APPLICATION OF UAS PHOTOGRAMMETRY AND GEOSPATIAL AI TECHNIQUES FOR PALM TREE DETECTION AND MAPPING.

A Thesis

by

PRATIKSHYA REGMI

BE, Kathmandu University, 2018

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

GEOSPATIAL SYSTEMS ENGINEERING

Texas A&M University-Corpus Christi
Corpus Christi, Texas

August 2023

APPLICATION OF UAS PHOTOGRAMMETRY AND GEOSPATIAL AI TECHNIQUES FOR PALM TREE DETECTION AND MAPPING.

A Thesis

by

PRATIKSHYA REGMI

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Michael J Starek, PhD
Chair

Tianxing Chu, PhD
Committee Member

Antonio Medrano, PhD
Committee Member

August 2023

ABSTRACT

Uncrewed aircraft systems (UAS), commonly known as drones, underwent significant advancements in recent years, particularly in the development of improved sensors and cameras that enabled high-resolution imagery and precise measurements. This study utilized a UAS to capture aerial imagery of Texas A & M University-Corpus Christi (TAMUCC) main campus, which was then processed using Structure-from-Motion (SfM) photogrammetric software to generate orthomosaic imagery. The primary purpose of this study was to utilize the orthomosaic imagery acquired from UAS to detect, map, and quantify the number of palm trees. Initially, three deep-learning models were trained using the same set of training samples. The model exhibiting the highest performance in terms of precision, recall, and F1-Score was selected as the optimal model. The model obtained through the fine-tuning of a pre-trained GIS-based model with additional training samples was identified as the optimal choice, yielding the following values: precision=0.88, recall=0.95, and F1-score=0.91. This model successfully detected a total of 1414 sabal palm trees within our study area. The chosen optimal model was employed to examine the impact of ground sampling distance (GSD) on the deep learning model. GSD values were varied, namely 5 cm, 10 cm, 20 cm, and 40 cm. The findings revealed that the model's performance deteriorated as the resolution decreased. Furthermore, the optimal model was subjected to an additional test using multi-temporal datasets with approximately the same GSD (1.5 cm). These datasets included one acquired a year prior to the model's training datasets, and another obtained three months after the training datasets. Remarkably, the results demonstrated that the model maintained a comparable level of accuracy across all three testing datasets. The obtained results were verified using ground truth values taken in a small portion of the study area. This study concludes that deep learning models for object detection exhibit superior performance when fine-tuned with training samples specific to the area of interest. Furthermore, it is evident that the optimal model's effectiveness diminishes significantly when the imagery resolution is reduced. Additionally, the performance of the deep learning model remains relatively consistent when applied to datasets acquired at different time frames, as long as

the resolution of the testing data remains the same. In summary, the application of deep learning demonstrates its efficacy, user-friendliness, and time-saving capabilities for object detection. This study shows how we can use UAS and deep learning to detect palm trees. It helps us develop better ways to monitor and manage palm trees.

DEDICATION

I dedicate this thesis to my parents, who have always believed in me and supported me in all of my endeavors. Their love and encouragement have been invaluable to me, and I am forever grateful to them.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr. Michael J. Starek, for his exceptional guidance, unwavering support, and continuous encouragement throughout the entire research and writing process. His expertise and insights were invaluable in shaping the direction of this study. I am also immensely grateful to the committee members of my research team, Dr. Tianxing Chu and Dr. Antonio Medrano, for their tireless efforts and significant contributions to the project, which greatly enriched the quality and depth of this work.

Furthermore, I wish to express my appreciation to Scott Meares, Associate Vice President for Operations-Finance And Administration, and Minerva Alaffa, Business Administrator-Administrative Operations, for their generous financial support of this study through the campus survey project. Their contribution played a crucial role in the successful completion of this research. I am equally grateful to the entire facility administration team for their unwavering assistance and support, which proved invaluable throughout this rewarding journey.

Last, but certainly not least, I am deeply grateful to my beloved family and dear friends for their unwavering love, support, and understanding throughout this challenging yet immensely rewarding journey. Their presence and encouragement provided me with the strength and motivation to overcome obstacles and reach this significant milestone in my academic pursuit.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Background

Uncrewed Aircraft System (UAS), commonly known as drones, have seen significant advancements in recent years. One of the main areas of advancement has been in the development of improved sensors and cameras, which have allowed for higher-resolution imagery and more accurate measurements (Colomina & de la Tecnologia (2008)). This has made UAS particularly useful for applications such as mapping, surveying, inspection, agriculture, etc (Gupta et al. (2013)).

In addition to the hardware improvements, there have also been significant advancements in software and algorithms for processing the data collected by UAS. One such example is Structure-from-Motion(SfM)/multi-view stereo (MVS) photogrammetry, which is a technique that uses multiple images taken from different viewpoints to create 3D models and maps of objects or environments (Starek & Wilkinson (2022)). SfM algorithms are able to automatically identify common features in the images and use them to align the images and create a 3D model. This process has been greatly improved in recent years, allowing for more accurate and efficient 3D modeling using UAS. Overall, the advancements in UAS-SfM photogrammetry have greatly increased the capabilities of these technologies and have led to their widespread adoption in a variety of industries and applications (Bäumker et al. (2013)).

The use of UAS, combined with SfM photogrammetry (referred to as UAS-SfM) has several benefits in surveying and mapping for geospatial data acquisition. Among these benefits, cost savings stand out as a primary advantage, as using UAS often proves to be more cost-effective than traditional surveying and mapping techniques, particularly at localized geographic scales. Additionally, UAS technology enables the inspection of hazardous environments or structures without exposing human personnel to risk, while also facilitating efficient coverage of large areas. Furthermore, these technologies have the capability to generate highly accurate 3D models, maps, videos, and other outputs, thus rendering them valuable for a diverse range of applications. Lastly, the application of UAS photogrammetry contributes to the reduction of environmental impact associated with geospatial data collection (Berra & Peppa (2020)).

One way to further improve the georeferencing accuracy of 3D models and maps created using UAS and SfM is to use Ground Control Points (GCPs). GCPs are physical markers placed on the ground at known locations, and their positions are precisely measured using survey-grade equipment. By including GCPs in the images taken by the drone, it is possible to improve the accuracy of the resulting 3D model by using the known positions of the GCPs as reference points. Using GCPs can help to overcome errors and distortions that may be present in the images, such as those caused by camera lens distortion or the movement of the drone. The use of GCPs can therefore further enhance the benefits of using UAS and SfM photogrammetry, such as cost savings, speed, and accuracy, making them even more useful for a wide range of applications (Awasthi et al. (2020)).

Direct georeferencing using Real Time Kinematics(RTK) or Post-Processing Kinematics(PPK) enabled onboard GNSS receivers on UAS provides precise positioning and accurate georeferencing. Integration of RTK/PPK on UAS allows the reconstruction of highly detailed and precise outputs without using GCPs (Famiglietti et al. (2021)). The evolution from the use of the indirect georeferencing method to direct georeferencing has come a long way with the advancement in UAS technology. Initially, the indirect georeferencing method used to be employed, and now it has evolved towards direct georeferencing. Direct georeferencing enabled with high-quality GNSS receivers gives very accurate photogrammetric products (Regmi et al. (2023)). Although the autonomous UAS equipped with GNSS receivers it gives mapping-grade data. To use UAS for accurate applications such as surveying, survey grade accuracy is required. So, RTK/PPK methods are enabled along with high-quality GNSS receivers. Although direct georeferencing using RTK/PPK enabled with GNSS gives survey-grade data we need to verify whether accuracy is in an acceptable range or not. To check the accuracy of photogrammetric products it is recommended to use GCPs as check points. They will be used to access the horizontal and vertical accuracy of products for Quality Assurance/Quality Control(QA/QC) of the obtained output (Starek & Wilkinson (2022)).

Segmentation, image classification, and object detection are fundamental tasks in computer vision. Segmentation involves partitioning an image into distinct regions or objects, enabling pixel-level understanding of the scene (Long et al. (2015)). Image classification aims to assign predefined labels or classes to an entire image, determining the presence of specific objects or scenes (Krizhevsky et al. (2017)). On the other hand, object detection focuses on localizing and classifying multiple objects within an image, providing bounding boxes and class labels for each detected object (Ren et al. (2015)). These techniques can be applied to photogrammetric products ie. orthomosaic imagery created using UAS-SfM methods for segmentation, image classification, and object detection (Yuheng & Hao (2017)). There are many potential applications for image segmentation and object detection in UAS-SfM-derived photogrammetric mapping products, including infrastructure inspection, agriculture, natural resources management, environmental monitoring, and disaster response (Hoeser & Kuenzer (2020); Mejias & Fitzgerald (2013)). These techniques can help to extract useful information from the 3D models and orthomosaics created using UAS-SfM photogrammetry, allowing for more effective decision-making and resource management (Hoeser & Kuenzer (2020)). Semantic segmentation is a computer vision task that involves assigning specific labels or categories to each pixel or object in an image while preserving the georeferencing location of the elements (L.-C. Chen et al. (2017)). Unlike object detection or classification, which focuses on identifying and categorizing objects, semantic segmentation provides a pixel-level understanding of the scene by assigning meaningful labels to individual pixels (Cordts et al. (2016)). This fine-grained level of analysis enables detailed and accurate mapping of objects and regions in the image, facilitating applications such as autonomous driving, remote sensing, and urban planning (LeCun et al. (2015)).

Machine learning and deep learning techniques can be used to improve the accuracy and efficiency of object detection in photogrammetric products created using UAS and SfM methods. These techniques involve training a computer program to recognize patterns and features in data, allowing it to automatically classify and identify objects in images (Singh (2019)). One way in which machine learning and deep learning can be used for object detection in UAS-SfM-derived photogrammetric products is by training a model to recognize specific objects of interest, such as infrastructure defects or types of vegetation. The model can then be applied to the images from the UAS and SfM photogrammetry to automatically locate and classify these objects.

In recent years, the utilization of deep learning techniques has experienced a significant surge across a wide range of applications, including the geospatial sector. Deep learning models have exhibited remarkable performance, surpassing traditional machine learning methods in numerous domains (LeCun et al. (2015)). These models, with their ability to learn hierarchical representations from data, have demonstrated superior capabilities in tasks such as image recognition, object detection, and natural language processing (Goodfellow et al. (2016)). The geospatial sector has also witnessed the benefits of deep learning, as it enables improved analysis and understanding of geospatial data, leading to enhanced mapping, land cover classification, and geospatial information extraction (Ma et al. (2019)). The increased adoption of deep learning signifies its potential to revolutionize geospatial data analysis and open new avenues for geospatial applications (Pashaei et al. (2020)).

1.2 Study purpose and objectives

The study purpose of this study is to utilize a high-resolution three-band (RGB) digital camera-equipped UAS to obtain aerial imagery of the Texas A&M University-Corpus Christi (TAMU-CC) campus for the purpose of semantically segmenting the imagery to classify and locate Texas sabal palm trees within UAS-SfM generated orthomosaic imagery. Sabal palm trees, scientifically known as Sabal Mexicana (Ramp (1989)), are native to South Texas and are an iconic part of the region's flora. These palm trees are characterized by their tall and slender trunks, topped with a crown of large fan-shaped leaves. They are well adapted to the subtropical climate of South Texas, thriving in moist and well-drained soil, particularly along riverbanks and coastal areas (Everitt et al. (2002)). Sabal palm is a common species of palm found in our study area. The aerial imagery obtained using UAS will be processed and the orthomosaic image will be analyzed in order to assess the effectiveness and efficiency of geospatial artificial intelligence (Geo-AI) tools for palm tree detection/classification using hyperspatial resolution (¡ decimeter) UAS imagery. In the end, this project seeks to determine the most effective and efficient deep-learning method for detecting palm trees using UAS-SfM orthomosaic imagery acquired over the campus area. This information will be valuable for various applications, such as maintenance and planning for campus landscaping.

Objectives:

1) Evaluate the effectiveness and efficiency of a commercial GIS-based deep learning model (toolkit built into the software) versus customizable deep learning models for detecting sabal palm trees. And select the optimal model which performs better.

2) Evaluate the effect of the image's ground sampling distance (GSD), which is a function of the UAS altitude above ground level and sensor resolution, on the overall performance of the optimal model obtained from objective 1 for palm tree detection.

3) Apply the most effective Geo-AI model(optimal model), based on results from objective 1, to UAS imagery acquired on a different flight date than the flight used for model development. This allows assessment of the model's generalizability and its utility for palm tree change detection via future UAS-SfM surveys.

1.3 Motivation

This study aligns with a project named the campus survey project. In this project, UAS-SfM surveys of the campus are performed on a biannual and quarterly basis to derive 2D/3D mapping products. The TAMUCC Historic Imagery Geoportal(https://arcg.is/1ir0Wy0), which is funded by TAMUCC Operations and implemented by the Measurement Analytics Lab (MANTIS) of the Conrad Blucher Institute for Surveying and Science, is also a product of this project. The primary goal of this initiative is to track changes on the TAMUCC campus in order to enhance and accelerate campus maintenance efforts.

The application of GeoAI techniques for semantic segmentation, classification, and detection of the trees in UAS imagery brings forth numerous advantages. Specifically, detecting palm trees within our area of interest serves as a valuable use case. These techniques play a pivotal role in enhancing efficiency by automating the process of locating and classifying objects, particularly in UAS orthomosaic images, eliminating the need for manual inspection. The scalability of object detection algorithms enables efficient and accurate analysis of extensive areas, while their versatility extends their applicability to various industries and applications. By leveraging these GeoAI techniques, we can unlock significant improvements in the efficiency and effectiveness of palm tree detection and mapping processes. In this study, the performance of the deep learning model was evaluated by downsampling high-resolution imagery and testing it on larger GSD values. This approach aimed to assess the model's efficiency over larger areas solely by adjusting the flying altitude while assuming consistent camera settings such as resolution and focal length.

This study can also be used to monitor changes in palm trees, loss of palm trees after storms or due to disease (if they fall over), and monitor planted palms. This information can be valuable for resource management and conservation efforts. In addition, palm tree detection can inform urban planning decisions on campus, such as the placement of new palm trees or the removal of existing ones. Finally, palm tree detection can be used to create maps and guides for tourists visiting the university, highlighting the locations and types of palm trees on campus. The results obtained from this study have several potential uses in a university setting. One application is in landscaping and maintenance, where palm tree detection can be used to identify and map the locations of palm trees on campus, allowing for more efficient landscaping and maintenance efforts. Another use is in emergency preparedness, where accurate mapping of palm tree locations can help to predict the potential damage that may be caused by natural disasters such as hurricanes eg. Hurricane Harvey and to plan accordingly.

1.4 Literature review

The growth and advancement of UAS technology have been remarkable in recent years. The combination of technological developments, cost reductions, and improved regulations has led to increased adoption of UAS across diverse sectors. The most common applications of UAS are in geospatial data acquisition, urban planning development (Vanderhorst et al. (2021)), agroforestry (Pádua et al. (2017)), crop biomass monitoring (Wang et al. (2021)), forest ecosystems monitoring (Tomljanović et al. (2022)), facility condition monitoring (Jeong et al. (2022)), etc. The utilization of UAS has proven beneficial in facility management, allowing for efficient inspection and monitoring of various infrastructure assets. UAS enables facility managers to assess the condition of buildings, roofs, and other structures, enhancing maintenance and decision-making processes (Yasin et al. (2016)).

Using SfM/MVS photogrammetry techniques on overlapping UAS images, it becomes possible to generate precise and detailed 3D point cloud data representing land cover and terrain characteristics within the captured scene (Sturdivant et al. (2017)). Conventional photogrammetry necessitates the use of expensive and precisely calibrated metric cameras, which limits the widespread application of UAS for mapping purposes. In contrast, SfM takes advantage of multiple overlapping images to extract three-dimensional object information, eliminating the requirement for precise camera calibration. By analyzing the camera movement and capturing different perspective views of the scene, SfM reconstructs the three-dimensional structure from the two-dimensional image sequences (Starek & Wilkinson (2022)). The SfM image processing workflow involves key steps such as inputting image sequences, extracting features using algorithms like the scale-invariant feature transform(SIFT), performing bundle block adjustment for error minimization and sparse point cloud generation, introducing GCPs and initial camera positions for georeferencing, and using MVS algorithms for densifying the point cloud (Strecha et al. (2012); Slocum & Parrish (2017)).

The output obtained from UAS-SfM photogrammetric method can be used to perform some tasks like image segmentation, object detection, land cover classification, change detection, etc. (Yao et al. (2019)). Object detection, a problem in computer vision that involves identifying and locating objects in images or videos, has been a subject of significant research in recent years. It is considered to be one of the most fundamental and challenging problems in computer vision, and its development over the past two decades can be seen as a representation of the overall progress of the field of computer vision Zou et al. (2019)). The commonly used application of object detection algorithms is pedestrian detection (Dollar et al. (2011)), video surveillance (Sharma & Lohan (2019)), tree detection (Yang et al. (2009)), fault detection, and condition monitoring (Reddy & Banerjee (1990)), etc.

Object detection algorithms are often used to identify trees in images or videos by looking for objects with characteristics that are typical of trees. The use of object detection algorithms for palm tree detection in high-resolution aerial imagery has become increasingly popular in recent years (Malek et al. (2014)). These methods are now widely used for this purpose due to their effectiveness and ease of use (Wuest et al. (2016)). This can involve training a machine learning model on a dataset of labeled images that include palm trees so that the model can learn to recognize the visual characteristics of palm trees. Once the model has been trained, it can be used to identify palm trees in new images by detecting objects that have similar characteristics to the palm trees in the training dataset. There are a number of different object detection algorithms that can be used for this purpose, including machine learning-based approaches such as convolutional neural networks (CNNs) (Mubin et al. (2019)) and traditional computer vision approaches such as Haar cascades (Rahmat et al. (2019)).

GeoAI refers to the integration of AI and machine learning techniques with geospatial data and GIS. It involves leveraging advanced algorithms and models to analyze and extract insights from spatial data. Traditional machine learning methods in GeoAI typically involve manual feature extraction and the application of statistical models for geospatial data analysis (Gehrmann et al. (2018)). In contrast, deep learning, a subset of machine learning, employs artificial neural networks to automatically learn hierarchical representations from data, eliminating the need for explicit feature engineering (LeCun et al. (2015)). Deep learning models, such as CNNs have shown superior performance in GeoAI tasks like image classification, object detection, and semantic segmentation. The ability of deep learning to handle complex spatial patterns and extract high-level features contributes to its prominent role in advancing GeoAI capabilities (Guo et al. (2020)). Specifically, in UAS semantic image segmentation, Geo-AI plays a crucial role in accurately delineating and categorizing objects and regions within UAS imagery, facilitating applications such as land cover mapping and environmental monitoring (Li & Hsu (2022)). The adoption of GeoAI has gained significant momentum in GIS software, such as Esri's ArcGIS Pro platform, enabling users to incorporate AI capabilities for spatial analysis, image classification, and predictive modeling (Janowicz et al. (2020)).

In 2016, Li et al. (2016) proposed a deep learning-based framework for oil palm tree detection and counting using high-resolution images for Malaysia. The trees in the study area were crowded and had overlapping crowns, which have not been well-studied in previous palm tree detection research. It detected more than 96% of the trees in the study area, compared to manual interpretation. This accuracy was higher than the accuracies of the other three tree detection methods used in a similar study. The study by Li et al. (2018) used a two-stage convolutional neural network-based oil palm detection method in high-resolution imagery and their proposed approach achieved a much higher F1 score of 94.99% compared to existing methods such as single-stage CNN, Support Vector Machine (SVM), Random forest and Artificial neural Network(ANN).

In the ongoing exploration of land cover classification and object detection through high-resolution remote sensing, recent studies have revealed the promising potential of deep learning methods. A study done in 2020, (Zhang et al. (2020)) demonstrated that these advanced models outperform traditional pixel-based methods, especially in the classification of varied vegetation types. Furthermore, these deep learning approaches have achieved remarkable accuracies of over 98% in object detection tasks, thereby substantially reducing the labor intensity associated with traditional methodologies (Zhang et al. (2020)). Another study carried out to classify landcover classification in high-resolution imagery addresses the challenge of scarce labeled data in the remote sensing domain. It employs two strategies: transfer learning (TL) with fine-tuning and remote sensing-specific data augmentation. Using these techniques on the UC Merced dataset, it achieves land-cover classification accuracies of 97.8 ± 2.3%, 97.6 ± 2.6%, and 98.5 ± 1.4% with CaffeNet, GoogLeNet, and ResNet, respectively (Scott et al. (2017)).

Some studies show that deep learning algorithms based on Faster-RCNN employed to build the model extracts from images and detect palm trees are more effective, accurate detection, and correctly count the number of palm trees from the UAS images (Liu et al. (2021)). A study shows although Faster-RCNN performs well on well-known datasets for general object detection, it does not show clear advantages over other classical machine learning-based methods for oil palm tree detection (Zheng et al. (2019)). In this study, the authors mentioned that they tailored the Region Proposal Network (RPN) and proposed simple empirical planting rules which achieved a higher average F1- score than the normal Faster-RCNN method. Despite achieving good results using artificial intelligence(AI) techniques for smart palm tree detection in many studies, the effective and efficient management of large-scale palm plantations still remains a challenge (Hajjaji et al. (2022)).

In recent years, countries like Indonesia, Malaysia, and Columbia which are the largest producers and exporters of palm oil (Potter (2015)) have been using artificial intelligence methods in UAS imagery in recent years to research palm trees and increase production. Recently, one study carried out by Khaled et al. (2022) shows AI was used for spectral classification to identify basal stem rot disease in oil palm using dielectric spectroscopy measurements. They investigated the feasibility of applying a genetic algorithm (GA) as a feature selection algorithm and their result showed that the best classification accuracy was achieved using this approach. The Spread of the RED Weevil (RPW) has become an existing threat to palm trees around the world (Hussain et al. (2013)). A study by Kang et al. (2022) proposed a novel remote sensing approach to recognize and monitor red palm weevil in date palm trees using a combination of vegetation indices, object detection, and semantic segmentation techniques.

Recent research indicates that the application of deep learning techniques for classification and detection using high-resolution aerial imagery, integrated into GIS-based software like ArcGIS Pro, is gaining momentum. This surge in interest can be attributed to its user-friendly interface, rapid processing speed, and the high-quality results it delivers (Abd-Elrahman et al. (2021)). The increasing adoption of these methods is driven by several compelling benefits. For one, the integration of deep learning with GIS is relatively seamless, reducing the barrier to entry for users who are not necessarily experts in machine learning (Ma et al. (2019)).

A study conducted in 2020 by Pashaei et al. (2020) shows the efficiency of deep learning, specifically deep CNNs, for land cover prediction tasks using hyper-spatial resolution images from UAS, even in complex ecosystems like coastal wetlands. After reviewing and training several recent deep CNN architectures on a limited labeled image set, it was found that some simpler models perform comparably or even better than more complex, deeper architectures, and with fewer training epochs. This result is particularly significant in the realm of RS applications where the availability of extensive training samples is typically limited (Pashaei et al. (2020)). Another study in individual banana tree delineation in high-resolution UAS imagery shows that the CNN algorithm outperformed SVM, object-based image analysis (OBIA), and Neural Network (NN) techniques (Kuikel et al. (2021)).

2. STUDY AREA AND DATASETS

2.1 Study area

The study took place at the main island campus of Texas A & M University-Corpus Christi (TAMUCC), located in the South Texas region, Figure 1 shows the location map of TAMUCC. The campus is situated in zone 9b on the plant hardiness zone map, which provides favorable conditions for the growth of palm trees (Cathey (1990)). Palm trees are known for their distinct visual appeal and diverse species that can flourish in hot weather environments. The primary aim of this study is to conduct a comprehensive count of the palm trees, namely the sabal palm, present on the campus, as they comprise a significant portion of the overall plant population within the campus grounds.

Figure 1: Study area: Texas A & M University- Corpus Christi.

2.2 Data collection

To obtain aerial imagery for this study, the WingtraOne drone (Figure 2), manufactured by the Swiss company Wingtra, was utilized. The WingtraOne drone is renowned for its advanced capabilities in capturing high-resolution aerial images with exceptional precision. Equipped with a 42-megapixel full-frame camera, the drone ensures the acquisition of detailed imagery for accurate analysis. The flight control was facilitated by the WingtraPilot application, which offers intuitive controls and enables precise mission planning and execution. Furthermore, for post-processing of the acquired imagery, the WingtraHub platform was employed, utilizing PPK techniques to enhance the geolocation accuracy of the imagery (Wingtra (n.d.). The combination of the WingtraOne drone, WingtraPilot application, and WingtraHub platform provided a comprehensive solution for acquiring and processing the aerial imagery required for this study (Wingtra (2022)). According

Figure 2: WingtraOne UAS platform (left) and captured UAS during the flight (right).

to the manufacturer's manual, the WingtraOne UAS can achieve high levels of accuracy, with a horizontal accuracy of 1 cm and vertical accuracy of 2-3 times that factor (Wingtra (2022)). This UAS features vertical take-off and landing (VTOL) capabilities, which allows it to take off and land vertically, reducing damage from belly landings and enabling operation in more confined spaces. It is powered by a pair of UN3481 compliant, 99 Wh Li-ion batteries, and has an automatic landing accuracy of less than 5 m. The UAS can handle wind speeds up to 8 m/s, but flying in stronger winds is not recommended. The WingtraOne has three levels of speed: operational cruise speed (16 m/s), sink cruise speed (3 m/s), and sink hover speed (2.5 m/s), and its current market price is around $30,000 without advanced add-ons (Wingtra (2022)). Table 1 shows the specifications of WingtraOne UAS.

Table 1: WingtraOne UAS specifications.

SN	Details
Classification	VTOL
Battery	Li-ion, UN3481 compliant
Weight (with batteries)	3.7 kg (8.1 lb)
Approximate flight time	59 minutes
Maximum winds sustained	43 km/h
Temperatures	-10 to +40 °C
Receiver	multi-frequency PPK GNSS receiver
Camera/Sensor	Sony RX1R II, 42 MP

For this study, the mission was planned by using WingtraPilot software with side and front overlaps of 80% and 75%, respectively. The transition height of UAS was 50m and the transition degree of 37°. The flight direction of UAS was 292°and the height above ground was 120m. The GSD setting at the flight planner was 1.6 cm/px. During this flight terrain following setting was disabled. Figure 3 shows the screenshot of the flight plan used for this research. Details about the flight are given in Table 2.

Table 2: Flight plan details.

SN	Details
Location	Texas A & M University-Corpus Christi
Transition Height	50m
Transition direction	37 degree
Height above ground	120 m
Ground sampling distance (GSD)	1.6 cm/px
Flight direction	292 degree
Flight time	57 minutes
Side overlap	80%
Front overlap	75%

Table 3 below displays the flight dates, platforms, flight height above ground level, and GSD values for each flight utilized in the research. Notably, a substantial portion of the study relied on the flight data collected on September 11, 2022, which served as the primary dataset for training

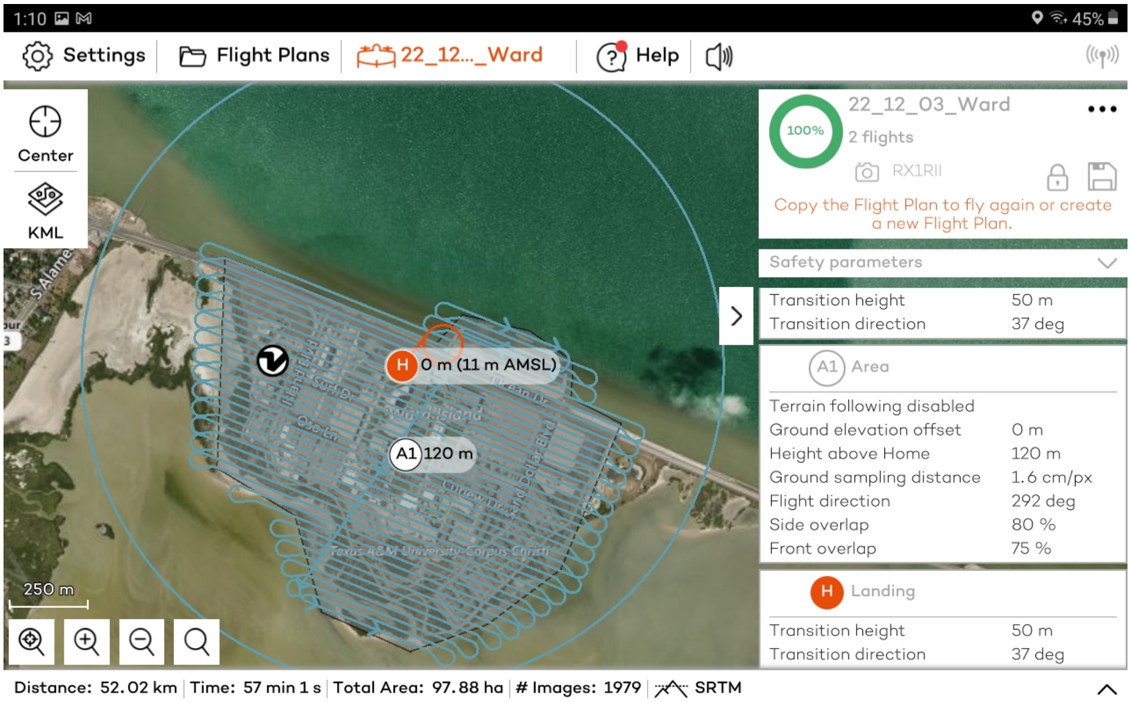

Figure 3: Screenshot of flight plan during image acquisition.

the models. The first and second objectives focused specifically on this flight data, while the third objective incorporated data from all three flights.

Table 3: Details of all UAS flights used for this study.

SN	Platform	Flight Date	Altitude(m)	GSD(cm/px)
1	Wingtra One Gen I	12/3/2022	120	1.6
2	***Wingtra One***	***9/11/2022***	***120***	***1.6***
3	Wingtra One	7/11/2021	120	1.6

The ground control data collection for this study involved conducting a survey using a Septentrio NR3 GNSS receiver. The Septentrio NR3 GNSS receiver is a high-precision global navigation satellite system (GNSS) device widely used for accurate positioning and data collection. It offers centimeter-level accuracy and is equipped with advanced features such as multi-constellation tracking and real-time kinematic (RTK) capabilities. The receiver operates on multiple satellite constellations, including GPS, GLONASS, Galileo, and BeiDou, ensuring robust and reliable positioning in various environments (Septentrio (n.d.)).

In this study, the collection of accurate ground truth GNSS data was accomplished using a high-precision survey-grade Septentrio NR3 GNSS receiver, as demonstrated in Figure 4. This particular GNSS receiver was chosen for its robustness, portability, and lightweight design, which make it a suitable option for fieldwork. Additionally, the device is equipped with quad-constellation, multi-frequency, and all-in-view Real-Time Kinematic (RTK) positioning, providing superior accuracy and enhancing anti-jamming and monitoring capabilities.

Figure 4: Field capture during GNSS survey.

Furthermore, the Septentrio NR3 GNSS receiver is equipped with features such as one-touch logging, an all-in-one base, rover operation, and 4G/LTE connectivity, which streamline data collection and transmission. The table 4 presents the technical specifications of the device used. A series of ground control targets were obtained and utilized as checkpoints to ensure the accuracy of the collected data. The rover pole was set at a height of 2 meters, and the epoch was set to 10 seconds, in order to maintain consistency and reliability throughout the data collection process. The base station was used for PPK image geotagging. The base station collected static GNSS data for two hours using regular RTK. And, the raw positioning data were transmitted to the Online Positioning User Service (OPUS) and underwent a correction in order to obtain the final location.

The data was collected on 9/30/2021. The utilization of this survey-grade GNSS receiver and careful data collection protocols ensured the acquisition of high-quality, accurate ground truth data in this study.

Table 4: GNSS receiver specifications.

SN	Details
Company	Septentrio
Model	Altus NR3
GNSS Technology	NavIC: L5
RTK-INS Horizontal accuracy	0.6 cm
RTK-INS Vertical accuracy	1cm
Certificates	CE, FCC6 Class B Part 15, ISO 9001-2015

2.3 Computing resources

In this study, the aerial imagery collected using UAS was processed using Agisoft Metashape Professional version 2.0.1 software. ArcGIS Pro Version 2.9.0 was utilized for visualization, map making, and the implementation of the GIS-based deep learning model. Python version 3.10.1 was employed for the customizable model. The hardware specifications for carrying out these tasks included an Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz 3.70 GHz processor, 64 GB of RAM, and an NVIDIA Quadro RTX 5000 GPU. These software tools and hardware specifications were instrumental in facilitating the successful execution of the research tasks and analysis.

3. METHODOLOGY

3.1 UAS image processing

The WingtraOne UAS platform was employed for conducting this study, utilizing three distinct flights. Table 5 below displays the dates of each flight used in the study. Notably, the majority of the study was conducted using the flight data collected on 09/11/2022, which served as the primary dataset for training the models. Objectives one and two were based on this specific flight data, while objective three incorporated all three flights.

Table 5: UAS flights used for this study

SN	Date
1	12/3/2022
2	***9/11/2022***
3	7/11/2021

Following the collection of raw image data from a UAS flight, post-processing procedures were conducted to generate orthomosaic imagery. The SfM photogrammetry technique was utilized in this study to process the collected images. SfM employs similar features present in overlapping images to extract a series of three-dimensional points, resulting in the generation of accurate orthomosaic imagery. Several user-friendly, open-source SfM photogrammetry processing software options are available for mapping data obtained from UAS. In this study, the Agisoft Metashape Professional software (Agisoft (n.d.)) was utilized for processing the collected data. This software is known for its reliability and accuracy in generating high-quality orthomosaics, making it a suitable choice for processing the data obtained in this study.

SfM is a powerful technique for generating orthomosaics from UAS imagery. The process begins with the collection of raw image data, which is then used by SfM photogrammetry software to extract 3D points from overlapping images. The software refines the camera positions through bundle adjustment, which improves accuracy. These points are then used to generate a dense point cloud, which is filtered to remove noise and outliers. Next, a digital surface model (DSM) is created by interpolating the point cloud to generate a continuous elevation model. The DSM is used to project the images and correct for terrain relief distortions, producing high-resolution orthomosaic imagery by mosaicking the images together. The orthomosaic is georeferenced to a specified coordinate system, and its output format is suitable for further analysis (Stanton et al. (2017)). Figure 5 shows the summary of the SfM workflow. The orthomosaic obtained from SfM is used later on for object detection.

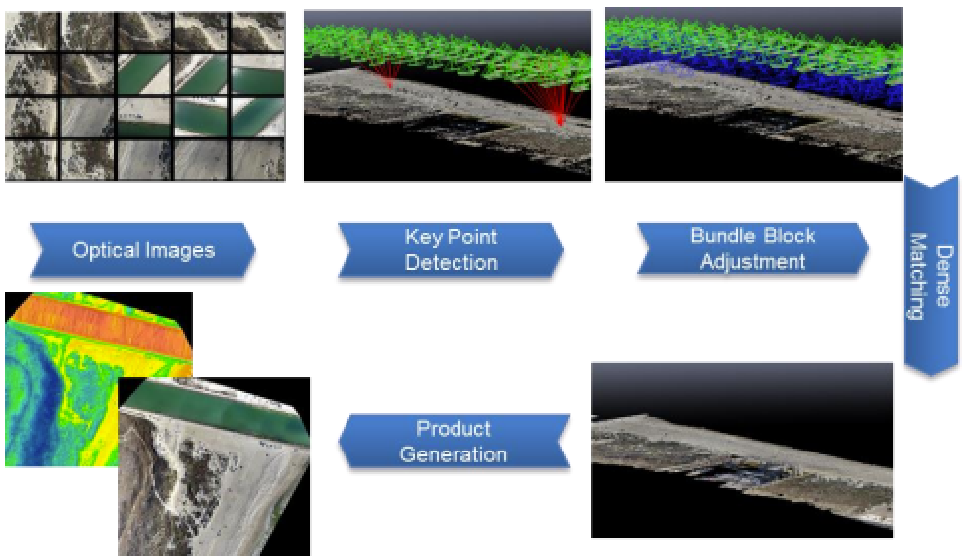

Figure 5: Summary of SfM workflow (Starek et al. (2019)).

3.2 Deep learning methods

In this study, two types of deep learning methods were employed to detect objects in a specific area of interest. One approach involved using a commercially available GIS-based pre-trained deep learning model, specifically Esri's pre-trained model for palm tree detection. The other method relied on the customizable model, namely maskRCNN.

While the commercialized GIS-based object models are user-friendly and require minimal modifications, they may not always be a perfect fit for our area of interest since they are trained on representative datasets. Although these models can be fine-tuned to improve their accuracy for specific applications, we may face limitations in modifying their architecture or parameters beyond what has been programmed into the tool.

In contrast, the customizable models offer greater flexibility and control over the model's parameters, making it possible to optimize the model's accuracy, recall, and precision for our study area. By tweaking the model's parameters and adding new code, more specific and detailed results were obtained that meet our needs. However, using these models may require more expertise and time to implement effectively.

3.2.1 Esri's pre-trained model for palm tree health detection

Esri's pre-trained palm tree health detection object detection model has emerged as a promising tool in the realm of remote sensing and precision agriculture, given its capacity to efficiently identify the health status of palm trees within large-scale plantations. This model, built on a deep learning architecture, is designed to detect and classify palm trees according to their health status, specifically identifying whether they are healthy, diseased, or dead (Esri (2022)).

The model's development is rooted in a robust deep learning framework that leverages convolutional neural networks (CNN) (Figure 6) and transfer learning. Transfer learning allows for the adaptation of pre-existing models trained on large-scale datasets to new applications with smaller datasets, making it possible to optimize model performance for specific use cases. The Esri model, in particular, was initially trained on a large dataset of images of palm trees, enabling it to accurately detect and classify palm tree health with high levels of accuracy (Esri (2022)). To achieve accurate

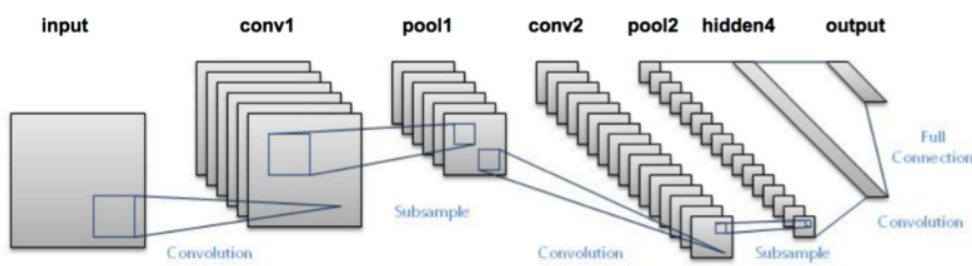

Figure 6: CNN Architecture.

palm tree detection, the model uses a set of pre-defined features, including color, texture, and shape. In addition, the model employs a region proposal network, which identifies potential palm tree locations in the input image, followed by a classification network that labels each detected palm tree according to its health status. The resulting output is a set of bounding boxes that indicate the location of each identified palm tree (Esri (2022)).

The utility of this pre-trained model has been demonstrated in a variety of scientific research contexts, including crop monitoring, disease management, and conservation biology. For instance, researchers have used the Esri model to analyze the health status of oil palm plantations, where accurate identification of diseased trees can help prevent the further spread of disease and minimize yield losses. Additionally, the model has been used to monitor the health of palm trees in natural habitats, where it can aid in conservation efforts and help prevent the spread of invasive species (Esri (2022)).

3.2.2 Fine tuned Esri's pre-trained model

In this study, a fine-tuning approach was employed to adapt a pre-trained deep-learning model for our area of interest, focusing on palm tree health detection. To this end, 350 new training samples were collected by using the Label Object tool in ArcGIS Pro, allowing us to label and annotate images of palm trees in our study area. Specifically, a new schema for the model was created by drawing bounding boxes around each palm tree using a circle drawing tool as shown in Figure 7 using ArcGIS Pro. These bounding boxes served as the training samples for the model, which was then fine-tuned with the new data to improve its accuracy in identifying palm trees in our area of interest.

Fine-tuning pre-existing deep learning models has emerged as an effective technique for optimizing their performance in specific applications. Although Esri's pre-trained object detection model has shown promise in identifying palm trees, fine-tuning can help enhance the model's accuracy and robustness for specific research questions and study areas. Here, the steps taken to fine-tune the Esri model are described for the investigation of palm tree health in our area of interest.

Initially, additional labeled images of palm trees within the study area were collected and annotated with bounding boxes to indicate the location and health status of each tree. Label Object tool was utilized in ArcGIS Pro to facilitate this annotation process, creating a new schema that was specific to our palm tree detection task.

Subsequently, the transfer learning capabilities of the Esri model were leveraged to fine-tune it with our new training samples. To this end, the final classification layer of the model was replaced

Figure 7: Training samples for Esri's deep learning model.

with a new layer having the same number of outputs but specific to our area of interest. The model was then retrained on the new dataset using stochastic gradient descent, adjusting the learning rate and number of epochs as necessary to optimize its performance.

3.2.3 Mask R-CNN

Mask R-CNN (Mask Region-based Convolutional Neural Network) is a powerful deep learning model that has made significant contributions to the field of object detection and instance segmentation. It was introduced by He et al. in their influential paper titled ”Mask R-CNN” published in 2017 (He et al. (2017)). The Mask R-CNN model builds upon the Faster R-CNN framework (Figure 8), which combines region proposal networks (RPN) and convolutional neural networks (CNN) for object detection. Mask R-CNN (Figure 9) extends this framework by incorporating an additional branch for pixel-level segmentation, enabling the accurate generation of instance-level masks along with object detection.

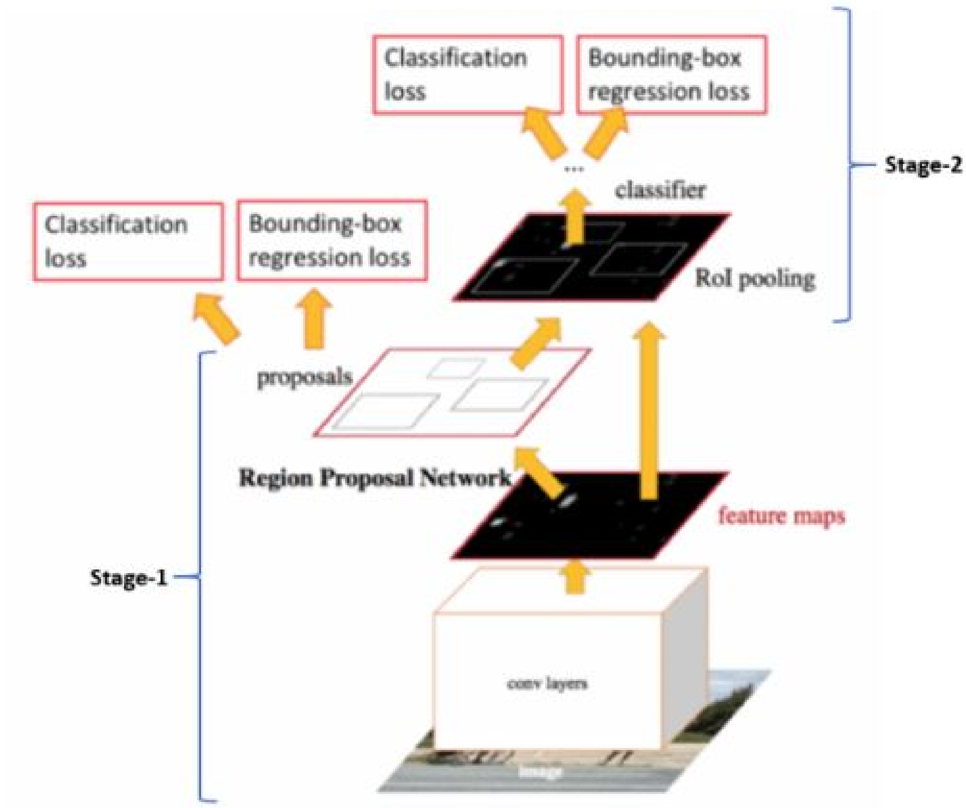

Figure 8: Faster R-CNN framework architecture

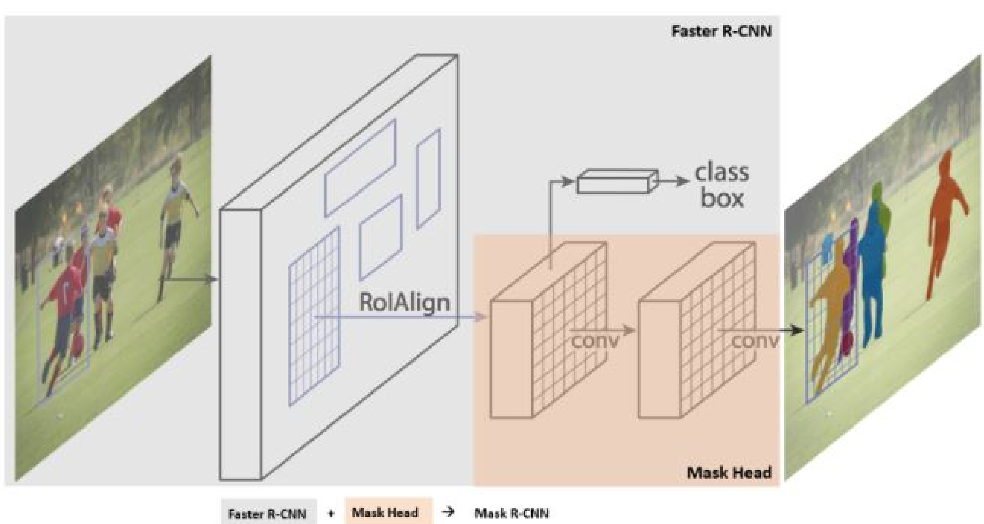

Figure 9: Mask R-CNN framework architecture

The architecture of Mask R-CNN consists of several key components. Firstly, it includes a backbone CNN, such as ResNet or VGG, which extracts high-level features from input images. These features capture important visual information necessary for accurate object detection and segmentation. The second component is the region proposal network (RPN). The RPN generates potential bounding box proposals by sliding a small network over the CNN feature map. These proposals are then refined and classified to identify object instances. The third component involves two sibling subnetworks. The first subnetwork performs object classification and bounding box regression to precisely identify the class labels and refined bounding box coordinates for each proposal. This subnetwork uses a region of interest (RoI) pooling operation to align the features with the proposals. The second subnetwork is responsible for generating a binary mask for each proposal, enabling pixel-level segmentation. This subnetwork applies RoIAlign, which preserves spatial information, to extract region-specific features. The features are then passed through a small fully convolutional network to predict the mask for each object instance (He et al. (2017)).

Mask R-CNN has achieved remarkable performance on various benchmark datasets, including the widely-used COCO (Common Objects in Context) dataset. It has outperformed previous methods in terms of accuracy and has become a fundamental tool in computer vision applications. Since its introduction, several studies have further improved and extended the Mask R-CNN framework. For example, Lin et al. proposed ”Panoptic FPN,” which combines Mask R-CNN with panoptic segmentation for unified scene understanding (Lin et al. (2017)). Moreover, Chen et al. introduced ”Cascade Mask R-CNN,” an improved version that utilizes a cascade of Mask R-CNN models to enhance accuracy and robustness (K. Chen et al. (2019)).

3.3 Data preparation

3.3.1 Evaluate the model effectiveness

The completion of this task relied on orthomosaic imagery obtained from the flight date of 09/22/2022. Initially, training samples were created using the label object tool within ArcGIS Pro. A total of 500 samples were generated, of which 350 were allocated for training and validation, while the remaining 150 were designated for testing purposes across all deep-learning models. It is important to note that for the pre-trained deep learning model, only the testing data were utilized since this model had already undergone training and validation by Esri.

Subsequently, the training and validation datasets were exported to generate image chips. The following Figure 10 shows examples of image patches created in this study. A total of 19,630 image patches, each with a resolution of 448x448 pixels, were extracted for training and validation. Following this, each deep learning model, except for the pre-trained one, underwent training and validation procedures. After training and validation, the trained models were employed to detect palm trees within the orthomosaic image from the same date. Subsequently, a non-maximum suppression technique was applied to eliminate detected objects with an overlap exceeding 0.5. Once the non-maximum suppression was performed, the obtained results were employed to compute accuracy for object detection. Esri's pre-trained model which was fine-tuned with extra training samples of our study area was selected as an optimal model for this study.

Figure 10: Examples of images patches

3.3.2 Evaluate the effect of the GSD

To conduct this study, the flight conducted on the same day as mentioned above, i.e., 09/22/2022 was utilized. The objective of the study was to assess the impact of different GSD values on the performance of the optimal model selected from objective one. To achieve this objective, the orthomosaic was resampled into multiple GSD values: 5 cm, 10 cm, 20 cm, and 40 cm. The bilinear interpolation method was employed for resampling, as it allows for determining new cell values based on the average of four neighboring cells, thereby facilitating data smoothing (Gupta et al. (2013)). It is important to note that only the GSD values were modified, while all other parameters remained constant for each GSD setting. By varying the GSD values, the study aimed to investigate the effect of resolution on the overall performance of the deep learning model.

To accomplish this objective, the same testing dataset used in objective one was employed. The model that exhibited optimal performance, trained using the original GSD value imagery in conjunction with Esri's pre-trained model, was utilized for palm tree detection. Subsequently, the detected results were evaluated to assess the model's performance for specific GSD values, generating an accuracy assessment table for each GSD setting. Consistency was ensured by employing the same set of 150 testing samples, from which performance values were derived. This systematic approach enabled a comprehensive evaluation of the model's performance across different GSD values, specifically for the task of palm tree detection.

3.3.3 Generalization of the optimal model performance on different flight dates

In the previous tasks, the imagery captured on 09/22/2022 served as the foundation for training, testing, and evaluating various models, ultimately identifying the optimal model across different GSD values. The optimal model was not trained on different time frame data. It was employed to see its performance on different flight dates. Based on these findings from the above two findings, the current task focuses on assessing the performance of the optimal model using data collected on different dates. Initially, the model's performance was evaluated using flight data captured a few months later on 12/03/2022, utilizing the same testing data as before. Subsequently, the same evaluation process was repeated with data collected approximately one year prior on 11/07/2021. By analyzing these datasets, changes in the number of detected palm trees over time were observed, allowing for insights into the dynamics of the palm tree population. Furthermore, the model's adaptability to varying temporal conditions was assessed by examining its performance across different flight datasets. This comprehensive analysis sheds light on the model's ability to detect palm trees over time and its flexibility in handling diverse temporal scenarios.

To evaluate the realism of the optimal model, a small portion of the study area was selected as shown in Figure 11. In this designated area, ground truth data regarding the number of palm trees were meticulously digitized through manual digitization, ensuring that no palm tree was overlooked within the defined region. The datasets captured on 09/22/2022 were employed to establish the ground truth. Subsequently, the optimal model was employed to detect palm trees within the same area for three different flights. This step aimed to verify whether the optimal model produced realistic results aligned with the ground truth data. By comparing the model's detections with the manually digitized ground truth, the study assessed the model's accuracy and realism in accurately identifying palm trees within the study area.

Figure 11: Area of campus selected for generalization of the optimal model on different flight dates

3.4 Accuracy assessment

Precision, recall, and F1 score are crucial metrics for accurately assessing the performance of object detection deep learning models. Precision measures the model's ability to correctly identify positive instances, recall evaluates its effectiveness in capturing all actual positives, and the F1 score provides a balanced measure considering both metrics. These metrics play a vital role in evaluating the accuracy of models across various domains, including medical imaging. Widely adopted in object detection evaluations, precision, recall, and F1 score offer a comprehensive understanding of the model's performance and facilitate informed decision-making. By considering different aspects of detection accuracy, these metrics provide reliable assessments of deep learning models for object detection tasks (Powers (2020)).

In object detection, TP (true positive), TN (true negative), FP (false positive), and FN (false negative) are terms used to assess the accuracy of the model's predictions. TP refers to correctly identified positive instances, meaning the model correctly detects and classifies an object as present. FN represents instances where an object is present, but the model fails to detect it. FP signifies instances where the model incorrectly identifies an object that is not actually present, resulting in a false alarm. TN, which stands for correctly identified negative instances, is not commonly used in object detection evaluation as it pertains to correctly identifying the absence of objects. Since the primary focus of object detection is to accurately detect positive instances, TP, FN, and FP are the key metrics for evaluating the model's performance in detecting objects of interest (Goutte & Gaussier (2005)).

Precision measures the model's capability to accurately identify positive instances among all instances it classifies as positive. It is calculated as TP divided by the sum of TP and FP (Goutte & Gaussier (2005)):

$$Precision = \frac{TP}{(TP+FP)}$$

Recall, also referred to as sensitivity, quantifies the model's ability to capture all true positive instances in the dataset. It is calculated as TP divided by the sum of TP and FN (Goutte & Gaussier (2005)):

$$Recall = \frac{TP}{(TP+FN)}$$

While precision emphasizes the accuracy of positive predictions, recall highlights the model's capacity to identify all relevant positive instances. The F1 score integrates both precision and recall into a unified metric, providing a balanced assessment of the model's overall performance. It is the harmonic mean of precision and recall, calculated using the formula (Goutte & Gaussier (2005)):

$$F1score = 2 * \frac{(precision * recall)}{(precision + recall)}$$

4. RESULTS AND DISCUSSION

4.1 Results for evaluation of model performance

4.1.1 Qualitative results

In the imagery captured on 09/22/2022, our evaluation involved three deep learning models: Esri's pre-trained model, Esri's fine-tuned model, and Mask R-CNN. Throughout the evaluation process, we computed the average precision scores for each model, providing a quantitative measure of their performance. Notably, the fine-tuned model achieved the highest average precision score of 0.85, indicating superior performance in detecting and localizing palm trees. Mask R-CNN exhibited a moderate average precision score of 0.75, showcasing its satisfactory performance. On the other hand, the pre-trained model demonstrated the lowest average precision score of 0.51 among the three models, highlighting its comparatively weaker performance. Figure 12 illustrates the Essential Model Details (EMD) files associated with each model to offer a deeper understanding of these models. These files encompass essential information such as the model name, specific parameters, and the corresponding average precision scores.


```
"MinCellSize": {
    "x": 0.05224900000000084,
    "y": 0.05224899999999096,
"MaxCellSize": {
    "x": 0.05224900000000084,
    "y": 0.05224899999999096,
"ModelName": "SingleShotDetector",
  "backend": "pytorch",
  "ModelParameters": {
    "backbone": "resnet34",
    "backend": "pytorch"
  },
  "average_precision_score": {
    "Palm": 0.51069356878
```

a)

```
"MinCellSize": {
    "x": 0.05224900000000084,
    "y": 0.05224899999999096,
"MaxCellSize": {
    "x": 0.05224900000000084,
    "y": 0.05224899999999096,
"LearningRate":
"slice('3.3113e-05', '3.3113e-04', None)",
  "ModelName": "SingleShotDetector",
  "backend": "pytorch",
  "ModelParameters": {
    "backbone": "resnet34",
    "backend": "pytorch"  },
  "average_precision_score": {
    "Palm": 0.88453854438640769
```

b)

```
"MinCellSize": {
    "x": 0.05224900000000084,
    "y": 0.05224899999999096,
"MaxCellSize": {
    "x": 0.05224900000000084,
    "y": 0.05224899999999096,
"LearningRate":
"slice('4.3652e-06', '4.3652e-05', None)",
  "ModelName": "MaskRCNN",
  "backend": "pytorch",
  "average_precision_score": {
    "Palm": 0.7170138890958494
  },
```

c)

Figure 12: Emd files for Esri's pre-trained model(a), Esri's fine-tuned model(b) and mask R-CNN model(c).

In Figure 13, the ground truth and prediction results obtained for all three models are presented. The observations for Esri's pre-trained model reveal instances where the model detected half palm trees as palm trees. Furthermore, the model exhibited false positives by identifying areas with only grass as palm tree locations. Additionally, it demonstrated incorrect classifications, labeling objects on the road as palm trees. Another notable misclassification occurred when the model identified bush areas as palm trees, as illustrated in Figure 13.

On the other hand, the fine-tuned model showed significant improvements. It accurately detected even partial palm trees, showcasing its ability to identify smaller portions of palm trees. This outcome underscores the effectiveness of the fine-tuning process in enhancing the model's performance. Furthermore, in the ground truth section, the model correctly refrained from detecting any palm trees in an area with grass that did not contain actual palm trees.

As for the mask R-CNN result, the model demonstrated commendable accuracy in detecting palm trees. The predicted locations are closely aligned with the ground truth values, as depicted in Figure 13. Notably, the model successfully detected both complete palm trees and partial portions of palm trees. However, limitations were observed when it came to distinguishing between palm trees and areas with grass. In certain instances, the model misclassified grassy areas as palm trees, as evident in Figure 13. These observations highlight the overall effectiveness of Mask R-CNN in accurately detecting palm trees, while also emphasizing the need for further refinement to address misclassifications and false positive results.

In addition to assessing the model's performance, another crucial factor in evaluating its effectiveness is training and validation loss. Figure 14 provides graphs that depict the training and validation loss trends across the processed batches. The blue curve represents the training loss, while the orange curve represents the validation loss. Analyzing the graph for Esri's pre-trained model, a noticeable misalignment between the validation loss and the training loss is observed. This misalignment has contributed to the generation of false positive results, indicating a deviation from the desired performance levels. Notably, the alignment between the two curves begins to improve after approximately 70,000 batches are processed, suggesting a gradual convergence towards optimal performance.

On the other hand, for Esri's fine-tuned model, the graph demonstrates a remarkable alignment

Figure 13: Ground truth and predicted results results for Esri's pre-trained model(a), Esri's fine-tuned model(b) and mask R-CNN model(c).

between the training and validation loss curves throughout the entire training process. This substantial improvement in performance is evident from the graph, signifying the effectiveness of fine-tuning. It is noteworthy that the training and validation loss quickly approach saturation after processing approximately 1,200 batches, indicating significant progress in model convergence and enhanced performance. Turning to the results for Mask R-CNN, the graph showcases a similar trend. Initially, the training loss is relatively high but gradually decreases as the number of batches increases, indicating the model's learning from the training data. As more training data is processed, the validation loss trend aligns more closely with the training loss trend, indicating the model's ability to generalize well to unseen data. After approximately 4,000 batches, both the validation and training loss reach a point of saturation, signifying the convergence of the model.

Figure 14 presents the palm trees detected using the respective models, providing a clear visual representation of the performance differences among all three models. Starting with the pre-trained

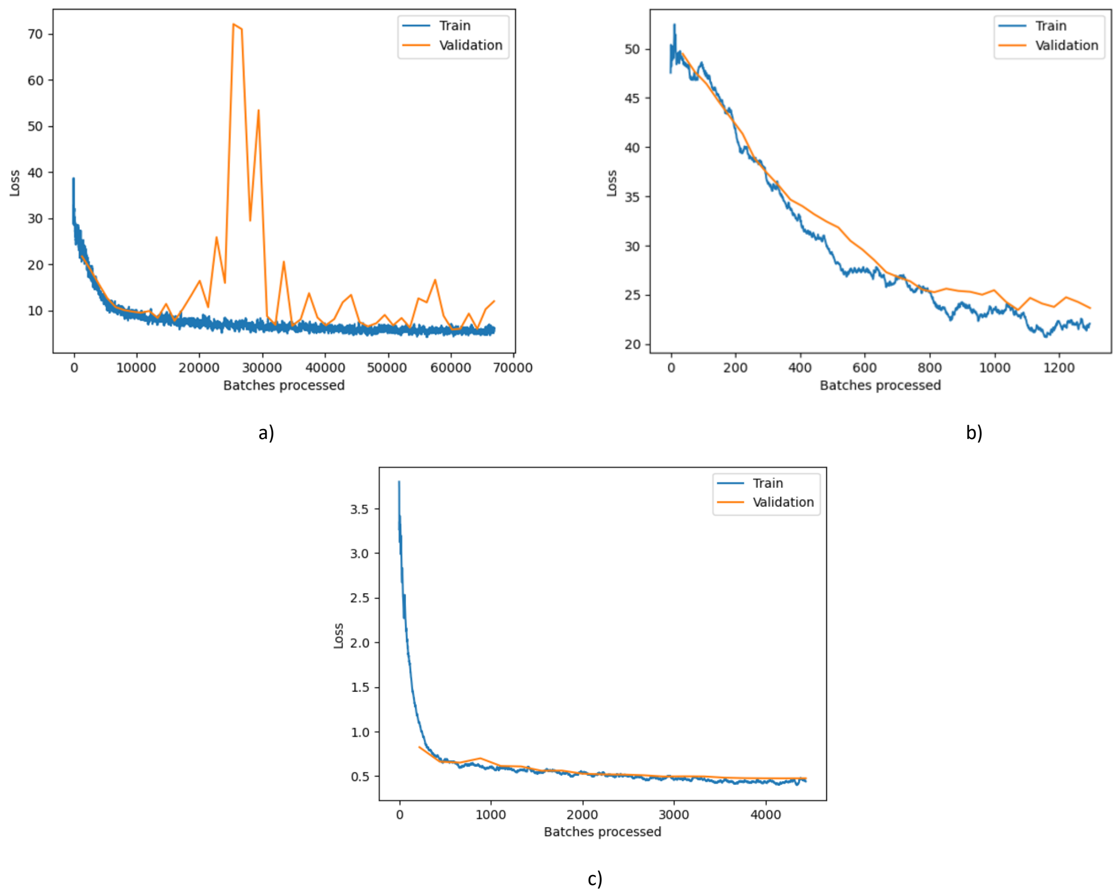

Figure 14: Graphs showing training loss and validation loss for Esri's pre-trained model(a), Esri's fine-tuned model(b), and mask R-CNN model(c).

model, the results indicate poor performance in palm tree detection. The model only identifies a limited number of palm trees, disregarding the majority of them. Additionally, Figure 14 reveals instances where the model mistakenly detects objects resembling palm trees, suggesting accidental detections. As a result, the pre-trained model fails to yield satisfactory outcomes within our specific area of interest. Hence, it becomes imperative to consider making changes or fine-tuning the model to enhance its performance and achieve more accurate results.

In contrast, the fine-tuned model showcases significant improvements in performance, surpassing its previous capabilities. The model demonstrates an impressive ability to accurately detect almost all palm trees. Although there are a few exceptions, where the model misclassifies certain portions of a parking lot as trees, overall, its performance is commendable. Figure 13 provides evidence of the successful detection of palm trees in the given area, reaffirming the effectiveness of the fine-tuning process in enhancing the model's performance and suitability for palm tree detection tasks.

Turning to Mask R-CNN, the results depicted in Figure 15 demonstrate the accurate detection of palm trees within the scene, with most of the palms being detected. The model exhibits a commendable ability to precisely identify palm trees. Notably, it showcases a superior feature in comparison to Esri's pre-trained model, as it provides more detailed and precise segmentation, as highlighted by the autonomous drawing of additional borders around the detected objects. However, one limitation of the model is the occasional occurrence of the same palm tree being detected as multiple distinct trees. This duplicative detection poses challenges in accurately quantifying the total count of palm trees. Figure 16 shows maps of the whole campus area showing detected palm trees from all three object detection models.

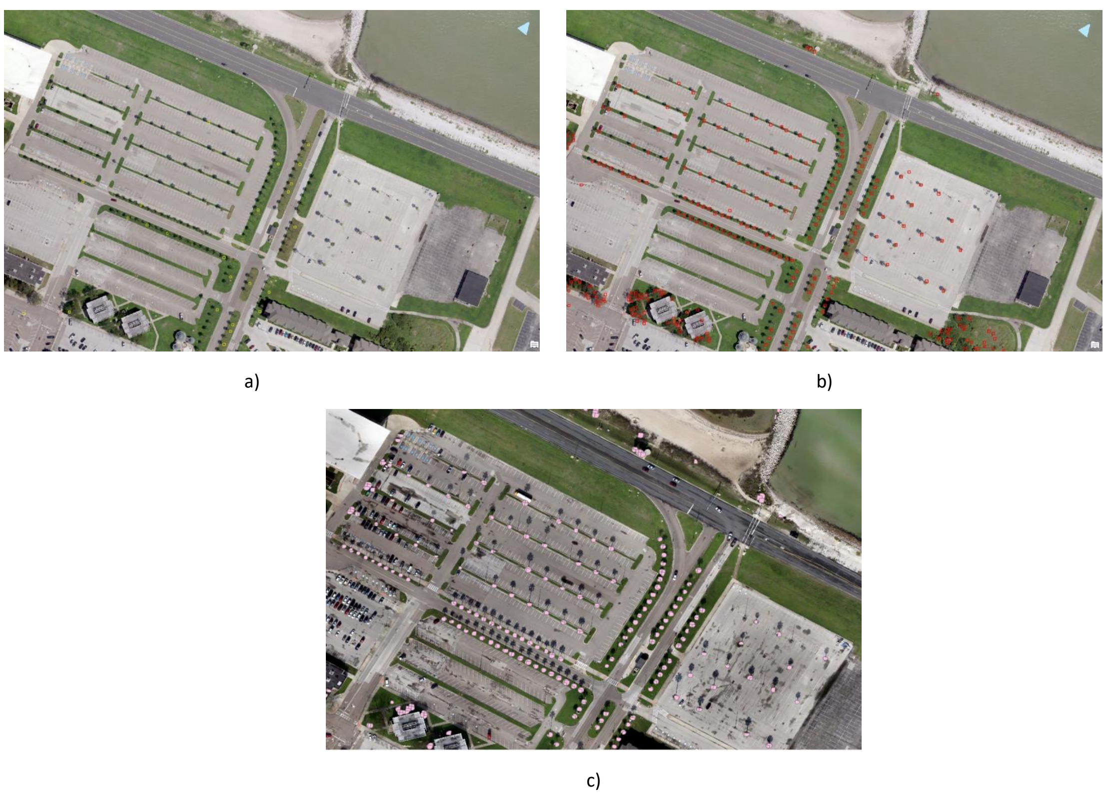

Figure 15: Detected palm trees for Esri's pre-trained model(a), Esri's fine-tuned model(b), and mask R-CNN model(c).

Figure 16: Map of detected palm trees for the whole study area for Esri's pre-trained model(a), Esri's fine-tuned model(b), and mask R-CNN model(c).

4.1.2 Quantitative results

The performance of the three deep-learning models was evaluated using quantitative metrics such as precision, recall, and F-score. Table 6 presents the obtained values for each model. Upon examining the table, it becomes evident that the fine-tuned model has achieved superior performance compared to the other two models. Mask R-CNN also demonstrates moderate results, while the pre-trained model performed poorly in all evaluated metrics. The notable contrast in performance among the models shows the effectiveness of fine-tuning in enhancing the model's capabilities.

Figure 17 presents a graph that is plotted using the values obtained from Table 4. The graph provides a visual representation of the performance comparison among the three models in terms of precision, recall, and F-score. Notably, Esri's fine-tuned model emerges as the top performer

Table 6: Accuracy assessment results for all three models

SN	Model	Precision	Recall	F1-Score
1	Pre-trained model	0.51	0.56	0.53
2	Fine-tuned model	0.85	0.94	0.89
3	Mask R-CNN model	0.75	0.93	0.83

across all aspects, exhibiting superior performance in precision, recall, and F-score. While Mask R-CNN demonstrates a comparable recall to the fine-tuned model, its performance in the other two parameters falls short. On the other hand, the pre-trained model consistently exhibits the lowest performance among the three models, underscoring its limitations in accurately detecting and classifying palm trees.

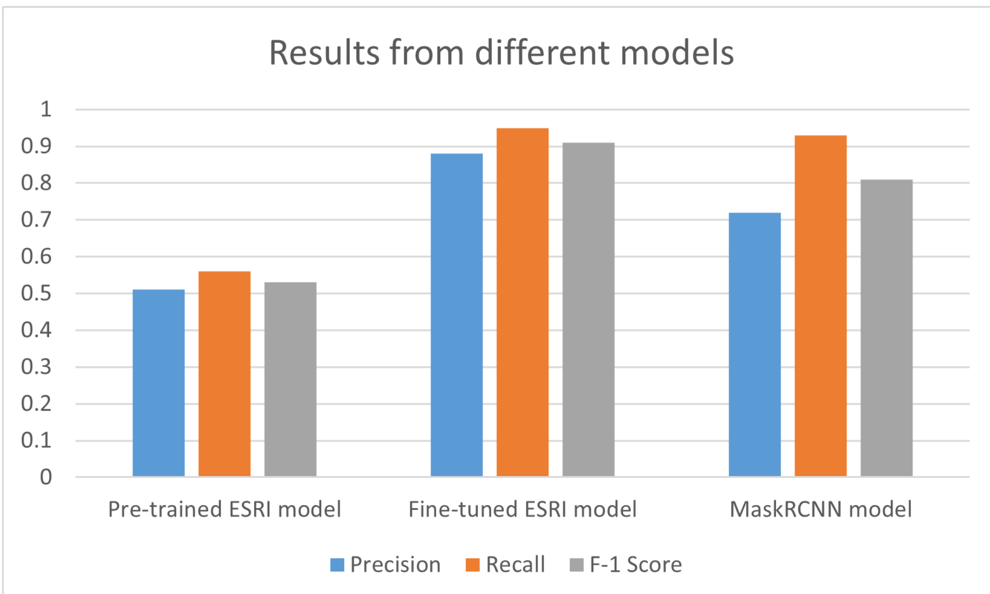

Figure 17: Graph showing value of precision, recall and f1-Score for all three deep learning models

After evaluating the results obtained from the three different methods, it is evident that the most effective and efficient deep learning model for palm tree detection is Esri's model after fine-tuning it with datasets specific to our area of interest and making parameter adjustments during object detection. This conclusion is based on the observation that the pre-trained model, although convenient to use, failed to deliver satisfactory results in our area of interest. Despite its suitability for the trained data, it did not meet our desired outcome. On the other hand, the model that was fine-tuned using training samples from our area of interest, along with adjustments to parameters such as threshold, maximum, and minimum cell size, demonstrated satisfactory performance.

Finally, the Mask R-CNN model proved to be a powerful approach. While it successfully detected most of the palm trees, it occasionally detected a single palm tree in multiple instances. The notable aspect of this model was its ability to draw precise boundaries for each palm tree based on its unique shape and it showed multiple palm trees in a single tree. However, this characteristic may not be suitable for accurate palm tree counting since it could lead to inflated counts. Considering all aspects, Esri's model after fine-tuning with our area-specific data stands out as the superior choice for palm tree detection in terms of overall performance.

After careful analysis, Esri's fine-tuned model has been selected as the optimal choice for detecting and counting palm trees on the island campus. In contrast, Esri's pre-trained model exhibited the lowest accuracy and weakest performance among the three models evaluated. However, after undergoing the fine-tuning process, the model achieved a significantly improved accuracy, surpassing the performance of the other models. Additionally, the fine-tuned model offers user-friendly features compared to Mask R-CNN, making it easier to use. It also demonstrates faster processing speed in comparison to Mask R-CNN. Even users with limited knowledge of object detection models can employ this model after conducting preliminary research.

Thus, the fine-tuned model is considered easy to use, fast, and efficient, providing the highest accuracy among the three models. Unlike Mask R-CNN, the fine-tuned model exhibits a lower frequency of miscounting a single palm as multiple instances. While it is important to note that the trained deep learning model is not perfect like a human, it remains the optimal model among the models tested in this study. Figure 16(b) illustrates a map of our campus with the detected palm trees, revealing a total of 1414 palm trees identified. Given these findings, we will continue utilizing Esri's fine-tuned model in further studies to test our results on GSD.

4.2 Results for optimal model performance on increasing GSD

4.2.1 Qualitative results

In the results section, we utilized Esri's fine-tuned model identified as the optimal choice in the previous analyses to evaluate the impact of various GSD values on its performance. Initially, the orthomosaic imagery had an original GSD value of 1.6 cm. To investigate different GSD values, we resampled our study area to encompass four additional GSD values. Specifically, the model was tested on GSD values of 5 cm, 10 cm, 20 cm, and 40 cm. Figure 18 presents the outcomes of the detected palm trees using the aforementioned fine-tuned model, using the original GSD value orthomosaic imagery. These results highlight the model's effectiveness in detecting palm trees across varying GSD values.

Figure 18(a) shows results obtained from the original GSD value, this result is the same as above. This was taken to visualize and the comparison with other GSD values. Figure 18(b) shows the detected palm trees for the orthomosaic with a 5 cm GSD value. The results indicate that the majority of palm trees were successfully detected, resembling the outcomes obtained with the original GSD value of 1.6 cm. However, there were a few areas where palm trees were not detected, and in some instances, non-palm trees were misclassified as palm trees. This discrepancy may arise because the model was trained on high-resolution data and is now being validated on lower-resolution data. Despite these exceptions, the model exhibited favorable performance at this GSD value overall. Therefore, we can conclude that the model performs well, even when the GSD value is increased by approximately five times compared to the original resolution.

In Figure 18(c), the results of detected palm trees on a 10 cm GSD value are presented. It is evident that while this model did not perform as well as it did on the original GSD value and the 5 cm GSD value, it still managed to detect a significant number of palm trees. However, it is important to note that the model also detected a considerable number of objects as trees that were not present in the ground truth data. This indicates that the model might have a tendency to produce false positive detections or exhibit challenges in accurately distinguishing between palm trees and other objects for this GSD value.

We applied the same model to evaluate the object detection performance on imagery with a 20 cm GSD value. However, the results demonstrated that the model was not effective for higher GSD values. Figure 18(d) visually presents the detected palm trees, revealing a minimal number of objects correctly identified as palm trees. Moreover, a majority of the objects classified as palm trees were found to be false positives. The reduced resolution in the imagery posed a significant challenge in accurately distinguishing palm trees from other objects. This outcome aligns with the expected behavior, as the model was trained using specific resolution training samples, making it difficult to adapt to different resolutions during feature detection.

Seeking further confirmation, we extended our investigation to an even lower resolution by testing the model on a 40 cm GSD value orthomosaic. Regrettably, the results shown in Figure 18(e) highlighted that the model detected very few true palm trees, with a notable presence of false positives. These consistent findings suggest that as the resolution of the imagery decreases, the model's performance gradually deteriorates. The limitations of the model became evident when applied to imagery with lower resolutions. The model's reduced accuracy in identifying palm trees at lower resolutions reinforces the importance of utilizing appropriate training data and considering the impact of varying resolutions on the model's performance in object detection tasks.

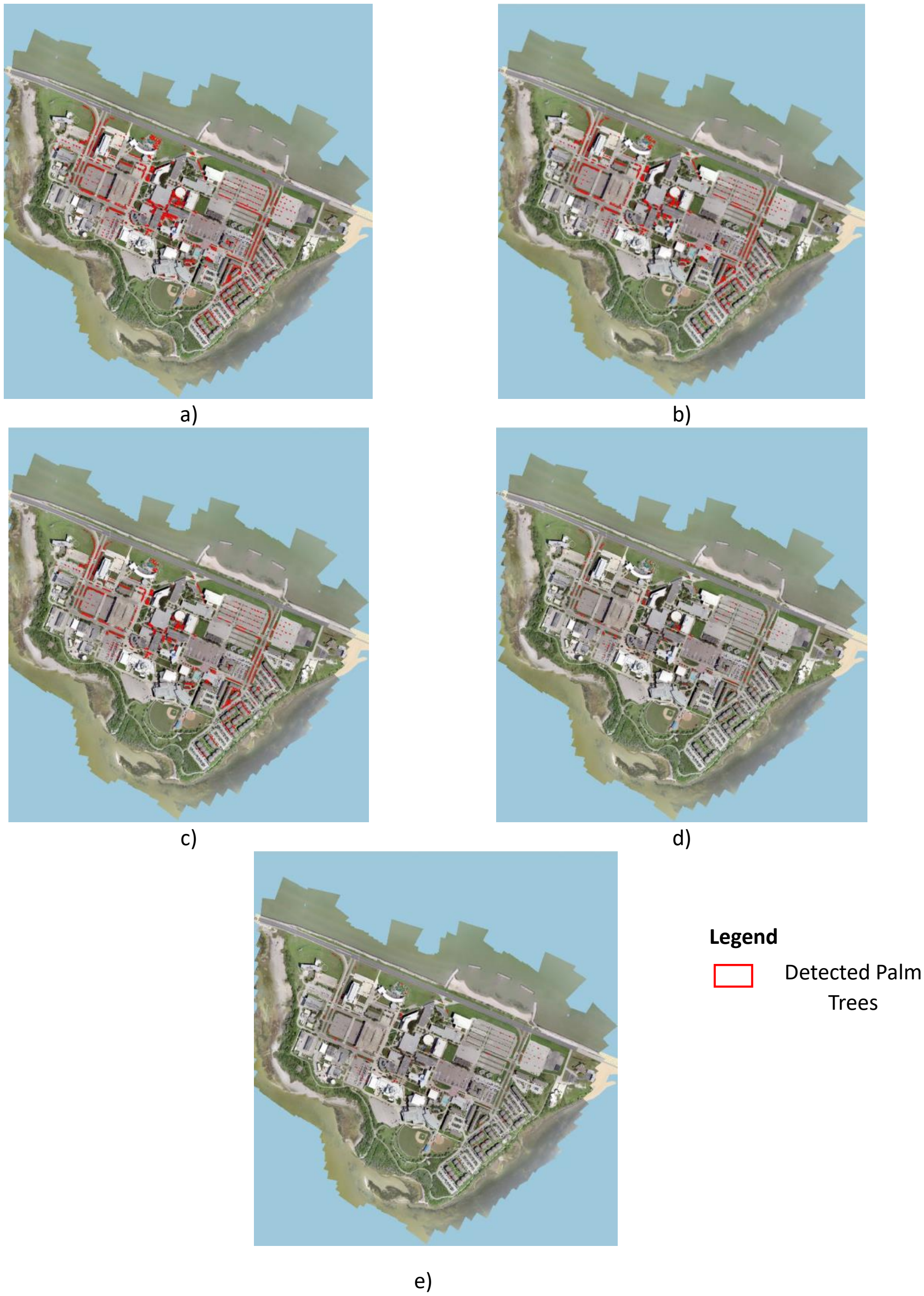

Figure 18: Detected palm trees for different GSD values using the optimal model: a) 1.6cm GSD; b)5cm GSD; c)10cm GSD; d)20cm GSD; e)40cm GSD

4.2.2 Quantitative results

Table 7 provides the precision, recall, and F-Score values obtained for each GSD value. Analysis of the table reveals that the model exhibits good performance up to a GSD decrease of approximately five times, and performs well up to a decrease of around 10 times. However, beyond these thresholds, the model's performance noticeably declines. In an effort to further assess its effectiveness, the model was tested on even higher GSD values. Unfortunately, the results indicated that the model is not effective for those GSD values. The findings emphasize the importance of considering the GSD value and its impact on the model's performance in object detection tasks.

Table 7: Accuracy assessment results for different GSD values

SN	GSD(cm)	Precision	Recall	F1-Score
1	Original(1.6)	0.88	0.95	0.91
2	5	0.91	0.88	0.89
3	10	0.91	0.85	0.87
4	20	0.42	0.12	0.18
5	40	0.11	0.03	0.05

Figure 19 displays a graph generated using the values obtained from Table 5, showcasing a visual representation of the performance comparison across various GSD values, including 1.6 cm, 5 cm, 10 cm, 20 cm, and 40 cm. Notably, the model exhibits satisfactory performance up to 10 cm GSD. However, its performance drastically declines at 20 cm and 40 cm GSD values. In fact, it can be concluded that the model fails to effectively operate at these higher GSD values.

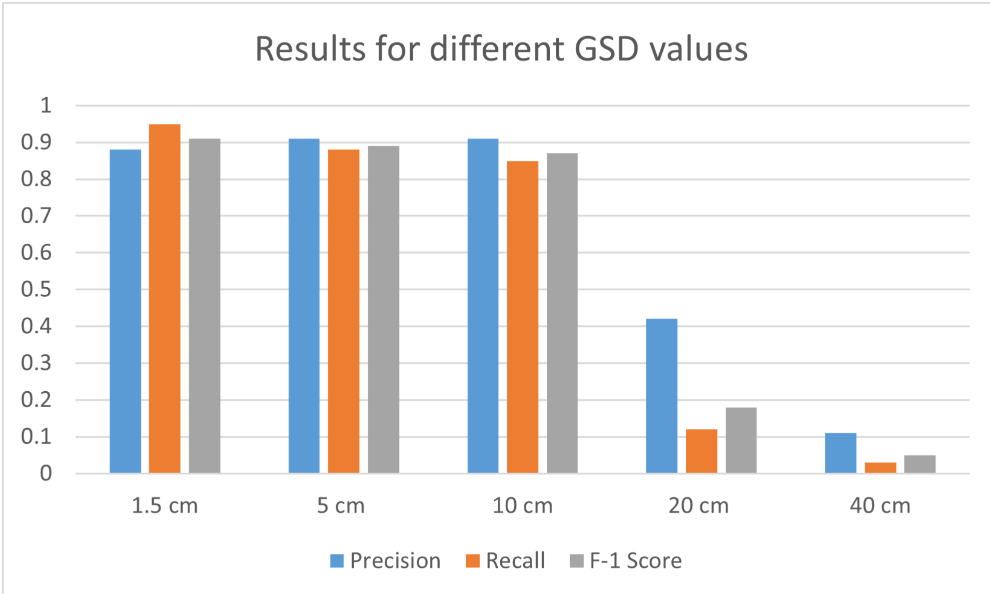

Figure 19: Graphs showing value of precision, recall and f1-Score for all GSD values

4.3 Results for optimal model performance on different UAS flights at same altitude/GSD

4.3.1 Qualitative results

In this phase of the study, we employed Esri's fine-tuned model, which emerged as the optimal model, as demonstrated in section 4.1. Table 8 presents the number of detected palm trees across three flight dates, offering insights into the model's performance. Furthermore, Figure 19 provides a graphical representation of the detected palm trees over these flight dates, facilitating visual comprehension of the results.

Table 8: Number of palm trees detected in datasets collected in three different flight dates

Date	**Number of detected palm trees**
12/3/2022	1256
9/11/2022	1414
7/11/2021	1589

The findings reveal a difference of 168 fewer detected trees in the datasets collected after the flight used for training the optimal model. Conversely, for the datasets gathered one year prior to the training datasets, the model detected an additional 175 trees. These results highlight the impact of training data on the model's performance and its ability to accurately detect palm trees in varying temporal contexts.

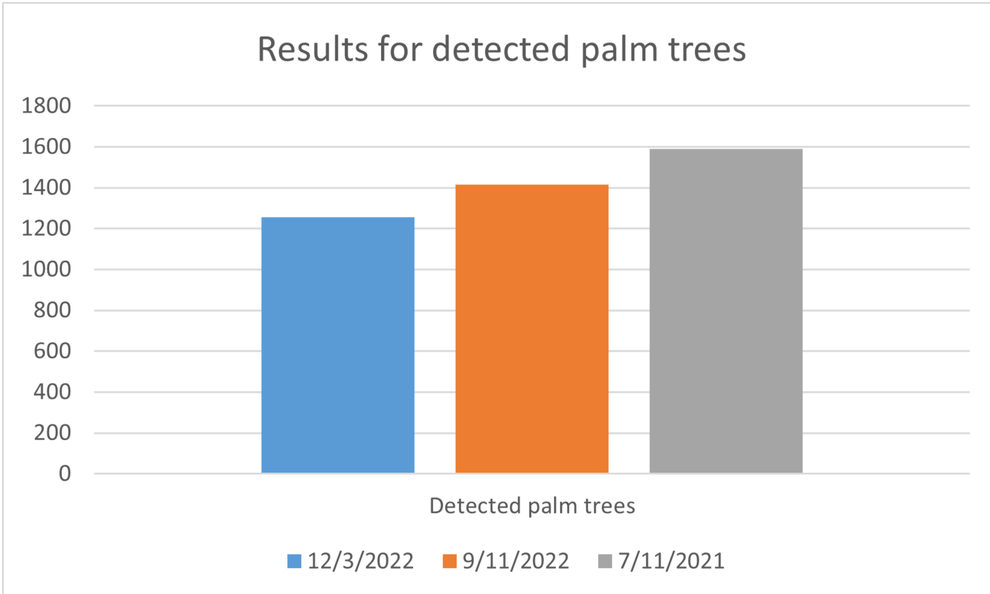

Figure 20: Graphs showing the number of palm trees detected on three different flight dates.

4.3.2 Quantitative results

Table 9 presents the precision, recall, and F-Score values obtained for all three flight dates. The results demonstrate comparable accuracy parameters across the different dates. However, when considering precision, recall, and F1-score, it becomes evident that the model tested on datasets collected a few months after the training datasets exhibit superior performance compared to the model used to detect objects in the datasets collected one year prior.

Table 9: Accuracy assessment results for different flight dates.

SN	Flight date	Precision	Recall	F1-Score
1	12/03/2022	0.89	0.93	0.90
2	9/11/2022	0.88	0.95	0.91
3	7/11/2021	0.87	0.92	0.89

To provide a visual representation of these results, Figure 21 showcases a graphical depiction of the performance obtained for the three different flight dates. The graph allows for a clearer understanding of the variations in model performance across the specified timeframes.

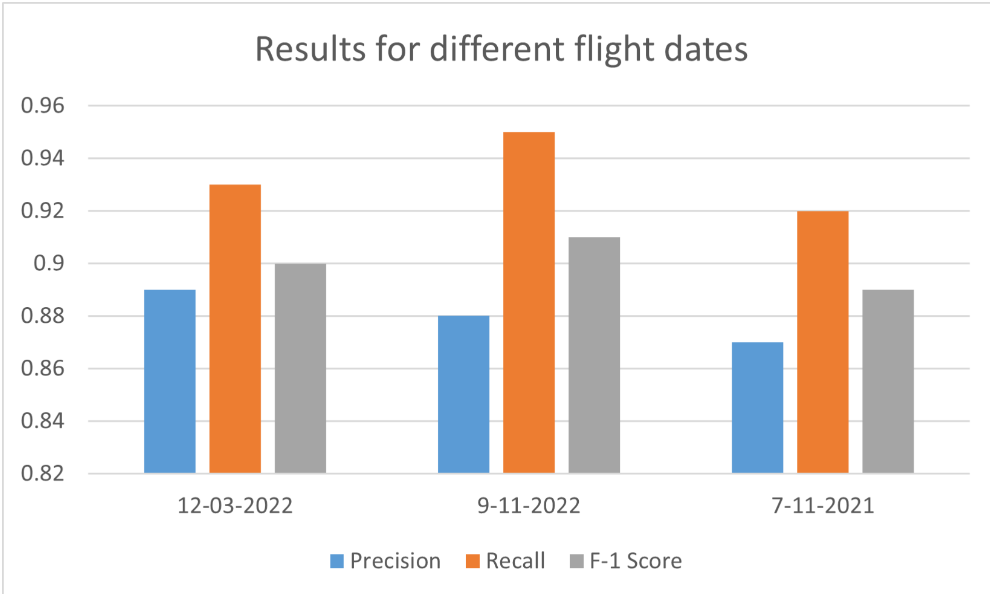

Figure 21: Graphs showing the value of precision, recall, and f1-Score for three different flight dates.

4.4 Verification of optimal model performance with the ground truth data

To verify the performance of the optimal model, ground truth data were utilized. A small portion of the study area designated for verification was carefully digitized. The optimal model was employed to detect palm trees within the same area. The findings are presented in Table 10, ground truth data, and the detected palm trees using the optimal model. It is important to acknowledge that no model is flawless, and slight deviations are expected. In this context, the detection of 2 extra palm trees is within a reasonable margin. Moreover, the results demonstrate a realistic trend as no additional palm trees have been planted after 2021. Furthermore, the model's detections align with the reduced number of palm trees observed in the imagery captured on later dates. Overall, these findings provide confidence in the model's performance and its ability to effectively detect and track changes in the palm tree population over time. Figure 22 visually demonstrates significant observations regarding the palm tree population within the study area. Within the yellow box in the imagery captured on 11/07/2021, a palm tree is clearly visible. However, in the subsequent imagery shown within the red box, which was taken at later dates, no palm tree exists in the same location. Furthermore, the yellow box highlights the presence of palm trees in both the imagery

Table 10: Ground truth data, predicted palm trees, and percentage of correctly detected palm trees for different flight dates.

SN	Flight date	Ground Truth Counted	Predicted Palms	% Correct
1	12/03/2022	186	184	98.92
2	9/11/2022	191	190	99.47
3	7/11/2021	194	191	98.45

from 11/07/2021 and 09/11/2022. However, in the imagery taken on 12/03/2022, depicted by the two red boxes, there are no palm trees at those specific locations. These findings suggest that certain palm trees were removed from the study area, potentially due to factors such as natural decay or intentional replacement with alternative features.

Figure 22: Change in the number of detected palm trees for three different flight dates in the small area designated for verification of the detected results with the ground truth.

4.5 Discussion

In this study, three deep learning models were tested: Esri's pre-trained model for palm tree health detection, Esri's model fine-tuned with our study area data, and the Mask-RCNN object detection model. The performance of these models was evaluated based on precision, recall, and F1-score. After carefully evaluating these models, we found that Esri's pre-trained model, when fine-tuned by adjusting the parameters and incorporating training samples from our specific area of interest, outperformed the other model. It exhibited superior results in terms of accuracy and efficiency. One of the key challenges encountered during the testing phase was determining the appropriate number of training samples for the model. Initially, testing with over seven hundred training samples led to overfitting, while using only fifty samples resulted in underfitting. To overcome this, extensive training and testing were conducted to identify the optimal number of training samples that would yield the best results for the model.

To assess the performance of the fine-tuned model across different GSD values, we initially worked with an orthomosaic that had an original GSD value of 1.6 cm. To explore the impact of varying GSD values, we resampled a portion of our study area, starting with a GSD value of 10 cm. This allowed us to evaluate the performance of the fine-tuned model, which had been identified as the best model among the three tested in this study. One of the primary challenges we faced during this task involved resampling the orthomosaic into various GSD values. This process proved to be time-consuming and resource-intensive. In future endeavors, it may be more efficient to consider utilizing different flight data instead of resampling the imagery. This approach would eliminate the need for resampling and potentially streamline the workflow, saving valuable time and resources.

Moreover, the performance of the optimal model was evaluated on datasets from different flight dates, examining the changes in the number of palm trees (i.e., removal or plantation) within the given time frame. When the optimal model was applied to the flight datasets captured three months after the training datasets, it detected fewer palm trees. There are two possible explanations for this observation: either there was significant plantation activity between the flights, resulting in new palm trees that the model failed to detect, or the model's performance was optimal for this particular dataset.

In summary, this study was subject to several limitations. First, training the deep learning model using high-resolution datasets proved to be time-consuming. Additionally, the model's limited flexibility in terms of modifications posed a constraint compared to more customizable models. Moreover, the process of resampling the orthomosaic into higher GSD values was found to be time-consuming. These limitations should be taken into consideration when interpreting the findings and planning future research in this area.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this project, a UAS was utilized to collect high-resolution imagery, while GNSS technology was employed to acquire accurate checkpoints. The primary objective of this study was to evaluate the performance of deep-learning models in detecting palm trees. To achieve this, the imagery of the Ward Island campus was captured and processed using the SfM technique. The resulting output, known as the orthomosaic, obtained from the SfM photogrammetric software, was then utilized for palm tree detection using deep learning methods.

Flights were conducted using the VTOL fixed-wing Wingtra One UAS to capture the necessary imagery. As mentioned earlier, the collected images were processed using SfM photogrammetric software, specifically the metashape software. This software played a crucial role in the processing and analysis of the imagery data, enabling the generation of accurate and detailed outputs for further investigation and evaluation.

The orthomosaic imagery obtained from the previous stage of this study served as the basis for creating training samples. The process involved utilizing a create label tool, where the boundary of each palm tree was delineated by drawing a circle around it. These training samples were strategically collected in a distributed manner to ensure the representation of all palm trees within the study area. Various types of training samples were carefully gathered, encompassing the full range of palm tree characteristics. Subsequently, the training samples underwent a training process, and the resulting models were thoroughly analyzed to assess their performance and effectiveness.

Upon analysis, it was determined that the fine-tuned model outperformed the other two models in several key aspects. This decision was based on various factors, including precision, recall, F1-score, and compatibility with new users. The three models tested in this study were Esri's pre-trained model for palm tree health detection, the fine-tuned Esri model using training data specific to our area of interest, and the customizable Mask R-CNN model.

Among the three models, the pre-trained model by Esri exhibited the poorest performance, while the Mask R-CNN model showed better results than the pre-trained model. However, the fine-tuned Esri model emerged as the top-performing model among the three. This superiority could be attributed to the fact that the Esri pre-trained model was initially trained on a broader range of palm tree data, while the fine-tuning process integrated training samples from our specific area of interest. Additionally, certain parameters were adjusted, such as the cell size and threshold value, to optimize the model's performance for our particular dataset. The fine-tuned Esri model showcased its capability to accurately detect palm trees within our study area, surpassing the performance of both the pre-trained Esri model and the Mask R-CNN model. These findings highlight the importance of fine-tuning deep learning models using localized training data, as it enables the model to adapt and excel in specific geographical contexts.

The optimal model obtained in the previous stage of this study was utilized to evaluate the impact of various GSD values, namely 5cm, 10cm, 20cm, and 40cm. The findings demonstrate that the model performs satisfactorily up to a resolution of 10cm. However, its performance deteriorates significantly at higher GSD values. This result highlights the importance of considering the resolution of imagery when utilizing the model. If the intention is to apply the model to lower-resolution imagery, it is necessary to retrain the model using training samples specifically obtained from lower-resolution imagery. This adjustment ensures that the model is optimized and aligned with the characteristics of the imagery it will be applied to.

Moreover, the results lead to the conclusion that deep learning models demonstrate superior performance when applied to testing samples possessing comparable resolution values. Models trained and evaluated on imagery with similar resolutions consistently produce more accurate and reliable outcomes. Hence, it is crucial to take into account the compatibility between the resolution of the training data and that of the testing data to ensure optimal model performance. This consideration greatly influences the model's ability to generalize effectively and make accurate predictions on new data. By aligning the resolutions of the training and testing datasets, the model can better adapt to the specific characteristics and features present in the target imagery, ultimately enhancing its overall performance.

The optimal model obtained from this study was evaluated using different temporal datasets.

As the resolution of all three datasets was approximately the same, the performance accuracy of the model was found to be comparable. Notably, the model exhibited similar performance values for the datasets collected three months apart. However, it displayed slightly lower accuracy values for the datasets collected more than a year apart. This discrepancy could be attributed to various factors, including the dissimilarity between the training data and the new datasets, differences in lighting conditions during image capture, variations in weather conditions, and distinct flight settings. Despite these considerations, the results indicate that the optimal model performs well on datasets with the same resolution, even when collected at different periods.

In conclusion, the utilization of UAS-derived orthomosaic imagery proved to be valuable for the detection of palm trees using deep learning models. The optimal model developed in this study exhibited promising performance in detecting sabal palm trees on the TAMUCC campus. This research successfully demonstrated the application of UAS-derived orthomosaic imagery in evaluating deep learning models, identifying the optimal model, and employing it to detect and map palm trees within our study area.

5.2 Future work

In future studies, it is recommended to broaden the scope of data sources by incorporating elevation data, such as a DSM, canopy height, etc. in addition to the orthomosaic obtained from processing UAS imagery. The inclusion of this information might improve the performance and generalizability, as it provides valuable information about the vertical distribution and structural characteristics of palm trees.

Expanding the range of tested models is also advisable. While this study focused on evaluating three specific models, there is room for improvement by considering a more diverse set of data. This includes not only using images from the orthomosaic but also incorporating training samples obtained by capturing palm trees from different angles. By diversifying the dataset, the model can be exposed to a wider range of variations and complexities, leading to improved performance and robustness.

Furthermore, it is important to explore and evaluate other deep learning models that offer greater customization options. While the GIS-based deep learning model used in this study provided satisfactory results, there may be alternative models that can offer enhanced performance by leveraging additional features and capabilities. Exploring and comparing different models can help identify the most suitable and effective approach for palm tree detection.

In this study, to test the results on GSD, the original orthomosaic imagery was resampled to get lower-resolution datasets. So, in future research, it is recommended to collect data from multiple flights and process them separately instead of relying on resampling the original orthomosaic imagery to obtain lower-resolution datasets. This approach would allow for the acquisition of diverse datasets with varying resolutions, better representing real-world scenarios. By utilizing separate flights for different resolution levels, the data collection process can be tailored to capture imagery specifically suited to each resolution, avoiding the potential loss of information or artifacts introduced by resampling.

For future research, it is recommended to consider various temporal settings when evaluating the model's performance. Accounting for factors such as lighting conditions, weather conditions, wind speed, and direction, consistent flight settings, the machine used for imagery processing, and maintaining the same settings during image processing might lead to improved model performance. By considering these factors, the model can be better aligned with the specific conditions encountered during different flights, thereby enhancing its overall effectiveness.

In conclusion, future research should consider incorporating elevation data such as canopy height, expanding the range of tested models, exploring alternative deep-learning models, taking different flight height data, and training the optimal model using a diverse set of testing data including the images taken from different camera angles on UAS, doing model performance test not only in sabal palms but also in other species of palm and considering weather conditions and consistent image processing settings and devices. By doing so, the accuracy, adaptability, generalizability, and overall performance of palm tree detection can be further enhanced, providing valuable insights for the effective management and monitoring of palm tree populations.

REFERENCES

Abd-Elrahman, A., Britt, K., & Liu, T. (2021). Deep learning classification of high-resolution drone images using the arcgis pro software: For374/fr444, 10/2021. *EDIS*, *2021*(5).

Agisoft. (n.d.). [web page]. St. Petersburg, Russia. Retrieved from `https://www.agisoft.com/features/professional-edition/`

Awasthi, B., Karki, S., Regmi, P., Dhami, D. S., Thapa, S., & Panday, U. S. (2020). Analyzing the effect of distribution pattern and number of gcps on overall accuracy of uav photogrammetric results. In *International conference on unmanned aerial system in geomatics* (pp. 339–354).

Bäumker, M., Przybilla, H., & Zurhorst, A. (2013). Enhancements in uav flight control and sensor orientation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(1/W2), 33–38.

Berra, E., & Peppa, M. (2020). Advances and challenges of uav sfm mvs photogrammetry and remote sensing: Short review. In *2020 ieee latin american grss & isprs remote sensing conference (lagirs)* (pp. 533–538).

Cathey, H. M. (1990). *Usda plant hardiness zone map* (Vol. 1475). US Department of Agriculture Washington, DC.

Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., . . . others (2019). Hybrid task cascade for instance segmentation. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 4974–4983).

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Colomina, I., & de la Tecnologia, P. M. (2008). Towards a new paradigm for high-resolution low-cost photogrammetryand remote sensing. In *Proceedings of the isprs xxi congress, beijing, china* (pp. 3–11).

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., . . . Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3213–3223).

Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2011). Pedestrian detection: An evaluation of the state of the art. *IEEE transactions on pattern analysis and machine intelligence*, *34*(4), 743–761.

Esri. (2022). [web page]. Retrieved from `https://learn.arcgis.com/en/projects/use-deep-learning-to-assess-palm-tree-health/`

Everitt, J. H., Drawe, D. L., Lonard, R. I., & Lonard, R. (2002). *Trees, shrubs & cacti of south texas*. Texas Tech University Press.

Famiglietti, N. A., Cecere, G., Grasso, C., Memmolo, A., & Vicari, A. (2021). A test on the potential of a low cost unmanned aerial vehicle rtk/ppk solution for precision positioning. *Sensors*, *21*(11), 3882.

Gehrmann, S., Dernoncourt, F., Li, Y., Carlson, E. T., Wu, J. T., Welt, J., ... others (2018). Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PloS one*, *13*(2), e0192360.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *Advances in information retrieval: 27th european conference on ir research, ecir 2005, santiago de compostela, spain, march 21-23, 2005. proceedings 27* (pp. 345–359).

Guo, Q., Su, Y., Hu, T., Guan, H., Jin, S., Zhang, J., ... others (2020). Lidar boosts 3d ecological observations and modelings: A review and perspective. *IEEE Geoscience and Remote Sensing Magazine*, *9*(1), 232–257.

Gupta, S. G., Ghonge, D. M., & Jawandhiya, P. M. (2013). Review of unmanned aircraft system (uas). *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume*, *2*.

Hajjaji, Y., Boulila, W., & Farah, I. R. (2022). Leveraging artificial intelligence techniques for smart palm tree detection: A decade systematic review. *Procedia Computer Science*, *207*, 2823–2832.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the ieee international conference on computer vision* (pp. 2961–2969).

Hoeser, T., & Kuenzer, C. (2020). Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. *Remote Sensing*, *12*(10), 1667.

Hussain, A., Rizwan-ul Haq, M., Al-Jabr, A. M., & Al-Ayied, H. Y. (2013). Managing invasive populations of red palm weevil: A worldwide perspective. *J. Food Agric. Environ*, *11*(2), 456–463.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). *Geoai: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond* (Vol. 34) (No. 4). Taylor & Francis.

Jeong, K., Kwon, J., Do, S. L., Lee, D., & Kim, S. (2022). A synthetic review of uas-based facility condition monitoring. *Drones*, *6*(12), 420.

Kang, Y., Chen, C., Cheng, F., & Zhang, J. (2022). A novel remote sensing approach to recognize and monitor red palm weevil in date palm trees. *arXiv preprint arXiv:2203.14476*.

Khaled, A. Y., Abd Aziz, S., Bejo, S. K., Mat Nawi, N., & Abu Seman, I. (2022). Artificial intelligence for spectral classification to identify the basal stem rot disease in oil palm using dielectric spectroscopy measurements. *Tropical Plant Pathology*, *47*(1), 140–151.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Kuikel, S., Upadhyay, B., Aryal, D., Bista, S., Awasthi, B., & Shrestha, S. (2021). Individual banana tree crown delineation using unmanned aerial vehicle (uav) images. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Li, W., Dong, R., Fu, H., & Yu, L. (2018). Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. *Remote Sensing*, *11*(1), 11.

Li, W., Fu, H., Yu, L., & Cracknell, A. (2016). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*, *9*(1), 22.

Li, W., & Hsu, C.-Y. (2022). Geoai for large-scale image analysis and machine vision: Recent progress of artificial intelligence in geography. *ISPRS International Journal of Geo-Information*, *11*(7), 385.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2117–2125).

Liu, X., Ghazali, K. H., Han, F., & Mohamed, I. I. (2021). Automatic detection of oil palm tree from uav images based on the deep learning method. *Applied Artificial Intelligence*, *35*(1), 13–24.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3431–3440).

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, *152*, 166–177.

Malek, S., Bazi, Y., Alajlan, N., AlHichri, H., & Melgani, F. (2014). Efficient framework for palm tree detection in uav images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(12), 4692–4703.

Mejias, L., & Fitzgerald, D. (2013). A multi-layered approach for site detection in uas emergency landing scenarios using geometry-based image segmentation. In *2013 international conference on unmanned aircraft systems (icuas)* (pp. 366–372).

Mubin, N. A., Nadarajoo, E., Shafri, H. Z. M., & Hamedianfar, A. (2019). Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. *International Journal of Remote Sensing*, *40*(19), 7500–7515.

Pádua, L., Vanko, J., Hruška, J., Adão, T., Sousa, J. J., Peres, E., & Morais, R. (2017). Uas, sensors, and data processing in agroforestry: A review towards practical applications. *International journal of remote sensing*, *38*(8-10), 2349–2391.

Pashaei, M., Kamangir, H., Starek, M. J., & Tissot, P. (2020). Review and evaluation of deep learning architectures for efficient land cover mapping with uas hyper-spatial imagery: A case study over a wetland. *Remote Sensing*, *12*(6), 959.

Potter, L. (2015). *Managing oil palm landscapes: A seven-country survey of the modern palm oil industry in southeast asia, latin america and west africa* (Vol. 122). CIFOR.

Powers, D. M. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Rahmat, R. F., Azzakirot, Y., & Lini, T. Z. (2019). Tree identification to calculate the amount of palm trees using haar-cascade classifier algorithm. In *2019 3rd international conference on electrical, telecommunication and computer engineering (elticom)* (pp. 36–39).

Ramp, P. F. (1989). *Natural history of sabal minor: demography, population genetics and reproductive ecology*. Tulane University, Graduate Program in Biomedical Sciences.

Reddy, A. L. N., & Banerjee, P. (1990). Algorithm-based fault detection for signal processing applications. *IEEE Transactions on Computers*, *39*(10), 1304–1308.

Regmi, P., STAREK, M. J., & BERRYHILL, J. (2023). Overview of the uncrewed aircraft system (uas) campus survey project at texas a&m university-corpus christi.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

Scott, G. J., England, M. R., Starms, W. A., Marcum, R. A., & Davis, C. H. (2017). Training deep convolutional neural networks for land–cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters*, *14*(4), 549–553.

Septentrio. (n.d.). [web page]. Retrieved from `https://www.septentrio.com/en/products/gnss-receivers/smart-antennas/altus-nr3`

Sharma, L., & Lohan, N. (2019). Performance analysis of moving object detection using bgs techniques in visual surveillance. *International Journal of Spatio-Temporal Data Science*, *1*(1), 22–53.

Singh, H. (2019). *Practical machine learning and image processing: For facial recognition, object detection, and pattern recognition using python*. Springer.

Slocum, R. K., & Parrish, C. E. (2017). Simulated imagery rendering workflow for uas-based photogrammetric 3d reconstruction accuracy assessments. *Remote Sensing*, *9*(4), 396.

Stanton, C., Starek, M. J., Elliott, N., Brewer, M., Maeda, M. M., & Chu, T. (2017). Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. *Journal of Applied Remote Sensing*, *11*(2), 026035–026035.

Starek, M. J., Gingras, M., & Jeffress, G. (2019). Application of unmanned aircraft systems for coastal mapping and resiliency. *Sustainable Development Goals Connectivity Dilemma (Open Access): Land and Geospatial Information for Urban and Rural Resilience*, 109–125.

Starek, M. J., & Wilkinson, B. W. (2022). Aerial surveying technology. in surveying and geomatics engineering: Principles, technologies, and applications. In (p. 341 - 392). Reston, VA: American Society of Civil Engineers.

Strecha, C., Küng, O., & Fua, P. (2012). *Automatic mapping from ultra-light uav imagery* (Tech. Rep.).

Sturdivant, E. J., Lentz, E. E., Thieler, E. R., Farris, A. S., Weber, K. M., Remsen, D. P., ... Henderson, R. E. (2017). Uas-sfm for coastal research: Geomorphic feature extraction and land cover classification from high-resolution elevation and optical imagery. *Remote Sensing*, *9*(10), 1020.

Tomljanović, K., Kolar, A., Ðuka, A., Franjević, M., Jurjević, L., Matak, I., ... Balenović, I. (2022). Application of uas for monitoring of forest ecosystems–a review of experience and knowledge. *Croat. J. For. Eng*, *43*, 487–504.

Vanderhorst, H. R., Suresh, S., Renukappa, S., & Heesom, D. (2021). Uas application for urban planning development. In *European conference on computing in construction* (Vol. 189, p. 196).

Wang, T., Liu, Y., Wang, M., Fan, Q., Tian, H., Qiao, X., & Li, Y. (2021). Applications of uas in crop biomass monitoring: A review. *Frontiers in Plant Science*, *12*, 616689.

Wingtra. (n.d.). [web page]. Retrieved from https://wingtra.com/mapping-drone-wingtraone/gen-ii/

Wingtra. (2022). [web page]. Retrieved from https://wingtra.com/mapping-drone-wingtraone/

Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, *4*(1), 23–45.

Yang, L., Wu, X., Praun, E., & Ma, X. (2009). Tree detection from aerial imagery. In *Proceedings of the 17th acm sigspatial international conference on advances in geographic information systems* (pp. 131–137).

Yao, H., Qin, R., & Chen, X. (2019). Unmanned aerial vehicle for remote sensing applications—a review. *Remote Sensing*, *11*(12), 1443.

Yasin, M. F. M., Zaidi, M. A., & Nawi, M. N. M. (2016). A review of small unmanned aircraft system (uas) advantages as a tool in condition survey works. In *Matec web of conferences* (Vol. 66, p. 00038).

Yuheng, S., & Hao, Y. (2017). Image segmentation algorithms overview. *arXiv preprint arXiv:1707.02051*.

Zhang, X., Han, L., Han, L., & Zhu, L. (2020). How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? *Remote Sensing*, *12*(3), 417.

Zheng, J., Li, W., Xia, M., Dong, R., Fu, H., & Yuan, S. (2019). Large-scale oil palm tree detection from high-resolution remote sensing images using faster-rcnn. In *Igarss 2019-2019 ieee international geoscience and remote sensing symposium* (pp. 1422–1425).

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055*.