance and loss indicators. Notably, measurements like Intersection over Union (IoU), accuracy, precision, recall, and F1 score show comparable upward patterns. However, it is clear that the validation set's performance measurements diverge significantly after a considerable improvement, indicating a noticeable shift in model performance.

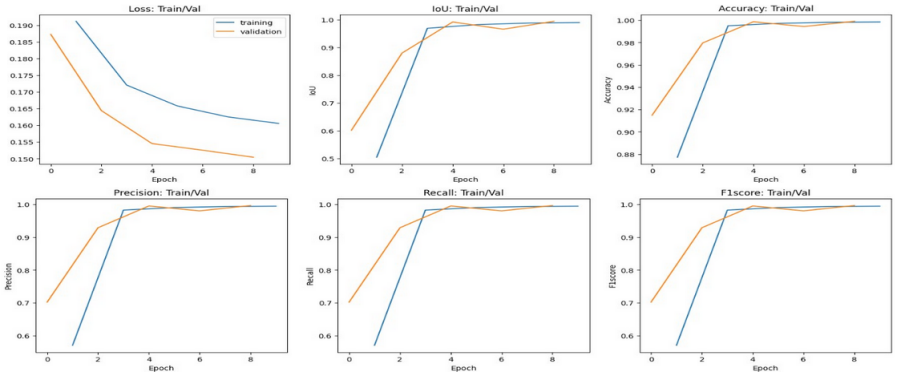

Figure 4.11: Figure 12: Loss, IoU, Accuracy, Precision, Recall and F1 Score Curves of DeepLabV3+ with respect to Epochs

The graphical data shows DeepLabV3+ outperforms U-net in the DeepGlove Landcover dataset, with persistent improvements in IoU, accuracy, precision, recall, and F1 score. The merging of atrous convolution and spatial pyramid pooling in DeepLabV3+ enables robust multi-scale context awareness. Dilated convolutions improve the receptive field without inflating parameters, allowing for fine delineation of complex landcover elements and giving it a significant performance advantage.

Variant	Accuracy	Precision	Recall	F1 Score
U-net	99.76%	99.17%	99.16%	99.16%
DeepLabV3+	99.96%	99.89%	99.87%	99.88%

Table 4.7: Evaluation metrics comparison

In this table, it is clear that DeepLabV3+ surpasses U-net across all performance criteria, showing a slightly greater overall performance.

Method	Training Time	Interface Time	Model Size
U-net	636 s	41 s	71 MB
DeepLabV3+	551 s	38 s	26.7 MB

Table 4.8: Model Complexity Comparison

Even in the model complexity comparison, DeepLabV3+ outperforms U-net marginally. Training the same dataset takes less time and memory.

4.5 Work Progress in SAM

The Segment Anything Model (SAM) [33] is an advanced tool for accurately segmenting satellite images and tree canopy. High accuracy, adaptability to various data, zero-shot learning, and user-friendliness are among SAM's characteristics. It maps tree cover effectively, assists agricultural land monitoring, and aids urban planning by segmenting elements such as houses and highways.

Figure 4.12: Result of segmentation done by SAM (Segment Anything Model) on a random image.

This is an example of a SAM (Segment Anything Model) segmentation work on a random image, where we can examine how well it succeeds at segmenting humans and other instances.

Still, We avoided using SAM (Segment Anything Model) for tree canopy segmentation tasks due to its low performance in this application. Despite our efforts, it did not match the accuracy and precision levels required for accurate tree canopy segmentation. As a result, various segmentation methodologies were investigated in order to produce more accurate and effective results in this setting.

Figure 4.13: Result of SAM on a tree canopy segmentation task.

This is an example of SAM (Segment Anything Model) output from a tree canopy segmentation task. It misses most of the trees because of the same color backdrop, and the polygons aren't very nice either. While SAM (Segment Anything Model) has shown efficacy in picture segmentation across a variety of images, it falls short in reliably segmenting tree canopies. The rich and subtle aspects of canopy photography provide hurdles for SAM, driving our quest for more acceptable options for this specific task.

Still, We avoided using SAM (Segment Anything Model) for tree canopy segmentation tasks due to its low performance in this application. Despite our efforts, it did not match the accuracy and precision levels required for accurate tree canopy segmentation. As a result, various segmentation methodologies were investigated in order to produce more accurate and effective results in this setting.

In the course of developing the SAM for multi-class land cover and land use segmentation, we encountered a notable challenge when implementing the softmax activation function, intended for multi-class tasks, which prompted a warning during model training. Despite the reduction in loss during training, the model exhibited a perplexing behavior during inference, consistently predicting entire images as a single class, resulting in a uniform segmentation mask across the entire scene.

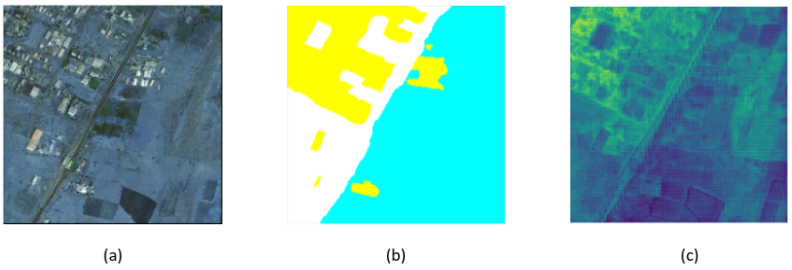

Figure 4.14: (a) Original Image, (b) Ground Truth, (c) Mask Predicted Mask

This is the predicted mask of Segment Anything Model (SAM) during our training the data. This anomaly, currently under investigation, is acknowledged as a work in progress. Our ongoing efforts focus on exploring the solution of this problem, fine tuning, and rigorous debugging to rectify this issue and improve the model's segmentation accuracy, with the aim of contributing valuable insights to the field of computer vision and deep learning for land cover and land use analysis.

Chapter 5

Conclusion

Based on the proposed methodology and expected outcomes presented, it can be concluded that the use of deep learning models for identifying tree canopies in challenging backgrounds UAV imagery holds significant potential.

The proposed study aims to develop and evaluate deep learning models for accurately identifying tree canopies in different backgrounds.

In our project, we trained the YOLOv5, Faster R-CNN, and Mask R-CNN models on different datasets. We faced challenges with data formats and had to convert annotations for compatibility. While YOLOv5 training proceeded well, we encountered difficulties with the other models on one dataset. We refined and optimized the models through training and evaluated their performance using various metrics. Comparisons were made based on speed, memory utilization, and detection accuracy. Our objective was to accurately detect and classify objects within the datasets using the strengths of each model.

In the remaining stages of the project, crucial tasks include fine-tuning and optimizing the models, evaluating their performance with metrics, expanding the dataset through augmentation, and comparing results to existing algorithms for insights on competitiveness and effectiveness. Moreover, the study will use a new dataset of high-resolution UAV imagery, which will be preprocessed and augmented to improve the performance of the deep-learning models.

The proposed study has significant implications for various applications, including agriculture forest management and urban planning. Accurate identification of tree canopy can help monitor and manage forests, identify green spaces in urban areas, and assess crop health and yield in agriculture.

We expanded our investigation beyond tree canopy segmentation to the larger realm of land cover research. Training models such as DeeplabV3+ and U-Net successfully demonstrated the promise of deep learning in this sector. However, we experienced difficulties in efficiently implementing SAM (Segment Anything Model) for this purpose. Recognizing SAM's repute for segmentation jobs, we intend to research it further and include it into our model. Using SAM's capabilities might greatly improve our capacity to properly segment and categorize land cover, widening the scope of our research and perhaps giving more complete results.

In conclusion, the proposed study has the potential to contribute to the field of remote sensing and deep learning by developing accurate and efficient models for identifying tree canopies in challenging backgrounds from UAV imagery.

Bibliography

[1] P. Cinat, S. F. Di Gennaro, A. Berton, and A. Matese, "Comparison of unsupervised algorithms for vineyard canopy segmentation from uav multispectral images," Remote Sensing, vol. 11, no. 9, p. 1023, 2019.

[2] J. A. C. Martins, K. Nogueira, L. P. Osco, F. D. G. Gomes, D. E. G. Furuya, W. N. Gonçalves, D. A. Sant'Ana, A. P. M. Ramos, V. Liesenberg, J. A. dos Santos et al., "Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning," Remote Sensing, vol. 13, no. 16, p. 3054, 2021.

[3] Z. Song, Z. Zhou, W. Wang, F. Gao, L. Fu, R. Li, and Y. Cui, "Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting," Computers and Electronics in Agriculture, vol. 181, p. 105933, 2021.

[4] X. Zhang, L. Fu, M. Karkee, M. D. Whiting, and Q. Zhang, "Canopy segmentation using resnet for mechanical harvesting of apples," IFAC-PapersOnLine, vol. 52, no. 30, pp. 300–305, 2019.

[5] J. Mo, Y. Lan, D. Yang, F. Wen, H. Qiu, X. Chen, and X. Deng, "Deep learning-based instance segmentation method of litchi canopy from uav-acquired images," Remote Sensing, vol. 13, no. 19, p. 3919, 2021.

[6] Z. Lu, L. Qi, H. Zhang, J. Wan, and J. Zhou, "Image segmentation of uav fruit tree canopy in a natural illumination environment," Agriculture, vol. 12, no. 7, p. 1039, 2022.

[7] P. Sicard, F. Coulibaly, M. Lameiro, V. Araminiene, A. De Marco, B. Sorrentino, A. Anav, J. Manzini, Y. Hoshika, B. B. Moura et al., "Object-based classification of urban plant species from very high-resolution satellite imagery," Urban Forestry & Urban Greening, vol. 81, p. 127866, 2023.

[8] H. Jemaa, W. Bouachir, B. Leblon, and N. Bouguila, "Computer vision system for detecting orchard trees from uav images," 2022.

[9] D. Verma, O. Mumm, and V. M. Carlow, "Identifying streetscape features using vhr imagery and deep learning applications," Remote Sensing, vol. 13, no. 17, p. 3363, 2021.

[10] B. G. Weinstein, S. Marconi, M. Aubry-Kientz, G. Vincent, H. Senyondo, and E. P. White, "Deepforest: A python package for rgb deep learning tree crown delineation," Methods in Ecology and Evolution, vol. 11, no. 12, pp. 1743–1751, 2020.

[11] T. Zhao, Y. Yang, H. Niu, D. Wang, and Y. Chen, "Comparing u-net convolutional network with mask r-cnn in the performances of pomegranate tree canopy segmentation," in Multispectral, hyperspectral, and ultraspectral remote sensing technology, techniques and applications VII, vol. 10780. SPIE, 2018, pp. 210–218.

[12] L. Qi, J. Zhou, J. Wan, Z. Yang, H. Zhang, and Z. Cheng, "Canopy recognition of cherry fruit tree based on segnet network model," in International Conference on Optics and Image Processing (ICOIP 2021), vol. 11915. SPIE, 2021, pp. 92–104.

[13] L. Zheng, D. Shi, and J. Zhang, "Segmentation of green vegetation of crop canopy images based on mean shift and fisher linear discriminant," Pattern Recognition Letters, vol. 31, no. 9, pp. 920–925, 2010.

[14] B. Cai, X. Li, and C. Ratti, "Quantifying urban canopy cover with deep convolutional neural networks," arXiv preprint arXiv:1912.02109, 2019.

[15] D. Lobo Torres, R. Queiroz Feitosa, P. Nigri Happ, L. Elena Cué La Rosa, J. Marcato Junior, J. Martins, P. Olã Bressan, W. N. Gonçalves, and V. Liesenberg, "Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution uav optical imagery," Sensors, vol. 20, no. 2, p. 563, 2020.

[16] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks," Remote Sensing, vol. 11, no. 11, p. 1309, 2019.

[17] A. A. d. Santos, J. Marcato Junior, M. S. Araújo, D. R. Di Martini, E. C. Tetila, H. L. Siqueira, C. Aoki, A. Eltner, E. T. Matsubara, H. Pistori et al., "Assessment of cnn-based methods for individual tree detection on images captured by rgb cameras attached to uavs," Sensors, vol. 19, no. 16, p. 3595, 2019.

[18] Y. Gan, Q. Wang, and A. Iio, "Tree crown detection and delineation in a temperate deciduous forest from uav rgb imagery using deep learning approaches: Effects of spatial resolution and species characteristics," Remote Sensing, vol. 15, no. 3, p. 778, 2023.

[19] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y.-A. Liou, and A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations—a review," Remote Sensing, vol. 12, no. 7, p. 1135, 2020.

[20] P. Ulmas and I. Liiv, "Segmentation of satellite imagery using u-net models for land cover classification," arXiv preprint arXiv:2003.02899, 2020.

[21] Z. Wang, S. Sun, X. Que, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[22] S. Natesan, C. Armenakis, and U. Vepakomma, "Individual tree species identification using dense convolutional network (densenet) on multitemporal rgb images from uav," Journal of Unmanned Vehicle Systems, vol. 8, no. 4, pp. 310–333, 2020.

[23] "Six ways drones are revolutionizing agriculture," www.technologyreview.com/2016/07/20/158748/six-ways-drones-are-revolutionizing-agriculture/, [Online: Last Accessed April 10, 2023].

[24] B. Weinstein, S. Marconi, and E. White, "Data for the neontreeevaluation benchmark," Jan. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.5914554

[25] Swedish Forest Agency, "Forest Damages – Larch Casebearer 1.0. ," https://lila.science/datasets/forest-damages-larch-casebearer/, 2021, National Forest Data Lab.

[26] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018.

[27] S. Jha, A. Dey, R. Kumar, and V. Kumar, "A novel approach on visual question answering by parameter prediction using faster region based convolutional neural network," IJIMAI, vol. 5, no. 5, pp. 30–37, 2019.

[28] "pyimagesearch," https://pyimagesearch.com/2018/11/19/mask-r-cnn-with-opencv/, [Online: Last Accessed June 30, 2023].

[29] "Overview of model structure about yolov5," https://github.com/ultralytics/yolov5/issues/280, [Online: Last Accessed June 30, 2023].

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015, pp. 234–241.

[31] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv:1706.05587, 2017.

[32] "Sam," https://learnopencv.com/segment-anything/, [Online: Last Accessed September 15, 2023].

[33] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo et al., "Segment anything," arXiv preprint arXiv:2304.02643, 2023.

[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[35] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," 2018.

[36] G. Jocher, "Ultralytics yolov5," 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[37] "Neontreeevaluation dataset," https://github.com/weecology/NeonTreeEvaluation, [Online: Last Accessed June 30, 2023].

[38] "Forest damages:larch casebearer," https://lila.science/datasets/forest-damages-larch-casebearer/, [Online: Last Accessed June 30, 2023].