UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

Universidad Autónoma
de Madrid

Master in Deep Learning for
Audio and Video Signal Processing

MASTER THESIS

CATALOGUING AND MONITORING VEGETATION IN URBAN ENVIRONMENTS USING GEO-POSITIONED IMAGES

Rafael López García
Advisor: Marcos Escudero Viñolo

December 2023

CATALOGUING AND MONITORING VEGETATION IN URBAN ENVIRONMENTS USING GEO-POSITIONED IMAGES

Rafael López García
Advisor: Marcos Escudero Viñolo

Dpto. Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
December 2023

This work has been funded by the SEGA-CV (TED2021-131643A-I00) and the HVD (PID2021-125051OB-I00) projects of the Ministerio de Ciencia e Innovación of the Spanish Government

Resumen

Esta Tésis de Máster explora el uso de herramientas avanzadas de procesamiento de imágenes para catalogar y caracterizar la vegetación en entornos urbanos. Se analizan secuencias capturadas por una cámara geoposicionada en un vehículo en movimiento, empleando algoritmos de segmentación semántica, reidentificación de imágenes, clasificación de especies vegetales, obtención de la densidad de vegetación. Esto permite obtener una categorización detallada de la vegetación, mejorando significativamente el conocimiento de su distribución y densidad en entornos urbanos. Esta segmentación avanzada es esencial para evaluar la salud del ecosistema urbano y para planificar intervenciones de gestión ambiental. La reidentificación de vegetación en imágenes capturadas en distintos instantes posibilita la monitorización a largo plazo de la evolución de la vegetación, permitiendo observar tendencias de crecimiento, respuesta a intervenciones urbanas y cambios estacionales. Además, se investiga el uso de un clasificador de especies vegetales, una herramienta clave para mejorar la especificidad del catálogo. Este clasificador no solo facilita la identificación y diferenciación de la vegetación, sino que también proporciona datos valiosos para estudios de biodiversidad y conservación. La obtención de la densidad de vegetación en función de la posición da información acerca de la distribución espacial y la abundancia de las especies vegetales en entornos urbanos. Este conocimiento es fundamental para entender la dinámica ecológica de estas áreas, incluyendo aspectos como la biodiversidad, la salud del ecosistema, y su capacidad para proporcionar mejoras en el entorno, como la regulación del clima, la mejora de la calidad del aire y el fomento de espacios recreativos y estéticamente agradables. Además, este análisis permite identificar zonas con deficiencia de vegetación, guiando así los esfuerzos de planificación urbana y gestión ambiental hacia una mejora en la cobertura vegetal, lo cual es crucial para el bienestar urbano y la sostenibilidad. El objetivo final es desarrollar una aplicación que procese la información derivada de estas técnicas de análisis, ofreciendo una herramienta útil y accesible para la gestión ambiental y la toma de decisiones informadas. Este proyecto, que busca crear un registro actualizado y completo de la vegetación en entornos urbanos como el campus de la UAM, incluye un catálogo de vegetación que integra datos espacio-temporales.

Palabras clave

Procesamiento de Imágenes, Aprendizaje Profundo, Vegetación Urbana, Segmentación Semántica, Monitorización Ecológica, Gestión Ambiental, Clasificación de Vegetación.

Abstract

This Master Thesis explores the use of advanced image processing tools for cataloging and characterizing vegetation in urban environments. Sequences captured by a geopositioned camera on a moving vehicle are analyzed using semantic segmentation algorithms, plant species classification, vegetation density determination. This allows for a detailed categorization of vegetation, significantly improving the understanding of its distribution and density in urban settings. Such advanced segmentation is essential for assessing the health of the urban ecosystem and planning environmental management interventions. The reidentification of vegetation in images captured at different times enable long-term monitoring of vegetation evolution, allowing observation of growth trends, responses to urban interventions, and seasonal changes. Additionally, the use of a plant species classifier is investigated, a key tool for enhancing the specificity of the catalog. This classifier not only facilitates the identification and differentiation of vegetation but also provides valuable data for biodiversity and conservation studies. Obtaining vegetation density based on position provides information about the spatial distribution and abundance of plant species in urban environments. This knowledge is crucial for understanding the ecological dynamics of these areas, including aspects such as biodiversity, ecosystem health, and their ability to provide enviroment improvements, such as climate regulation, air quality improvement, and the promotion of recreational and aesthetically pleasing spaces. Moreover, this analysis allows for the identification of areas with vegetation deficiency, thus guiding urban planning and environmental management efforts towards improving vegetation coverage, which is crucial for urban well-being and sustainability. The ultimate goal is to develop an application that processes information derived from these analytical techniques, offering a useful and accessible tool for environmental management and informed decision-making. This project, which aims to create an updated and comprehensive vegetation register in urban environments like the UAM campus, includes a vegetation catalog that integrates spatio-temporal data.

Keywords

Image Processing, Deep Learning, Urban Vegetation, Semantic Segmentation, Ecological Monitoring, Environmental Management, Vegetation Classification.

Acknoledgements

Firstly, I would like to thank my mentor, Marcos Escudero Viñolo, for giving me the opportunity to work on something I am passionate about, and for being always available whenever I needed, helping and supporting me at all times. His mentorship has been one of the best experiences of my entire Master.

I would also like to thank my family for their support and belief in me from the beginning, without whom I could not have reached this point.

Finally, to my grandmother, I know that wherever you are flying, you will be watching over us.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The primary motivation behind this project comes from the need for an updated and comprehensive record of vegetation in urban environments, specifically on the campus of the Universidad Autónoma de Madrid. This necessity aligns with the increasing environmental challenges faced by urban areas, where effective management and monitoring of vegetation are crucial for sustainability and urban quality of life.

The importance of urban vegetation extends beyond aesthetic enhancement; it plays a crucial role in ecological and social functions, such as air quality improvement, heat island effect mitigation, and fostering mental and physical well-being in the university community. This project underscores these multifaceted benefits, emphasizing the ecological and social value of green spaces.

Technologically, the use of advanced image processing tools, including semantic segmentation algorithms, vegetation elements reidentification, and plant species classification, offers an innovative and automated methodology for detailed cataloging and characterization of urban vegetation.

The contribution of this project to urban sustainability provides a model for organizations aiming to integrate green space management into their sustainability goals. By creating a vegetation catalog that incorporates temporal data, this project facilitates long-term monitoring and informed decision-making, essential for future conservation strategies and urban planning.

Furthermore, the interdisciplinary nature of this work, involving collaborations across various departments and expertise, highlights the intersection of technology, ecology, urban planning, and education, raising awareness about biodiversity and sustainability among students and staff.

In summary, this thesis contributes to enhancing the quality of urban life by leveraging innovative techniques in the study of plant biodiversity. It highlights the importance of integrating ecological considerations into the fabric of urban planning and management, demonstrating the potential of technology in addressing environmental challenges in urban landscapes.

1.2 Objetives

The main objective of this Master's Thesis is to catalog and monitor vegetation in urban environments using geo-positioned images. This aims to utilize advanced image

processing tools for a detailed and automatic study of vegetation in urban contexts, such as the campus of the Universidad Autónoma de Madrid, providing an updated and comprehensive registry for environmental management and enhanced understanding of urban vegetation distribution and density.

This final objective can be divided into a series of partial objectives, which have allowed for the segmentation of the work and the gathering of necessary information to reach the final conclusions. These secondary objectives have been:

- Study and analysis of the related work in image processing for vegetation cataloging: Conducting an exhaustive review of literature and existing work in the field of image processing applied to vegetation, identifying strengths and limitations of current methods to build a solid foundation for our development.

- Design and development of a system for segmentation, reidentification, classification, and determination of vegetation density: Detailing the design and development process of the proposed system, including semantic segmentation algorithms, plant species classification, and vegetation density determination, explaining the integration of these components for accurate cataloging and effective system functionality.

- Evaluation and comparative analysis of the performance of the proposed system: Presenting the methods and results of tests conducted to evaluate the effectiveness of the developed system, with the purpose of conducting a comparison and determining the accuracy and feasibility of the system in a real-world context applications.

1.3 Report structure

This report has the following chapters:

- chapter 1 Introduction: sets the stage by introducing the challenges and significance of cataloging and monitoring urban vegetation, with a focus on the Universidad Autónoma de Madrid campus. It outlines the ecological and social importance of urban vegetation and sets the objectives and scope of the study.

- chapter 2 Related work: comprehensive literature review and analysis of existing image processing methodologies for vegetation cataloging are presented. This chapter highlights the strengths and limitations of current techniques, providing a foundation for the proposed system.

- chapter 3 Design and development: details the design and development of the proposed system, including semantic segmentation algorithms, plant species classification, and vegetation density determination. It explains how these components are integrated for effective cataloging and system functionality.

- chapter 4 Evaluation: methods and results of tests conducted to evaluate the system's effectiveness are presented in this chapter. It includes performance and analysis, assessing the system's accuracy and feasibility in a real-world setting.

- chapter 5 Conclusions and future work: summarizes key findings, highlighting the study's contributions and limitations. It discusses the implications for environmental management and vegetation monitoring in urban areas and suggests future research directions.

Chapter 2

Related work

2.1 Deep Learning in Image Processing

2.1.1 Fundamentals and General Applications

Deep Learning [9], a subset of machine learning, has emerged as a pilar in modern image processing. By employing artificial neural networks, this approach autonomously learns to perform complex image analysis and recognition tasks. Among various neural network architectures, Convolutional Neural Networks (CNNs) [10] and Visual Transformers [2] are particularly notable for their efficiency and accuracy in processing visual data.

CNN Designed specifically for visual data processing, CNNs are distinguished by their ability to detect patterns and features in images. Their architecture mirrors the way the human visual system processes information, identifying patterns ranging from simple to complex across multiple layers. What sets CNNs apart is their ability to learn features directly from the data, eliminating the need for manual feature extraction. This means CNNs can adapt to a wide range of image processing tasks, from object identification to image segmentation. Training a CNN involves using large datasets of labeled images. Through backpropagation and optimization algorithms, the network adjusts its weights to minimize the error in its predictions. This training process allows the CNN to improve its accuracy over time.

The basic layers of a CNN can be seen in Figure 2.1. They are:

- Convolutional Layers: Each convolutional layer applies a set of filters to the image to detect specific features like edges, textures, or shapes. These filters are automatically adjusted during training to capture crucial aspects of the images.

- Pooling Layers: Following each convolutional layer, there is usually a pooling layer that reduces the spatial dimension of the data, helping to decrease computation and prevent overfitting.

- Fully Connected Layers: At the end of the network, fully connected layers combine all features learned by previous layers to perform classification or regression tasks.

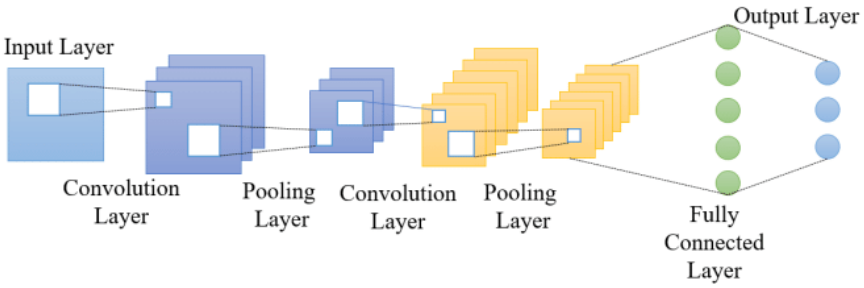


Figure 2.1: Figure showing the basic architecture of a CNN. Extracted from [1].

Deep Learning applications in image processing are diverse and impactful, cutting across various fields. In the realm of computer vision, these techniques have revolutionized how machines interpret and interact with visual information. Deep learning algorithms, particularly CNNs, have become integral multiple areas, as facial recognition, object detection or image segmentation. In facial recognition, deep learning enables the creation of applications ranging from security authentication to user identification in consumer devices. In object detection, deep learning helps in recognizing and locating objects within images, which is crucial for applications like autonomous vehicles, surveillance systems, and robotics. Another example is image segmentation, which involves dividing an image into multiple segments to make the image more meaningful and easier to analyze. Deep learning algorithms have enabled precise segmentation, useful in fields like satellite image analysis and medical image diagnostics.

Visual Transformers Visual Transformers (ViTs) are a class of machine learning models that adapt the Transformer architecture, originally designed for natural language processing tasks, for visual data processing. The essence of ViTs lies in treating images as sequences of pixels or patches, similar to how the Transformer architecture treats text as sequences of tokens.

In traditional Convolutional Neural Networks (CNNs), the image is processed through a series of convolutional layers, pooling layers, and fully connected layers. The convolutional layers are adept at capturing local features and maintaining the spatial hierarchy of the image data. However, one of the limitations of CNNs is the local receptive fields of the convolution operations, which can impede the model's ability to capture long-range dependencies within the image.

The ViT addresses this limitation by employing a Transformer architecture which is inherently designed to capture long-range dependencies. The Transformer utilizes self-attention mechanisms that weigh the influence of different parts of the input data, allowing it to consider the entire context of the input sequence, whether it be text for language models or patches of an image for ViTs.

Figure 2.2 shows the components present in the basic Visual Transformers structure. This components are the following:

- Patch and Position Embedding: An image is divided into fixed-size patches, which are then flattened and linearly projected into a sequence of vectors. Each vector is combined with a position embedding to retain the notion of order, as Transformers do not have any inherent notion of sequence order.

- Class Token: A learnable embedding, known as the 'class token' (usually denoted

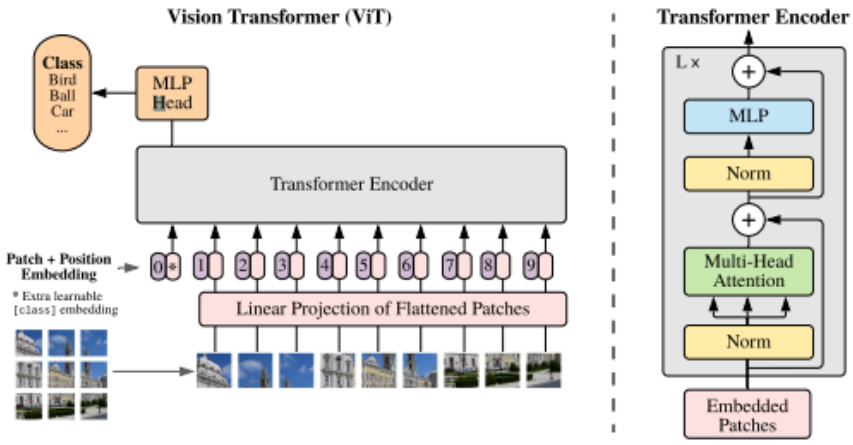


Figure 2.2: Figure showing the basic components of a Visual Transformer. Extracted from [2].

as [CLS]), is prepended to the sequence of embedded patches. The state of this token at the output of the Transformer encoder is used for classification tasks.

- Transformer Encoder: The sequence of embeddings (patches + class token) is passed through multiple layers of the Transformer encoder. Each layer has two main components:
  - Multi-Head Attention: This is where the self-attention mechanism comes into play, allowing the model to focus on different parts of the image simultaneously.
  - MLP (Multilayer Perceptron): This is a small feedforward neural network applied to each position separately and identically.
- Normalization and Residual Connections: Each sub-layer in the Transformer encoder, i.e., the attention and MLP layers, includes a residual connection followed by layer normalization. This design helps in training deeper models by mitigating the vanishing gradient problem.
- MLP Head: After the last Transformer layer, the state of the class token is passed through an MLP head to produce the final classification output.

The ViT represents a shift from inductive biases specific to image data, such as locality and translation invariance in CNNs, towards a more general-purpose architecture that can learn to recognize patterns in data, whether it be text or images, based on the context provided by self-attention mechanisms. This has been a significant development in the field of computer vision, allowing for more flexible models that can be applied to a variety of tasks beyond just classification, such as object detection and semantic segmentation.

Despite their effectiveness, CNNs and Visual Transformers face challenges like requiring large datasets and computational resources. Techniques like data augmentation and transfer learning have emerged to address these issues, allowing for more efficient training and better generalization to new data.

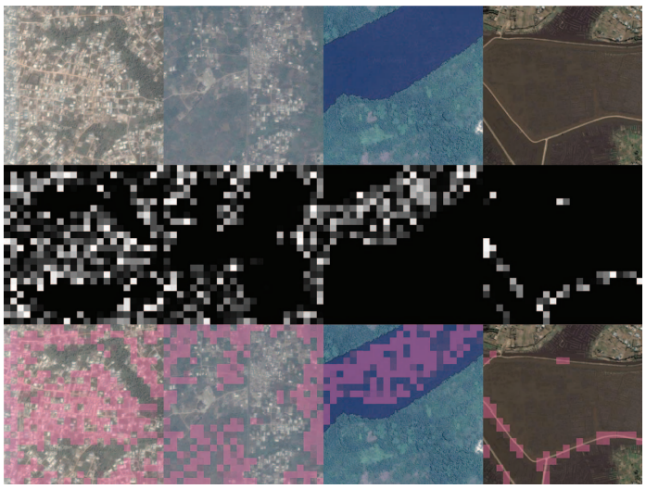


Figure 2.3: Figure showing the features extracted from an aerial imaging trained CNN. Extracted from [3].

2.1.2 Recent Advances in Urban Enviroment imaging

Urban environment imaging, a crucial aspect of modern urban planning and management, has been significantly enhanced by recent advances in Deep Learning. These advancements offer profound insights into urban landscapes, contributing to better planning, management, and sustainability. The advancements in image processing powered by Deep Learning have facilitated the monitoring and management of urban growth and transformations. This is particularly important in rapidly urbanizing regions, where keeping pace with changes is crucial for sustainable development.

High-Resolution Imaging and Analysis The use of high-resolution satellite and aerial imagery, processed through Deep Learning models, has revolutionized urban planning. These technologies enable detailed analysis of urban layouts, infrastructure, and land use. Deep Learning algorithms, especially CNNs, are now capable of processing high-resolution images to identify features like roads, buildings, green spaces, and water bodies with high accuracy. For instance, through the application of various convolutional filters, different elements such as urban areas, non-urban zones, bodies of water, and road networks can be distinctly identified and analyzed.

Figure 2.3, extracted from [3], illustrates this concept effectively. The first row presents original daytime satellite images sourced from Google Static Maps, showcasing a variety of urban and natural landscapes. Subsequently, the second row reveals the activation maps generated by the CNN model. Each map corresponds to a specific filter designed to detect particular features—urban textures, vegetation, water, and roadways are denoted by distinct activations, which are visually represented by the color pink. Finally, the third row fuses these activation maps with the original images, accentuating the detected features. These overlaid maps vividly demonstrate where the CNN filters are most activated, providing a clear visual representation of the model's feature detection capabilities.

Semantic Segmentation in Urban Landscapes Semantic segmentation using Deep Learning has become a key tool in understanding urban environments. It involves classifying each pixel in an image into categories, such as buildings, roads, or

Figure 2.4: Figure showing the segmentation obtained by SAM. Extracted from [4].

vegetation, providing detailed insights into the urban fabric. This technique is vital for urban monitoring, land-use planning, and environmental assessment, helping cities to become more sustainable and livable.

A prime example of these advancements is the Segment Anything Model (SAM). SAM is an advanced framework that aids in the precise segmentation of images. In Figure 2.4, we observe the output of SAM applied to an urban street scene. The image is partitioned into color-coded segments, each representing a different object class. SAM's ability to segregate and label complex environments is demonstrated here, showcasing its potential to significantly contribute to the use of this kind of technology in multiple areas.

Urban Vegetation and Green Space Analysis Recent image processing advancements have enabled the effective assessment of urban green spaces, essential for maintaining healthy ecosystems and planning urban landscapes. Utilizing Google Street View, researchers have developed a modified Green View Index to measure the quality and impact of street-level greenery [11]. This method offers a new perspective for urban planners to evaluate and manage green spaces within cities. The approach shows promise for improving urban landscape planning by providing objective and detailed views of urban vegetation.

3D Urban Modeling and Depth Estimation 3D modeling of urban environments using Deep Learning provides comprehensive spatial understanding, which is vital for urban design and disaster management. Deep Learning-based depth estimation techniques are being employed to generate 3D models from 2D images, offering valuable insights into the urban landscape.

While these advancements are promising, challenges like data privacy, computational requirements, and accuracy in diverse urban settings persist. Ongoing research is directed towards making these technologies more efficient, inclusive, and accessible.

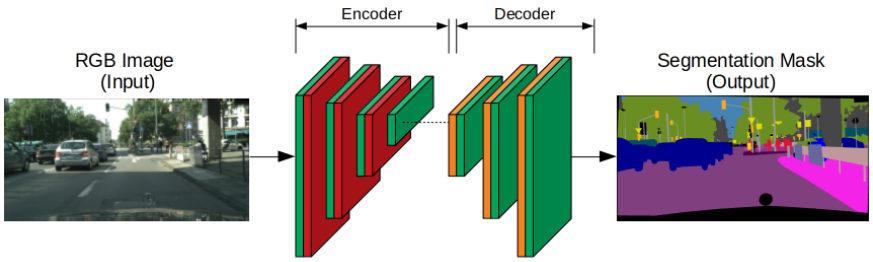


Figure 2.5: Figure showing the architecture of a segmentation model. Extracted from [5].

2.2 Semantic Segmentation with Deep Learning

2.2.1 Fundamentals, Techniques and Algorithms

Semantic segmentation is a fundamental task in the field of computer vision that involves assigning a semantic label to each pixel in an image. This task is essential for a wide variety of applications, including autonomous driving, object detection, medical imaging, precision agriculture, and robotics. As research in computer vision advances, deep learning techniques have proven to be highly effective in addressing semantic segmentation.

Deep Learning Architectures for Semantic Segmentation Initially, an encoder and a decoder formed the fundamental architecture used in picture segmentation. By using filters, the encoder extracts features from the picture. The final output, which is often a segmentation mask with the object's contour, is produced by the decoder. An example of the basic architecture of this kind of network is shown in Figure 2.5.

This structure, or a variation on it, is present in the majority of the architectures for this kind of technique. Some of the principal models include the following:

- **Deep Convolutional Networks (CNNs):** Deep Convolutional Networks, or CNNs, have been the backbone of many successful approaches in semantic segmentation. These networks consist of convolutional layers that learn low to high-level patterns in images. Examples of popular architectures include U-Net [12], Fully Convolutional Network (FCN) [13], and SegNet [14].

- **Efficient Neural Networks (ENet):** ENet [15] is an example of a lightweight network architecture designed specifically for real-time semantic segmentation. It uses a set of more efficient convolutional layers and dimensionality reduction strategies to achieve accurate results with fewer computational resources.

- **Dilated Convolutional Neural Networks (DCNNs):** DCNNs are an extension of traditional CNNs that use dilated convolutional layers to increase the receptive field without exponentially increasing the computational burden. These networks are useful for capturing contextual information in images and enhancing semantic segmentation accuracy.

- **Segment Anything:** this project created by Facebook [4] represents a significant advancement in the field of image segmentation, leveraging Deep Learning to create a foundation model for this purpose. This model is unique in its ability

to understand and segment any object in any image or video, a capability known as zero-shot transfer. SAM's design allows it to adapt to various tasks and domains without the need for additional training, making it highly versatile and broadly applicable. A more detailed explanation of the model will be made in Section 3.3, as this is the model used for the segmentation in this project.

2.2.2 Applications in Urban Environments

Semantic segmentation with deep learning has proven to be a powerful tool in a wide range of applications in urban environments. In this section, we will explore some applications that illustrate how semantic segmentation is used to address different challenges in urban areas.

Autonomous Driving One of the most prominent applications of semantic segmentation in urban environments is autonomous driving. Autonomous vehicles heavily rely on the ability to understand and segment their surroundings to make safe and efficient decisions. Semantic segmentation is used to identify and classify objects on the road, such as pedestrians, vehicles, traffic signs, traffic lights, and lanes. This allows autonomous driving systems to make informed decisions, avoid obstacles, and follow traffic rules [16].

Urban Planning and Traffic Management Semantic segmentation has also been applied in urban planning and traffic management. By analyzing images and videos from traffic cameras and real-time surveillance systems, cities can monitor traffic flow and identify congestion [17]. The information obtained from semantic segmentation helps make informed decisions about road design, traffic policy implementation, and emergency management.

Infrastructure Maintenance In urban environments, infrastructure plays a crucial role in citizens' quality of life. Semantic segmentation has been used to inspect and maintain critical infrastructure, such as bridges, buildings, sewer systems, and water systems [18]. Drones equipped with cameras and semantic segmentation algorithms can identify damage, corrosion, and other maintenance issues, enabling early and efficient intervention.

2.3 Image Reidentification

2.3.1 General Idea and Techniques

Image Reidentification (ReID) is an area within computer vision that seeks to identify and match individuals or objects across different images or video frames. The fundamental objective is to maintain the identity consistency of objects in diverse visual scenes, despite variations in angle, illumination, occlusion, and other environmental factors. This process involves extracting distinctive features from the subjects that are invariant to such changes. The applications of image reidentification are broad, extending from security surveillance systems, where the technology enables the tracking of individuals across different camera feeds, to smart retail solutions, where it aids in the analysis of customer behavior.

Traditionally, methods for image ReID relied heavily on feature engineering, where hand-crafted features were extracted to represent the visual appearance of subjects. However, with the advent of deep learning, there has been a paradigm shift towards learning feature representations. Convolutional Neural Networks (CNNs) have been the backbone of such approaches, benefiting from their ability to learn hierarchical representations. More recently, the introduction of Transformer-based models has marked a significant advancement in the field. Transformers, initially developed for natural language processing tasks, have been adapted to handle image data by treating images as sequences of patches. This allows for capturing long-range dependencies and global contexts, offering promising results in reidentification tasks [19].

2.3.2 Reidentification based on Image Warping

Reidentification based on Image Warping is a technique in computer vision that aims to improve the accuracy of matching individuals across different images by addressing discrepancies in pose, viewpoint, and even camera modalities. Image warping involves transforming different images of the same object or individual to a common perspective or alignment, enabling more effective comparison and matching of features. Several steps are performed for image reidentification:

- **Feature Extraction:** Extracting discriminative features from images is a fundamental step in image reidentification. Commonly used features include color histograms, texture descriptors, deep learning-based embeddings, and keypoints-based descriptors such as SIFT (Scale-Invariant Feature Transform) or ORB (Oriented FAST and Rotated BRIEF).

- **Feature Matching:** Once features are extracted, matching algorithms are employed to find correspondences between features in different images. Matching can be based on various similarity metrics, including Euclidean distance, cosine similarity, or more advanced techniques like the nearest-neighbor search in high-dimensional feature spaces.

- **Geometric Transformations:** To account for changes in viewpoint, scale, or orientation, geometric transformations are applied to align the matched features. The transformation that is often used for reidentification is the Homography transformation, which models perspective changes between images. The Homography matrix relates the coordinates of points in one image to the coordinates of their corresponding points in another image.

Homography estimation plays an important role in image reidentification, and is the technique used for this thesis. Some studies have shown that the application of homography estimation together with other techniques, as deep convolutional neural network [20], achieve great results for image reidentification.

The Homography matrix, denoted as H, is a 3x3 transformation matrix that relates the coordinates of points in one image to the coordinates of their corresponding points in another image. In the context of vegetation tracking in urban environments, Homography estimation allows for the correction of perspective distortions caused by changes in camera viewpoint and position.

(a) First perspective. (b) Second Perspective (c) Image 2.6a with Image 2.6b perspective

Figure 2.6: Figure showing an example of how Homography Estimation works.

The estimation of the Homography matrix typically involves selecting a set of corresponding points in both images, where these points represent the same physical locations or objects. Once these correspondences are established, various methods can be used to estimate the Homography matrix. One common approach is the Direct Linear Transform (DLT) algorithm, which is based on solving a system of linear equations derived from the point correspondences. A visual example can be seen in the Figure 2.6.

Homography estimation is crucial for accurate vegetation reidentification because it allows for the transformation of vegetation regions from one image to the coordinate space of another image, ensuring that the same vegetation instances are consistently tracked and identified across different time intervals.

2.4 Image Clasification

2.4.1 Clasification Methods and Algorithms

Image classification is a fundamental task in computer vision that involves assigning a label or category to an input image based on its visual content. Over the years, significant progress has been made in the field of image classification, driven by advancements in deep learning.

In the context of urban vegetation analysis, vegetation image classification plays a critical role in characterizing and cataloging plant species within urban environments. This specialized area of image classification focuses on identifying and differentiating various types of vegetation based on their visual characteristics.

Continuing with the advancements in the field of image classification, it is crucial to highlight the significant evolution of state-of-the-art models in vegetation image classification in recent years. The current leading methods in this field employ advanced deep learning architectures, encompassing both convolutional and transformer-based approaches, for the automatic recognition of plant species based on images. These approaches have been evaluated and benchmarked, utilizing some of the largest and publicly available datasets for plant recognition [21].

Cutting-edge vision transformer models have achieved great accuracy rates, significantly outperforming previous benchmarks. On the other hand, the latest convolutional neural network models have shown a substantial reduction in error rates. Furthermore, performance-enhancing techniques such as class priority adaptation, image augmentations, learning rate scheduling, and optimized loss functions have further increased the efficacy of these models.

Among these innovative solutions, platforms like Pl@ntNet [22] represent a notable development. Pl@ntNet, which is the model used in this project, is a collaborative tool that leverages the power of image-based machine learning to identify plant species. Utilizing a vast database of plant images and citizen science contributions, it applies advanced classification algorithms to assist users in identifying various plant species from photos. This approach not only enhances public engagement in biodiversity studies but also contributes valuable data to the scientific community.

These advancements in vegetation image classification not only significantly improve the ability to catalog and monitor vegetation in urban environments but also contribute to broader efforts in environmental management and sustainable urban planning. With the rapid development of these technologies, the accuracy and efficiency in classifying and tracking urban vegetation are expected to continue improving, which will have a significant impact on biodiversity conservation and urban ecosystem management.

2.4.2 Challenges in Vegetation Image Classification

Vegetation image classification presents unique challenges due to the complex and diverse nature of plant species in urban settings. The fine-grained recognition of plants from images is recognized as a challenging task due to the diverse appearance and complex structure of plants, along with high intra-class variability and small inter-class differences [23]. Moreover, traditional classification approaches face difficulties in accurately extracting vegetation covers from aerial imagery, as urban vegetation categories have complex characteristics [24]. This complexity is compounded by the need for land cover classification to focus on chlorophyll-rich vegetation detection, which is vital for urban growth monitoring and planning, autonomous navigation, drone mapping, and biodiversity conservation [25].

To overcome these difficulties in vegetation image classification for urban environments, it is necessary to develop and apply different strategies. The use of advanced image preprocessing techniques, can help mitigate the issue of inter and intra-class variability. Additionally, the design of balanced datasets and the implementation of techniques like weighted sampling or class rebalancing can address the problem of data imbalance.

To contend with seasonal changes, classification models can be trained with data collected during different times of the year, ensuring that the model is capable of recognizing plant species in all their growth stages and phenological states. Integrating metadata, such as geographical location and the date of image capture, can also provide valuable insights that enhance classification accuracy.

Advancements in deep learning techniques and the development of more sophisticated models, such as deep neural networks and hybrid models that combine different types of network architectures, offer new opportunities to address these challenges. Furthermore, the growing interest and participation in open projects, like Pl@ntNet, provide a valuable source of diverse and labeled data that can be used to train and improve classification models.

2.5 Studying Vegetation Density Through Imaging

2.5.1 Techniques and Tools

The assessment of vegetation density in urban environments is a crucial component of urban planning and environmental management. Understanding the spatial distribution and abundance of plant species within cities has become increasingly important due to the growing impact of urbanization on ecosystems. Vegetation within urban areas serves multifaceted functions, from providing essential ecosystem services such as air purification and temperature regulation to enhancing the aesthetics of the urban landscape. Monitoring and characterizing vegetation density through advanced imaging techniques have emerged as powerful tools in addressing these challenges. The traditional methods of collecting vegetation data, such as field surveys, are often limited in scope, time-consuming, and labor-intensive. This has led to the exploration of innovative and technologically-driven approaches to assess vegetation density [26]. Advanced imaging techniques have emerged as powerful tools for the systematic, high-resolution, and large-scale characterization of urban vegetation. These techniques enable researchers to capture detailed information about the density, distribution, and health of vegetation in ways that were not previously possible. The assessment of vegetation density within urban environments has greatly benefited from the continuous evolution of imaging techniques and tools. These technologies have opened new horizons for researchers and practitioners seeking to understand and manage urban ecosystems.

Remote Sensing and Satellite Imagery Traditional remote sensing techniques have been instrumental in assessing vegetation cover at large scales. Satellite imagery, particularly from missions like Landsat and Sentinel, provides a macroscopic view of urban vegetation dynamics over extensive geographical areas. These data sources have been invaluable for regional and global studies, offering insights into broad trends in urban greenery.

However, satellite imagery often suffers from limitations in spatial and temporal resolution [27]. The coarse spatial resolution may not capture fine-scale urban features, while the infrequent revisit times can miss rapid changes in vegetation patterns.

Aerial Photography and Drones Aerial photography, enabled by drones, has revolutionized vegetation assessment in urban environments. Drones equipped with high-resolution cameras can capture detailed images of urban landscapes with a level of granularity that was previously unattainable. This enables researchers to closely examine vegetation structure, health, and distribution at the local level [6]. Figure 2.7 shows how drone images can be used to obtained the vegetation elements of a specific area.

The advantages of drones include their flexibility in data acquisition, ability to cover small and intricate areas, and rapid deployment. Researchers can tailor drone missions to specific urban sites of interest, making them ideal for detailed urban ecology studies.

Figure 2.7: Figure showing the vegetetation obtained from a drone. Extracted from [6].

2.6 Cataloguing and Monitoring Vegetation in Urban Areas

2.6.1 Impact and Relevance for Urban Management

The cataloging and monitoring of vegetation in urban environments have become a topic of significant relevance in environmental research and urban planning in recent decades. Rapid urbanization and the expansion of metropolitan areas have led to a decline in natural vegetation, resulting in a range of environmental and public health challenges. The ability to catalog and monitor vegetation in urban areas is essential for understanding and effectively addressing these issues. Urban vegetation plays a crucial role in enhancing the quality of life in cities. It provides not only aesthetic benefits but also has significant impacts on air quality, urban temperature, biodiversity, and human health. Additionally, vegetation contributes to climate change mitigation by absorbing carbon dioxide (CO2) and reducing local temperatures through shading and evapotranspiration. Historically, cataloging and monitoring of urban vegetation have been conducted through on-site visual inspections, surveys, and analysis of aerial imagery. However, with advancements in deep learnig and image processing techniques, more accurate and efficient methods have been developed.

Several studies have investigated the cataloging and monitoring of vegetation in urban areas using various geospatial techniques. The study "Effects of Urban Vegetation on Urban Air Quality" [28], highlights the significant role of vegetation in improving air quality and combating global warming. It discusses the benefits of tree planting in mitigating urban heat island effects, sequestering carbon dioxide, and trapping air pollutants on leaves. The study also notes the importance of choosing appropriate plant species, as some emit biogenic volatile organic compounds (BVOCs) affecting air quality. Another example is the study "Role of Urban Vegetation in Air Phytoremediation" [29], which demonstrates the significant impact of urban vegetation in reducing air pollution. It reports reductions in concentrations of particulate matter, nitrogen oxides, and sulfur dioxide, highlighting the effectiveness of urban green spaces in improving air quality. The study, "The Nexus Between Vegetation, Urban Air Quality, and Public Health" [30] suggests that urban greening strategies, such as vegetation screens and barriers, can mitigate urban heat and air pollution, reducing pollution-related deaths. This study emphasizes the need for strict laws and monitoring programs in major cities

for effective pollution and vegetation management.

In summary, the cataloguing and monitoring of vegetation in urban environments are fundamental elements in contemporary environmental research and urban planning. Advancements in deep learning and image processing techniques have revolutionized the way we catalog and monitor vegetation in urban areas, enabling greater precision and efficiency in our research. As evidenced by the cited examples of studies, research in this field yields valuable insights into how vegetation can combat air pollution, reduce urban heat effects, and enhance residents' quality of life.

2.6.2 Challenges and Future Trends

The task of cataloguing and monitoring urban vegetation, faces several challenges and presents exciting future trends.

Challenges:

- **Data Quality and Availability:** One of the primary challenges is the availability of high-quality, up-to-date data [31]. Despite advances in remote sensing, the resolution and frequency of data collection can vary, impacting the accuracy of vegetation cataloguing.
- **Integration of Multi-source Data:** Integrating data from various sources, such as satellite imagery, aerial photography, and ground surveys, presents challenges in terms of compatibility and standardization.
- **Urban Complexity:** The complex urban environment, with its diverse range of surfaces and materials, poses challenges for accurately detecting and monitoring vegetation using remote sensing techniques [32].
- **Climate Change Impact:** Changing climatic conditions can rapidly alter urban vegetation patterns [33], making monitoring a moving target.

Future Trends:

- **Artificial Intelligence and Machine Learning:** AI and machine learning algorithms are expected to play a significant role in processing and analyzing large datasets, enabling more accurate and efficient vegetation monitoring.
- **Citizen Science:** The involvement of the public through citizen science projects can augment traditional data collection methods, providing ground-level observations and increasing community engagement [22].
- **Urban Planning Integration:** An increased integration of vegetation monitoring data into urban planning and policy-making processes is anticipated, emphasizing the role of green spaces in sustainable urban development [34].
- **Climate Resilient Vegetation:** Focus on planting and monitoring climate-resilient vegetation to adapt to changing environmental conditions in urban areas [35].

In conclusion, while challenges such as data quality and urban complexity persist, advancements in technology and methodology offer promising solutions. The integration of sophisticated remote sensing, AI, and community involvement are shaping the future of cataloguing and monitoring urban vegetation. This evolution is crucial for ensuring sustainable urban development and enhancing the quality of life in urban areas.

Chapter 3

Design and development

3.1 System Pipeline Overview

This section provides a detailed description of the workflow for the system designed to monitor and catalog vegetation in urban environments using geo-positioned images. Employing images captured from a moving vehicle, the workflow explores advanced image processing techniques focusing on semantic segmentation, species reidentification, vegetation classification, and determination of vegetation density. This workflow is crucial for understanding the distribution and density of urban vegetation and planning environmental management interventions. This pipeline can be seen in Figure 3.1, and consists of the following steps:

1. **Data Acquisition:** Data was acquired by equipping a vehicle with a geo-positioned camera, following a predetermined route during different seasons (March and May). This approach allowed the capture of seasonal variations in vegetation and facilitated the reidentification of species. The selection of routes and seasons was key to obtaining a diverse and representative dataset, essential for the creation of the vegetation catalog.

2. **Video Processing:** The captured videos were processed using advanced image processing techniques. Specific frames were extracted from the videos, selected based on their relevance to subsequent analysis stages.

3. **Semantic Segmentation:** Semantic segmentation was performed. This stage was computationally demanding, leading to a detailed study of algorithm parameters to optimize precision and efficiency. Accurate segmentation was critical for the reidentification and classification stages. This process was time and resource-intensive, taking approximately 15 days to complete segmentation for around 20,000 frames. The possibility of processing a representative sample of the videos in limited time situations was considered a viable alternative.

4. **Image Reidentification:** Vegetation reidentification in images captured at different time intervals was performed using homography estimation algorithms. This step was crucial for refining the segmentation and facilitating subsequent classification. Different combinations of estimations were studied to improve the accuracy of the process.

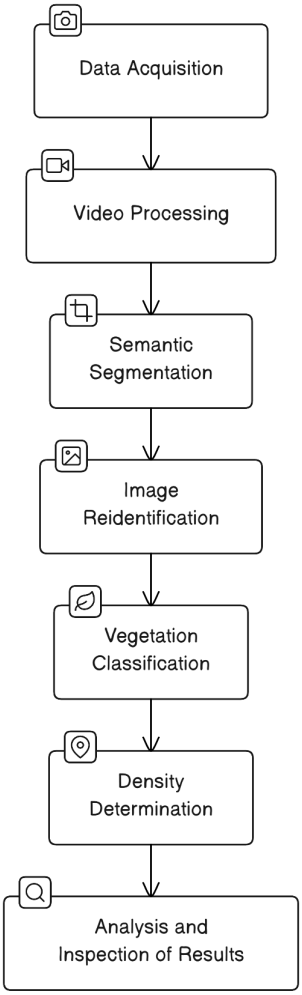


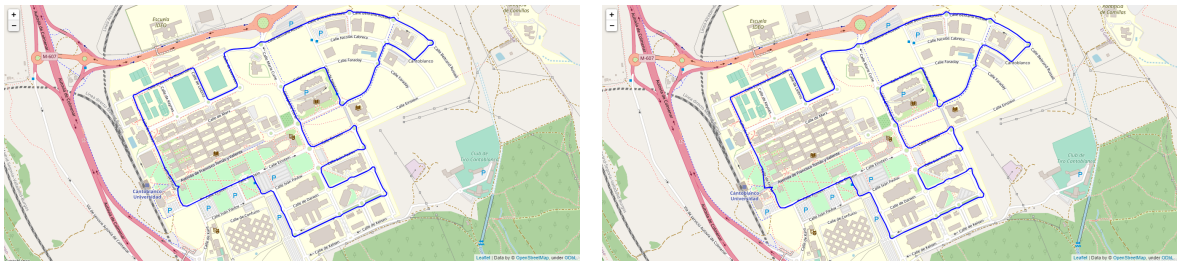
Figure 3.1: Figure showing the system pipeline.

5. **Vegetation Classification:** Vegetation classification was based on identifying the genus of species. Probabilities associated with each class were analyzed, and those belonging to the same genus were summed to identify the most probable one. This method was preferred over classification based solely on the species with the highest score, due to its higher reliability.

6. **Density Determination:** The vegetation density was determined by associating frames by their location and comparing the vegetation present in them. This method allowed for precise differentiation of density, resulting in detailed maps of the routes with associated vegetation density.

7. **Analysis and Inspection of Results:** The analysis of the results was primarily visual, allowing for a comprehensive understanding of the system's functioning. This visual inspection revealed details about the effectiveness of each process stage and provided valuable insights for future improvements.

In summary, the workflow described in this section shows the developed system for cataloging and monitoring vegetation in urban environments. This system allows us to create multiple applications, that are really useful in the context of vegetation in urban enviroments.



(a) March route map (b) May route map

Figure 3.2: Figure showing the two routes taken to collect the videos.

3.2 Data Acquisition and Video Processing

3.2.1 Data Collection

This study is based on the utilization of a dataset comprised of two videos captured on the campus of the Autonomous University of Madrid (UAM) using a GoPro camera. These videos serve as a fundamental source of information for the research, enabling an in-depth exploration of vegetation in urban environments from a novel perspective. The study location, UAM's campus, is adecuate because of its diversity of urban landscapes, making it an ideal environment for analyzing vegetation in different contexts.

Both videos were recorded from a moving vehicle following the same route at two different temporal moments: one in March and another in May. The selection of these specific dates was strategic, as it allowed us to capture seasonal variation in vegetation. The duration of each video is approximately 15 minutes, providing a substantial amount of data for analysis.

3.2.2 Video Processing

The processing of the videos was carried out in multiple stages to ensure data quality and consistency. Firstly, the videos were split into individual frames, resulting in a sequence of snapshots representing different moments along the route. Each frame contains visual information about the surrounding environment.

A crucial feature of the videos recorded with the GoPro camera is that they provide geospatial data associated with each frame. Thanks to this information, the precise location of each frame in terms of GPS coordinates was obtained, allowing for accurate georeferencing of the data.

The next step in the data processing was the construction of a map that visualizes the route taken by the vehicle. This map reveals how the frames are distributed along the route, which is essential for understanding the spatial coverage of data acquisition and identifying areas of interest.

In Figure 3.2, we present two maps representing the routes taken in March and May, respectively. These maps clearly illustrate that the routes are nearly identical, following the streets of the Autonomous University of Madrid (UAM) campus. The similarity in the routes is crucial for ensuring an effective comparison of vegetation at different time points, as it allows us to analyze the same geographical areas in different seasons of the year. This increases the robustness of our results by considering seasonal variation in vegetation and its response to different environmental conditions over time.

(a) March frame example (b) May frame example

Figure 3.3: Figure showing an example of the frame association.

3.2.3 Frame Association

Once GPS coordinates for the frames in both videos were obtained, a fundamental challenge arose: the temporal and spatial alignment of frames from the two videos. Since the videos did not have the same temporal duration due to possible interruptions during recording, a matching process was required to associate each frame from one video with its nearest counterpart in the other video.

To address this challenge, different distance metrics were explored, including Euclidean distance and Haversine distance, the latter taking into account the curvature of the Earth. After comprehensive analysis, it was determined that Euclidean distance provided similar results with greater computational efficiency, making it the preferred choice.

Frame association was achieved using the Hungarian algorithm, an optimal assignment technique that allowed for establishing relationships between frames from the two videos in such a way that each frame in one video was paired with its nearest frame in the other video. This ensured proper alignment and enabled consistent frame comparisons.

However, a new challenge emerged: the fact that the two videos did not have the same number of frames. Once a frame was associated with its nearest counterpart, it could not be reused, posing issues in situations where the nearest frame had a greater distance than desired, so it made the comparations between those frames unuseful as they were not representing the same location. To overcome this hurdle, a distance filter was implemented, allowing frames to be paired even if they had already been used, as long as the distance was less than a predefined threshold. This ensured a continuous and coherent correspondence between frames from both videos.

Figure 3.3 shows a visual representation of the frame association process and the alignment of frames between the two videos. This process facilitates the comparison of images captured at different time points and also enables a examination of the differences that exist between them. By visually inspecting the results of the frame association, it is possible to get an understanding on how the urban vegetation landscape evolves over time.

3.2.4 Dataset Outcome

With frames properly aligned and paired, the resulting dataset was prepared for subsequent analysis. Each geospatially referenced and temporally aligned frame became a

valuable resource that would support all aspects of the research. These data will serve as the foundation upon which the image processing techniques and vegetation analysis described in the subsequent chapters of this thesis will be applied. It's important to note that this dataset is not only valuable for our current research but also possesses the potential for future expansion and enhancement. This expansion can occur on two main fronts. Firstly, by including data from additional time periods beyond the initial captures in March and May, we can increase the dataset's temporal diversity, providing a more comprehensive understanding of vegetation dynamics. This will enable us to explore seasonal variations, long-term trends, and responses to environmental changes more comprehensively. Secondly, the dataset can be expanded geographically by capturing data from additional areas within the campus. This broader coverage will facilitate the creation of a comprehensive catalog of the existing vegetation throughout the entire campus. Such a catalog would be instrumental not only for ongoing maintenance but also for other research groups interested in conducting studies to enhance sustainability and the overall quality of the campus environment. The dataset created in this process stands as an resource not only for this particular research but also as a tool that can benefit future investigations in the field of urban ecology and environmental management.

3.3 Semantic Segmentation

3.3.1 Segmentation in the Context of Vegetation Cataloging and Monitoring

In this section, we present the implementation and utilization of semantic segmentation techniques within the context of this thesis project, focusing on cataloging and monitoring vegetation in urban environments using geopositioned images. Semantic segmentation plays a fundamental role in achieving the project's goals by providing detailed information about the distribution and characteristics of vegetation in urban settings.

3.3.2 Segmentation Algorithm

In our quest to catalog and monitor urban vegetation, the initial exploration led us to consider the "Segment Anything" algorithm [4] developed by Facebook. This algorithm has gained recognition for its capabilities in segmenting objects within images and has become one of the prominent choices in the field of image segmentation.

Facebook's "Segment Anything" project introduces a new task, dataset, and model for image segmentation, aiming to democratize this field. The Segment Anything Model (SAM) and the Segment Anything 1-Billion mask dataset (SA-1B), the largest segmentation dataset to date, are available for research under an open license.

SAM is a foundational model for image segmentation, inspired by prompting techniques used in natural language processing models. It can generate masks for any object in images or videos, including those not encountered during training. SAM's ability to adapt to new image "domains" without additional training, a feature known as zero-shot transfer, makes it suitable for a wide range of applications.

SAM generalizes two traditional segmentation approaches: interactive and automatic segmentation. It can efficiently perform both tasks thanks to its adaptable

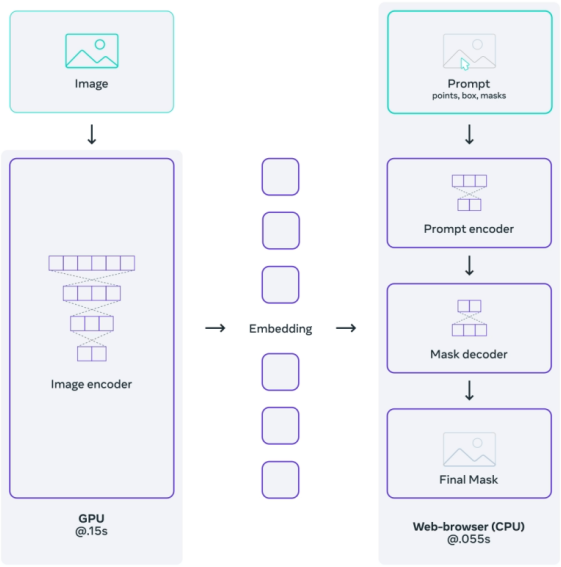


Figure 3.4: Figure showing the pipeline of SAM. Extracted from [4].

interface. Trained on a diverse, high-quality dataset of over 1 billion masks, SAM can generalize to new types of objects and images beyond its training, reducing the need for users to collect their own segmentation data and fine-tune the model for their use case.

Figure 3.4 shows the workflow of SAM model. Internally, SAM uses an image encoder to produce a unique embedding for the image, while a lightweight encoder converts any prompt into an embedding vector in real time. These two information sources are then combined in a lightweight decoder that predicts segmentation masks. After computing the image embedding, SAM can produce a segment in just 50 milliseconds with any prompt in a web browser.

However, as we delved deeper into our research objectives, we realized that the algorithm, while proficient in segmenting images, lacked a crucial element that our project demanded: semantic understanding of the segmented regions. While this algorithm could skillfully segment the vegetation within images, it did so without assigning semantic labels to the resulting segments. In other words, it could identify and isolate various elements of the image but lacked the capability to categorize them into meaningful classes (needed to identify vegetation).

The absence of semantic information made it impossible to distinguish between the different elemnts, including vegetation, which was fundamental to our research objectives. Without this semantic context, the segmented regions remained abstract clusters of pixels rather than meaningful representations of urban vegetation. Consequently, to fulfill our need for semantic context in vegetation segmentation, we sought out an advanced solution. We discovered the 'Semantic Segment Anything' algorithm [7], an iteration of SAM designed to not only perform segmentation but also provide the crucial semantic information our project required. This algorithm maintains the robust segmentation capabilities of SAM, while enhancing it with the ability to assign semantic labels to the segmented regions. This integration allows us to distinguish between different types of urban elements, aligning precisely with our research objectives and enabling a comprehensive analysis of urban ecosystems

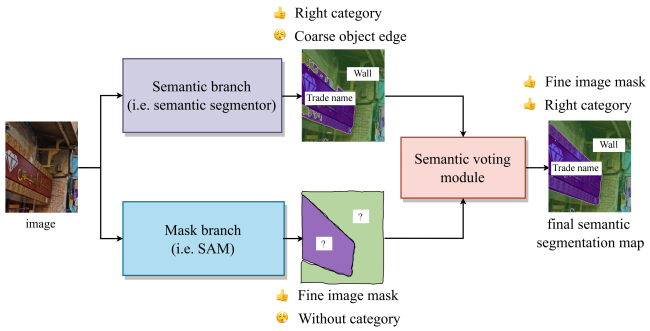


Figure 3.5: Figure showing the structure of SSA. Extracted from [7].

3.3.3 Introducing Semantic Information

The Semantic Segment Anything (SSA) project builds upon the Segment Anything Model (SAM) to address a critical gap in semantic segmentation. While SAM offers precise object segmentation and SA-1B stands as a vast dataset, they lack semantic categorization for masks. SSA integrates an automated dense open-vocabulary annotation engine, the SSA-engine, which provides rich semantic category annotations, reducing the need for manual annotation efforts.

SSA facilitates the use of SAM for semantic segmentation tasks, enabling seamless integration with existing semantic segmenters. This means users can improve their models' generalization and mask precision without retraining or fine-tuning SAM. The framework consists of two main branches and a voting module. This can be seen in Figure 3.5.

- **Mask Branch:** Utilizes SAM to provide clear boundary masks.
- **Semantic Branch:** Implements a semantic segmenter that classifies each pixel's category, which can be customized according to the user's needs.
- **Semantic Voting Module:** Determines the mask's category by cropping out pixel categories and selecting the most frequent (top-1) category as the classification result.

The SSA-engine, shown in Figure 3.6, consists of a close-set semantic segmentor for basic category information, an open-vocabulary classifier for diverse label extraction using image captioning, and a final decision module that filters the most reasonable predictions to determine the mask's category.

1. **Close-set Semantic Segmentor:** Beginning with the input image, a close-set semantic segmentor (depicted in green) processes the image to provide a semantic segmentation map. This segmentor, trained on datasets like COCO and ADE20K, assigns basic category labels to various regions of the image.
2. **Open-vocabulary Classifier:** For each mask produced by the segmentor, a corresponding image patch is cropped. An open-vocabulary classifier (in blue, such as BLIP) then describes these patches, extracting nouns or phrases as candidate categories, offering a broader range of labels.
3. **Top-k Classes and Final Decision:** The semantic segmentation map and the open-vocabulary classifier's output feed into the final decision module. A class

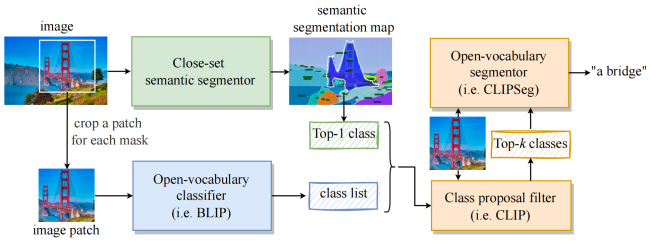


Figure 3.6: Figure showing the pipleine of the SSA engine. Extracted from [7].

proposal filter (like CLIP, shown in orange) filters the top-k class predictions to refine the selection. The top-1 class from these predictions is then assigned to the corresponding mask on the image, completing the annotation process.

This combination of close-set and open-vocabulary segmentation enhances SAM's capabilities, offering a potent tool for dense categorization and substantial improvements in the accuracy of semantic annotations. This innovation positions SSA-engine as a foundational element for training sophisticated visual perception models and fine-grained CLIP models, filling the void in SA-1B's fine-grained semantic labeling.

3.3.4 Application of the Model to Collected Data

The semantic segmentation algorithm was executed on each frame, resulting in the creation of segmented images. Importantly, alongside each segmented image, an associated file containing comprehensive segmentation information was generated. This file contains valuable data pertaining to the class or category assigned to each segmented region within the image, enabling us to distinguish between different types of classes (including vegetation). Additionally, this model returns each segmentation element separately. This greatly facilitates subsequent processing, as the images are already separated by vegetation elements and not just as a complete mask of vegetation without differentiation between elements. The output of the semantic segmentation process serves as a resource for subsequent tasks undertaken in this project.

3.4 Image Reidentification

3.4.1 Reidentification for Vegetation Cataloguing and Monitoring

Reidentifying vegetation at various times and frames is essential in the field of urban vegetation monitoring. In addition to making it possible to observe the condition and evolution of vegetation in great detail, this approach offers important insights into the biological dynamics of urban areas. Researchers and urban planners can better understand how vegetation reacts to environmental and urban changes by being able to recognize and contrast the same plant at different times. The growth, health, and reaction of plants to urban interventions and seasonal variations can be tracked by examining photos taken at various dates. When planning environmental management actions, like planting trees or creating green areas, this information is essential for assessing the condition of the urban ecosystem.

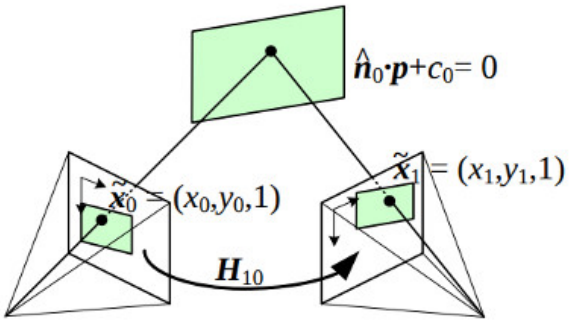


Figure 3.7: Figure showing the operation of the estimation of homography. Extracted from [8].

3.4.2 Estimation of Homography

Homography provides a mathematical way to describe how an image of a flat surface can change when viewed from different angles and distances. This is particularly relevant in applications where perspective shifts are significant, such as in aerial photography, robotics and urban vegetation monitoring.

Homography is based on projective geometry. It is represented by a 3x3 matrix that transforms the coordinates of points in one plane (or image) to corresponding points in another plane. The key idea is that if we have a set of points in one image and these same points in another image, we can use homography to relate these two sets of points.

To compute the homography matrix, at least four pairs of corresponding points are needed, assuming that the points are not collinear. This matrix, when multiplied with the coordinate vector of a point in one image, yields the corresponding point in the other image. A diagram of how the homography estimation works is shown in Figure 3.7.

The practical computation of homography in this project involves several steps. First, key points or features are detected in both images using a feature detection algorithm. The algorithm chosed was SIFT (Scale-Invariant Feature Transform). This algorithm is designed to find points that are unique and easily recognizable in different views of the same scene.

Once these key points are identified, the next step is to match these points between the two images. This is done using a feature descriptor which encode the appearance of the feature at each key point, allowing for comparisons between features in different images. We use a Fast Library for Approximate Nearest Neighbors (FLANN) based matcher to find potential correspondences between keypoints in two images. The matcher compares the descriptors of keypoints in both images and identifies potential matches.

After matching, a set of corresponding point pairs is established. Not all potential matches are valid correspondences. We apply a distance ratio test to filter out spurious matches and retain only those with a sufficiently low distance ratio. This ensures that the matches are robust and accurate. With the filtered keypoint correspondences, we proceed to estimate the homography matrix using the Random Sample Consensus (RANSAC) algorithm. RANSAC is a robust estimation technique that can handle out-

liers and noise in the data. The resulting homography matrix represents the geometric transformation between the two images.

Once we have the homography matrix, we can warp one of the images to align with the perspective of the other. This alignment facilitates the reidentification of vegetation elements by transforming one image into the same coordinate system as the other.

3.4.3 Application in Reidentification of Vegetation

With the images aligned, we can now match vegetation instances between the images more accurately. This matching process relies on the transformed keypoints, which correspond to the same vegetation elements in both images. The matched keypoints allow us to establish a correspondence between vegetation instances, even across different time intervals and frames.

Once an image is warped to align with the perspective of another, the next step is to compare the segmentations that were previously obtained. For this purpose, an Intersection over Union (IoU) matrix is used. The IoU matrix is a fundamental tool in evaluating the accuracy of segmentations in image processing. This matrix measures the degree of overlap between two segmentations, providing a quantitative value for the accuracy of the fit.

The IoU matrix is calculated by comparing the intersection area of the segmentations with their union area. Each element in the matrix represents the IoU value for a pair of segmentations, one from the original image and the other from the transformed image. Higher IoU values indicate a greater correspondence between the segmentations, suggesting that they represent the same vegetation element.

Subsequently, we employ the Hungarian algorithm to find optimal assignments between segmentations. This algorithm is a classic method in combinatorial optimization, allowing the solving of assignment problems in polynomial time. By applying the Hungarian algorithm, we aim to match each segmentation in one image with its most similar counterpart in the other image, based on the IoU values.

The result of this process is a set of correspondences between segmentations of the two images, which allows us to accurately reidentify vegetation elements. By comparing the segmentations of both images from the same perspective, we can observe changes in vegetation, such as plant growth, the emergence of new species, or the loss of vegetation due to urban or environmental factors.

To achieve a stronger vegetation reidentification, we have developed an approach that combines multiple frame reidentification of vegetation features. The methodology employed not only improves the precision of our findings but also guarantees increased resilience in the recognition and monitoring of vegetation occurrences over an extended period. We conducted reidentification by comparing the March frame with the May frame, as well as the March frame with its next frame and the May frame with its next frame. This multi-frame strategy allows for a deeper understanding and more accurate localization of vegetative elements, taking into account temporal variations and the subtle differences between successive images. In addition, it allows our previously obtained segmentations to be refined, since by taking into account several frames, those segmentations that were erroneous or unreliable are discarded, since for a segmentation to be taken into account it has to be present in several frames.

3.5 Vegetation Classification

3.5.1 Vegetation Classification in Urban Enviroments

Following the acquisition and refinement of geopositioned image segmentations captured in urban environments, the next step in our research is the classification of these segmentations. This phase is crucial for identifying and differentiating plant species, which enriches our catalog of urban vegetation.

3.5.2 Plant Species Classification Model

To undertake this task, we have chosen Pl@ntNet, an advanced and specialized tool for the identification of plant species. Pl@ntNet operates by analyzing images and provides a classification based on the visible characteristics of the plant. This application stands out for its extensive database and its ability to deliver precise and reliable results. It is also a collaborative project, in which users upload their images, which can be validated, allowing a continuous expansion of the database in a cooperative way. It works as follows:

1. **Image Capture and Submission:** Users capture images of plants and send them to Pl@ntNet through the app or website.

2. **Image Preprocessing:** Upon receiving an image, Pl@ntNet processes it to enhance quality and extract relevant features. This can include adjustments in lighting, cropping to focus on the plant, and noise reduction.

3. **Feature Extraction:** The system uses algorithms to identify and extract key features of the plant in the image, such as leaf shape, vein patterns, flower color, etc.

4. **Comparison with Database:** Pl@ntNet compares these features with its extensive database, which contains detailed information and images of thousands of plant species.

5. **Classification:** Pl@ntNet's machine learning model, trained with a large set of labeled image data, classifies the plant by comparing its features with those of known plants. Pl@ntNet doesn't provide clear information regarding the specifics of its model and architecture. Consequently, the exact details of how it operates and its underlying structure remain uncertain.

6. **Results and Feedback:** Pl@ntNet presents the user with the most probable identification results, typically as a set of species along with their respective confidence 'scores'. Users can confirm or reject these identifications, providing valuable feedback that is used to continuously improve the model.

In the context of this project, the Pl@ntNet Application Programming Interface (API) was used to automate the process of plant species classification. The API provides a programmable interface to the advanced image recognition capabilities of Pl@ntNet, allowing for seamless integration into our automated vegetation cataloging system.

The classification process begins by sending each of our refined segmentations to Pl@ntNet via its API. In return, we receive a 'score' or rating for each segmentation, indicating the likelihood of the segmentation belonging to the classes available in Pl@ntNet's database. This score helps us to determine more accurately the species of plant present in each image segment. In the work, we have focused on the genus of the existing vegetation. Consequently, when we receive the answer from Pl@ntnet, we do a filtering that consists in adding up all the scores of the species that belong to the same genus. Once we have all the species grouped by genus, we associate that element with the genus whose score was the highest. This way we get a more reliable result than if we were simply looking at the highest score and classifying by species instead of genus.

A significant improvement in our research process has been the increase in our capacity to make requests to Pl@ntNet's API. Thanks to a collaboration with the Pl@ntNet team, we have managed to expand the number of requests we can make, which has been crucial in handling the volume of data generated by our geopositioned images. This expansion allows us to classify a larger number of segmentations efficiently, improving the quality and extent of our urban vegetation catalog.

3.6 Density Determination

3.6.1 Measuring Vegetation Density in Urban Environments Using Geo-Positioned Images

In this section, we address the critical importance of measuring vegetation density in urban environments. Our initial methodology focused on using geo-positioned images captured from a moving vehicle. These images were processed using semantic segmentation algorithms to identify and classify vegetation, so they were ready to calculate the density in them. However, we faced an obstacle: the segmentation did not adequately consider the actual density and lushness of plant species. The segmentation tended to generalize vegetation without distinguishing between species of different densities, resulting in an inaccurate representation of the actual vegetative density in the images.

To overcome this challenge, we implemented the Otsu binarization technique. This technique significantly improved our ability to differentiate between vegetation and background in our images. Applying Otsu's mask to the segmentation cutouts allowed us to obtain a more accurate representation of vegetal density. This was crucial for our goal of accurately measuring urban vegetation density. Figure 3.8 shows a pair of images on the same location but different season. Thanks to the use of otsu binarization, it is possible to obtain more accurate results, since there is a better differentiation between the background and the vegetation part of the image. Without this processing, the parts of vegetation with different leafiness would give a similar result in terms of density. It is not possible to obtain such a precise differentiation between the background and the vegetation only using semantic semgmentation in this context.

Subsequently, we developed a method for calculating vegetation density. We selected a specific frame from the captured videos that showed the maximum presence of vegetation, using this as a reference for density calculations in other frames. Density was calculated as the proportion of vegetation pixels in each frame relative to the reference frame. This approach allowed us not only to measure the density of vegetation but also to make temporal and seasonal comparisons.

(a) March Original Image

(b) March Otsu Binarized Image

(c) May Original Image

(d) May Otsu Binarized Image

Figure 3.8: Figure showing the otsu filter applied to the segmentation image in the two seasons.

The density data thus obtained were represented on a map, providing a clear view of vegetation distribution in the study area. This map not only revealed vegetation density but also highlighted variations over time, allowing us to observe the effects of urban interventions and seasonal changes in vegetation. Although our study has limitations, such as dependence on image quality and lighting conditions, it offers a solid foundation for future research and applications in urban environmental management.

3.7 Final Applications

3.7.1 Introduction to Final Applications

This thesis has so far examined the difficulties and complications involved in using geo-positioned images for vegetation monitoring and cataloging in urban contexts. As a result of our investigation, we have created two useful applications that are firmly based on the conclusions and techniques presented in earlier chapters. These applications serve as practical examples of how our research is applied in the real world and provide useful resources for sustainable urban planning and environmental management.

3.7.2 Catalog of Vegetation Elements

The first application we have developed focuses on the advanced cataloging of vegetation. Using the semantic segmentation detailed previously, this application distinguishes vegetation from other urban elements. Through advanced processing of images captured by geo-positioned cameras on moving vehicles, we achieve the identification

of each vegetation element. Once the vegetation is segmented, the application employs re-identification techniques to associate vegetation elements across different temporal and visual frames. This is achieved by assigning a unique identifier to each vegetation element, allowing for detailed tracking and effective monitoring over time. This process not only facilitates cataloging but also enhances the ability to monitor and understand growth trends, responses to urban interventions, and seasonal changes. Then, the classification of plant species is performed based on the pattern recognition algorithms commented previously. This allows for a detailed identification of the vegetation and also contributes to the creation of a comprehensive catalog, which is of great value for biodiversity and conservation studies.

3.7.3 Vegetation Density Estimation

The second application developed aims to determine vegetation density based on temporal and spatial variables. This tool is capable of analyzing how vegetation behaves according to its location and the time of year, providing valuable insights into the most suitable species for different areas of the campus and how these maintain their conditions throughout the year. This analysis of vegetation density allows, for example, to identify areas where vegetation is sparse, which in turn can indicate a lack of irrigation or care. These data are crucial for urban planning and environmental management efforts, as they enable decision-makers to make informed choices to improve vegetation coverage and, thereby, urban life quality.

3.8 System Integration

In the course of this research, one of the most significant undertakings has been the integration of various complex systems and models. Each component, from data acquisition to density determination, involves algorithms and processing techniques that are individually complex. The challenge has been to integrate these diverse components into a cohesive system.

1. **Interoperability of Different Models:** The core challenge in system integration lay in the interoperability of different models. Each model, whether it's for semantic segmentation, image reidentification, vegetation classification, or density determination, operates on distinct principles and produces outputs in unique formats. Bridging these differences to ensure a smooth flow of data and processes was critical.

2. **Data Flow and Processing Pipeline:** A systematic approach was adopted to manage the data flow and processing pipeline. The integration started with the establishment of a standardized data format that could be utilized across different models. This standardization ensured that the output of one process could be effectively used as input for the next without extensive reformatting or adaptation.

3. **Synchronization of Temporal and Spatial Data:** Given the project's reliance on temporal and spatial data, synchronizing these aspects across different

models was crucial. Special attention was paid to maintaining consistency in geo-positioning and temporal markers, ensuring that data processed through different models could be accurately correlated and analyzed.

4. **Scalability and Flexibility of the System:** The system was designed to be scalable and flexible, allowing for future enhancements and adaptations. This aspect was particularly important given the evolving nature of the research and the potential for incorporating new technologies or methodologies.

The integration of various complex systems in this research represented a significant challenge due to the variability and intricacies of the different models. However, a cohesive system was developed, that can integrate all the models and technologies for the development of the different applications.

Chapter 4

Evaluation

4.1 Introduction

In the previous chapter, we detailed the development of the system. This chapter will present the experimental enviroment and elaborate on the various stages involved in the implementation of the system. We will then proceed to conduct tests to evaluate the system's performance. The chapter will conclude with a thorough analysis of the results obtained from these tests.

4.2 Experimental Enviroment

This section briefly describes the environments used for the implementation of the work.

Conda: In this project, we have employed Conda as our package manager and virtual environment manager. Conda is a crucial tool for the efficient management of multiple packages and libraries required in data science and machine learning projects. Its ability to create isolated virtual environments allows for the installation and management of different package versions and dependencies, avoiding conflicts between them. This feature is essential in our project, as we integrate a variety of models and applications, each with its own library requirements and version needs. Conda facilitates the management of these complexities, ensuring consistency and reproducibility of our development and execution environment.

PyCharm: For code development, we have chosen PyCharm as our integrated development environment (IDE). Developed by JetBrains, PyCharm is a Python-specific IDE, known for its efficiency and support capabilities for professional development. It offers advanced functionalities such as code analysis, graphical debugging, integration with version control systems, and an intuitive interface for project and virtual environment management. Its use in our project contributes to greater efficiency in code writing, debugging, and maintenance, which are crucial for the success of complex projects like ours.

NVIDIA GeForce GTX 1080 Ti: The project demands considerable computational capacity, especially for image processing and the execution of advanced machine

learning algorithms. Therefore, we have opted to use the NVIDIA GeForce GTX 1080 Ti graphics card. It offers exceptional performance in parallel processing, which is fundamental for accelerating the intensive calculations involved in semantic segmentation, plant species classification, and estimation of vegetation density. Its ability to handle large volumes of data and execute complex algorithms quickly is vital for our project, allowing us to efficiently and effectively process image sequences.

4.3 First Application: Catalog of Vegetation Elements

As previously discussed, the first application will consist of creating a catalog of the vegetation present on the university campus. To achieve this, the explained system will be used, where a series of techniques have been applied to successfully catalog the species of vegetation using videos obtained with a camera mounted on a car. To better understand the results, they have been differentiated according to the processes that have led to their identification.

4.3.1 Semantic Segmentation

Setup The first step has been the acquisition of semantic segmentation, as in order to determine the types of vegetation present in the images, the first task is to differentiate them from other elements. Using the model described in Section 3.3, this task has been successfully accomplished. A study of the model parameters was conducted to determine the best way to use the model. The parameters used can be found on 4.1. The model, referred to as sam, is the segmentation model used. The points per side parameter, set to 32, specifies the number of points on each side used for segmentation, indicating a detailed approach to defining the segmentation frame. The pred IoU threshold, set at 0.86, is the threshold for the Intersection Over Union (IOU) metric; masks exceeding this IOU value are considered successful predictions. The stability score threshold is set at 0.92, indicating that only masks with a stability score above this threshold are considered reliable. Crop layers set to 0 suggests no additional neural network layers are used for cropping during mask generation, keeping to the default configuration. The crop points downscale factor at 2 reduces the resolution of the segmentation points, which can speed up the process. The min mask region area parameter, with a value of 100, ensures that only mask regions larger than this area (in square pixels) are considered, ignoring smaller regions. Lastly, the output mode set to 'coco_rle' specifies that the masks will be generated in the Run-Length Encoding format, commonly used in the COCO dataset for image segmentation. These settings collectively balance the segmentation process's accuracy, processing speed, and memory usage.

Results Output Once the parameters were determined, it was possible to obtain the segmentation results. The outcome comprised two files: a JSON file containing all the segmentation information and a binarized image representing the obtained mask. Thanks to the model and parameters used, it is possible to achieve segmentation by elements, meaning there isn't a single mask for all the vegetation in the image, but rather different elements within the mask can be differentiated. An example of the

Parameter	Value
Model	SAM
Points per side	32
Pred IoU threshold	0.86
Stability score threshold	0.92
Crop layers	0
Crop points downscale factor	2
Min mask region area	100
Output code	coco_rle

Table 4.1: Paramters used for the Semantic Segmentation model.

(a) Original Image. (b) Masked Image

Figure 4.1: Figure showing the mask obtained from the segmentation algorithm.

masked obtained from the original image, can be observed on Figure 4.1. It can be observed that each element is given a value on the grayscale to differentiate them from one another. Based on a visual evaluation, it can be concluded that the segmentation for this case is quite accurate, as it achieves a separation of vegetation with high precision. Depending on the specific frame, the results may be slightly better or worse, but in general, a very appropriate segmentation is obtained. Therefore, we conclude that this model is suitable for the task at hand.

4.3.2 Image Reidentification

Setup Subsequently, species re-identification has been carried out. For this purpose, the homography estimation method explained in Section 3.4 was used. Through this, it is possible to view one image from the perspective of another, thus overlaying the segmentations of both images. Once overlaid, the Intersection over Union (IoU) metric is used to determine those segmentations that belong to the same element in both images. An anaylsis of the more adecuate value for the IoU threshold revealed that that a value of 0.4 was the better fit for this application.

Results Output After obtaining the matchings, it is possible to re-identify vegetation elements both in different frames of the same video and in different videos when their frames have been associated.

Table 4.2 presents the quantitative outcome of the system across different stages of analysis. The table shows the number of vegetation elements identified in the months

Stage	Number of Vegetation Elements
March	843.714
May	852.039
Reidentificated	149.589

Table 4.2: Number of vegetation elements obtained in the diferent analysis made for the application.

of March and May, as well as the count of reidentified elements throughout the study period.

The data reveals an increase in the number of vegetation elements from March to May, with counts of 843,714 and 852,039, respectively. This increment may be attributed to several factors. One plausible explanation is the enhanced detectability of vegetation elements in May due to increased foliage density, which makes the elements more discernible to our image processing algorithms. Additionally, the disparity in numbers could also stem from differences in recording conditions, such as lighting or weather variations, which may affect the visibility and consequently the detection of vegetation.

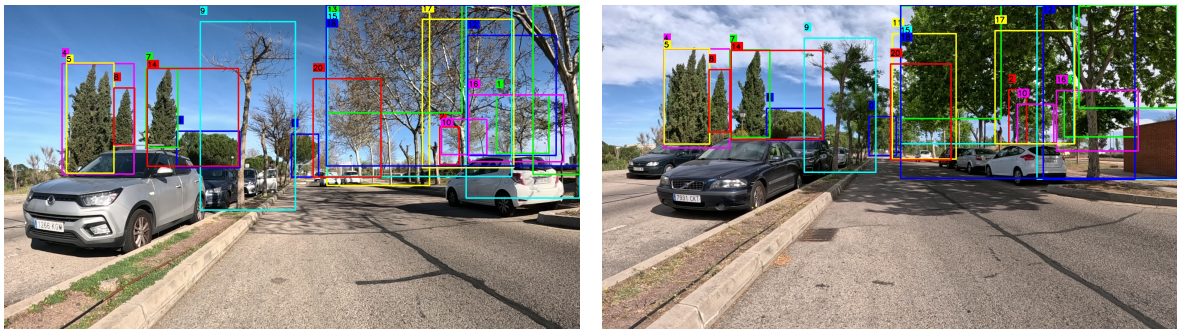
A significant observation from the table is the discrepancy between the total number of elements identified in the individual months and the considerably lower number of reidentified elements, which is 149,589. This marked reduction is attributable to the criteria set for reidentification. As previously detailed, reidentification is obtained upon the detection of vegetation elements in subsequent frames within the same video and across equivalent frames in the other month's video. This process filters out transient or less robust elements, ensuring that only those with a strong presence in both temporal and visual frames are cataloged. This mechanism serves to eliminate redundancies and reinforces the reliability of the cataloging process by focusing on persistent and identifiable vegetation.

Figure 4.2 shows the relationship between the elements identified in two associated frames. It can be observed how the elements found in March are re-identified in May. It can be seen that the re-identification is done correctly. It is noticeable that certain elements overlap others, but it has been decided to keep all elements and not remove redundancies as we consider it important to preserve as much information as possible to achieve greater robustness of the system.

In addition, our reidentification system appears to obtain correct results when applied to both evergreen and deciduous species. Evergreens, such as the one depicted in Figure 4.3, present little change between the images captured in March and May. This consistency likely aids the system's algorithms in maintaining high reidentification rates for these species, as the Intersection over Union (IoU) metric confirms their continued presence.

Deciduous species, illustrated in Figure 4.4, exhibit more pronounced seasonal transformation. These species, which may be fully leafed in one season and completely bare in another, represent a more substantial challenge for the reidentification process. Yet, the system seems capable of overcoming this challenge by leveraging the reidentification model developed. The ability of the system to reidentify vegetation elements despite the seasonal changes suggests that the underlying algorithms can discern between the stable aspects of the vegetation's structure.

Thanks to this model, we can obtain information about the different vegetation



(a) March Reidentificated Elements (b) May Reidentificated Elements

Figure 4.2: Figure showing the reidentification obtained between March and May.

(a) March (b) May

Figure 4.3: Figure showing an example of the reidentification obtained between evergreen elements.

elements on the campus over time, allowing for a more reliable catalog. Moreover, it enables the conduct of various types of studies, focusing on the changes that the elements exhibit depending on the season or the changes observed based on their location.

4.3.3 Image Clasification

Setup As previously mentioned in Section 3.5, the Pl@ntnet model will be used for classification. To do this, we will utilize the image segments obtained through image segmentation, subsequently filtered with re-identification to enhance their robustness, using only the elements obtained after reidentification for the classification. Once we have these segments, they are passed to the Pl@ntnet API, which returns the scores for all possible species for that segment. Aiming for greater reliability, our focus has been on the genus of the vegetation. This approach not only considers the highest value according to the species but also contrasts it with various possible classes and adds the scores of all classes within the same genus.

Results Output With this, we can now determine the genus associated with each segment and also know its location thanks to the geolocation information of the frame.

(a) March (b) May

Figure 4.4: Figure showing an example of the reidentification obtained between deciduous elements.

Element	Class	Score
Element 7	Cupressus	0.52
Element 16	Platanus	0.72
Element 9	Gravillea	0.11

Table 4.3: Clasification scores obtained for the elements represented in Figure 4.6

In Figure 4.5, all vegetation elements found for a specific frame can be observed. Additionally, in Figures 4.6a, 4.6b and 4.6c , some of these elements are represented individually as segments. These segments are what are passed to the API, which then returns a score for the class associated with each image.

Examining Table 4.3, it can observed the classes and the scores obtained for each of the elements individually showcased in the Figure 4.6. This provides an insight into the model's behavior. For elements 7 and 16, a high score is obtained, indicating good reliability for these segments. We can also observe how the model performs with different types of images, in this case, the first being the tree canopy and the second the trunk, demonstrating the model's versatility. However, for element 9, the score is significantly lower, suggesting less reliability for this particular segment. This indicates that while the model generally achieves good results in classification, it is not infallible for all elements found in the frames. This inconsistency largely stems from the methodology of extracting segments from a whole image. The application's core functionality is predicated on analyzing distinct, individual images of vegetation, not crops of a bigger image. Therefore, in our work, various techniques have been implemented to enhance reliability, such as the reidentification of vegetation. This approach ensures that this result is not the only one available for that vegetation element. Consequently, if a more reliable result is obtained in another instance, we can prefer that one. Thus, we have treated the robustness of the system as one of its principal aspects.

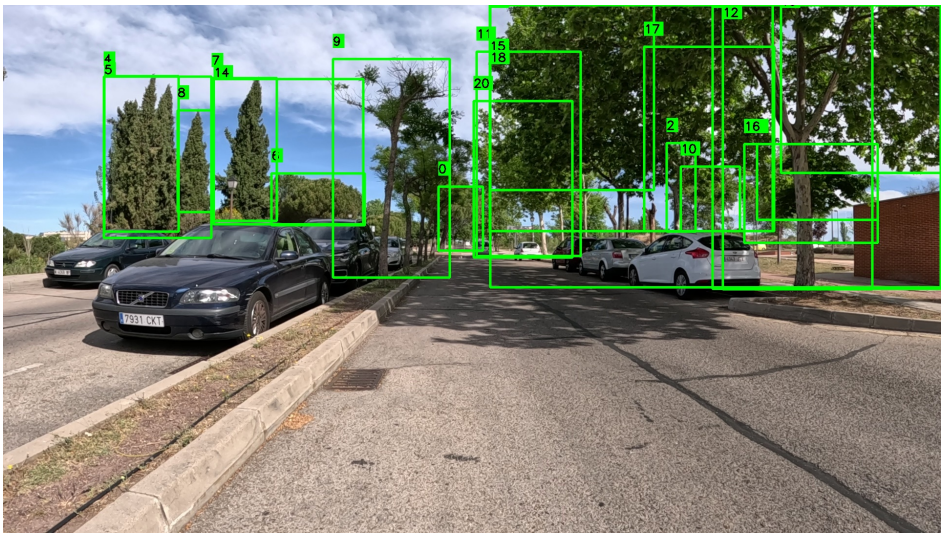


Figure 4.5: Figure showing the elements to be classified.

(a) Element 7

(b) Element 16

(c) Element 9

Figure 4.6: Figure showing 3 elemtents extracted from 4.5.

4.3.4 Overall Application Outcome

After obtaining the intermediate results from our initial application, we have observed how it performs for the main tasks developed. Now, we are poised to construct the final application, aimed at cataloging the existing vegetation on the campus. This will be built upon the various stages previously completed. Following the segmentation, reidentification, and classification, we now have all the necessary results for creating the catalog. To do this, we will gather all identified elements and group (thanks to reidentification) those belonging to the same entity. Each vegetation element on the campus will be assigned a unique ID, and this ID will be attributed to elements reidentified as the same. This will enable us to create a catalog where each element is listed with its ID, location, and a series of images corresponding to the segments associated with that ID, serving as a visual verification of the system's results. Once all elements are grouped under their respective IDs, we determine their class. This is done by accumulating the scores of all elements under that ID and retaining the class with the maximum value. This approach allows us to create a hierarchical catalog that can be easily visualized. An example of how this catalog functions can be found in the

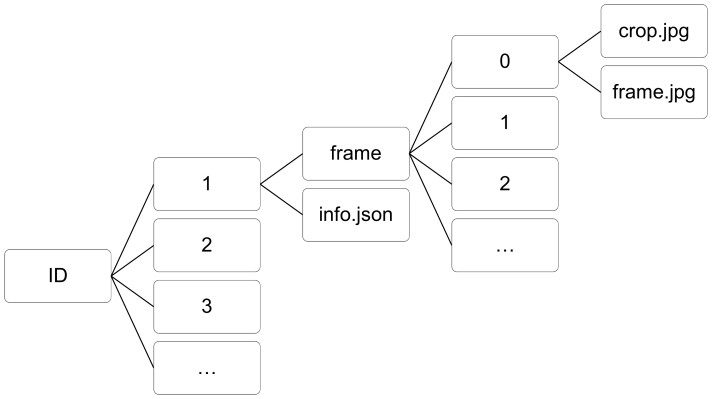


Figure 4.7: Figure showing the hierarchy of the catalog.

following github. This hierarchy consists of a primary directory where the IDs of all found elements are located. Within each ID, there are all the segments representing that ID, as well as the frame from which they have been extracted and a file with information about the class, location, and score of each crop. Figure 4.7 shows the jararchy of the catalog.

Figure 4.8 illustrates the 'crop.jpg' and 'frame.jpg' files for two elements with the same ID. The use of the crop and whole image allow for the observation of the element both individually and within the context of the entire image, facilitating its realistic localization.

The creation of this catalog provides a useful tool for accessing information about the campus vegetation. It also facilitates future studies due to its comprehensive data. Furthermore, as a dynamic and versatile tool, it can be modified in the future to add more information or to apply other techniques aimed at acquiring new elements.

4.3.5 Application Results

As previously discussed, we have collaborated with the university's Department of Botany to interpret the results from our vegetation cataloging application. They have provided us with a catalog of vegetation elements for a specific segment of our study area, particularly the initial part of the route covered by our videos. The route provided by them can be apreciated in Figure 4.9. This collaboration has been crucial for validating the application's output against an authoritative source of botanical information.

To facilitate a more granular analysis, the provided part of the route has been subdivided into five distinct segments. For each segment, the department has supplied a breakdown of the existing vegetation elements and their respective proportions along the route. This segmentation allows for an understanding of the distribution and prevalence of various plant species within the route.

It should be noted, however, that the probabilities assigned by the application do not add up to one. This is because the application may detect erroneous or duplicate elements, meaning not all classified outputs correspond to the main genus, which accounts for the discrepancy in percentages. Moreover, it's important to highlight that since the analysis is based on elements captured by a moving camera, variables such as

(a) March Frame image with selected crop

(b) March

(c) March Frame image with selected crop

(d) May

Figure 4.8: Figure showing an example of images found in the catalog with a given ID.

the vehicle's speed, position, and angle can result in duplications that alter the final result. Furthermore, it is necessary to consider that the accuracy achieved by the image classifier is not always highly reliable. This is primarily due to the use of excerpts extracted from a global image, as the application is designed to yield results on a specific image of vegetation. Therefore, it is important to acknowledge that these results do not represent conclusive findings, but rather provide an initial insight into the behavior of the system. For this reason, future endeavors will maintain collaboration with the Department of Botany to conduct further analyses and enhance the presentation and validation of the results. Consequently, this analysis has been conducted to observe the general behavior of the system, and as the Department of Botany provides more information for analysis, we will be able to yield results that are more representative.

The initial segment of the route encompasses the intersection of "C/ Marie Curie" with "C/ Faraday". The results from this specific section can be seen in Table 4.4. The table is organized to show the percentages of various botanical genus identified by both the botanical experts and our application. The second segment of the route extends from "C/ Faraday" to "C/ Newton". The results for this segment are presented in Table 4.5. The third segment of the route corresponds from "C/ Newton" to "C/ Francisco Tomás". The findings for this portion are detailed in Table 4.6. For the fourth segment, which runs from "C/ Francisco Tomás" to "C/ Nicolás Cabrera", the obtained results are shown in Table 4.7. The fifth and final segment encompasses Nicolás Cabrera Street. The results for this section are represented in Table 4.8. All these tables reveal

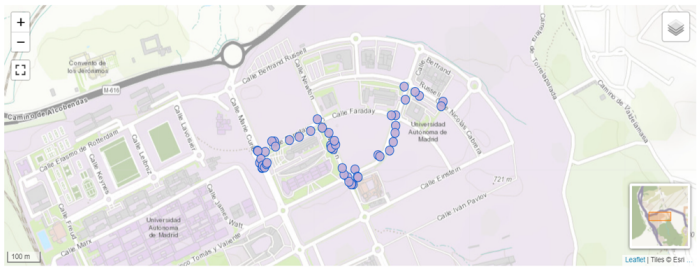


Figure 4.9: Figure showing the route with the vegetation elements provided by the Botanic Department.

Class % for the first segement			
Botanic Genus	**Botanic %**	**Application Genus**	**Aplication %**
Pinus	0.54	Pinus	0.38
Malus	0.17	Platanus	0.21
Platanus	0.08	Cupressus	0.08
Sophora	0.08	Ulmus	0.04
Cupressus	0.04	Quercus	0.04
Quercus	0.04	Acer	0.03
Retama	0.04	Populus	0.02

Table 4.4: Classes and porcentages obtained for the botanic catalog and application scores obtained for the first segment of the route.

that while the application successfully detects several major classes of vegetation, there are notable differences in the exact percentages and some genus identifications when compared to the botanical catalog. This points to the need for continued refinement of the application's algorithms to enhance its accuracy and reliability in urban vegetation monitoring.

Upon examining the results across the various segments of the route, it is evident that the model identifies the some vegetation classes. However, the discrepancies in the percentages of identified species and the occasional misclassification suggest that there is a need for further refinement of the model.

Continued collaboration with the Department of Botany is essential in this refinement process. Their expert insights and additional information will be needed in enhancing the accuracy of the model. By providing more detailed botanical data, they can assist in fine-tuning the application's classification algorithms. Moreover, allowing the botanical experts to review the catalog created by the application can lead to more representative results. Instead of solely relying on the existing percentages along the route, this collaboration would enable a more comprehensive and accurate representation of urban vegetation. Apotential enhancement for the collaboration would be the acquisition of vegetation elements based on their GPS coordinates. This would allow the use of data associated with each element captured by the application to verify if there is any corresponding element at those coordinates. Additionally, compiling a list of all species and genera observable along the route and informing the classification model that only those types of elements are present could be beneficial. Such improvements in the information provided by the Department of Botany could aid in refining the results' accuracy and enhancing the overall performance of the application.

Class % for the second segement			
Botanic Genus	**Botanic %**	**Application Genus**	**Aplication %**
Celtis	0.38	Pinus	0.19
Acer	0.25	Desconocido	0.14
Ulmus	0.19	Ulmus	0.05
Pinus	0.13	Quercus	0.05
Populus	0.06	Acer	0.04

Table 4.5: Classes and porcentages obtained for the botanic catalog and application scores obtained for the second segment of the route.

Class % for the third segement			
Botanic Genus	**Botanic %**	**Application Genus**	**Aplication %**
Platanus	0.57	Prunus	0.12
Ligustrum	0.20	Desconocido	0.11
Prunus	0.13	Hydrocotyle	0.09
Robinia	0.08	Acer	0.08
Pinus	0.03	Fallopia	0.08

Table 4.6: Classes and porcentages obtained for the botanic catalog and application scores obtained for the third segment of the route.

Class % for the fourth segement			
Botanic Genus	**Botanic %**	**Application Genus**	**Aplication %**
Acer	0.85	Acer	0.16
Albizia	0.09	Albizia	0.09
Retama	0.03	Desconocido	0.05
Ulmus	0.02	Pinus	0.04
Cupressus	0.02	Genista	0.04
Platanus	0.02	Tilia	0.04
Picea	0.02	Hydrocotyle	0.04

Table 4.7: Classes and porcentages obtained for the botanic catalog and application scores obtained for the fourth segment of the route.

Class % for the fifth segement			
Botanic Genus	**Botanic %**	**Application Genus**	**Aplication %**
Albizia	0.55	Albizia	0.24
Celtis	0.45	Ulmus	0.13

Table 4.8: Classes and porcentages obtained for the botanic catalog and application scores obtained for the fifth segment of the route.

4.4 Second Application: Vegetation Density Estimation

The second application developed provides a detailed analysis of vegetation density using temporal and spatial data. Through this tool, we gain a better understanding of how vegetation varies based on its location and the time of the year. This understanding is essential for environmental management and urban planning, allowing for the identification of the most suitable species for different areas of the campus and how they maintain their conditions throughout the year.

4.4.1 Setup

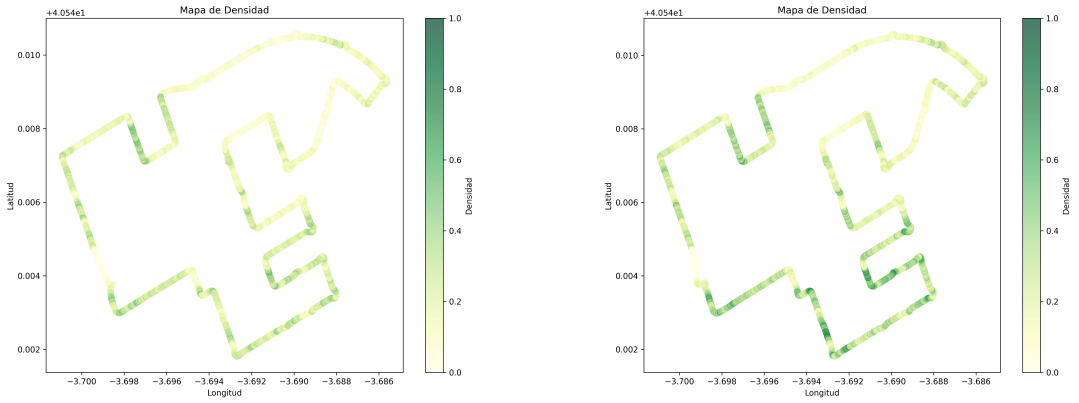
To develop the application, the Otsu binarization technique was implemented. This technique enhanced the differentiation between vegetation and background in the images. This commenced with the geo-positioned images, which were loaded into our framework. Each image was first converted to grayscale, a necessary step for the subsequent application of the Otsu method. The grayscale conversion simplifies the image data, reducing it to a format that emphasizes luminance while discarding color information, thus making the Otsu algorithm more effective. Following this, the core of the Otsu technique was applied. This involved calculating an optimal threshold value, a process automated by the Otsu algorithm. The algorithm determines this threshold by analyzing the grayscale image and selecting a value that best separates the foreground (vegetation) from the background. This threshold defines the binary image - pixels above this threshold are considered vegetation, while those below are classified as background.

4.4.2 Results Output

Figure 4.10 shows a comparation between the two videos, representing the density in March and May. In the graph you can see the route, obtained from the gps coordinates of the video frames, and you can see in the bar on the right side of both maps the relationship between color intensity and density. The more intense the green color, the higher the density at that coordinate. Although it may not seem a very clear difference, it can be seen how the map of the Figure 4.10b, representing May, has in general higher values than the map of the Figure 4.10a, representing March. Furthermore, if more detailed information about the density in a specific location is required, access to this information is facilitated through the geoposition data we have for the different frames.

This analysis of vegetation density allows us to identify areas where vegetation is sparse, which in turn can indicate a lack of irrigation or care. These data are crucial for urban planning and environmental management efforts, as they enable decision-makers to make informed choices to improve vegetation coverage and, thereby, urban life quality.

In addition to the density maps, statistical metrics were employed to further explain the vegetation density observed in the videos. These metrics included the mean, median, standard deviation, and the 95th percentile. The mean provided an average density value, offering an overall sense of vegetation presence, while the median offered a middle-ground perspective, unaffected by extreme values. The max value shows the biggest density for the two seasons. The the 95th percentile gave insight into the up-



(a) March density map (b) May density map

Figure 4.10: Figure showing the vegetation density for the two routes in different seasons.

Density Metrics Results		
Metric	**March**	**May**
Mean	0.089684	0.118288
Median	0.059020	0.083523
Max	0.701997	1.000000
95th percentile	0.286425	0.368832
STD	0.094797	0.120158

Table 4.9: Statistical results obtained for the vegetation density of the diferent seasons.

per extremes of vegetation density, highlighting areas of particularly dense vegetation. Lastly, standard deviation shed light on the variability of vegetation density within the images, indicating the extent of density fluctuation. Table 4.9 shows the metrics obtained.

The analysis of the vegetation density metrics provides evidence of seasonal variation in urban vegetation density. One observation is the higher vegetation density in May compared to March. This is evidenced by the mean density value of 0.118288 in May, which is higher than the March mean of 0.089684. This increase suggests a denser vegetative cover in May, likely due to the growth and flourishing of vegetation during the spring season. The median values further corroborate this seasonal change. In May, the median vegetation density is 0.083523, compared to March's 0.059020, supporting the observation of generally denser vegetation in May. The maximum and 95th percentile values in May also are higher than those in March, with May reaching a maximum density of 1.000000 compared to March's 0.701997. Similarly, the 95th percentile in May is 0.368832, higher than March's 0.286425. These metrics indicate not only an overall increase in vegetation density but also highlight that the densest areas of vegetation are significantly more pronounced in May. The standard deviation in May (0.120158) is higher than in March (0.094797), suggesting a greater variability in vegetation density during the later month. This could be attributed to the varied growth patterns and blooming of different plant species in the spring.

The graph presented in Figure 4.11, provides a visual comparison of vegetation

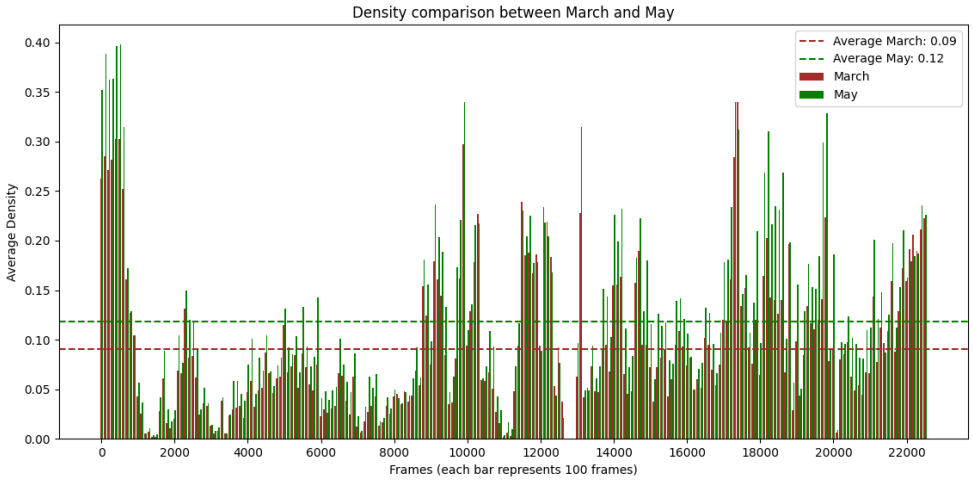


Figure 4.11: Figure showing the density comparation between the two months studied.

density across the two distinct months, March and May. The vertical axis represents the average density of vegetation, while the horizontal axis shows the total number of frames captured in the videos, grouped in sets of 100 frames per bar. This grouping method is done as the number of frames is too big to show a bar for each of the frames and offers a clearer visualization of density trends over the sequential frames.

Upon examining the graph, it can be seen that the vegetation density in May is, on average, higher than that in March. This general trend can be observed in the height of the bars, where May consistently shows greater density values. The horizontal dashed lines represent the average density for each month, with May displaying a higher mean density of 0.12 compared to March's 0.09. These averages indicate the overall density of vegetation and confirm the visual data presented by the individual bars.

However, it is important to note that in certain sections of the graph, March exhibits higher density values than May. This could be attributed to the presence of specific plant species that do not exhibit significant variability between these months. The consistency in density for these species across seasons may result in sporadic instances where March's density surpasses that of May.

The disparities in density between March and May are likely influenced by seasonal growth patterns. In temperate climates, May typically marks a period of growth, which can explain the overall higher density values. The variations in density also provide insights into the phenological stages of different species and their responses to the changing seasons.

The graph depicted in Figure 4.12, presents the variation in vegetation density between March and May. The vertical axis quantifies the variation as a difference between the densities of the two months, with positive values indicating higher density in March and negative values indicating higher density in May. The horizontal axis groups the total frames from the videos, with each bar representing an aggregate of 100 frames, the same technique as in the previous graph.

The horizontal dashed line across the graph indicates the average difference in density, which favors May, as evidenced by its position below the zero line. This average difference of -0.03 suggests that, overall, vegetation density is greater in May than in March.

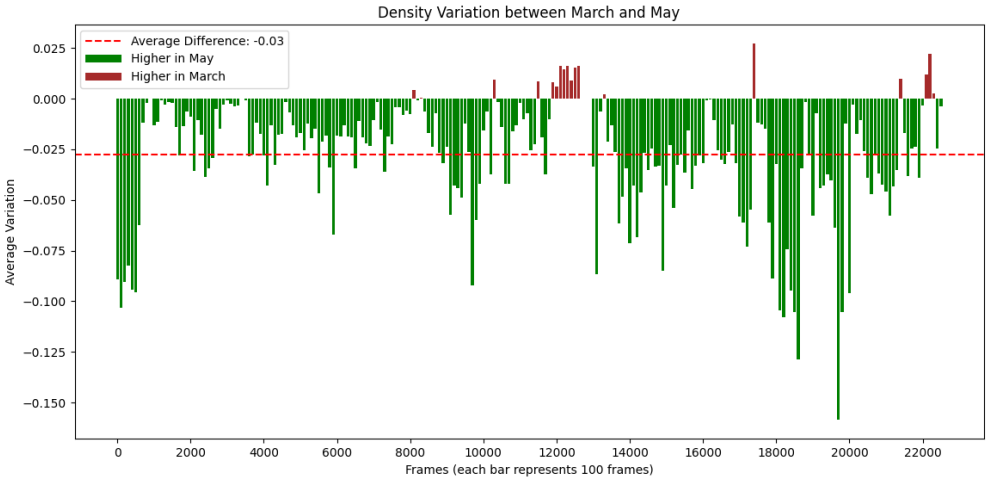


Figure 4.12: Figure showing the density variation between the two months studied.

A detailed examination of the graph reveals that the majority of the data points display negative values, confirming that in most frames, May exhibits higher vegetation density compared to March. However, the variation is not markedly high, indicating that while May generally has more dense vegetation, the difference from March is not extreme.

This pattern of variation is consistent with seasonal growth expectations, where vegetation in May is typically more abundant following the spring growth period. The predominance of negative values across the graph supports this seasonal trend. The presence of positive values at certain points suggests that there are instances where March's density exceeds that of May, which may be due to specific environmental conditions or the phenological state of certain plant species during the earlier month.

In summary, the results obtained with the second application highlight the importance of monitoring and managing vegetation in urban environments. The data collected offer a clear view of how vegetation evolves over time, providing a solid foundation for future interventions and improvements in the environmental management of the campus.

Chapter 5

Conclusions and future work

5.1 Conclusions

In this Master's thesis, a system has been developed for the cataloging and monitoring of vegetation in urban environments.

This has allowed us to draw the following conclusions, which address the objectives set at the beginning of the project.

- The study and analysis of the related work has been extremely useful, as it has allowed us to identify the strengths and limitations of current methods. This analysis enabled us to establish a solid foundation for our development, incorporating the most advanced techniques and overcoming previously identified challenges. Thus, we have contributed to expanding knowledge in this field, proposing innovative and efficient solutions for image processing of vegetation in urban environments.

- The system we designed and developed successfully integrates semantic segmentation, reidentification, plant species classification, and vegetation density determination. Each component was developed and optimized to function in unison, resulting in a system capable of accurately and effectively cataloging vegetation. This system not only demonstrates new applications for environmental management in urban areas.

- Although not completely proved, the intitial exploration of the system pipeline shows promising capabilities for an application in a real-world context. The analysis of the first application has enabled an understanding of how the system behaves in cataloging and monitoring real data, yielding results that lay the foundation for the creation of a vegetation catalog on the university campus. Additionally, it provides insight into the strengths and weaknesses of current models for such tasks. The second application has provided important information about vegetation density, which is very useful for conducting studies on urban environment vegetation.

All these conclusions lead us to the next section, which proposes possible future work that can be undertaken to continue the study conducted in this thesis.

5.2 Future work

Following the completion of this research, the results have led to a series of future proposals that could be implemented in the studied system for its enhancement.

An important prospect for future work is to continue collaborating with botanical departments. Such collaborations can provide several benefits. Firstly, botanical experts can offer deeper insights into the types of vegetation present in urban environments. Their expertise would be invaluable in verifying and refining the catalog created in this study. Secondly, the collaboration could lead to more accurate and reliable results, as the botanical department's knowledge would help fine-tune the classification and identification processes.

Another continuation for the study is to include video sequences captured at other temporal moments. This expansion of the temporal database will allow for the analysis of vegetation behavior throughout the different seasons. This extended temporal perspective is crucial for understanding the seasonal dynamics of urban vegetation, such as growth patterns, responses to urban interventions, and inherent seasonal changes.

Another possible area for future research is the experimentation with new algorithms and technologies. Specifically, the exploration of advanced algorithms for depth estimation could allow for more precise localization of vegetation in the urban environment. This improvement in localization accuracy would facilitate better urban landscape planning and management, allowing for more specific and effective interventions.

Finally, developing new applications that use the data collected and analyzed in this study would be interesting. For instance, monitoring the lack of irrigation or adaptability of certain urban areas through continuous analysis of vegetation density and condition. Additionally, integrating data on temperature and air quality in certain urban areas, combined with the vegetation catalog, could provide valuable insights into which types of vegetation are most suitable for improving sustainability and well-being in urban environments.

Bibliography

[1] H. Gu, Y. Wang, S. Hong, and G. Gui, "Blind channel identification aided generalized automatic modulation recognition based on deep learning," *IEEE Access*, vol. 7, pp. 110722–110729, 2019. ix, 6

[2] Y. Liu, Y. Zhang, Y. Wang, F. Hou, J. Yuan, J. Tian, Y. Zhang, Z. Shi, J. Fan, and Z. He, "A survey of visual transformers," *IEEE Transactions on Neural Networks and Learning Systems*, 2023. ix, 5, 7

[3] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016. ix, 8

[4] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023. ix, 9, 10, 23, 24

[5] A. Acharya, "Guide to image segmentation in computer vision: Best practices." https://encord.com/blog/image-segmentation-for-computer-vision-best-practice-guide/, November 2022. Accessed: 2023-11-22. ix, 10

[6] Y.-G. Han, S.-H. Jung, and O. Kwon, "How to utilize vegetation survey using drone image and image analysis software," *Journal of Ecology and Environment*, vol. 41, pp. 1–6, 2017. ix, 15, 16

[7] J. Chen, Z. Yang, and L. Zhang, "Semantic segment anything." https://github.com/fudan-zvg/Semantic-Segment-Anything, 2023. ix, 24, 25, 26

[8] OpenCV, "Homography examples using opencv." https://docs.opencv.org/4.x/d9/dab/tutorial_homography.html, 2023. Accessed: 16/11/2023. ix, 27

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015. 5

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012. 5

[11] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing street-level urban greenery using google street view and a modified green view index," *Urban Forestry & Urban Greening*, vol. 14, no. 3, pp. 675–685, 2015. 9

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015. 10

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015. 10

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017. 10

[15] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016. 10

[16] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2020. 11

[17] S. M. Azimi, P. Fischer, M. Körner, and P. Reinartz, "Aerial lanenet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2920–2938, 2018. 11

[18] S. Shim, J. Kim, G.-C. Cho, and S.-W. Lee, "Multiscale and adversarial learning-based semi-supervised semantic segmentation approach for crack detection in concrete structures," *IEEE Access*, vol. 8, pp. 170939–170950, 2020. 11

[19] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "Transreid: Transformer-based object re-identification," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 15013–15022, 2021. 12

[20] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Deep image homography estimation," *arXiv preprint arXiv:1606.03798*, 2016. 12

[21] L. Picek, M. Šulc, Y. Patel, and J. Matas, "Plant recognition by ai: Deep neural nets, transformers, and knn in deep embeddings," *Frontiers in Plant Science*, p. 2788, 2022. 13

[22] H. Goëau, P. Bonnet, A. Joly, V. Bakić, J. Barbe, I. Yahiaoui, S. Selmi, J. Carré, D. Barthélémy, N. Boujemaa, *et al.*, "Pl@ ntnet mobile app," in *Proceedings of the 21st ACM international conference on Multimedia*, pp. 423–424, 2013. 14, 17

[23] M. Šulc and J. Matas, "Fine-grained recognition of plants from images," *Plant Methods*, vol. 13, pp. 1–14, 2017. 14

[24] A. Abdollahi and B. Pradhan, "Urban vegetation mapping from aerial imagery using explainable ai (xai)," *Sensors*, vol. 21, no. 14, p. 4738, 2021. 14

[25] B. Ayhan, C. Kwan, B. Budavari, L. Kwan, Y. Lu, D. Perez, J. Li, D. Skarlatos, and M. Vlachos, "Vegetation detection using deep learning and conventional methods," *Remote Sensing*, vol. 12, no. 15, p. 2502, 2020. 14

[26] Y. Chen, G. Sanesi, X. Li, W. Y. Chen, and R. Lafortezza, "Remote sensing and urban green infrastructure: A synthesis of current applications and new advances," *Urban Remote Sensing: Monitoring, Synthesis, and Modeling in the Urban Environment*, pp. 447–468, 2021. 15

[27] N. Pettorelli, W. F. Laurance, T. G. O'Brien, M. Wegmann, H. Nagendra, and W. Turner, "Satellite remote sensing for applied ecologists: opportunities and challenges," *Journal of Applied Ecology*, vol. 51, no. 4, pp. 839–848, 2014. 15

[28] D. Y. Leung, J. K. Tsui, F. Chen, W.-K. Yip, L. L. Vrijmoed, and C.-H. Liu, "Effects of urban vegetation on urban air quality," *Landscape Research*, vol. 36, no. 2, pp. 173–188, 2011. 16

[29] C. Gong, C. Xian, T. Wu, J. Liu, and Z. Ouyang, "Role of urban vegetation in air phytoremediation: differences between scientific research and environmental management perspectives," *npj Urban Sustainability*, vol. 3, no. 1, p. 24, 2023. 16

[30] M. K. Khan, K. Naeem, C. Huo, and Z. Hussain, "The nexus between vegetation, urban air quality, and public health: an empirical study of lahore," *Frontiers in Public Health*, vol. 10, p. 842125, 2022. 16

[31] Q. Zhuang, Z. Shao, J. Gong, D. Li, X. Huang, Y. Zhang, X. Xu, C. Dang, J. Chen, O. Altan, *et al.*, "Modeling carbon storage in urban vegetation: Progress, challenges, and opportunities," *International Journal of Applied Earth Observation and Geoinformation*, vol. 114, p. 103058, 2022. 17

[32] A. Krtalić, D. Linardić, and R. Pernar, "Framework for spatial and temporal monitoring of urban forest and vegetation conditions: Case study zagreb, croatia," *Sustainability*, vol. 13, no. 11, p. 6055, 2021. 17

[33] F. A. Yannelli, M. Bazzichetto, T. Conradi, Z. Pattison, B. O. Andrade, Q. A. Anibaba, G. Bonari, S. Chelli, M. Ćuk, G. Damasceno, *et al.*, "Fifteen emerging challenges and opportunities for vegetation science: A horizon scan by early career researchers," *Journal of Vegetation Science*, vol. 33, no. 1, p. e13119, 2022. 17

[34] T. T. Nguyen, P. Barber, R. Harper, T. V. K. Linh, and B. Dell, "Vegetation trends associated with urban development: The role of golf courses," *PLoS one*, vol. 15, no. 2, p. e0228090, 2020. 17

[35] Z. Uçar, A. E. Akay, and E. Bilici, "Towards green smart cities: Importance of urban forestry and urban vegetation," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives*, 2020. 17