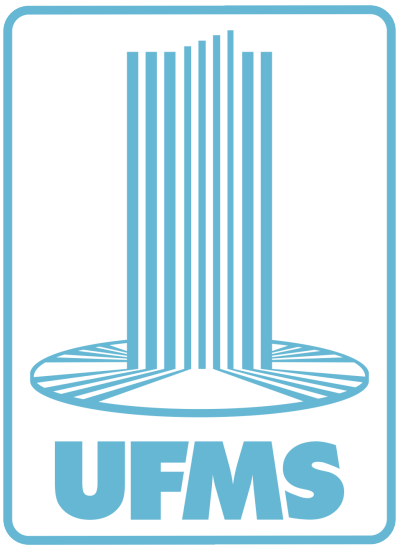
José Augusto Correa Martins



UNIVERSIDADE FEDERAL
DE MATO GROSSO DO SUL

Remote sensing and Deep Learning applied to vegetation mapping

Campo Grande/Mato Grosso do Sul, Brazil

1 de fevereiro de 2023

José Augusto Correa Martins

Remote sensing and Deep Learning applied to vegetation mapping

Doctoral dissertation submitted to the Graduate Program in Environmental Technologies - Academic Doctorate, as a requirement for doctoral qualification, under the guidance of Prof. Dr. José Marcato Junior. co-supervision of Prof. Dr. Paulo Tarso S. Oliveira and Prof. Dr. Wesley Nunes Gonçalves. The present work is carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001

Federal university of Mato Grosso do Sul – UFMS

Postgraduated program in Environmental technologies

Supervisor: Prof. Dr. José Marcato Junior

Campo Grande/Mato Grosso do Sul, Brazil

1 de fevereiro de 2023

José Augusto Correa Martins

Remote sensing and Deep Learning applied to vegetation mappingJosé Augusto Correa Martins. – Campo Grande/Mato Grosso do Sul, Brazil, 1 de fevereiro de 2023-

111p. : il. (algumas color.) ; 30 cm.

Supervisor: Prof. Dr. José Marcato Junior

Doctoral dissertation (Doctorade) – Federal university of Mato Grosso do Sul – UFMS

Postgraduated program in Environmental technologies , 1 de fevereiro de 2023.

1. visão computacional. 2. Sensoriamento remoto. 3. aprendizagem de máquina. 4. aprendizagem profunda. 5. Cidades inteligentes. 6. florestas urbanas. 7. Segurança hídrica.

I. José Marcato Junior. II. Universidade Federal de Mato Grosso do Sul. III. Programa de Pós-graduação em Tecnologias Ambientais IV. Detecção e classificação de flora usando imagens de sensoriamento remoto e técnicas de visão computacional. V. Desenvolvimento de métrica Global de análise de produção de alimentos e conservação florestal.

José Augusto Correa Martins

Remote sensing and Deep Learning applied to vegetation mapping

Doctoral dissertation submitted to the Graduate Program in Environmental Technologies - Academic Doctorate, as a requirement for doctoral qualification, under the guidance of Prof. Dr. José Marcato Junior. co-supervision of Prof. Dr. Paulo Tarso S. Oliveira and Prof. Dr. Wesley Nunes Gonçalves. The present work is carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001

Prof. Dr. José Marcato Junior
Main Advisor

Prof[a]. Dr[a]. Ana Paula Marques
member of the examining board

Prof. Dr. Veraldo Liesenberg
member of the examining board

Prof. Dr Fabio de Oliveira Roque
member of the examining board

Prof. Dr. Diogo Nunes Gonçalves
member of the examining board

Campo Grande/Mato Grosso do Sul, Brazil
1 de fevereiro de 2023

This work is dedicated to
my parents Josiberto Martins and Alaide Correa for their support, patience and understanding during my absences in recent years
and my sisters Juliana and Maria Augusta who are my foundation

Acknowledgements

This is the process of preparing a researcher who can contribute to society.

I thank my supervisor, **Prof. Dr. José Marcato Junior**, first for agreeing to guide this doctoral research, also for his patience, dedication, advice and incentives during the course.

I also thank the supervisor of the Environmental Technologies program, **Prof(a). Isabel Kaufmann de Almeida**, for always being attentive and participative during my learning trajectory in the program, answering calls and requests in a timely manner when necessary.

To my advisors and extra partners: **Prof. Paulo Tarso; Prof. Hemerson Pistori; Prof. Wesley Nunes Goncalves; Prof. Veraldo Liesenberg; Prof. Anette Eltner.**

To the National Center for Scientific and Technological Development **CNPq**, to the Coordination of Superior Level Staff Improvement **CAPES** and the Foundation to support the development of education, science and technology in the state of Mato Grosso do Sul **Fundect**, for the Doctoral scholarships in the Environmental Technologies program.

Finally, I thank the Institution **Universidade Federal de Mato Grosso do Sul** for providing the infrastructure, environments and connections necessary to carry out the research.

"Inventions are the most important product of a human creative mind.
This ultimate purpose is the complete mastery of the mind over
the material world, the harnessing of human nature
to attend human needs"
(Nikola Tesla).

Abstract

The fast development of human civilization imposed a recent and big environmental impact on the planet earth and the collective of life on Earth to support. The dawn of civilization is a very recent impact on the geologic time scale of the planet. Human needs to have a dimension of the effect of their actions inside the areas where it lives and in other natural environments to know their environmental impacts and consequences. The will to give an objective answer for this topic guided this research work. This doctoral dissertation presents the results of three years of research as a Doctoral student in the Environmental Technologies program at UFMS (Federal University of Mato Grosso do Sul). During my research, remote sensing and deep learning were the leading scientific fields I studied. The applications of the conjunction of these sciences to analyze the vegetation composition of urban and natural environments in the form of wetlands. We achieved exciting results in applying these techniques, i.e., an F1-score of 91% and an IoU of 73% for urban vegetation segmentation. Moreover, achieving a maximum 97% of F1-score for a specific plant species and 88% average for the whole dataset of 11 wetland plant species. The advances of the experiments conducted during the Doctorate program comprehend a broad range of sensors, from Unmanned aerial vehicles (UAV) that produce centimeter-level data to sensors capable of producing large earth mosaics. We also worked with a broad range of Deep Learning techniques to develop vegetation models. This research work and development can technologically assist the community in improving the understanding of the natural environment that we live in, leading to more resilient, sustainable, and healthy earth environmental systems

Keywords: computer vision; remote sensing; deep learning; semantic segmentation; ecology; smart cities.

Resumo

O rápido desenvolvimento da civilização humana impôs um recente e grande impacto ambiental ao planeta Terra e ao coletivo da vida terrestre. O alvorecer da civilização é um impacto muito recente na escala de tempo geológico do planeta. O ser humano precisa ter uma dimensão do efeito de suas ações dentro das áreas onde vive e em outros ambientes naturais para conhecer seus impactos e consequências ambientais. A vontade de dar uma resposta objetiva a este tema orientou este trabalho de pesquisa. Esta dissertação de doutorado apresenta os resultados de três anos de pesquisa como aluno de doutorado no programa de Tecnologias Ambientais da UFMS (Universidade Federal de Mato Grosso do Sul). Durante minha pesquisa, sensoriamento remoto e aprendizado profundo foram os principais campos científicos que estudei. As aplicações da conjunção dessas ciências para analisar a composição da vegetação de ambientes urbanos e naturais na forma de zonas úmidas. Obtivemos resultados empolgantes na aplicação dessas técnicas, ou seja, um F1-score de 91% e um IoU de 73% para a segmentação da vegetação urbana. Além disso, alcançando um máximo de 97% de pontuação F1 para uma espécie de planta específica e 88% de média para todo o conjunto de dados de 11 espécies de plantas de zonas úmidas. Os avanços dos experimentos realizados durante o doutorado abrangem uma ampla gama de sensores, desde veículos aéreos não tripulados (VANT) que produzem dados centimétricos até sensores capazes de produzir grandes mosaicos da terra. Também trabalhamos com uma ampla gama de técnicas de Deep Learning para desenvolver modelos computacionais de de vegetação. Este trabalho de pesquisa e desenvolvimento pode ajudar tecnologicamente a comunidade a melhorar a compreensão do ambiente natural em que vivemos, levando a sistemas ambientais terrestres mais resilientes, sustentáveis e saudáveis.

Palavras-chave: visão computacional; sensoriamento remoto; aprendizado profundo; segmentação semântica; ecologia; cidades inteligentes.

List of Figures

Contents

1 Introduction

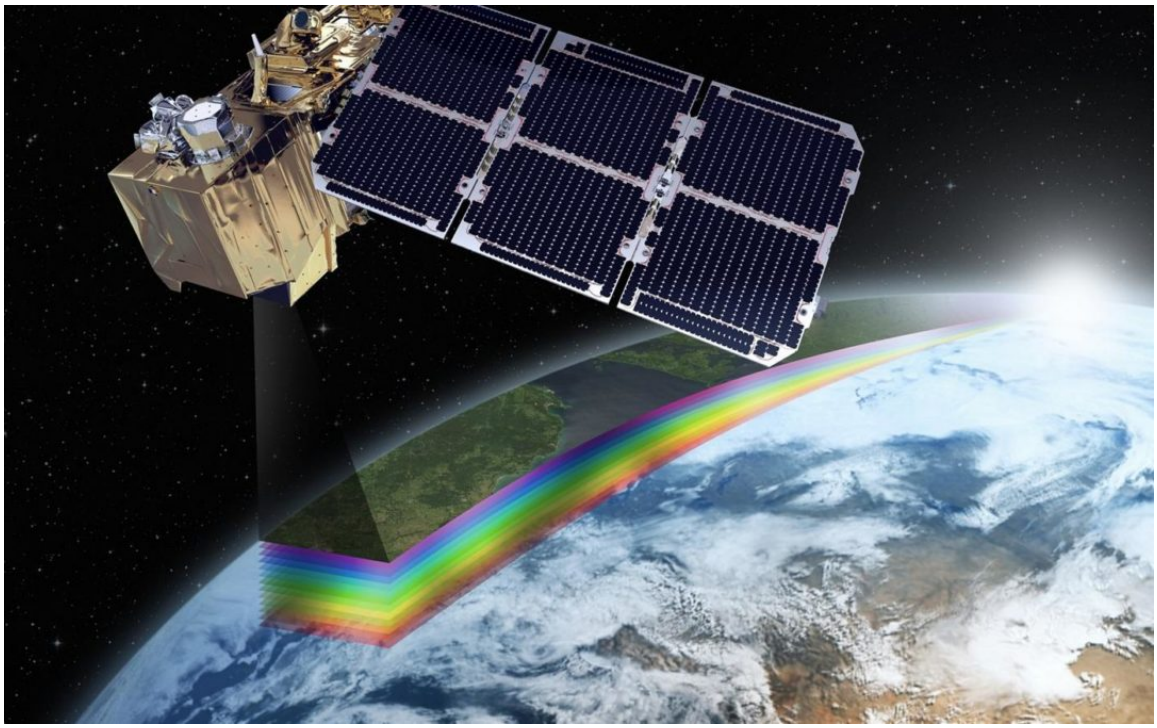


Figure 1 – *Illustration of a satellite remote sensor obtaining data from the surface of the Earth. Credits to 'https://www.orbitaleos.com/remote-sensing-technologies/'. image obtained via website in may/2022*

Water and vegetated environments is one of the most abundant essential elements that make part of the Earth (CHANG, 2006). They are a way that life finds support to inhabit this planet, making them essential to secure the sustainability of life. To develop a greater understanding of this element and their abundance or richness in a given location, on the crust surface of the planet, it is necessary to create maps, being remote sensing as an alternative to provide data even in the most difficult remote areas, such as top of mountains, isolated islands rocky terrains and others areas (ESCH et al., 2017; GORELICK et al., 2017). According to (LILLESAND; KIEFER; CHIPMAN, 2015), remote sensing is *"the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation"*. For the act of reading this doctoral dissertationis, you are applying remote sensing. Your eyes when passing through a page are collecting "data" in the form of electromagnetic impulses created by a source of light that make contact with the paper sheet page or screen and reflect till your eyes, depending on the color that it encounters it have a different intensity of absorption, resulting in a different intensity of reflected light that reaches till yous. The “data” your eyes acquire are light reflected from the surface of the page or screen, forming letters or some more colorful

are pictures and images, creating a recognizable pattern in your mind that forms words, sentences, and phrases that you are capable of understanding and extracting interpretable information; this can be called an interpretation process. In many aspects, remote sensing can be through as the process of reading and interpreting data. Using a great diversity of sensors, we collect data that can be analyzed to obtain information about objects, areas, and phenomena of interest. The remotely collected data can be of many forms, including variations in temperature, acoustic waves, forces, or electromagnetic energy distributions.

1.1 Scientific Background and Motivation

To map accurately and efficiently water and forests, efficient detection methods are needed. Data on the extent and dynamics of these environments are crucial for a better understanding of natural and anthropogenic processes and impacts on the environment, such as the alteration in quality, distribution, and circulation of water, the loss or increase in biodiversity, and populations. With remote sensing data, we are able to create mathematical models that predict, interpret or quantify specific processes and impacts, with possible approximations and analysis. The data available is abundant and rich enough for us to make pretty accurate assumptions about the environment and its actors. Despite significant progress in data-science engineering, efficient processing of higher-level products remains inherently non-automatable, creating barriers for the general public to access and understand the natural and anthropogenic processes. Understanding these processes is becoming time to time more important because water resources and forest resources are under growing pressure from economic sectors such as agriculture, energy, industry, tourism, and domestic use. Furthermore, water availability is decreasing, driving more regions into water, food, and energy insecurity (D'ODORICO et al., 2018). Recent international agendas on climate change and the environment demand objective information on planetary land and water surface conditions and changes to study the drivers behind them.

The United Nations Department of Economic and Social Affairs (BAN, 2016) define seventeen challenges to be achieved by 2030, and many of them directly or indirectly require up-to-date and high-resolution information on forests and cities. To name some: 1) Promote sustainable use of terrestrial ecosystems; 2) Make cities and human settlements sustainable; 3) Implementation and revitalize the global partnership for sustainable development.

The IPCC report for 2021 (PÖRTNER et al., 2022) identifies knowledge gaps in observations and our understanding of water, forest, and population dynamics. Furthermore, given the diversity and volume of EO data accessible today, these gaps typically indicate a lack of appropriate algorithms to interpret the available data. The raw data must be correctly engineered to derive higher-level variables that a broader, sometimes non-academic audience may understand. And one of the

This will bring to the reader a briefly revision of the current bibliography of remote sensing and deep learning related to environmental applications. Earth surface observation and interpretation technologies, in conjunction with field data collection, are used to measure, monitor, and model the tangible components that are parts of the cycles of natural and anthropogenic ecosystems (WENG, 2012). We are witnessing a time of intense human evolution result of much research, development, and sharing of technologies. As society demands, technological advances occur in the area of remote sensors and computer vision, creating new possibilities and interactions. Some blooming science fields that involve the combined use of remote sensors and computer vision are: ***Autonomous Driving*** (ALBUS, 2002; HUANG et al., 2018; SALLAB et al., 2017; CAESAR et al., 2020); ***Voice Assistant*** (FERNANDES; OLIVEIRA, 2021; MCLEAN; OSEI-FRIMPONG, 2019); ***Computation of tropospheric delay*** (ASKNE; NORDIUS, 1987; Memarian Sorkhabi; ASGARI; AMIRI-SIMKOOEI, 2021; LIM; BAE, 2021). And for the theme of this research, we are going to focus on ***Land Cover Mapping*** (CIHLAR, 2000; FRIEDL et al., 2002; BARTHOLOME; BELWARD, 2005; WULDER et al., 2018; CALDERÓN-LOOR; HADJIKAKOU; BRYAN, 2021; MARTINS et al., 2021).

Modern remote sensors can generate high spatial, temporal and spectral resolution data, making it possible to identify and quantify terrestrial objects, areas, and phenomenons. Perceiving both its abundance or lack and its changes over time and creating possibilities for investigating the causes of these changes, creating possible future scenarios and debates, bringing to light good management and sustainability strategies.

Satellites and other sensors are collecting massive amounts of data for many years, resulting in multi-petabyte datasets of "raw" data. However, it has only been in the last decade, thanks to recent advances in cloud computing, that we have begun to turn these enormous amounts of data into useful knowledge. The age of satellite remote sensing for earth observations began with the launch of Sputnik 1 on October, 1957, the first man-made satellite, by the former Soviet Union (CRACKNELL; VAROTSOS, 2007), sputnik send back radio signals, which scientists used to study the ionosphere. On July, 1972 – Landsat 1 the 1st civilian Earth observation satellite with cameras RBV and MSS was launched (BOLAND, 1976). After that numerous Earth-observing satellites have come, and with the necessity of global scale vegetation monitoring, comes the AVHRR (CRACKNELL, 1997). The commercial character of these enormous archives of satellite data was a major obstacle for their examination and use. This changed with the open access to the open of the Brazilian Data Catalog for HRCC and WFI data from CBERS (FERREIRA; CÂMARA, 2008) and followed by the Landsat mission datasets by NASA in 2008 (ZHU et al., 2019). Events which represented a significant change in understanding of the management of data, enabling a more profound investigations in many fields that are possible to advance and investigate with remote sensing data one of them being the use of EO for vegetation monitoring. Furthermore, more time was needed to produce the first

global-scale mosaics research results, allowing the databases of forests to be thoroughly studied and analyzed by the general audience in the form of a Land-cover classification map by the work of Hansen (HANSEN et al., 2013).

Plants-related sciences like plant morphology, Botany, and Agricultural Sciences, care a lot about the shapes of leaves, petals, and complete plants, forests, and agricultural lands. They can distinguish between species, monitor plant health and growth. The growing interest in biodiversity changes related to several reasons, always related to many linked events that have many different natures, but one of the very predominant events is the economical pressure on a forest or vegetated area, intending to increase production or to just clear cut the space for the selling of the natural resources present in that area. With the increasing availability of digital images (data) combined with the increasing power (Computer Vision Algorithms (CVA)) and mechanisms (Hardware) to process this data, this topic is timely and always presents a lot of new opportunities.

1.2 Questions

Many attempts have been made to create accurate regional and global scale maps of forest cover and plant species existence and coverage. For efficient and accurate land-cover data extraction detection from Earth Observation (EO) data, we need to address issues of local and global objectives, accuracy, and applicability, as well as provide access for a broad range of users. In the present study, this will be done by answering the following questions using for the time, advanced remote sensing and deep learning tools:

- How to accurately segment and quantify vegetation inside cities?
- How to differentiate species of plants inside a very bio-diverse environment such as wetlands?
- How to accurately monitor vegetation inside large areas?

1.3 Contributions

The contributions can be summarized as follows:

- Develop procedures to explore the advances in the field of computer vision applied to vegetation mapping.
- Create and use datasets with a broad range of sensors with the finality of vegetation mapping

- Awaken the interest of managers in the use of technology in the task of accurately mapping and quantifying the environment,and possibly while using the proposed methods to help in the development of effective policies related to the environment.

1.4 Outline

The format chosen for the presentation of this doctoral dissertation is of **Articles**. So the organization of the document will be. Chapter 2 reviews the relevant literature and existing methods used to detect vegetation inside different environments. Chapters 3, 4 and 5 present the articles developed during the doctorate program. Chapter 6 presents discussions and conclusions and directions for future works.

2 Brief overview of remote sensing for vegetation monitoring

This chapter gives an overview of available remote sensing approaches for processing remote sensor imagery, also a overview of the freely available datasets that were used in this study will also be provided, also explanations of the methodologies applied in this study, that guide interpretations of the environment.

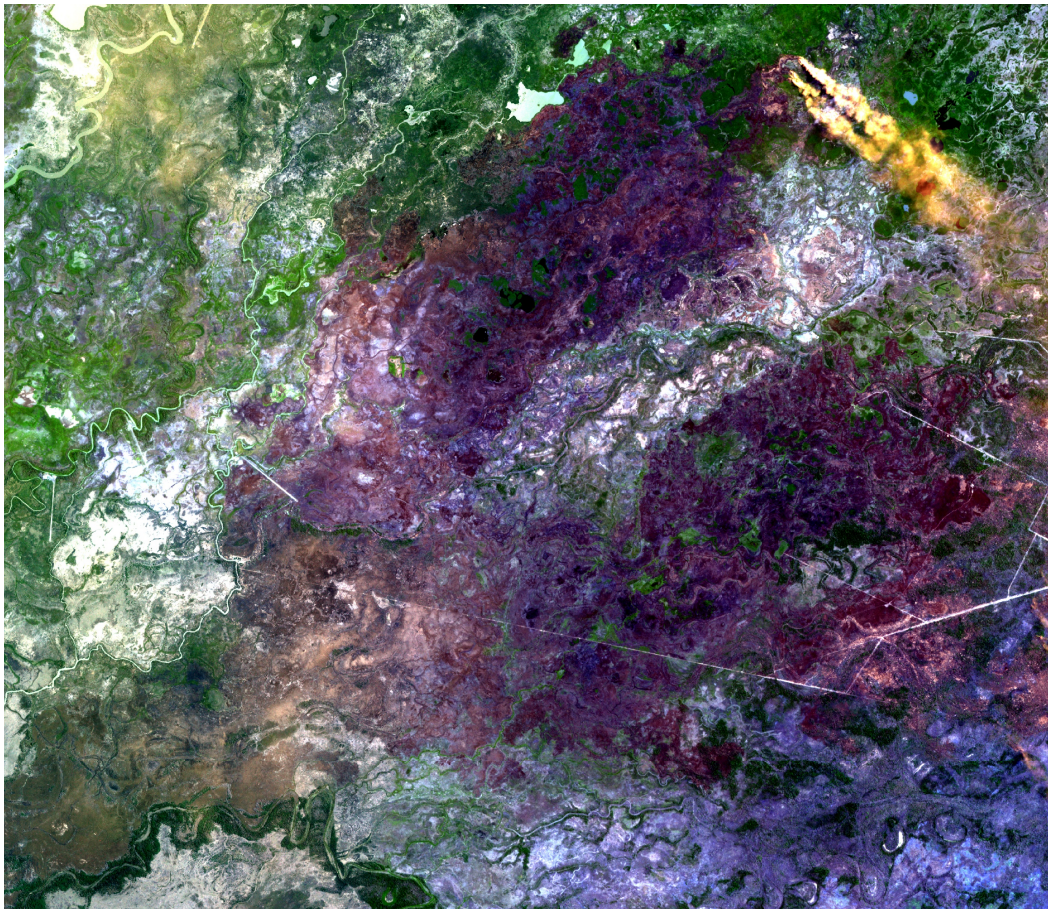


Figure 2 – Aerial view of part of desecrated land in Pantanal Brazil - 09/21. Image composed by the Autor via Planet API using GEE.

Keywords: literature overview, datasets, multi-spectral satellite imagery, drone imagery

2.1 Introduction

With the concept that digital images are formed by capturing electromagnetic energy propagated through the atmosphere via a sensor. ***Remote sensing*** is the science of detecting, recording, and interpreting these electromagnetic energy data (LILLESAND; KIEFER; CHIPMAN, 2015). ***Deep learning*** is the science of programming computers so they can learn from data, ***Artificial Neural Networks (ANNs)*** or ***Deep Learning*** is a model inspired by the networks of biological neurons found in our brains (GÉRON, 2019). In later years, the implementation of artificial neural networks in computer science focused less on replicating actual biological structures and processes and instead developed its own statistical, and computational logic (LILLESAND; KIEFER; CHIPMAN, 2015). This work brings land surface classification strategies using *Remote sensing* images such as satellite, airplane, and drone images, in combination with computer vision techniques such as *Deep Learning*. In this way, making possible to discretize large volumes of data efficiently and create accurate maps and quantification of urban forests, individual trees, and other plants. A general workflow technique performed to map an urban area with Unmanned Aerial Vehicles, and deep learning techniques are presented in Figure 3. It is an oversimplification, and many nuances encompass every part of the process. Creating and processing and deep Learning model is not a simple task; it needs much dedication and knowledge of the operator, and the same can be said for the data collection phases.

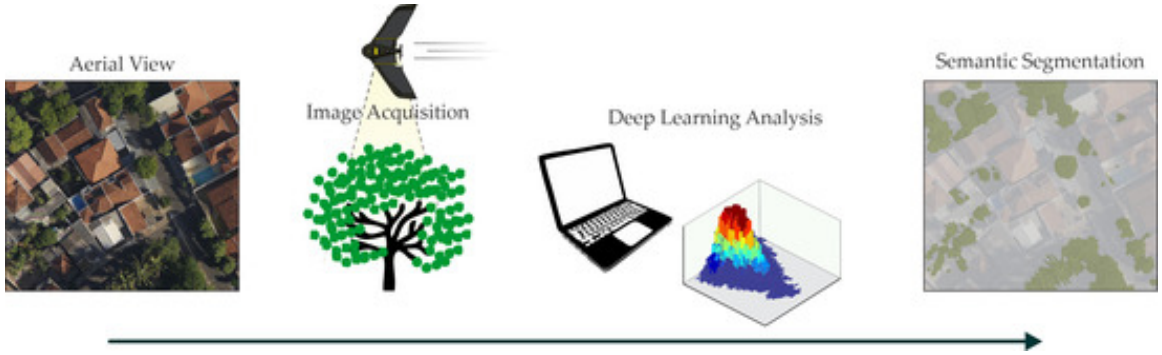


Figure 3 – Basic workflow of a urban forest deep learning mapping. Figure taken from (MARTINS et al., 2021).

Image acquisition: There are a diversity of sensors or *digital cameras* that are capable of capturing electromagnetic energy data. The core sources of remote sensed data that we will approach in this research are: satellites, airplanes, and drones.

"Typically digital cameras are a two-dimensional array of silicon semiconductors consisting of either charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) detectors. Each array detector (or photosite location) senses the energy radiating from one pixel in the image field. When this energy strikes the surface of the detector, a small electrical charge is produced, with the charge's magnitude proportional to the

scene brightness within the pixel. The charge is converted to a voltage value, which is, in turn, converted to a digital brightness value. This process results in the creation of digital brightness value ***data*** *for each photosite pixel in the array" (LILLESAND; KIEFER; CHIPMAN, 2015).*

Image processing: certain preprocessing operations are performed on the raw **data** of an image prior to its use in any further interpretation. These operations can be made to correct flaws in the data, like noise removal, radiometric correction, geometric correction, subsetting, mosaicking, and layer stacking. Furthermore, we can enhance the data by performing operations like Contrast manipulation, Spatial feature manipulation, and Multi-image manipulation.

Deep learning: The ***data*** value reaches the computer and needs to be interpreted. Deep learning is one of the sciences that are capable of doing it, and creating a quantification of a determined feature of the provided data. There are an uncountable number of ways and configurations that we can address to make a Deep-learning model. The current state of the art of image processing techniques involves Deep Learning in the form of ***Convolutional Neural Networks*** proposed by (LECUN; BENGIO et al., 1995), and transformer based architectures (VASWANI et al., 2017). *"Neural networks are "self-training" in that they adaptively construct linkages between a given pattern of input data and particular outputs" (LILLESAND; KIEFER; CHIPMAN, 2015).*

Agricultural expansion and climate variability are significant events of disturbance in natural areas (DAVIDSON et al., 2012). Moreover, these events of what sometimes may appear to be a localized environmental transformation have detectable and in-detectable effects on near and far community health, availability of resources, and development. The phenomenon known as climate change on earth is caused by in-numerous previous anthropological actions that they we do not know the real environmental implications (CLINE et al., 1992; VITOUSEK, 1994; DELWORTH; KNUTSON, 2000; NORDHAUS; BOYER, 2003; KERR, 2007; CHOI; GAO; JIANG, 2020).

One of humanity's urgent present and future challenges is the *global urbanization*. The actual dimensions and impacts of this phenomenon are not yet fully understood (ESCH et al., 2017), but it is established that the urbanization phenomena cause impacts on global environmental change (GRIMMOND, 2007). For us to have a perspective of the explosive growth of urban settlements, in 1950, the rural population was twice the size in the number of individuals as the urban population, and around 2008 the urban population exceeded the rural population for the first time in history (ESCH et al., 2017), and this phenomenon is increasing over time in many regions of the world. Also, the earth's population is increasing, the *highlights of the United Nations - World Population Prospects -, World Population Prospects of 2019* (ECONOMIC; AFFAIRS, 2019), the global population could grow to around 8.5 billion in 2030, 9.7 billion in 2050, and 10.9 billion in 2100 (ECONOMIC;

AFFAIRS, 2019). Therefore, we need densely inhabited urban environments to be prepared to provide an adequate quality of life for their residents in tune with actions that provide sustainability for the environment. Related to this, one of the themes explored in this research is **Smart Cities**. Smart Cities is a concept and field of research that search for engineering methods that make cities and human settlements inclusive, safe, resilient, and sustainable(ECONOMIC; AFFAIRS, 2019). Capable of efficiently overcoming challenges external and internal of the most diverse natures. We can mention some of these natures: economic, sanitary, environmental, and climatic (WENG, 2012). For this, the smart city must develop or implement innovative methods and techniques that support the effective management of the various components of the urban ecosystem and its resources and other assets to mitigate the adverse impacts caused by the urbanization phenomena (WENG, 2012).

Furthermore, encompassing the theme of Smart Cities, we will address the theme of **Urban Forests**, which refers to all forms of vegetation land covers that grow naturally or artificially in and around human habitats (BLUM, 2017). So reaching for one of the objectives of this research is the quantification and mapping of urban forests, presenting tools and methods for a better understanding of these regions and their respective impact on their ecosystem or, in other words, their cities. Moreover, creating a work that brings proximity to biological sciences and computer vision, in this way, provides a better understanding of the city environment. The motivation for the analysis of this challenge is that vegetation in cities performs a significant number of "services" , these services vary from place to place, but they include air purification, carbon sequestration, temperature regulation, noise reduction, sustaining the local biodiversity, storm-water drainage, and scenic and recreational opportunities (BLUM, 2017). Moreover, they also may offer "disservices" such as the release of volatile organic compounds and particulate matter in the form of pollen and the compromise of urban structures such as wiring and buildings (BLUM, 2017). Therefore, an essential part of *Urban forests* research resides in the mapping and understanding of these environments and their components and the computation of the services, interaction, and impacts with the urban environment that will assist us in realizing adequate management of this so unique and precious type of land cover.

With their large size, wilderness atmosphere, and fresh air, cool and shaded urban forests are great places for people to unwind from stress, pressure, and hectic activities. The presence of woods can be beneficial to the health of those who live nearby. This could be due to the continuous replenishment of oxygen in forested areas and the reduction of dust and air pollution. A greenbelt of trees can reduce city and traffic noise while providing a beautiful contemplation scenario.

The number of satellites orbiting Earth is growing, and their technical capabilities are constantly improving, resulting in improved spatial, temporal, spectral, and radiometric

resolutions. The number of EO satellites launched into Earth's orbit and the quantity of photos observed has increased rapidly in recent years. We owe that to the growth of the satellite industry, the shrinking satellite sizes, and the costs associated with their delivery to orbit. This also created new scientific challenges, necessitating the development of more efficient and reliable data processing methods. With technology and methodological advancements, the focus of urban forest research can shift to a planetary-scale study, allowing for a better understanding of Earth's natural and anthropogenic processes. This is a recent technological possibility that was only feasible recently (HANSEN et al., 2013).

We utilize satellites for planetary-scale analysis, and we have the technological advances of UAV systems for a more refined analysis of our environment. They are especially suited for many civilian applications, notably in environmental monitoring, resource management, and infrastructure management (LALIBERTE et al., 2010).

2.2 Challenges of vegetation detection

Process-based vegetation models are widely used to predict local and global ecosystem dynamics and climate change impacts. Due to their complexity, they require careful parameterization and evaluation to ensure that projections are accurate and reliable.

While vegetation detection from cloud-free images appears to be trivial, it remains a very challenging task when working with real-world images and when teaching a Neural Network of what is or is not vegetation. In this case, agricultural and grassy fields may be misinterpreted as vegetation. We also face many challenges when we intend to work with a specific kind of plant or vegetation, that the features present are very similar to the other objects and background present in the image. Additionally, errors of commission (false positives) can be observed in areas with shadows due to topographic conditions or the presence of clouds. One of the things that we can observe while working with images is that the main challenges of vegetation detection start with the own creation of the datasets and the obstacles that we may face.

2.2.1 Spectral and radiometric resolutions

When working with multi-mission satellite data, the radiometric resolution is another factor that may influence surface vegetation detection. Radiometric resolution refers to the sensor's sensitivity to incoming radiation, which is characterized by the minimum and maximum radiance values and the number of bits used to store measured values (DONCHYTS et al., 2017). The actual radiometric resolution for Landsat TM, ETM+, and ASTER sensors is 8 bits. It was modified to 12-bit for Landsat 8 and Sentinel-2 (stored as 16-bit). However, the effect of 8-bit radiometric resolution may influence thresholds used

to detect vegetation. The UAV used in the study had a FAT32 method of storing data an update from the 16bit method to a 32bit, meaning it's better for threshold detection.

The spectral resolution of a sensor is related to its capacity to capture the incoming radiance and store it in bands. The high areas in Figure 4 represent where that sensor can capture data, and each band corresponds to a space between those. The finer the spectral resolution, the narrower the wavelength range for a particular channel or band.

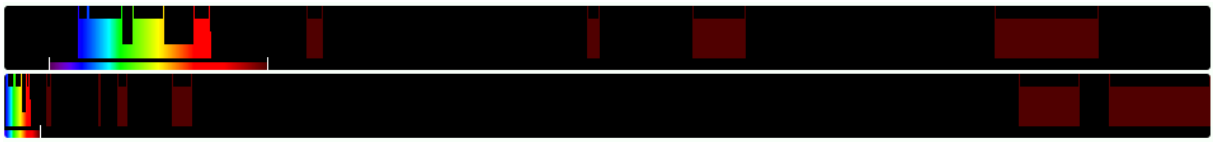


Figure 4 – Illustration of the spectral resolution range from the Landsat 8 sensor. source: <https://www.indexdatabase.de>.

In Figure 5 we have a reflectance illustration of how water, soil and vegetation behave for each spectral band. The higher the value of reflectance the higher the value the band will capture. That´s an visual explanation also for the NDVI formula present in Equation 2.1, where we have a high value for NIR and a small value for red.

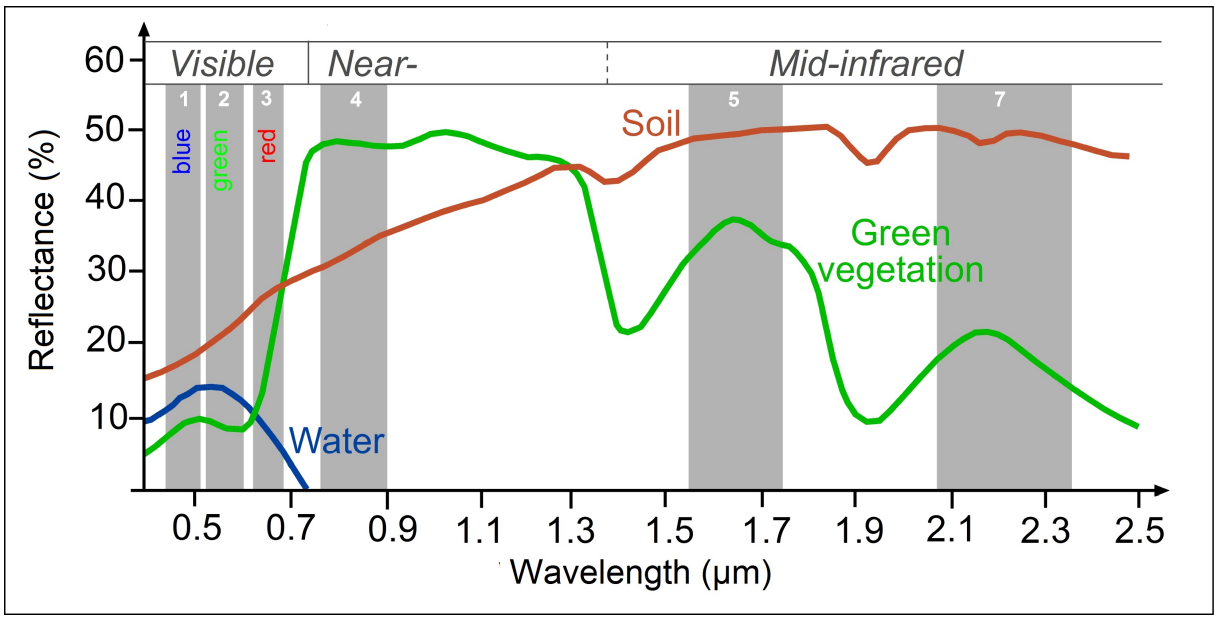


Figure 5 – Reflectance of water, soil and vegetation in different wavelengths and Landsat TM channels 1 (0.45-0.52 m), 2 (0.52-0.60 m), 3 (0.63-0.69 m), 4 (0.76-0.90 m), 5 (1.55-1.75 m) and 7 (2.08-2.35 m). source: <https://seos-project.eu>.

2.2.2 Spatial and temporal resolution

For the sensors studied in the present research, the spatial resolution of the sensors varies between 1cm from the UAV to 4.77m from the satellite for optical bands. Figure 6 shows a few examples of cloud-free satellite images over the same area and comes to illustrate how the spatial resolution is present for different sensors and how we need to use the spatial resolution adequate for each objective.

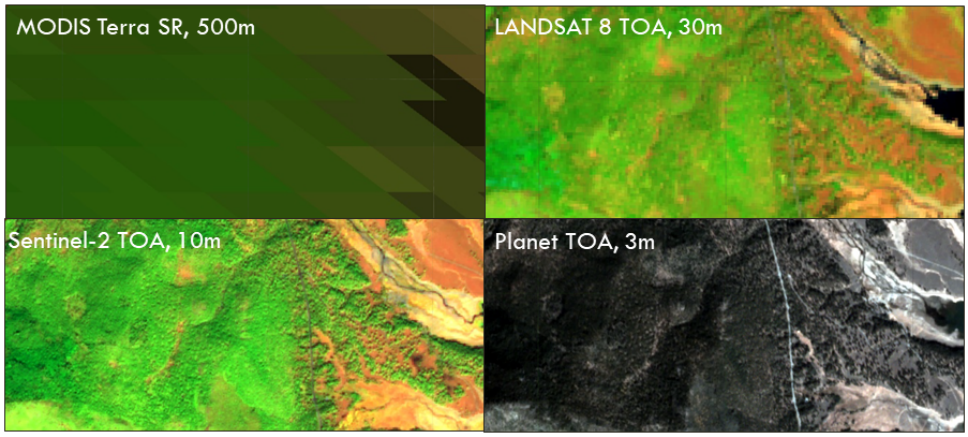


Figure 6 – Examples of images with different spatial resolutions, from four different sensors source: <https://code.earthengine.google.com/93e8cd71bbc2e58584f04645e8690279>. code adapted from (DONCHYTS et al., 2017)

A temporal resolution mismatch between observations and the actual changes occurring on the land surface may significantly influence the applicability of satellite data for multi-temporal analysis. Temporal resolution is the time a satellite takes to orbit and return to the same place and get another image. Temporal resolution plays a fundamental role in detecting changes over time. When dealing with UAV data, the user can more freely determine the temporal resolution.

2.2.3 Vegetation wrongly classified due to natural and anthropogenic processes

Detection and correction of effects caused by clouds and aerosols are one of the most studied topics of optical remote sensing (DONCHYTS et al., 2017). Clouds, haze, fog, and other substances cause absorption and scattering of the solar electromagnetic flux on its way to the Earth's surface and then on the way back to the satellite sensor. Higher-level surface reflectance (SR) products are already available from several satellite missions, including Landsat, MODIS, and PROBA-V, where images are corrected for atmospheric influences. Some data providers also offer a quality assessment (QA) band that shows whether a pixel is cloud-covered. Even if the generated photos are more appealing and closely resemble

the desired spectral signatures of various land cover types, using atmospherically adjusted images from several satellite missions may lead to an image mismatch.

Observed vegetation pixels may have noise, making it challenging to classify vegetation using remotely sensed images. It often consists of a combination of spectral signatures produced by various elements seen at a specific time and place. Physical limitations of the satellite sensors, such as their spectrum and spatiotemporal resolution, as well as systematic mistakes in the data processing pipeline, are additional variables that contribute to spectral mixes. Numerous non-stationary events are to blame for the spectral signatures of observed vegetation, making it difficult to anticipate with any degree of accuracy.

2.3 Data acquisition and cloud servers

Every day an immense amount of data is generated from all kinds of remote sensors, such as satellites, aircraft, and UAVs. These data need to be processed and interpreted to understand a broader quantity of users. The activity of downloading and processing these datasets in a standard desk computer is often unfeasible due to data storage and processing limitations that these kinds of systems have. This kind of problem is addressed when we use cloud computing. Many advances related to Big Data applied to Remote sensing, and Deep learning occurred in recent years. One of these moments of advance was the opening to the public of the Platform Google Earth Engine (GORELICK et al., 2017), which is a platform for planetary-scale scientific research of geographical datasets. It is possible to develop algorithms and use a powerful computer to process and analyze vast quantities of data, creating visual interpretations that a broader audience can understand. According to Gorelick N. *Earth Engine consists of a multi-petabyte analysis-ready data catalog combined with a high-performance, intrinsically parallel computation service. It is accessed and controlled through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment (IDE) that enables rapid prototyping and visualization of results.* We also have other available cloud computing options such as Sentinel Hub (SENTINEL, 2022), Open Data Cube (KILLOUGH, 2018), System for Earth Observation Data Access, Processing, and Analysis for Land Monitoring SEPAL (SEPAL, 2022).

The present research is related to a broad range of sensors. Including an Aircraft sensor with 10cm spatial resolution and Red, Green, and Blue bands, an UAV sensor with $\approx$ 1cm of spatial resolution and Red, Green, and Blue bands. And a satellite sensor with 4.77m resolution and Red, Green, Blue, and NIR bands (PLANET, 2017).

2.4 Methods of vegetation detection from images

Specialized algorithms, frequently modified to recognize specific characteristics and find abnormalities, are necessary for multi-spectral satellite sensors and openly available data in raw form satellite images. The primary causes for which satellite data. Processing is challenging because satellite photos show a complex mixture of natural and artificial activities, many of which are dynamic and interfere with one another. The technical restrictions of the sensors and data processing pipelines frequently constrain the observations produced. Before any useful information can be gleaned from the satellite data, several types of noise or gaps in the measurements caused by the presence of clouds, aerosols, complex topography circumstances, and technical sensor constraints must be resolved.

Assessments have been made during the last decade to establish global scale forest coverage (HANSEN et al., 2013; SEXTON et al., 2013; KIM et al., 2014; SONG et al., 2018; VANCUTSEM et al., 2021). As technologies advances, we have new capabilities and opportunities to Map and quantify our forestry resources and perceive how it is changing as time passes. Accurate and fully automated detection of forest coverage from multi-temporal satellite data at high spatial resolution from EO data is a complex task. For efficient and accurate forest detection, we need also to address issues of global objectivity, and applicability and provide access for a broad range of users.

Spectral signatures of vegetation in most clear-sky observations are very distinctive. They can easily be detected even using only Top of atmosphere (TOA) data products. TOA reflectance data provides information about the reflectance of the Earth's surface as seen from space, which is used for applications such as monitoring vegetation health and change, tracking land-use changes, and detecting mineral deposits (AGREEMENT, 2015).

In order to map land cover, monitor crop quality, and identify changes in land use, surface reflectance data is utilized to provide information on the reflectance of the Earth's surface as seen from the ground. By comparing TOA and surface reflectance data, researchers can learn more about the Earth's surface and make more precise observations about changes over time. TOA and surface reflectance data can be combined with other data sources, such as temperature and precipitation data (AGREEMENT, 2015).

However, the sensor's limited spectral, spatial, and radiometric resolution and many other factors may significantly influence the accuracy of the detected vegetation. Existing methods for vegetation detection from multi-spectral satellite data use that vegetation significantly reflects most radiation at near-infrared wavelengths. This fact makes it easy to detect vegetation employing spectral indices, such as the Normalized Difference Vegetation Index (NDVI) (JR et al., 1974).

$$NDVI = (pRed - pNIR)/(pRed + pNIR) \tag{2.1}$$

Where pRed and pNIR correspond to the spectral reflectance of the Red band and of the NIR band, the index values vary between -1 and 1, with vegetation appearing mostly when the index value is greater than zero. The NDVI is present as a reference for index interpretation of data, but many more indexes exist for a broad range of applications. Such as the leaf area index (LAI), proposed by (WANG et al., 2007), which is used to measure processes that occur on the surface of the plants, such as photosynthesis, evaporation, transpiration, and to estimate terrestrial ecosystem net primary production, but also is a crucial parameter in models of vegetation and carbon-circling. Besides these, we have a wide variety of indexes for a broad range of applications (HENRICH et al., 2012).

Detecting vegetation from remote sensor images is simple, but doing this for noisy images is challenging. When applied to real-world satellite imagery, most existing methods require manual tuning and, in general, need to be significantly adjusted to be used for planetary-scale analysis. No method works perfectly for all land use types and atmospheric conditions. Most of the processes that occur in the same area and are observed by the satellite sensor are random, with unknown distribution, and very hard to model using existing methods. However, some recent efforts in applying more advanced statistical learning methods, such as Bayesian Networks (MELLO et al., 2010), Conditional Random Fields (WALLACH, 2004), Markov Random Fields (CLIFFORD, 1990), Deep Learning (GOODFELLOW; BENGIO; COURVILLE, 2016) and transformers (VASWANI et al., 2017) look very promising.

2.5 Deep Learning

Several frameworks,such as KERAS (GÉRON, 2019), TensorFlow (ABADI et al., 2015), PyTorch (PASZKE et al., 2017) have emerged as a result of the growth in the capacity of GPUs to perform calculations more quickly than the CPU over the past years. Deep neural networks, also known as "Deep Learning," have recently gained popularity. When working with images, a neural network is trained to carry out a task by extracting the features of these images.

Therefore, when working with computer vision, it is very usual to use Convolutional Neural Networks, which are specific networks to work with image datasets based on the characteristics extracted from the images, can infer something. Among the networks intended for use are the convolutional neural networks (LECUN; BENGIO et al., 1995; SIMONYAN; ZISSERMAN, 2014; KATTENBORN; EICHEL; FASSNACHT, 2019). These network architectures have been used in image and video processing to identify objects in

images by layering convolutional images of various sizes. The operation structure of each design varies depending on the dataset and computing power. Since these convolutional architectures can be applied to both small and big image bases, the network can recognize objects more accurately when the sample size for each object is larger but also increasing the computational cost of processing.

Recently a new architecture is being used for image processing that are called Transformers. This is a type of neural network architecture that was introduced in the paper "Attention Is All You Need" by Google researchers in 2017 (VASWANI et al., 2017). The architecture is based on the idea of self-attention, where the network is able to weigh the importance of different parts of the input when making a prediction. This allows the network to effectively process sequential data, such as natural language, where the order of the words is important. Transformers networks have been used in a variety of remote sensing applications, including vegetation mapping. In this context, the goal is typically to use satellite or aerial imagery to map the distribution and type of vegetation in a given area.

2.6 Unmanned Aerial Vehicles (UAVs)

The advent of the accessibility of UAVs and satellite imagery to a broader audience created a scenario of new possibilities for obtaining EO data and developing technologies. Moreover, do not bring up some of the previously discussed problems, such as clouds and object interference. Not long ago, when UAVs were not accessible to a broader public in terms of costs and knowledge of the technology and the utilities of UAVs. Commercial drones take flight supported by decades of military research and development. The first is devoted to developing SAGE. The first continental comprehensive computerized and automated enemy detection system was created during the 1950s by the U.S. military to detect long-range Soviet atomic bombers (PACKER; REEVES, 2013). A UAV is the prominent part of a whole system that is necessary to fly the aircraft. The fact that no pilot is physically present in the airplane does not imply that it can fly independently.

UAVs provide flexibility in data collecting since users may program flights according to their needs, and they are less expensive than other systems. UAVs also provide high levels of detail when collecting data, also the capacity to embed RGB, multispectral, hyperspectral, thermal, and LiDAR sensors, and the capacity to collect data from hard-to-reach locations,in addition to those sensors built into UAVs are capable of producing data from various viewpoints and altitudes. Along with other factors, these traits enable cameras to capture images with a more comprehensive dynamic range than other sensing systems. This enables the same object to be seen from various perspectives, affecting not only its spatial and spectral information but also its form, texture, pattern, geometry,

illumination, and other attributes. This makes multi-domain detection difficult. Therefore, research shows that Deep Learning (DL) is the most popular remedy for addressing these drawbacks.

In areas where field reconnaissance is challenging, vegetation surveys utilizing UAV are feasible and capable of an exact community division, as presented in Chapter 4. The UAV method is efficient and will help to advance research techniques in the future. It may also lower research expenditures by reducing the number of field survey days and staff required.

2.7 Smart cities

Urbanization displays a massively increasing global trend. According to data from the UN (ECONOMIC; AFFAIRS, 2019), more than half of the world's population habit urban areas, and by 2050, it is projected that 68% of the world's population to be urban. The cities can be interpreted as a complex system incorporating people and diffusion and exchanging work, assets, capital, and information. One of the essential aspects of a city is the vegetation; they provide the city's residents with several environmental and social services, supporting the development and improving inhabitants' quality of life (La Rosa; WIESMANN, 2013; JENNINGS V.; YUN, 2016; ARANTES et al., 2021; JIM; CHEN, 2009; CHEN; WANG, 2013). The types of services that the urban forests provide can be cited as regulating and maintaining local climatic conditions by reducing the formation of urban heat islands, providing habitat for local biodiversity, provisioning resources (e.g., wood, food, and biomass), cultural and historical values and also scenic landscapes (BARÓ et al., 2014; MCHUGH et al., 2015; KARDAN et al., 2015; FENG; LIU; GONG, 2015; ALONZO et al., 2016).

One of the significant problems associated with crescent urbanization is the loss of vegetated areas within the city, which can profoundly impact the urban environment. The loss of vegetation can lead to increased temperatures, reduced air quality, and decreased biodiversity (BARÓ et al., 2014). It can also contribute to the urban heat island effect, increasing energy consumption and exacerbating the impacts of extreme weather events. Additionally, the loss of vegetation can reduce the ability of the urban landscape to absorb and filter rainwater, leading to increased runoff and flooding (**??**). Furthermore, the loss of vegetation can reduce the ability of the urban landscape to provide ecosystem services such as carbon sequestration, pollination, and recreation opportunitie (BARÓ et al., 2014). To address these problems, urban planners and designers must consider the importance of maintaining and enhancing vegetated areas within the city by integrating green spaces, rooftop gardens, and other forms of urban vegetation into the built environment. Rapid urbanization severely hinders city dwellers' sustainable growth and living standards.

Smart cities create comprehensive solutions for urban ecosystems using data gathered from many sensors. As a result of advances in the internet of things (IoT) and Deep learning technology, interest in urban studies and applications has significantly increased. Deep learning approaches, which are sophisticated machine learning methodologies, offer an interactive framework that makes data mining and knowledge discovery tasks easier, particularly in computer vision and natural language processing. Advanced technology has already been applied to monitoring, assessing, and analyzing environmental and natural resource management.

With this thesis we want to address the Smart Cities vegetation problem by using remote sensing and deep learning techniques to quantify the vegetated areas of a city. Chapter 3 brings an example of how we can connect Deep learning and Smart cities for environmental monitoring.

CHAPTER 3

Paper 1: Semantic Segmentation of Tree-Canopy in Urban Environment with Pixel-Wise Deep Learning

Manuscript published in *Remote sensing* journal

remote sensing

MDPI

Technical Note

Semantic Segmentation of Tree-Canopy in Urban Environment with Pixel-Wise Deep Learning

José Augusto Correa Martins [1], Keiller Nogueira [2], Lucas Prado Osco [3], Felipe David Georges Gomes [4], Danielle Elis Garcia Furuya [4], Wesley Nunes Gonçalves [1], Diego André Sant'Ana [5], Ana Paula Marques Ramos [4,6,*], Veraldo Liesenberg [7], Jefersson Alex dos Santos [8], Paulo Tarso Sanches de Oliveira [1] and José Marcato Junior [1]

1 Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande 79070-900, Brazil; jose.a@ufms.br (J.A.C.M.); wesley.goncalves@ufms.br (W.N.G.); paulo.t.oliveira@ufms.br (P.T.S.d.O.); jose.marcato@ufms.br (J.M.J.)
2 Computing Science and Mathematics Division, University of Stirling, Stirling FK9 4LA, UK; keiller.nogueira@stir.ac.uk
3 Faculty of Engineering and Architecture and Urbanism, University of Western São Paulo, Rodovia Raposo Tavares, km 572, Bairro Limoeiro 19067-175, Brazil; lucasosco@unoeste.br
4 Environment and Regional Development Program, University of Western São Paulo, Rodovia Raposo Tavares, km 572, Bairro Limoeiro 19067-175, Brazil; felipedgg@yahoo.com.br (F.D.G.G.); daniellegarciafuruya@gmail.com (D.E.G.F.)
5 Environmental Science and Sustainability, INOVISÃO Universidade Católica Dom Bosco, Av. Tamandaré, 6000, Campo Grande 79117-900, Brazil; diego.santana@ifms.edu.br
6 Agronomy Program, University of Western São Paulo, Rodovia Raposo Tavares, km 572, Bairro Limoeiro 19067-175, Brazil
7 Forest Engineering Department, Santa Catarina State University, Avenida Luiz de Camões 2090, Lages 88520-000, Brazil; veraldo.liesenberg@udesc.br
8 Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil; jefersson@dcc.ufmg.br
* Correspondence: anaramos@unoeste.br

Abstract: Urban forests are an important part of any city, given that they provide several environmental benefits, such as improving urban drainage, climate regulation, public health, biodiversity, and others. However, tree detection in cities is challenging, given the irregular shape, size, occlusion, and complexity of urban areas. With the advance of environmental technologies, deep learning segmentation mapping methods can map urban forests accurately. We applied a region-based CNN object instance segmentation algorithm for the semantic segmentation of tree canopies in urban environments based on aerial RGB imagery. To the best of our knowledge, no study investigated the performance of deep learning-based methods for segmentation tasks inside the Cerrado biome, specifically for urban tree segmentation. Five state-of-the-art architectures were evaluated, namely: Fully Convolutional Network; U-Net; SegNet; Dynamic Dilated Convolution Network and DeepLabV3+. The experimental analysis showed the effectiveness of these methods reporting results such as pixel accuracy of 96,35%, an average accuracy of 91.25%, F1-score of 91.40%, Kappa of 82.80% and IoU of 73.89%. We also determined the inference time needed per area, and the deep learning methods investigated after the training proved to be suitable to solve this task, providing fast and effective solutions with inference time varying from 0.042 to 0.153 minutes per hectare. We conclude that the semantic segmentation of trees inside urban environments is highly achievable with deep neural networks. This information could be of high importance to decision-making and may contribute to the management of urban systems. It should be also important to mention that the dataset used in this work is available on our website.

Keywords: remote sensing; image segmentation; sustainability; convolutional neural network; urban environment

Citation: Martins, J.A.C.; Nogueira, K.; Osco, L.P.; Gomes, F.D.G.; Furuya, D.E.G.; Gonçalves, W.N.; Sant'Ana, D.A.; Ramos, A.P.M.; Liesenberg, V.; dos Santos, J.A.; et al. Semantic Segmentation of Tree-Canopy in Urban Environment with Pixel-Wise Deep Learning. *Remote Sens.* **2021**, *13*, 3054. https://doi.org/10.3390/rs13163054

Academic Editor: Bailang Yu

Received: 6 May 2021
Accepted: 16 July 2021
Published: 4 August 2021

1. Introduction

Urbanization displays a massively increasing global trend. According to data of the UN [1], more than half of the world's population habit urban areas, and by 2050, it is projected to 68% of the world's population to be urban. The cities can be interpreted as a complex system that incorporates people and diffusion and exchange of work, assets, capital, and information. One of the essential aspects of a city is the vegetation; they provide the residents of the city with several environmental and social services, in this way supporting the development and improving inhabitants quality of life [2–6]. The types of services that the urban forests provide can be cited as regulating and maintaining local climatic conditions by reducing the formation of urban heat islands, provide habitat for local biodiversity, provisioning resources (e.g., wood, food, and biomass), cultural and historical values and also scenic landscapes [7–11]. According to [12], urban forests can be categorized in three primary forms, such as (i) forest remnants occurring either in the urban perimeter or in the urban-rural interface and contains a large number of trees, (ii) green areas over a given landscape that presented different tree species devoted to meet social, aesthetic and architectural benefits, ecological needs and even economic benefits and (iii) street trees that are any trees along public roads, whether on sidewalks or in flowerbeds. Urban growth is usually associated with forest remnants suppression leading to ecosystem stress and biodiversity losses [13]. In the urban environment, vegetation suppression is mainly correlated to the urban process of growth and often results in impervious areas and brings other negative impacts for the environment, such as urban biodiversity loss and changes in the hydrodynamics of the cities [14]. Therefore, mapping urban forests are essential in order to propose strategies that optimize citizen's quality of life, city hydrodynamics, and biodiversity by preserving and improving this valuable ecosystem [2,15].

The capability of detecting individuals and groups of trees is essential for many applications in forest monitoring according to [16], such as resource inventories, wildlife habitat mapping; biodiversity assessment; and threat and stress management. Nevertheless, the task of mapping trees, especially for an urban context, is a crucial procedure for environmental planning [10,17]. The urban tree segmentation task refers to the automated classification of each pixel of a given image into a tree or background, and that is a challenging problem. As highlighted before, urban forests are a particular form of forest with unusual characteristics as they can be isolated, densely, or sparsely distributed and mixed with other urban features [18]. All these characteristics make automated urban forest mapping a complex computational task. Therefore, it requires high spatial resolution images and a robust machine learning process to differentiate them as objects. The consolidated approach for mapping trees using remote sensing refers to the use of data acquired from sensors embedded in satellite, airplane, or Unmanned Aerial Vehicles (UAV) [19–24]. Remote sensing can provide valuable data at different acquisition levels to support policies related to urban forest mappings such as forest health, regulation, climate change mitigation, and long-term sustainability. As a result, continuous remote sensing tracking of forest patterns allows for cost-effective periodical assessment of vegetated areas [25], supporting decision making.

Artificial intelligence and the remote sensing field are allies of an extended period. As a subgroup of the machine learning area, deep neural networks improved performance for extracting information from images. Deep learning-based methods are evolving continuously, and their high performances being confirmed in several fields of applications [26–33]. As an example of advances in this field of study, we will briefly describe some works. Ref. [34] proposed a methodological approach for detecting individual fruits using a pixel-wise segmentation method based on Mask R-CNN and multi-modal deep learning models. It used RGB and HSV images, and the results revealed a more precision score when deep learning models are trained with RGB and HSV datasets altogether. Ref. [35] opens opportunities for a better understanding of the on-ground feature mapping by using simplified deep learning methods. By adopting high spatial resolution remote sensing data and models of

dynamic multi-context segmentation approach, based on convolutional networks, Ref. [35] research showed improvements in pixel-wise classification accuracy when compared to state-of-the-art deep learning methods. One research [30] evaluated five methods based on deep fully convolutional networks using high-spatial-resolution RGB images to map a specific threatened tree species finding out an overall accuracy ranging from 88.9% to 96.7%. The research of Zamboni et al. [36] propose the evaluation of novel methods for single tree crown detection. A total of 21 methods were investigated, including anchor-based (one and two-stage) and anchor-free state-of-the-art deep-learning methods. Here, the authors focused on generating bounding boxes in each tree, not in the tree segmentation.

The accurate segmentation of urban forests is a relevant matter, as it aims to support management decisions related to urban environmental planning. So this work intends to provide a low-cost and effective method of mapping trees inside the urban areas. Our study area is a metropolitan region from Brazil, inserted in Cerrado Biome called Campo Grande. The trees of the metropolitan region are very diverse, having as many as 61 cataloged tree species that are common encountered in the city center and avenues [37]. Cerrado is a nature hotspot that is rich in biodiversity, has endemic species (species of plants or animals that only occur in that Biome), and whose maintenance is threatened. Therefore, they are places that need more attention from conservation programs. The Cerrado, together with the Atlantic Forest, are the two areas considered as hotspots in Brazil. The expansion of agriculture, especially in the Brazilian Cerrado, which has an ideal climate for cultivating various crops, generated pressure to suppress natural areas of this Biome, increasingly threatening its existence. The monitoring through technology allows for more effective planning of actions and acts quickly to curb vegetation removal without permission.

Therefore, our originality comes from the dataset and the exertion of deep learning algorithms to improve the nature conservation efforts in this region, and our primary contribution is related to the investigation of state-of-the-art semantic segmentation methods to detect trees in urban areas. For this task, we used data with a high spatial resolution (Ground Sample Distance (GSD) of 10 cm) inside an urban area inside the Cerrado biome and used five state-of-the-art deep learning architectures to process the data. Deep learning-based approaches designed to tackle this task receive, as input, an image and return as output, another image, generally with the exact size of the input data with each pixel associated with one class. We choose to use deep learning because they present a better performance in semantic segmentation and scene interpretation tasks over traditional machine learning and trained professionals in many fields of science as demonstrated by [38–42].

The rest of this paper is organized as follows: Section 2 provides detailed info of the urban area and methods applied; Section 3 explored the results; Section 4 argues the implications of the results of the methods applied in this research and in Section 5 we conclude the paper and provides some futures directions for this and other studies.

2. Materials and Methods

Our workflow was divided into five main stages (Figure 1). In (a) we have the RGB-imagery acquisition by an Airplane flight performed in the metropolitan region of Campo Grande, State of Mato Grosso do Sul, Brazil, in the heart of the Cerrado Biome. (b) presents a geometric correction of the images and orthophotos generation. In (c), annotation of the trees in the orthophotos, and preparation of the data by splitting it into test and training subsets. In (d), evaluation of five state-of-the-art deep neural networks selected for the proposed task. Finally, in (e), comparison of the performance of the methods.

2.1. Data Acquisition and Image Processing

The imagery-dataset was acquired with an airplane flight performed in Campo Grande, Mato Grosso do Sul. This flight took place in 2013 and created a total of 1394 RGB orthoimages with 5619 × 5946 pixels each, with dimensions of 561.9 × 594.6 m and a ground sampling distance of 10 cm. It is important to note that the images are not

overlapped. Inside our study area, the urban forests are randomly distributed and mixed between the constructions (Figure 2). Inside the study area, there are a substantial diversity of tree species, many of them are not natives trees, as the region is a neighborhood area where citizens tend to plant the trees following some local set of rules as of the size of the tree, but besides that, they tend to plant trees that pleases then the most.

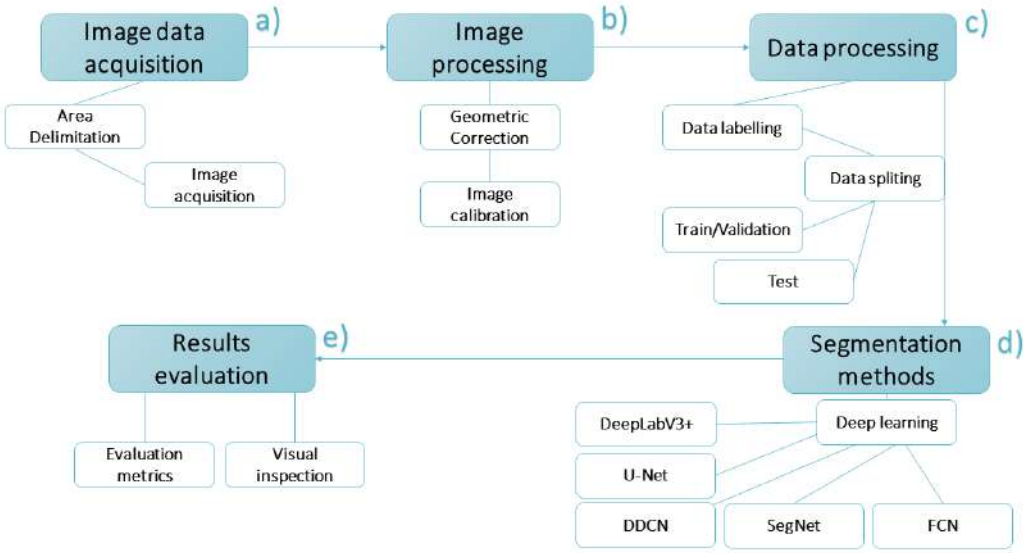


Figure 1. Workflow summarizing the fundamental steps of the conducted approach. Adapted from [43].

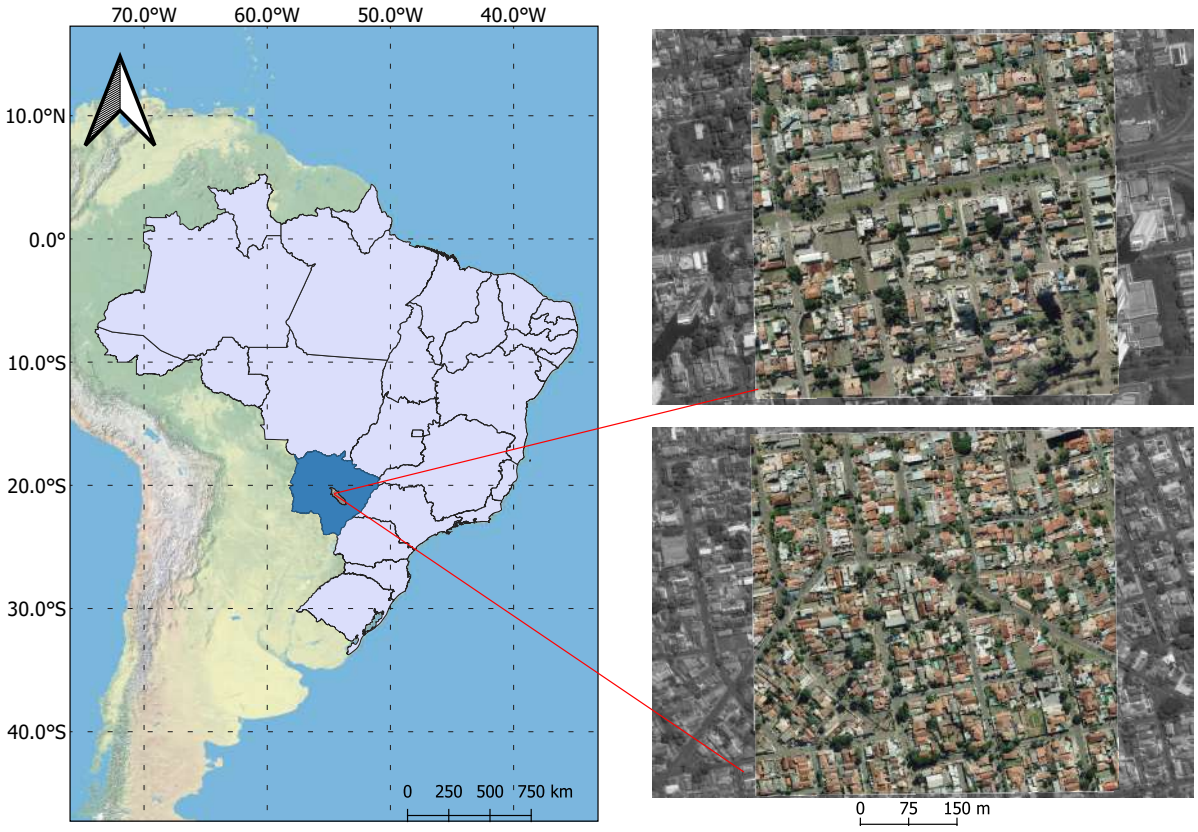


Figure 2. Overall visualization of the study area, the location is inside the metropolitan region of Campo Grande, Mato Grosso do Sul, Brazil. And also in color a visualization of the two base images used to make the labels of the tree canopies.

For the classification quality assessment of the deep learning algorithms, we select two images from this dataset and manually labeled all the trees presented in them, mapping all the tree cover in these two images. The images have 33.4 hectares (ha) each and a tree cover of 4.8 ha in one image and 3.8 ha in the other. All tress are randomly distributed in the area, making a ratio of 12.8% of the dataset composed of our target object and the rest being background, that is composed of roads, buildings, cars, houses, and many other elements that compose a typical urban environment, e.g., paths, districts, edges, nodes, and landmarks).

2.2. Semantic Segmentation Methods and Experimental Setup

This section presents five state-of-the-art deep learning architectures evaluated in our study case. These architecture were the same applied in the work of [43], our previous work with precision agriculture, but now we want to attend the urban forest context. To better visualize and analyze CNN's architectures, we organize the CNN structure illustrations with colored blocks. Each block represents a processing layer with the data as described in Figure 3.

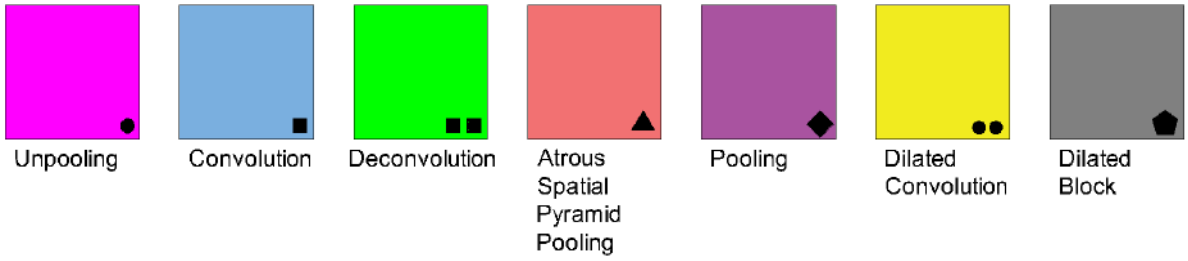


Figure 3. Building blocks used to illustrate the different CNNs layers.

Fully Convolutional Network (FCN): FCN architecture is presented in Figure 4. It was proposed by [44]. This deep network creates a classification map with a set of convolutional layers returning a spatially reduced result. After that, it applies deconvolution layers to upsample the initial classification and produce a dense prediction, restoring the image's original resolution.

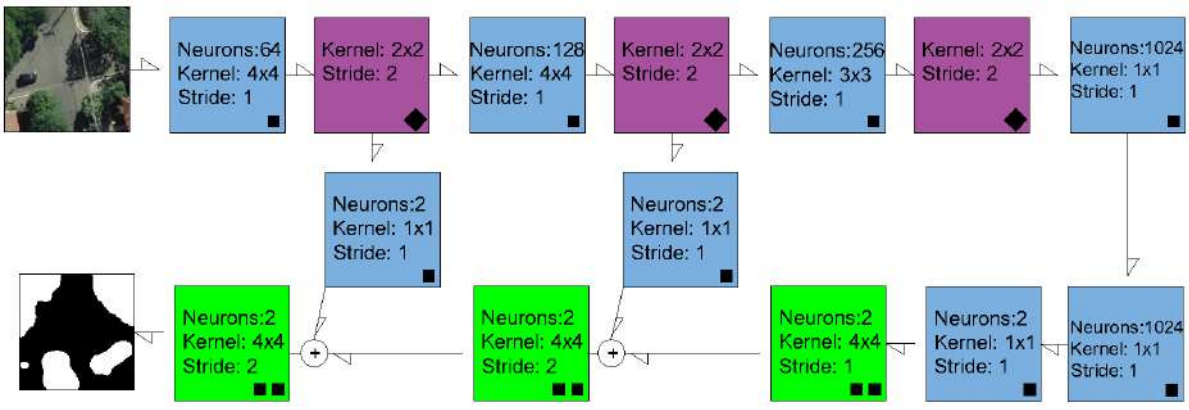


Figure 4. Fully Convolutional Network (FCN) architecture. Adapted from [43,44].

U-Net and SegNet: According to [43], the U-Net architecture was the first network to propose an encoder-decoder architecture to perform semantic segmentation tasks. This deep network was created by [45] to segment biomedical images. To generate an initial prediction map, are used the encoder and max-pooling layers with feature extraction. The encoder consists of a stack of convolution, and the decoder comprises convolutions, deconvolutions, and unpooling layers in a symmetrical expanding path, using deconvolution filters to up-sample the feature maps. An illustration of this architecture is presented

in (Figure 5). SegNet is also an encoder-decoder network path like U-Net, but with the replacement of the deconvolution layers by unpooling operations to increase the spatial resolution of the initial prediction map generated by the encoder. Ref. [46] proposed this architecture based on the VVG 16 network developed by [47]. A scheme of deep network SegNet is shown in (Figure 6).

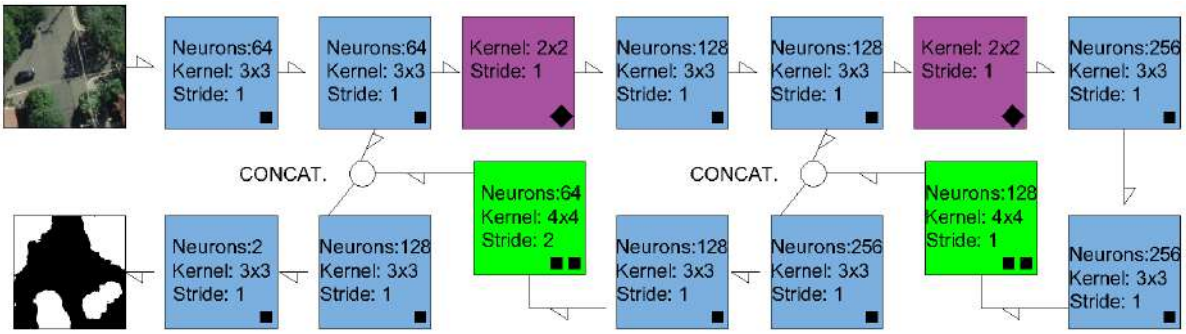


Figure 5. U-Net architecture. Adapted from [43,45].

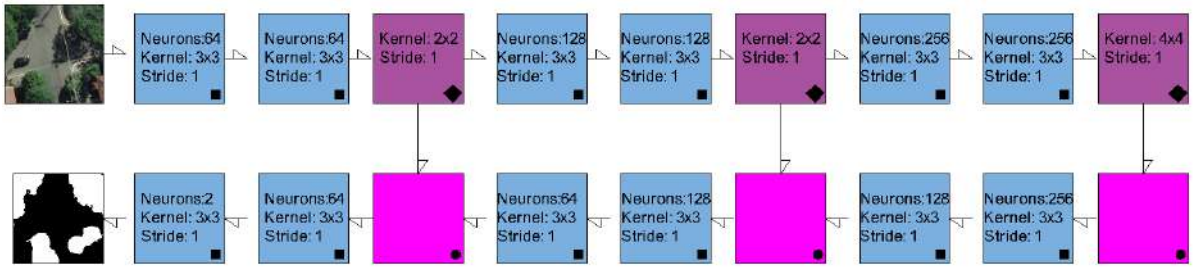


Figure 6. SegNet architecture. Adapted from [43,46].

DeepLabV3+: The DeepLabV3+ [48] starts with three blocks composed of two convolutions and one polling layer that performs the feature extraction and an initial prediction map. These features are then processed by a particular layer, called Atrous Spatial Pyramid Pooling (ASPP) introduced in [49]. This technique involves employing atrous convolution in parallel to extract features at multiple scales and alleviate the loss of spatial information due to prior pooling or convolutions with striding operations. The data is then processed with features extracted from the first pooling layer and refined by one extra convolutional layer. Then three convolutional layers process the concatenated segments upsampled by a bilinear interpolation producing the final prediction map. For more details, see Figure 7 [48].

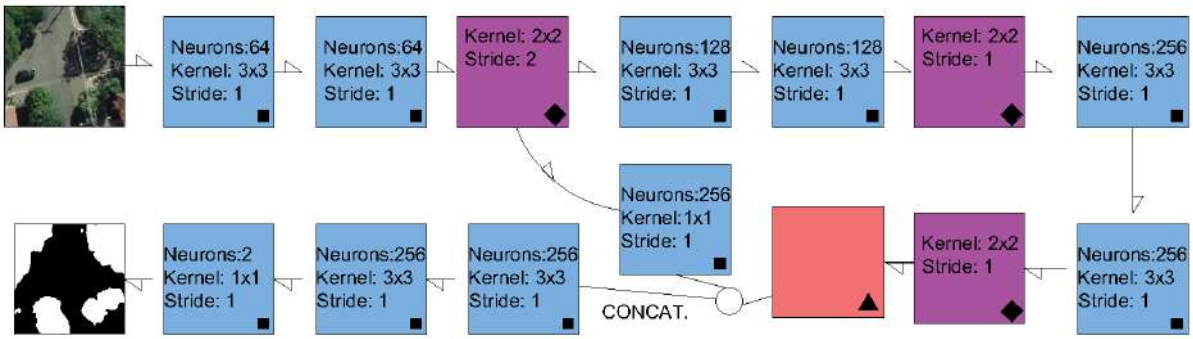


Figure 7. DeepLabV3+ architecture. Adapted from [43,48].

Dynamic Dilated Convolutional Network (DDCN): Proposed by [35], the DDCN is designed to preserve the input image resolution. Ref. [43] describe this network in more detail. However, in summary, the Dynamic Dilated Convolutional Network uses a multi-scale training strategy that implements dynamically-generated input images to converge a dilated model that does not downsample the input data due to a specific configuration of stride and padding. The model has eight dilated blocks; each block comprises a dilated convolution and a pooling layer; the blocks are followed by a standard convolutional layer responsible for the final prediction map. In each iteration of the training procedure, a dimension is randomly selected from this distribution and used to create a new batch. The model captures multi-scale information by processing these batches with a pre-determined size, advancing in the processing phase. The network selects, based on scores obtained during the training phase for each evaluated input size, the best image resolution. Then the DDCN processes the testing images using batches composed of the images with the best-evaluated size (Figure 8).

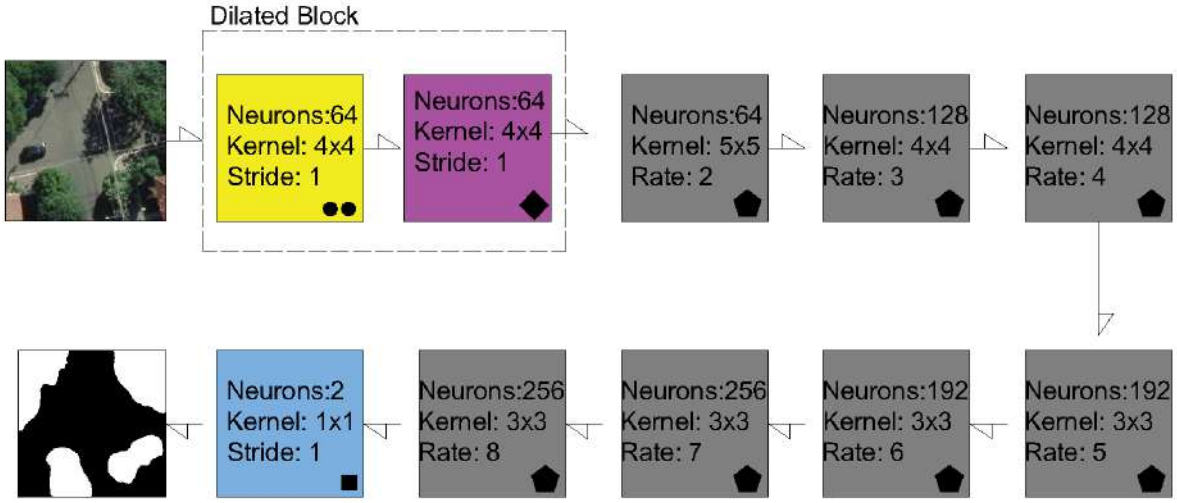


Figure 8. Dynamic Dilated Convolutional Network (DDCN) architecture. Adapted from [35,43].

2.2.1. Experimental details

We adopted the same training/trial protocol for all CNNs. All networks used in this research have been trained from scratch (i.e., without pre-trained weights from other datasets, like ImageNet, for example). We used 1938 input patch sizes of 256 × 256 pixels, with 388 patches for the test, 1162 for train, and 388 for validation, a distribution of approximately 20%, 60%, and 20% respectively. It is important to note that additional input patch sizes were tested in an experimental phase, but the results did not change considerably and only increased the training time. All approaches used the same set of hyperparameters during training, which was defined based on previous analyses. Specifically, the learning rate, weight decay, momentum, and iterations were 0.01, 0.005, 0.9, and 200,000, respectively. The model is optimized using stochastic gradient descent (SGD). To assist the convergence process and prevent overfitting (Overfitting is a concept used to refer to a model that adjusts too well to the training data, but it does not generalize to the unseen before dataset, i.e., a test dataset), after 50,000 iterations, the learning rate was reduced following an exponential decay parameter of 0.5 by an SGD scheduler. Aside from this, we used rotation, noise, and flip (as in [50]) for data augmentation, and we were capable of augmenting the dataset by six times. With the data augmentation technique, we can make the CNN classification more robust and generalize better. In Figure 9, we can see the schematic diagram for the evaluation process.

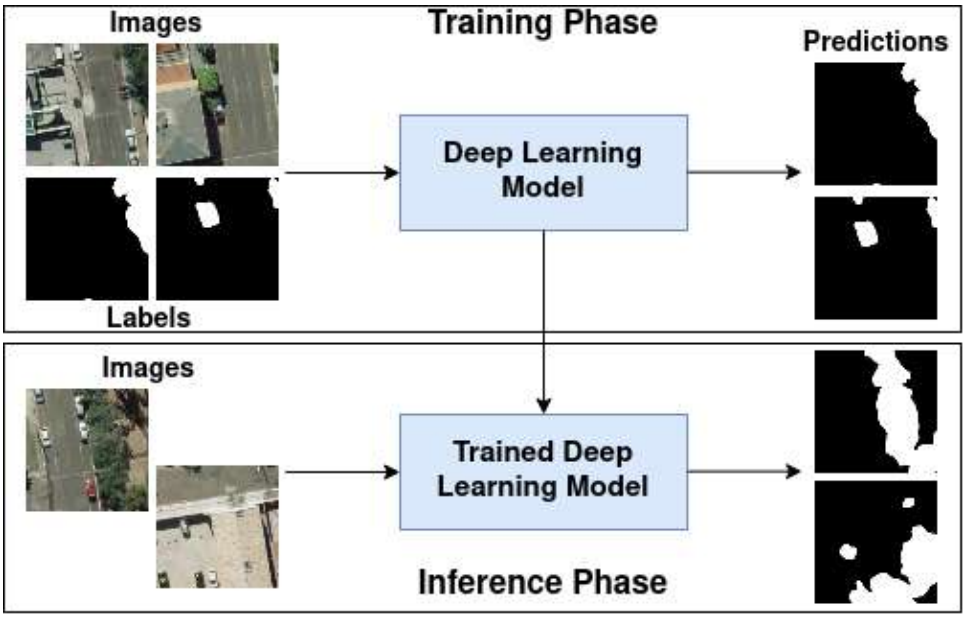


Figure 9. Schematic diagram of the evaluation process.

All deep learning-based models exploited in this work were implemented using the TensorFlow [51], a Python framework conceived to allow efficient analysis and implementation of deep learning with Graphics Processing Units (GPUs). All experiments conducted here were performed on a 64-bit Intel i7-8700K@3.70GHz CPU workstation, 64 GB memory, and NVIDIA® GTX 1080 GPU with 12Gb of memory, under a 10.0 CUDA version. Debian 4.195.98-1 version was used as the operating system.

2.2.2. Evaluation Metrics

The networks were evaluated using five different classification metrics: pixel accuracy, average Accuracy, F1-score, Kappa, IoU/Jaccard. The variables TP, TN, FP, FN stand for the number of true positives, true negatives, false positives and false negatives, respectively. In our analysis, *positives* and *negatives* refer to the pixels assigned by the underlying classifier to the trees and background classes, respectively. Such *positives* and *negatives* are *true* or $false$, depends on whether or not they agree with the ground truth, respectively.

The Pixel accuracy is given by:

$$Pix.Acc. = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The average accuracy (Av. Acc.) is the mean accuracy result given by class, in this case we have 2 classes tree and background.

The F1-score is given by:

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

where P and R stand for Precision and Recall, respectively, and are given by the ratios present in Equations (3) and (4):

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

The Cohen's Kappa statistic (K) proposed by [52], tells you how much better your classifier is performing over the performance of a classifier that guesses at random according

to the frequency of each class. The mathematical formulation of this metric is given by a balance between a positive and negative response, as follows:

$$Ppositive = \frac{TP + FN}{TP + TN + FP + FN} * \frac{TP + FP}{TP + TN + FP + FN} \quad (5)$$

$$Pnegative = \frac{TN + FN}{TP + TN + FP + FN} * \frac{TN + FP}{TP + TN + FP + FN} \quad (6)$$

$$Pe = Ppositive + Pnegative \quad (7)$$

$$K = \frac{Pix.Acc. - Pe}{1 - Pe} \quad (8)$$

The Union Intersect (IoU), also known as the Jaccard Index, is frequently used as a precision metric for semantic segmentation tasks [53,54]. In the Reference and the Prediction mask, IoU is indicated by a ratio of the number of pixels in both masks to the total number of pixels in:

$$IoU = \frac{|Reference \cap Prediction|}{|Reference \cup Prediction|} \quad (9)$$

3. Results

In this section, we present the results of the experimental evaluation of the selected semantic segmentation approaches. In (Section 3.1) we made a quantitative analysis with the metrics described in (Section 2.2.2), the (Section 3.3) presents a visual analysis of the segmentation outcomes, and in (Section 3.2), we assess the computational efficiency of each method.

3.1. Performance Evaluation

The evaluated deep learning methods returned similar results for the proposed approach. Ranging from 96.18% to 95.56% for pixel accuracy, 91.25% to 88.80% for Av. Acc., 91.40% to 89.91% for F1-score, 82.20% to 79.83% for Kappa and 73.89% to 70.01% for IoU. Results of the classification for the test set are presented in Table 1.

Table 1. Classifiers' results for each evaluation metrics using the test set of patches, while classifying tree canopies.

Set	Network	Pix. Acc.	Av. Acc.	F1-Score	Kappa	IoU
Test.	FCN	0.9614	0.9008	0.9123	0.8247	0.7342
	SegNet	0.9607	0.9125	0.9130	0.8260	0.7370
	U-Net	0.9597	0.9082	0.9104	0.8208	0.7301
	DDCN	0.9556	0.8880	0.8991	0.7983	0.7001
	DeepLabV3+	0.9618	0.9059	0.9140	0.8280	0.7389

A slight difference indicates that DeepLabV3+ was the best classifier method from a quantitative perspective, returning the best available results for the chosen metrics. In the recent few years, the architecture DeepLabv3+ has been regarded as state-of-the-art in semantic segmentation. So it is no surprise that it achieved the best performance among all tested architectures in our experiments, both in terms of absolute average accuracies as in terms of variability for the test set. Nevertheless, it also is accurate to say by analyzing the results presented in Table 1 that all of the five state-of-the-art networks are capable of segmenting trees inside a Cerrado urban environment in a satisfactory way with the proposed imagery dataset. These deep neural networks can separate tree covered-area

from other objects inside an urban environment while maintaining the original resolution of the image input is an important characteristic, making it possible to extract valuable information that could be used to support urban planning strategies.

Figure 10 presents the confusion matrix of all the CNN methods used in this research, which was used to derive the metrics presented in Table 1.

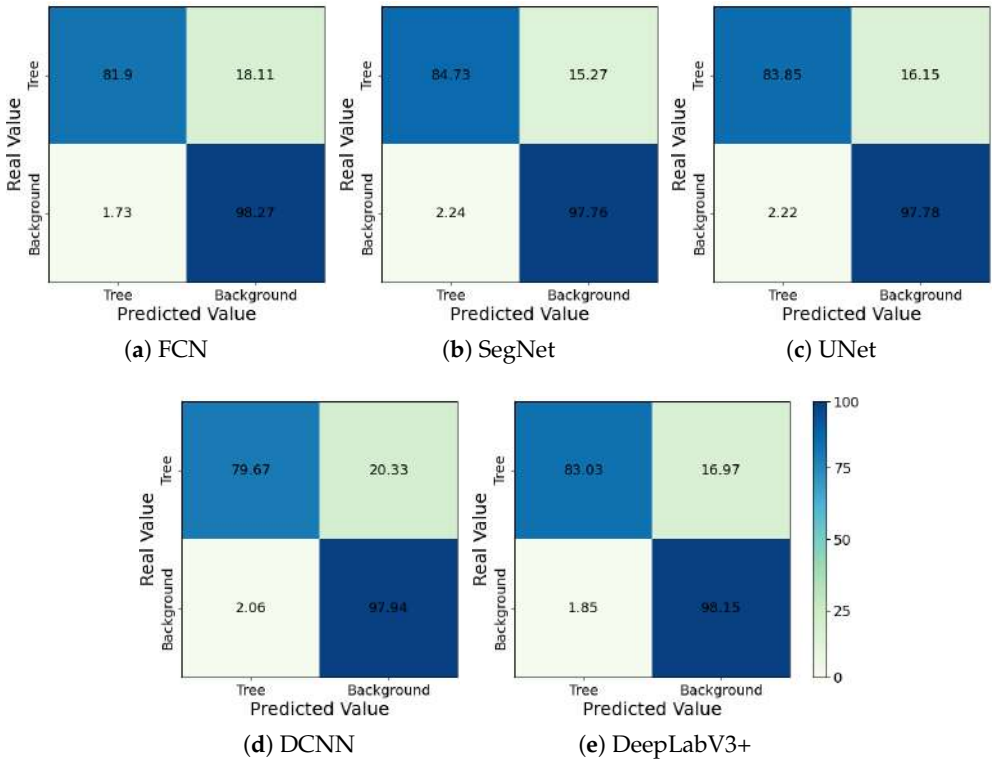


Figure 10. Confusion Matrix for all the analysed ConvNets.

Finally, we present in Figure 11 the accuracy and loss curves. We can see that the FCN, UNet, and SegNet performed similarly with stable slight variation and close to the minimum loss value after 100,000 iterations or half the way of the learning process established. The DCNN presented some increase in loss after 100,000 iterations. After the learning rate was reduced in the lasts 50,000 iterations by the SGD scheduler, it reached its minimum value, and the loss fluctuations were reduced. Moreover, the DeepLabV3+ became stable after 100,000 iterations with the minimum loss observed.

3.2. Computational Complexity

This section compares the methods in terms of computational efficiency and computational load for training and inference. Table 2 presents the average training and inference times measured on the hardware infrastructure described in Section 2.2. Considering that the methods were trained with the same optimizer and learning rates, these results are highly correlated with the network depth and the selected batch size. For instance, the DDCN network is deeper than the others, and the consequence is that it took longer than the other networks for training and inference.

The most significant variation of performance is concerning the number of parameters and with training and inference time. Despite being the best architecture in performance, According to Table 2, DeepLabv3+ needed more parameters than the other architectures, about 2.75 times more parameters than the U-Net, the least requiring one. The need for a more significant number of parameters often implies a higher demand for training samples that our dataset or another dataset may not have met that the methods present in this research paper may be applied, possibly causing the DeepLabV3+ architecture to perform below its potential.

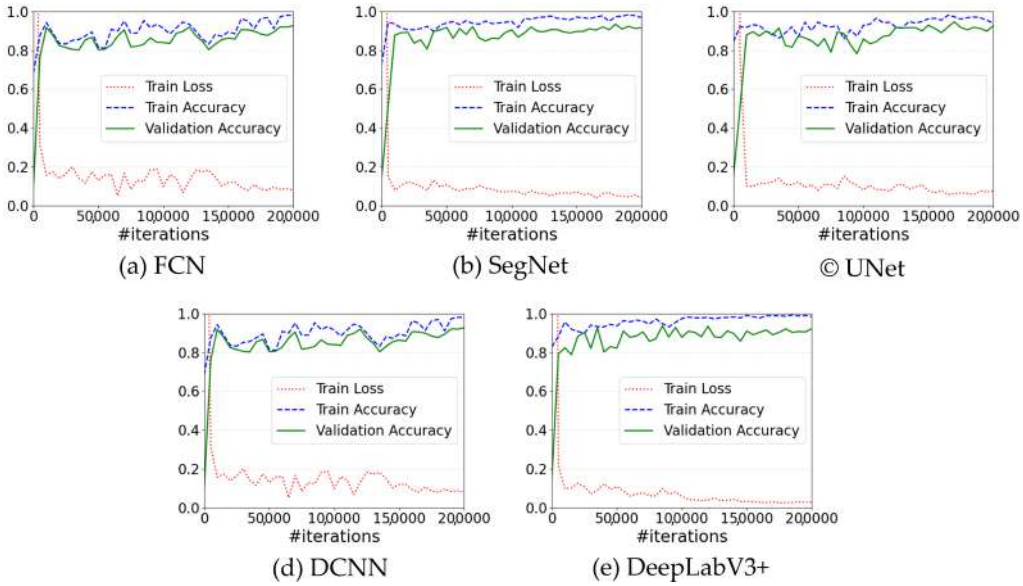


Figure 11. Convergence of the evaluated networks.

Table 2. Number of Parameters and Processing Time of the proposed approaches. The training time represents the results for the test set of each method. The inference time stands for the time taken by each model to make predictions for each image.

Method	FCN	U-Net	SegNet	DeepLabV3+	DDCN
Number of Parameters (in millions)	3.83	1.86	2.32	5.16	2.08
Training Time (GPU hours)	485	450	472	486	500
Inference Time (GPU min.)	1.4	1	1.1	1.4	5.1
Inference Time (CPU min.)	1.9	1.3	1.5	1.9	6.2
Inference Time (GPU min./ha)	0.042	0.030	0.033	0.042	0.153
Inference Time (CPU min./ha)	0.057	0.039	0.045	0.057	0.186

3.3. Visual Analysis

Some features maps, learned by the convolutional layers, are presented in Figure 12. Specifically, this image presents low-, mid- and high-level feature maps learned by the first, a middle, and the last layers of the networks, respectively. We can see the each CNN performs very differently from one another.

Figures 13 and 14 show an example of the results for the chosen methods with cropped images from our dataset that came to represent common occurrences of vegetation inside an urban environment. References are presented on the left row of the cropped images in the first line of both Figures, the annotated label is in the second line, and each subsequent line presents the segmentation produced by the CNN. We chose these images to represent the visual segmentation of the dataset because they are different in their content and represent common situations that CNNs may encounter while segmenting a tree in urban areas.

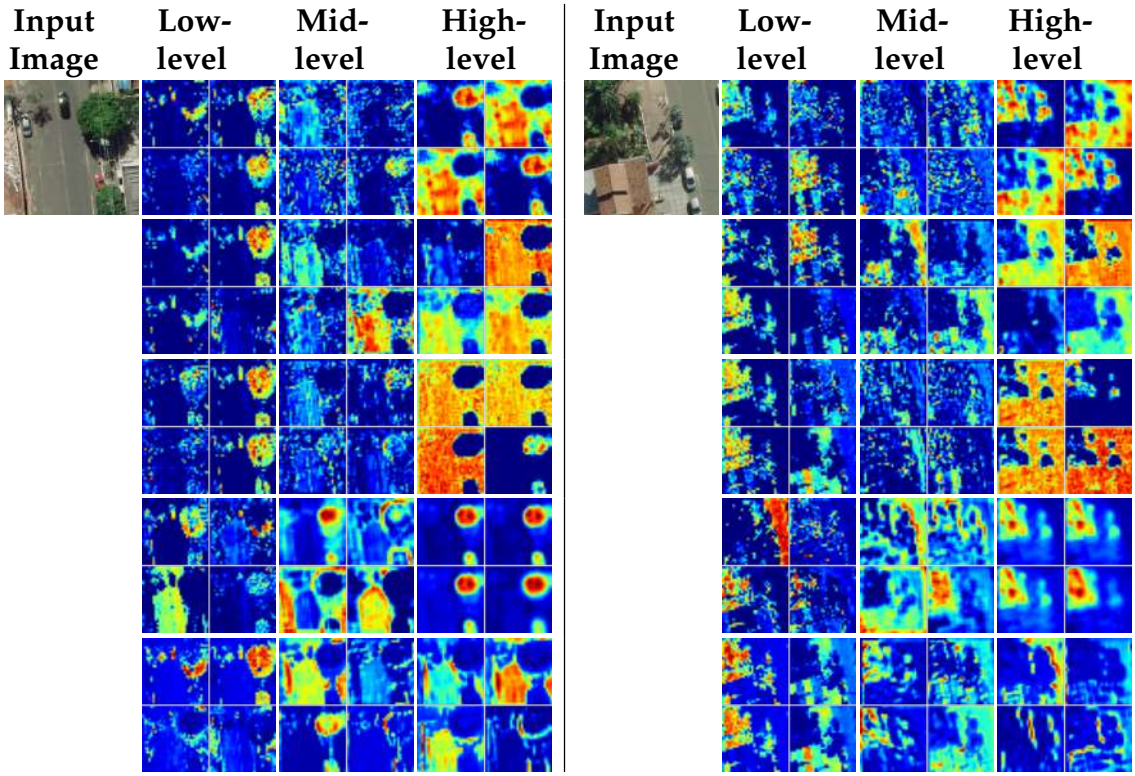


Figure 12. Input images and some produced (upsampled) feature maps extracted from initial, mid, and end layers of the networks. From top to bottom: FCN [44], UNet [45], Segnet [46], DeepLabV3+ [48], and DCNN [35]. The high level column presents an approximation of the final result of CNN classification.

In row (a) of Figure 13, there is a parking lot background. Inside this parking lot, we have one large group of trees with sparse distribution, one singular tree in the parking lot, one tree truncated in half by the picture frame, and a shadowed area at the top of the image. The white spots, or false negatives, compose a large area of the photograph, meaning that the CNNs missed a lot of our intended objective, classifying as background much of the trees in this area. However, it is essential to note that shadows were not a problem for any CNNs in this particular case. Another issue was related to the shape of the object. The SegNet, for example, returned a worse response related to it. All the CNNs localized the truncated tree. The errors mainly occurred at the canopies' edges, meaning that the CNNs missed information regarding vegetation borders. In row (b) of Figure 13, the image possesses a greater variety of backgrounds; a patch of grass, a patch of sidewalk, a patch of road, a car, and a patch of house, and more shadows. Furthermore, there are two large groups of trees with tension lines crossing them in this image, and even a single tree near one of the groups and a small one crop at the top of the image. This region, in particular, highlighted some of the flaws in the segmentation results. The DeepLabV3+ and the DDCN returned higher FN, especially in the shadowed areas. Still, most of the FN cases of this patch are inside the tree group at the bottom, meaning that the CNNs had a hard time separating singular trees in this context. In row (c) of Figure 13, there is a variety of backgrounds as diverse as in row (b), but a uniform distribution of trees inside the image, with four medium size and contiguous groups of canopies. Here, the CNN's presented fewer mistakes than on the other examples, as most of the errors were FP in the edges of large groups of trees; with this example, we see that the CNNs perform better in patches with groups of trees no isolated trees nearby.

In row (a) of Figure 14, we have large tree canopies and a road with a rooftop background, and all the CNNs performed reasonably well with large canopies segmentation except for U-Net and the FCN that faced a difficult time segmenting the shadowed areas in this case. In row (b) of Figure 14, we have a large patch of trees that are beneath a residential

area, and the networks performed similarly in this case of one large group of trees even for the false positives case of the shadows, but we can see that in this case, DeepLabV3+ was the most assertive CNN. In row (c) of Figure 14, we have a small patch of trees located beneath houses and cars. As discussed previously, we have the most significant part of the errors with the isolated and small trees, shadowed areas, and grassed and bush areas.

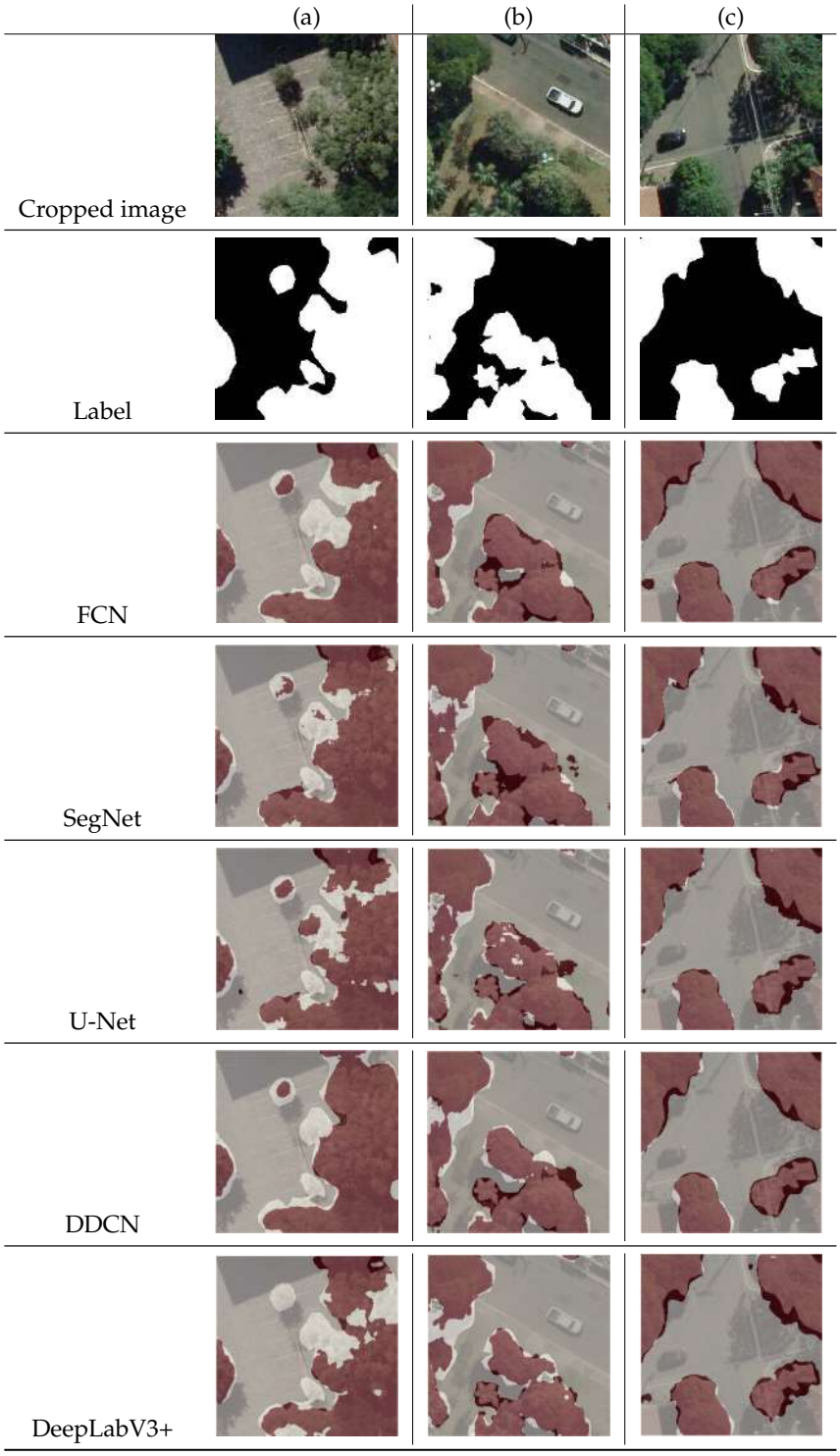


Figure 13. Cropped images, labeled canopies and the resulting map for the cropped image with the evaluated CNNs. The light red areas are the true positives (TP), the soft grey areas are the true negatives (TN), the white spots are the false negatives (FN), and the dark red areas are the false positives (FP).

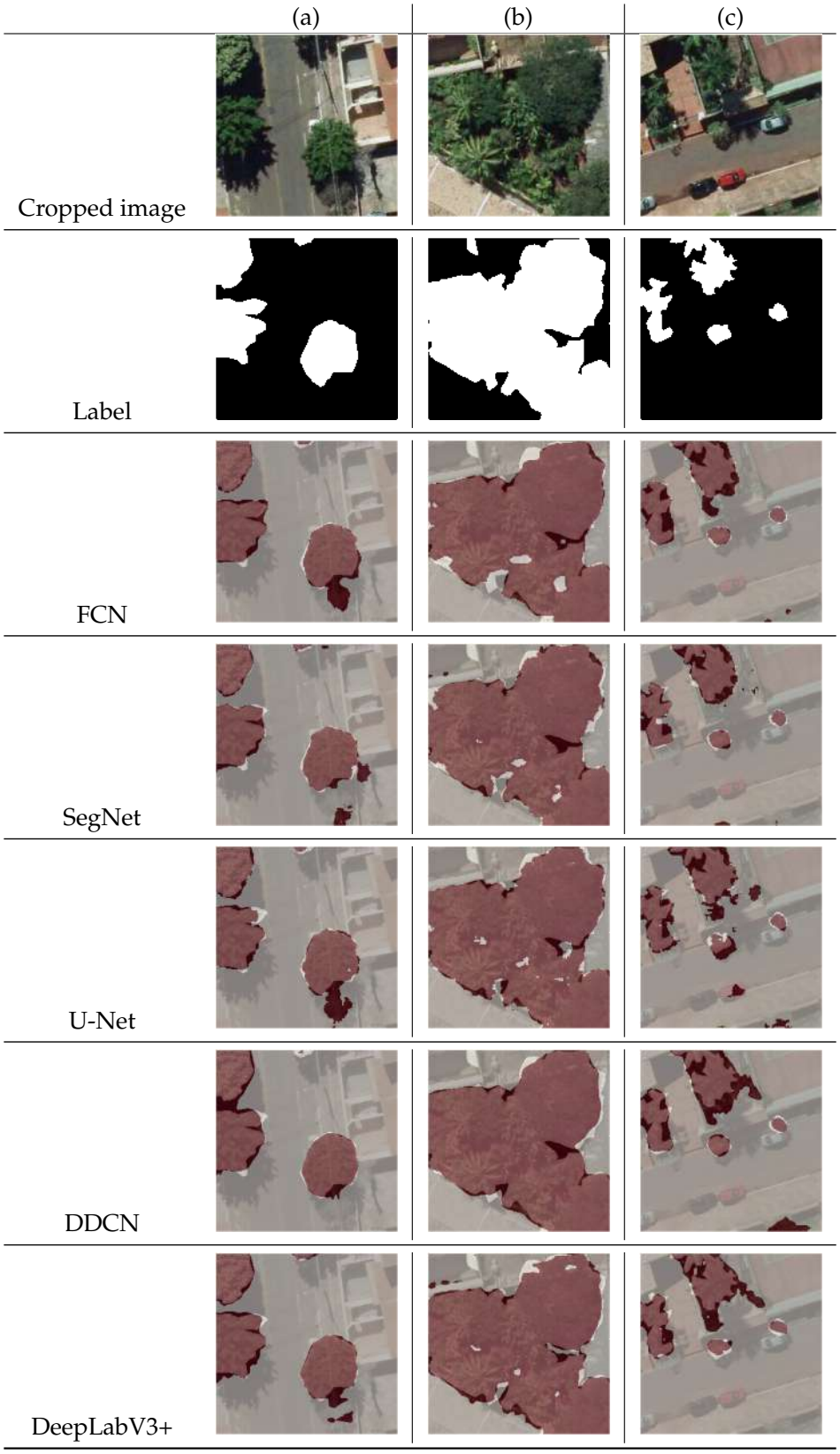


Figure 14. Cropped images, labeled canopies and the resulting map for the cropped image with the evaluated CNNs. The light red areas are the true positives (TP), the soft grey areas are the true negatives (TN), the white spots are the false negatives (FN), and the dark red areas are the false positives (FP).

4. Discussion

A set of state-of-the-art deep learning semantic segmentation approaches was applied to map urban forests in high-resolution RGB imagery in the urban context. Our results indicated that the investigated methods performed fairly similarly in this task, returning a pixel accuracy between 96.18% (DeepLabV3+) and 95.56% (DDCN), an average accuracy between 96.07% (DeepLabV3+) and 95.56% (DDCN), a F1-score between 91.40% (DeepLabV3+) and 89.91% (DDCN), a Kappa 82.80% (DeepLabV3+) and 79.83% (DDCN) and IoU between 73.89% (DeepLabV3+) and 70.01% (DDCN). Visually evaluating the classification map obtained with each network, it is difficult to emphasize an overall better method. Nonetheless, we have a quantitative advantage for the DeepLabV3+. This CNN also presented a satisfactory visual result with little noise and false-positives rates regardless of tree detection. Despite the quantitative results, the DDCN also presented a smooth visual result.

Most of the evaluated methods returned proximal inference time for both GPU and CPU tests, except for the DDCN method [35], which took around four times the amount of time needed to perform the same task. U-Net consistently presented lower inference times, being the fastest method among all for training and prediction. However, an estimation of this inference time per area demonstrates how rapidly these neural networks can segment trees in the given data set once they are trained. This information is vital for precision image segmentation tasks since this response could be incorporated into decision-making strategies regarding area size and priority. It should also be noted that the times informed here are considering the system used to train these methods, see Section 2.2.1.

For a practical approach, the final use of this method is the application in a city to detect vegetated areas. For our example images, we have a range of pixel accuracy of 96.18% to 95.56%. These remaining percentage not correctly classified in our test dataset is explained in Section 3.3 as the CNNs are known for producing errors in image-boundaries [55,56]. Moreover, most of the problems faced by the investigated deep networks are related to shadowy areas, isolated and small trees, and the grass and bush areas inside the images. Things for the CNN and even for the human operator can be a little ambiguous. Some solutions for these kinds of problems are the labeled dataset be manually segmented by more than one operator, creating a more detailed and overlapped map of the trees inside an area, we also can augment the dataset using techniques, such as the presented in the Section 2.2.1. The analysis of the classification results shown in Table 1 demonstrates that the deep neural networks are lacking accuracy if compared to our annotations; the result of the IoU metric demonstrates this. Regardless, the visual output of the segmentation methods is rather noisy in some of the patches. Shadows were a problem, even more, when mixed inside large groups of trees; when they are further away from our object of interest, CNNs have fewer issues in segmenting them. However, small isolated trees were a higher challenge for all the networks to deal with, especially when they are inside the same patch as large groups of trees; all the CNNs tend to miss-classify and even ignore them. The grass and bush areas were more of a problem when they were near large groups of trees, and the CNNs might interpret them as a continuation of the tree patches and classify them as trees, creating false positives.

Possible solutions to contour this problem are to use a higher resolution RGB imagery dataset and use different kinds of datasets in conjunction. For example, we can use RGB and LiDAR fusion to create an exact 3D point cloud; these approaches are a hot topic in autonomous driving vehicles [57,58], for the reason that they have an excellent capacity for accurate and fast scenery reconstruction. Another approach is the use of image segmentation methods with multi-spectral, and high spatial resolution sensors, the exploration of the spectral index for the detection and analysis of vegetation is a mature topic in remote sensing [17,22,59] and it can also be implemented with Deep learning approaches inside urban areas for semantic segmentation tasks.

5. Conclusions

We evaluated five state-of-the-art Convolutional Neural Networks for the semantic segmentation of urban forests using airborne high spatial-resolution RGB images. The architectures tested were: Fully Convolutional Network, U-Net, SegNet, Dynamic Dilated Convolution Network, and DeepLabV3+. The experimental analysis showed the effectiveness of the methods reporting a very proximal value for all the classifiers, with a mean for Pixel accuracy of 96.11% average pixel accuracy of 90.70%, F1-score of 91.27%, Kappa of 82.54% and IoU of 73.56%. With the best results being from the DeepLabV+3 architecture. Our research confirmed that CNNs could segment urban trees from high spatial resolution remote sensing imagery in the urban context. However, the networks still possess some limitations in the urban environment.

For future directions and development, we intend to improve the develop methods to improve the IoU mainly because it was our worst-performing metric. For this, we suggest creating a supplementary labeling phase created by a different human operator of the same area and merging the labels. As we have more correct labels, human error cases will be minimized in the labeling phase, proving a better feature map for the CNNs to work. The labeling phase of a large and complex image is a classical case of estimation error, and to overpass it, another stage of the labeling process is then suggested as an alternative.

We further intend to test the generalization and the transferability of the CNN architectures on datasets from different regions, as the urban landscape are diverse in composition applying the concept of domain adaptation. Future studies should also involve differentiate tree species, and to perform data fusion with multiple sensors data, such as Li-DAR or multi-spectral data in addition to the optical RGB data.We also suggest exploring instance segmentation (detection and segmentation) architectures such as Mask RCNN, Detectron2, FCIS, BlendMask, and YOLACT. These approaches are also of interest to urban planning applications, because they can differentiate the urban landscape by object, contributing to the tree classification by species.

Author Contributions: Methodology, W.N.G., J.M.J., P.T.S.d.O., L.P.O. and J.A.C.M.; software, J.A.d.S. and K.N.; validation, W.N.G., J.M.J., J.A.d.S., K.N., A.P.M.R. and V.L.; formal analysis, V.L., K.N. and J.M.J.; investigation, L.P.O. and J.A.C.M.; resources, V.L. and J.M.J.; data curation, W.N.G., J.M.J. and J.A.C.M.; writing—original draft preparation, J.A.C.M. and L.P.O.; writing—review and editing, V.L., A.P.M.R., F.D.G.G., D.A.S., D.E.G.F., K.N. and J.M.J.; visualization, J.A.C.M. and K.N.; supervision, W.N.G., V.L. and J.M.J.; project administration, W.N.G. and J.M.J.; funding acquisition, J.M.J., V.L. and W.N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Brazilian National Council for Scientific and Technological Development (CNPq) p: (310517/2020-6, 303559/2019-5, 313887/2018-7, 433783/2018-4 and 304173/2016-9), the Coordination for the Improvement of Higher Education Personnel (CAPES) Print p: (88881.211850/2018-01), and the Foundation to Support the Development of Education, Science and Technology of the State of Mato Grosso do Sul (FUNDECT) p: (59/300.066/2015 and 59/300.095/2015).

Data Availability Statement: The data presented in this study are openly available, access through link: https://sites.google.com/view/geomatics-and-computer-vision/home/datasets (accessed on 16 July 2021).

Acknowledgments: We would like to thank the Graduate Program of Environmental Technologies of the Federal University of Mato Grosso do Sul (UFMS) to support the doctoral dissertation of the first author and the Coordination for the Improvement of Higher Education Personnel (CAPES). We also would like to thank the editors and three reviewers for providing constructive concerns and suggestions. Such feedback helped us improve the quality of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Urbanization Prospects—Population Division—United Nations. 2018. Available online: https://population.un.org/wup/Publications/Files/WUP2018-Highlights.pdf (accessed on 16 July 2021).
2. La Rosa, D.; Wiesmann, D. Land cover and impervious surface extraction using parametric and non-parametric algorithms from the open-source software R: An application to sustainable urban planning in Sicily. *GISci. Remote Sens.* **2013**. [CrossRef]
3. Jennings, V.L.L.; Yun, J. Advancing Sustainability through Urban Green Space: Cultural Ecosystem Services, Equity, and Social Determinants of Health. *Int. J. Environ. Res. Public Health* **2016**, *13*, 196. [CrossRef] [PubMed]
4. Arantes, B.L.; Castro, N.R.; Gilio, L.; Polizel, J.L.; da Silva Filho, D.F. Urban forest and per capita income in the mega-city of Sao Paulo, Brazil: A spatial pattern analysis. *Cities* **2021**, *111*, 103099. [CrossRef]
5. Jim, C.; Chen, W.Y. Ecosystem services and valuation of urban forests in China. *Cities* **2009**, *26*, 187–194. [CrossRef]
6. Chen, W.Y.; Wang, D.T. Urban forest development in China: Natural endowment or socioeconomic product. *Cities* **2013**, *35*, 62–68. [CrossRef]
7. Baró, F.; Chaparro, L.; Gómez-Baggethun, E.; Langemeyer, J.; Nowak, D.J.; Terradas, J. Contribution of ecosystem services to air quality and climate change mitigation policies: The case of urban forests in Barcelona, Spain. *Ambio* **2014**. [CrossRef]
8. McHugh, N.; Edmondson, J.L.; Gaston, K.J.; Leake, J.R.; O'Sullivan, O.S. Modelling short-rotation coppice and tree planting for urban carbon management—A citywide analysis. *J. Appl. Ecol.* **2015**. [CrossRef]
9. Kardan, O.; Gozdyra, P.; Misic, B.; Moola, F.; Palmer, L.J.; Paus, T.; Berman, M.G. Neighborhood greenspace and health in a large urban center. *Sci. Rep.* **2015**. [CrossRef] [PubMed]
10. Feng, Q.; Liu, J.; Gong, J. UAV Remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sens.* **2015**, *7*, 1074–1094. [CrossRef]
11. Alonzo, M.; McFadden, J.P.; Nowak, D.J.; Roberts, D.A. Mapping urban forest structure and function using hyperspectral imagery and lidar data. *Urban For. Urban Green.* **2016**, *17*, 135–147. [CrossRef]
12. Liisa, T.; Stephan, P.; Klaus, S.; de Vries S. *Benefits and Uses of Urban Forests and Trees*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 81–114. [CrossRef]
13. Song, X.P.; Hansen, M.; Stehman, S.; Potapov, P.; Tyukavina, A.; Vermote, E.; Townshend, J. Global land change from 1982 to 2016. *Nature* **2018**, 639–643. [CrossRef] [PubMed]
14. McGrane, S.J. GImpacts of urbanisation on hydrological and water quality dynamics, and urban water management: A review. *Hydrol. Sci. J.* **2015**, *61*, 2295–2311. [CrossRef]
15. Schneider, A.; Friedl, M.A.; Potere, D. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sens. Environ.* **2010**, *114*, 1733–1746. [CrossRef]
16. Fassnacht, F.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **2016**, *186*, 64–87. [CrossRef]
17. Onishi, M.; Ise, T. Automatic classification of trees using a UAV onboard camera and deep learning. *arXiv* **2018**, arXiv:1804.10390.
18. Jensen, R.R.; Hardin, P.J.; Bekker, M.; Farnes, D.S.; Lulla, V.; Hardin, A. Modeling urban leaf area index with AISA+ hyperspectral data. *Appl. Geogr.* **2009**, *29*, 320–332. [CrossRef]
19. Lausch, A.; Erasmi, S.; King, D.J.; Magdon, P.; Heurich, M. Understanding forest health with Remote sensing-Part II-A review of approaches and data models. *Remote Sens.* **2017**, *9*, 129. [CrossRef]
20. Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**. [CrossRef]
21. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [CrossRef]
22. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens.* **2017**, *9*, 1110. [CrossRef]
23. Arfaoui, A. *Unmanned Aerial Vehicle: Review of Onboard Sensors, Application Fields, Open Problems and Research Issues*; Technical Report; 2017. Available online: https://www.researchgate.net/publication/315076314_Unmanned_Aerial_Vehicle_Review_of_Onboard_Sensors_Application_Fields_Open_Problems_and_Research_Issues (accessed on 16 July 2021).
24. Shojanoori, R.; Shafri, H.Z. Review on the use of remote sensing for urban forest monitoring. *Arboric. Urban For.* **2016**, *42*, 400–417.
25. Alonzo, M.; Bookhagen, B.; Roberts, D. Urban tree species mapping using hyperspectral and LiDAR data fusion. *Remote Sens. Environ.* **2014**, *148*, 70–83. [CrossRef]
26. Osco, L.P.; Ramos, A.P.M.; Pereira, D.R.; Moriya, É.A.S.; Imai, N.N.; Matsubara, E.T.; Estrabis, N.; de Souza, M.; Junior, J.M.; Gonçalves, W.N.; et al. Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. *Remote Sens.* **2019**, *11*, 2925. [CrossRef]
27. Osco, L.P.; de Arruda, M.S.; Marcato Junior, J.; da Silva, N.B.; Ramos, A.P.M.; Moryia, É.A.S.; Imai, N.N.; Pereira, D.R.; Creste, J.E.; Matsubara, E.T.; et al. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**. [CrossRef]
28. Martins, J.; Junior, J.M.; Menezes, G.; Pistori, H.; Sant'Ana, D.; Goncalves, W. Image Segmentation and Classification with SLIC Superpixel and Convolutional Neural Network in Forest Context. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 6543–6546. [CrossRef]

29. dos Santos Ferreira, A.; Matte Freitas, D.; Gonçalves da Silva, G.; Pistori, H.; Theophilo Folhes, M. Weed detection in soybean crops using ConvNets. *Comput. Electron. Agric.* **2017**, *143*, 314–324. [CrossRef]
30. Torres, D.L.; Feitosa, R.Q.; Happ, P.N.; La Rosa, L.E.C.; Junior, J.M.; Martins, J.; Bressan, P.O.; Gonçalves, W.N.; Liesenberg, V. Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. *Sensors* **2020**, *20*, 563. [CrossRef]
31. Zhang, Q.; Xu, J.; Xu, L.; Guo, H. Deep Convolutional Neural Networks for Forest Fire Detection. In Proceedings of the 2016 International Forum on Management, Education and Information Technology Application, Guangzhou, China, 30–31 January 2016; pp. 568–575. [CrossRef]
32. Bazi, Y.; Melgani, F. Convolutional SVM Networks for Object Detection in UAV Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3107–3118. [CrossRef]
33. Zhao, X.; Yuan, Y.; Song, M.; Ding, Y.; Lin, F.; Liang, D. Use of Unmanned Aerial Vehicle Imagery and Deep Learning UNet to Extract Rice Lodging. *Sensors* **2019**, *19*, 3859. [CrossRef]
34. Ganesh, P.; Volle, K.; Burks, T.F.; Mehta, S.S. Deep Orange: Mask R-CNN based Orange Detection and Segmentation. *IFAC-PapersOnLine* **2019**. [CrossRef]
35. Nogueira, K.; Dalla Mura, M.; Chanussot, J.; Schwartz, W.R.; Dos Santos, J.A. Dynamic multicontext segmentation of remote sensing images based on convolutional networks. *IEEE Trans. Geosci. Remote Sens.* **2019**. [CrossRef]
36. Zamboni, P.; Junior, J.M.; Silva, J.d.A.; Miyoshi, G.T.; Matsubara, E.T.; Nogueira, K.; Gonçalves, W.N. Benchmarking Anchor-Based and Anchor-Free State-of-the-Art Deep Learning Methods for Individual Tree Detection in RGB High-Resolution Images. *Remote Sens.* **2021**, *13*, 2482. [CrossRef]
37. Pestana, L.; Alves, F.; Sartori, Â. Espécies arbóreas da arborização urbana do centro do município de campo grande, mato grosso do sul, brasil. *Rev. Soc. Bras. Arborização Urbana* **2019**, *6*, 1–21. [CrossRef]
38. Ososkov, G.; Goncharov, P. Shallow and deep learning for image classification. *Opt. Mem. Neural Netw.* **2017**, *26*, 221–248. [CrossRef]
39. Walsh, J.; O' Mahony, N.; Campbell, S.; Carvalho, A.; Krpalkova, L.; Velasco-Hernandez, G.; Harapanahalli, S.; Riordan, D. Deep Learning vs. Traditional Computer Vision. *Tradit. Comput. Vis.* **2019**. [CrossRef]
40. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C.; et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *Lancet Digit. Health* **2019**, *1*, e271–e297. [CrossRef]
41. Bui, D.T.; Tsangaratos, P.; Nguyen, V.T.; Liem, N.V.; Trinh, P.T. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *CATENA* **2020**, *188*, 104426. [CrossRef]
42. Sujatha, R.; Chatterjee, J.M.; Jhanjhi, N.; Brohi, S.N. Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocess. Microsyst.* **2021**, *80*, 103615. [CrossRef]
43. Osco, L.P.; Nogueira, K.; Ramos, A.P.M.; Pinheiro, M.M.F.; Furuya, D.E.G.; Gonçalves, W.N.; de Castro Jorge, L.A.; Junior, J.M.; dos Santos, J.A. Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. *Precis. Agric.* **2021**. [CrossRef]
44. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015. [CrossRef]
45. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015. [CrossRef]
46. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
47. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
48. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**. [CrossRef] [PubMed]
49. Chen, S.W.; Shivakumar, S.S.; Dcunha, S.; Das, J.; Okon, E.; Qu, C.; Taylor, C.J.; Kumar, V. Counting Apples and Oranges with Deep Learning: A Data-Driven Approach. *IEEE Robot. Autom. Lett.* **2017**. [CrossRef]
50. Volpi, M.; Tuia, D. Dense Semantic Labeling of Subdecimeter Resolution Images with Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 881–893. [CrossRef]
51. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 16 July 2021).
52. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
53. Wu, Z.; Gao, Y.; Li, L.; Xue, J.; Li, Y. Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. *Connect. Sci.* **2019**. [CrossRef]
54. Berman, M.; Triki, A.R.; Blaschko, M.B. The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
55. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

56. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 16 July 2021).
57. Madawy, K.E.; Rashed, H.; Sallab, A.E.; Nasr, O.; Kamel, H.; Yogamani, S. Rgb and lidar fusion based 3d semantic segmentation for autonomous driving. *arXiv* **2019**, arXiv:1906.00208.
58. Zhao, X.; Sun, P.; Xu, Z.; Min, H.; Yu, H. Fusion of 3D LIDAR and Camera Data for Object Detection in Autonomous Vehicle Applications. *IEEE Sens. J.* **2020**, *20*, 4901–4913. [CrossRef]
59. Zheng, G.; Moskal, L.M. Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors* **2009**, *9*, 2719–2745. [CrossRef] [PubMed]

CHAPTER 4

Paper 2: Identifying plant species in kettle holes using UAS images and deep learning techniques

Manuscript published in *Remote Sensing in Ecology and Conservation* journal

Identifying plant species in kettle holes using UAV images and deep learning techniques

José Augusto Correa Martins[1] | José Marcato Junior[1] | Marlene Pätzig[2] | Diego André Sant'Ana[3,4] | Hemerson Pistori[3] | Veraldo Liesenberg[5] | Anette Eltner[6]

1 Universidade Federal de Mato Grosso do Campo Sul, Brazil
2 Provisioning of Biodiversity in Agricultural Systems, Leibniz Centre for Agricultural Landscape Research (ZALF) e.V., Germany
3 Universidade Católica Dom Bosco, Campo Grande, Brazil
4 Instituto Federal de Mato Grosso do Sul, Aquidauana, Brazil
5 Department of Forest Engineering, Santa Catarina State University (UDESC), Lages, Santa Catarina, Brazil
6 Institute of Photogrammetry and Remote Sensing, Technische Universität Dresden, Germany

Abstract
The use of uncrewed aerial vehicle (UAV) to map the environment increased significantly in the last decade enabling a finer assessment of the land cover. However, creating accurate maps of the environment is still a complex and costly task. Deep learning is a new generation of artificial neural network research that, combined with remote sensing techniques, allows a refined understanding of our environment and can help to solve challenging land cover mapping issues. This research focuses on the vegetation segmentation of kettle holes. Kettle holes are small, pond-like, depressional wetlands. Quantifying the vegetation and
and biodiversity present in this environment is essential to assess the health of the ecosystems services. A machine learning workflow has been developed, integrating a superpixel segmentation algorithm to build a robust dataset, which is followed by a set of deep learning architectures to classify eleven plant classes present in kettle holes. The best architecture for this task was Xception, which achieved an average F1-score of 85\% in the detection and segmentation of the species. The application of solely 318 samples per class enabled a successful mapping in the complex wetland environment, indicating an important direction for future health assessments in such landscapes.

Keywords: Plant species segmentation, deep learning, uncrewed aerial vehicle (UAV), superpixels, image segmentation, wetland

1 | INTRODUCTION

Uncrewed aerial vehicle (UAV) in combination with automatic image classification offer a great tool to monitor vegetation, plant diversity, and related plant traits in high spatial and temporal resolution (Anderson and Gaston (2013); Bendig et al. (2015); Sotille et al. (2020); Van Iersel et al. (2018)). This technical-digital combination benefits small-scale ecosystems, transition zones, and habitat patches with high spatio-temporal complexity. The possibility of monitoring such landscape features created a growing demand of discretization of ecological features and regions and provided new potential for conservation planning and political management of environmental areas of interest (Foster et al.

(2003); Azpiroz et al. (2012); Dobson et al. (1997)). One of these landscape elements that have high ecosystem importance are kettle holes (Figure 1). Kettle holes can be described as depression wetlands that are mostly smaller than one hectare and mainly occur at high densities in the agricultural landscapes of the young moraine areas (Kalettka and Rudat (2006); Pätzig et al. (2012)). They provide a wide range of ecosystem services, such as the improvement of the hydrological cycle, flood control, chemical condition of fresh waters, biotic remediation of wastes acting as an biological filter and shelter for local biodiversity (Vasić et al. (2020); Pätzig et al. (2012)). Most of the kettle holes experience a seasonal wet-dry cycle and exhibit high potential for the presence of large biological species diversity. Thus, due to their position within farmed areas and their high perimeter-area ratio, they are highly vulnerable and severely impacted by intensive agricultural practices occurring in their surroundings. They cover up to 5% of the arable land of Germany (Pätzig et al. (2012); Vahrson and Frielinghaus (1998)). There is a demand for a better understanding of the impacts of such activities and for a proposition of solutions for a better conservation of these habitats by a continuation of sustainable agricultural activities. Furthermore, the functioning and ecosystem services and disservices of kettle holes as a result of complex interaction of abiotic and biotic processes within kettle holes and among their surroundings needs to be comprehended.

Regular high-resolution monitoring allows a better understanding of the functioning of these small-scale ecosystems and promote the development of effective conservation and management measures. Photogrammetry workflows that allow an automated high precision detection of dominant plants species are valuable tools to investigate the ecosystem health. Much of the kettle hole area is most of the time not flooded, especially during the peak vegetation period at the end of the summer. Therefore, a large percentage of the area is covered with emerged plant species such as amphibious or terrestrial plants and can be detected with aerial images.

Efficient workflows for automated vegetation segmentation with UAV imagery already exist and present promising results (Casado et al. (2015); Cruzan et al. (2016); Torres-Sánchez et al. (2014)). They are often based on RGB and multi-spectral images in combination with shallow or Deep learning (DL) techniques (Turner et al. (2018); Chabot et al. (2018); Cen et al. (2019)). DL is a set of techniques that allow computers to learn complex patterns from data. These techniques can adapt to changing environments and enable ongoing improvement to understand the scene. With the advance of mapping and computer technologies, DL methods are being tested to identify patterns in the most varied fields of science Shen et al. (2017); Dos Santos Ferreira et al. (2017); Torres et al. (2020); Martins et al. (2021). Thus far, a large amount of training data is needed to train a DL based network for subsequent classification Goodfellow et al. (2016). However, the manual labelling of all pixels to extract image

features for training is troublesome when dealing with imagery from kettle holes due to the visual complexity of the features. Even when using high spatial resolution datasets, a human operator has difficulties distinguishing between species of plants and assigning each pixel to a class. Therefore, only some parts of each training image will be labelled. We used an approach combining unsupervised superpixel segmentation with subsequent convolutional neural networks (CNNs) to circumvent this challenge as the complex dataset of the natural ecosystem has an abundance of nuances and details. Superpixels group image regions with similar perceptual features into clustered regions (Achanta et al. (2012)). Each superpixel is treated as a training dataset. Thereby, creating a labelled dataset that the CNN can use to find structured patterns without the need of labelling each image in its entirety.

We classify data from kettle holes to support the understanding and the composition of the ecosystem by performing a detailed detection of the flora present to characterize the species in the ecosystem. Our approach paves the way to the creation of databases that help to identify environmental factors that influence the health of the ecosystem.

2 |MATERIALS AND METHODS

Our workflow can be summarized the following: data acquisition, manual image labeling, superpixel clustering, training the CNNs based on the superpixel images and final segmentation of the image with trained CNN models (Figure 2).

2.1 |Dataset

The data acquisition was performed in the landscape laboratory AgroScapeLab Quillow (ASLQ, Leibniz Centre for Agricultural Landscape Research (ZALF)) named after the river Quillow that covers a catchment area of about 170 km². The region is located in the Central European lowlands, shaped into a hilly landscape by the last ice age. About three-quarters of the landscape is used for agriculture, the climate is sub-humid with a negative climatic water balance ranging from 1992 till 2019 during which an average of 658 mm of water evaporated and an average of 573 mm precipitation fell per year. The mean annual temperature was 8.8°C (Pätzig and Düker (2021)). UAV images were collected from 19 of more than 1500 kettle holes in the region. The selected kettle holes are very diverse in their morphological and hydrological characteristics and their classification was done according to (Kalettka and Rudat (2006)).

TABLE 1 Examined plant species classes (including dead plants) and the associated plant life forms and the

number of samples, i.e. superpixels, from each class in the unbalanced and balanced superpixel dataset.

Dominant plant species classes included (Class)	Plant life form	Unbalanced	Balanced
Carex riparia	*Helophytes*	13626 (16.52%)	318 (9.09%)
Cirsium arvense (L.) Scop.,	*Hemicryptophytes (nitrophilous perennials)*	318 (0.39%)	318 (9.09%)
Dead plants	*Helophytes etc. (no woody plants)*	1450 (1.76%)	318 (9.09%)
Oenanthe aquatica (L.) Poiret	*Amphibian plants*	2559 (3.10%)	318 (9.09%)
Others	*Other plants, soil, water, stones, crops, etc.*	22712 (27.53%)	318 (9.09%)
Phalaris arundinacea	*Helophytes*	7532 (9.13%)	318 (9.09%)
Phragmites australis (Cav.) Trin. ex Steud.	*Helophytes*	16103 (19.52%)	318 (9.09%)
Salix alba	*Phanerophytes (woody plants)*	3180 (3.85%)	318 (9.09%)
Salix cinerea	*Phanerophytes (woody plants)*	9749 (11.82%)	318 (9.09%)
Typha latifolia	*Helophytes*	468 (0.57%)	318 (9.09%)
Urtica dioica L. s. l.,	*Hemicryptophytes (nitrophilous perennials)*	4806 (5.83%)	318 (9.09%)

As a result of the high variability in hydrogeomorphological conditions and water quality the kettle holes are very singular in their biodiversity composition making each one unique and covering a wide range of dominant vegetation types (Table 1 and Figure 3). For the purpose of this work we investigated nine local dominant plant species classes commonly present in kettle holes with the inclusion of dead plant biomass belonging to different plant life forms occurring at each kettle hole. The plant life-forms ranged from amphibian plants to woody plants (Phanerophytes; Appendix A4). The Helophytes tended to be the most numerous plant life form, as they often dominated the surface of the kettle holes, especially in the dry season as in our study period.

The size of the selected kettle holes covered a range of 0.038 ha to about 1.2 ha. Concerning their water permanence, some kettle holes had been dry for an extended period of time and were considered episodic, while others were classified as permanent. However, due to prolonged droughts since 2018, all examined permanent kettle holes became semi-permanent, holding water most of the time but potentially drying up, especially at the end of the summer.

UAV imagery was obtained using a DJI Phantom 4 RTK carrying a 20 Megapixel CMOS sensor with a fixed focal lens of 8.8 mm. The sensor captured RGB images. The camera was attached to a gimbal that helped compensating for system vibrations due to the rotor movement and for pitch and roll movements of the aircraft

due to wind. The mission planning was done with the integrated software DJI GS RTK App of the P4 RTK. We used the 2D photogrammetry flight plan for ten kettle holes and the 3D photogrammetry multi-oriented flight plan for nine kettle holes at a flight altitude between about 25 to 35 m above ground leading to a mean ground sampling distance (GSD) of 9 mm, each image of the dataset has 5472 x 3648 pixels. The flight campaign was either at cloud-free times or with uniform cloud cover. The flights were realized at around 10:00 to 13:00 in broad daylight.

2.2 |Manual labeling and Superpixels

We labelled manually nine plant species classes, one class others and one class dead plants (Table 1). Including dead plants to the dataset was done to improve the quality of the classification and to enable potential health assessment of the ecosystem. The labeling of plant species classes was done by manually drawing polygons around homogeneous plant species areas with the VGG Image Annotator (Dutta and Zisserman (2019)). We randomly selected images from different kettle holes to create a detailed dataset for the segmentation.

Because the UAV imagery of kettle holes was displaying an environmental scenery, it was a highly complex challenge to manually label every pixel of the image to a particular class. Thus, instead we took only patches of the image of our objects of interest.

Afterwards, we used the superpixel algorithm to divide pixels into perceptually meaningful regions by considering the pixel grid's abstract structure. Eventually, we used only labelled superpixels as training images. Superpixels are an unsupervised classification approach that captures image redundancy to provide a simpler primitive for computing image attributes and to considerably simplify subsequent image processing tasks (Achanta et al. (2012)). The superpixel dataset, was created with the Simple Linear Iterative Clustering (SLIC) algorithm (Achanta et al. (2012)). The SLIC generates clusters of pixels with similar attributes such as color, texture and shape. We used the software Pynovisão (Dos Santos Ferreira et al. (2017)) and considered 405 attributes (Table 2). The attribute extraction can be understood as mathematical operations performed in the abstract binary data of the digital image to group regions with similarities. The chosen attributes for extraction were based on previous approaches (Costa et al. (2019)), and improved with the implementation of the K-curvature extraction algorithm (Abu Bakar et al. (2015)). The algorithm uses K-means to cluster similar pixels and thereby separate the image into small pieces, i.e., superpixels. We used a SLIC configuration with a K-value of 4000; K corresponds to the approximate number of segments that will separate the given image. The superpixel construction size depends

on the K value and the image size. Other configurations were sigma 5 and compactness 10. Sigma smooths each image channel and compactness balances the proximity of pixel color and space with higher values favoring space proximity and therefore resulting in squarer superpixel shapes. The final size of each superpixel was in the range of about 50 to 100 pixels (Figure 4). The size varied depending on the color attributes as they influence the border delimitation.

TABLE 2 Attributes and corresponding number of extractors used in the SLIC approach

Feature extracted	Method	Quantity of features	Reference
Color	Red, green and blue (RGB)	12	(Swain and Ballard (1991))
	Hue saturation value (HSV)	12	
	CIELAB color space	12	
Texture	Gray-level co-occurrence matrix (GLCM)	36	(Soh and Tsatsoulis (1999))
	Gabor filter	160	(Feichtinger and Zimmermann (1998))
	Local binary patterns (LBP)	18	(Van Klaveren et al. (1999))
Shape and Grandient	Hu Moments	7	(Hu (1962))
	Central Moments	10	
	Histogram of oriented gradientes (HOG)	128	(Triggs et al. (1999))
	K-curvature	10	(Abu Bakar et al. (2015))

In the final step of the training data generation, we used the previously annotated plant species classes to extract their annotation coordinates and their corresponding class label. The class assignment of a superpixel was then performed considering that information. Thus, the annotated class that dominated each cluster, i.e. superpixel, was chosen as the representing class considering a probabilistic threshold of 50%. For instance, if a superpixel region overlapped with two annotations such as: *Typha latifolia* and dead plants having 51% and 49% of overlap, respectively, the superpixel was classified as *Typha latifolia*. The attribute information of that superpixel was then stored to be later fed into the CNN training, making the machine understand these features as a representation of *Typha latifolia*. The SLIC approach was used across the entire annotated image dataset to separate segments of each class.

Sampling the labelled images with the superpixel approach led to a very unbalanced dataset (Table 1) as most of the superpixels (63.57%) were spread into only 3 classes: Others (27.53%), *Phragmites australis* (19.52%) and *Carex riparia* (16.52%). This might lead to a bias in the DL algorithms. Therefore, we also created a balanced dataset by randomly selecting 318 superpixels from each class using the undersampling technique. Thereby, 318 referred to the smallest class in the unbalanced dataset; *Cirsium arvense*. Thus, a high number of labels of the plant classes were created but to ensure a balanced labeled dataset not every identified region was selected. Four samples from each of the eleven superpixels classes are shown in Figure 5, where each superpixel covers an approximated area of 9.2 by 9.2 cm^2.

2.3 Deep Learning

Three state-of-the-art (SOTA) CNNs have been used for the image segmentation. The chosen CNNs were NasNetMobile (Zoph et al. (2018)), EfficientNet (Tan and Le (2019)) and Xception (Chollet (2017)). They had been chosen due to their SOTA results in previous tasks and different learning strategies.

A CNN architecture is made up of an input layer, which is in this study are the superpixel. Following the input we have the hidden layers, which are layers that process the data to logically extract information. There are many ways to configure the hidden layers and we can stack many of them to make the network 'deeper'. The number of hidden layers depends on the architecture of the network. The hidden layers are learning mathematical functions to obtain predictions from the input data and to create an output that is tangent to human's way of understanding and making sense of data entries. The final output layer concludes the network; in our study it was a raster in which each pixel was assigned to a class.

2.3.1 |NASNet Mobile

NASNet is a novel class of algorithms that searches for an adequate network architecture for the problems at hand. It uses a reinforcement learning search method to optimize the architecture configurations (Zoph and Le (2016)). For this research, we used the mobile implementation NASNetMobile, which is less resource intensive than NASNet. It searches for the architecture's building blocks on small datasets and then transfers them to larger datasets. The "NASNet search space" enables the transferability of knowledge from smaller datasets to bigger ones.

The controller recurrent neural network (RNN) samples parallel or 'child' networks, which are trained to converge to a target accuracy on a held-out validation set. These accuracies create a gradient and update the controller that will improve the architecture as processing goes forward. The structures of the cells are searched within a search space (Appendix A1). The controller RNN selects an operation from a set of operations, which were selected based on their prevalence in the DL literature (Zoph et al. (2018)), to apply to the hidden states. These operations are, e.g., max poolings of different sizes and depthwise-separable convolutions of different sizes. The model is supposed to find the best architecture of the CNN related to the dataset and the computation processing capabilities.

2.3.2 |EfficientNet

CNNs are usually developed at a determined resource quantity and manually scaled for better results if more resources are available (Tan and Le (2019)). EfficientNet uses a compound coefficient to automatically scale the network depth, width, and resolution dimensions to use all the available resources of the machine. Thus,

differently sized CNNs will be generated depending on the data and hardware used. The main difficulty of scaling the model is that the optimal depth, width or resolution depend on each other, and the values change under different resource constraints. Following observations related to scaling up dimensions were made (Tan and Le (2019): Scaling up any dimension of network width, depth, or resolution improves accuracy. But the accuracy gain diminishes for bigger models. In order to pursue better accuracy and efficiency, it is critical to balance all dimensions of the network during CNN scaling. The network achieved very good results with Image-Net datasets (Deng et al. (2009) while claiming to be smaller and faster than previous CNNs. EfficientNet achieved SOTA accuracies on CIFAR (Calik and Demirci (2018) and four other transfer learning datasets (Appendix A2).

2.3.3 |Xception

Xception explores inception modules to leverage depth-wise separable and regular convolutions because it is assumed that cross-channel and spatial correlations are decoupled. Less computational power is needed because fewer operations are required to perform the convolutions. A depth-wise separable convolution, commonly called "separable convolution" in DL frameworks such as TensorFlow and Keras, can be understood as an Inception module with a maximally large number of towers (Chollet (2017)). The towers are defined by a pooling phase followed by convolutions (Chollet (2017)). More specifically, inception modules have been replaced by depth-wise separable convolutions i.e., a spatial convolution performed independently over each channel of an input, followed by a point-wise convolution, i.e., a 1x1 convolution, projecting the channel's output by the depth-wise convolution onto a new channel space (Chollet (2017)). The Xception architecture has 36 convolutional layers forming the feature extraction base of the network (Appendix A3). In short, the Xception architecture is a linear stack of depth-wise separable convolution layers with residual connections (Chollet (2017)). The Xception presented gains in performance on the ImageNet dataset (Deng et al. (2009)) if compared to other structures such as Inception V3 (Szegedy et al. (2015)).

2.4 |Hyperparameters and Optimizers

The networks used in this study exhibit a different number of parameters (Table 3), which control the learning process. Parameters, in general, are weights that are learned during training. They come in the form of matrices that contribute to the model's predictive power and are changed during the back-propagation process. This change is governed by the chosen algorithms, i.e., the optimization strategy.

TABLE 3 Number of trainable parameters of the 3 CNNs used

Architectures	Trainable parameters
EfficientNet	43293709
NasNetMobile	30994717
Xception	72450857

After having learned the convolutional cells, several hyperparameters, which are not learned, may be explored to build a final network for a given task (Table 4). In this study, the networks had been trained with 1000 and 100 epochs considering the balanced and unbalanced datasets, respectively. The number of epochs is lower for the unbalanced case due to hardware limitations because the number of training instances is more than one magnitude higher than for the balanced case. The remaining parameter values are the same for both balanced and unbalanced datasets. A 10% patience was used for early stopping (*patience*, i.e., the training would stop if, after 10% of the total number of epochs, no alteration in the validation loss was observed). Data augmentation has been implemented through random flips, random zooms (max=10%), random horizontal and vertical shifts (max=50%) and random rotations (max=90°).

TABLE 4 Hyperparameter set used in this study (same for all 3 CNNs)

Hyperparameter	Value
Training Epochs	1000(B), 100(U)
Early Stop Patience	10%
Early Stop Monitor	Loss
Loss function	Softmax
Checkpoint Saving	True
Initial Learning Rate	0.01
Validation Split	20%
Neurons FC Layer	512
Dropout FC Layer	50%
Data augmentation	Yes
Cross-Validation data technique	5-Fold
Transfer learning	ImageNet
Fine tuning	True

For the optimization of the CNN, three adaptive strategies had been tested. An optimizer is used to update the weights in the search of the smallest loss value. Thereby, the network is said to be learning information. The general idea of the optimization is to tweak the parameters iteratively, changing the learning rate to minimize the loss function.

The optimizers chosen in this study were: Adagrad (Duchi et al. (2011)); Adadelta (Zeiler (2012)); Adam (Kingma and Ba (2014)). They are adaptive gradient descent-based algorithms. Adaptive learning rate optimizers have consistently shown better results than the standard, non-adaptive strategies when it is not possible to fine-tune a specific learning rate schedule (Bera and Shrivastava (2020)). We chose these optimizers because they perform well when the data is sparse, and they are capable of adapting well to the data.

Adagrad is an algorithm for gradient-based optimization that adapts the learning rate to the parameters. For parameters associated with often occurring features, smaller updates (i.e., low learning rates) are used, while for parameters associated with uncommon features, more significant updates (i.e., high learning rates) are used. As a result, it is well-suited to handle sparse data and there is no need to tune the learning rate manually. One of the disadvantages of Adagrad is that, sometimes the learning rate tends to become infinitesimally small, resulting in the algorithm losing its capacity of learning; this happens because Adagrad accumulates all past gradients in the denominator. Adadelta, is a less aggressive extension of Adagrad. Instead of accumulating all past squared gradients, it restricts the window of accumulated past gradients to some fixed size variation. This way, Adadelta continues learning even when many updates have been done. Adam is an alternative way for calculating adaptive learning rates for each parameter. Aside from retaining an exponentially decreasing average of previous events, it also keeps an exponentially decaying average of the past gradients.

2.5 |Metrics and experimental setup

We calculated the confusion matrix to evaluate the classification performance, and from it, we derived the precision P (eq. 1), recall R (eq. 2), and F1-score (eq. 3). TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives, respectively. In our analysis, positives and negatives refer to the pixels correctly and falsely assigned to the corresponding class by the underlying classifier. Such positives and negatives are true or false, depending on whether or not they agree with the assigned ground-truth class.

The Precision (P) is given by: $P = \frac{TP}{TP+FP}$ (1)

The Recall (R) is given by: $R = \frac{TP}{TP+FN}$ (2)

The F1-score is given by: $F1_{score} = 2\frac{PR}{P+R}$ (3)

Three CNNs were trained using three optimizers. The balanced and unbalanced datasets were randomly divided into five-folds, separated into one test set, and using the remaining four folds for training. The resampling strategy, called 5-fold stratified cross-validation, is commonly used to evaluate machine learning algorithms (Wong

and Yeh (2019); Wilson et al. (2020)). In the case of the unbalanced dataset, the random division considered the different number of samples of each class and hence preserving the class size ratio in each fold. Precision, recall, and F1-score metrics were used to measure the performance of each algorithm over the 5-folds test set.

3 |RESULTS

The result assessment of the processing used the superpixel images as a base. The best performance of classification for the unbalanced dataset is achieved by Xception using the Adagrad optimizer, which is indicated by the highest average precision (0.83), recall (0.75), and F1-score (0.77) (Table 5). However, EfficientNet has a lower interquartile range (IQR) also using the Adagrad optimizer (Figure 5). In the case of the balanced dataset, the Adagrad optimizer provides the best overall performance (precision, recall, and F1-score) for the EfficientNet and Xception architectures; the latter displays a higher median and smaller IQR. The F1-score for Xception is 0.85 and a good balance between precision (0.86) and recall (0.85) is achieved (Table 5). For the NASNetMobile, the adadelt optimizer reveals a higher median without outliers. However, the results for NASNetMobile are generally inferior to the other two CNNs (Figure 6). As expected, the results using a balanced dataset are superior to the unbalanced dataset, but the IQR for the EfficientNet using Adagrad is lower than for Xception in the unbalanced case.

TABLE 6 Average precision, recall and F1-score median values for all the three DL architectures and three optimizers using the unbalanced and balanced dataset

	Dataset							
	Unbalanced				Balanced			
Architecture	**Optimizer**	**Precis.**	**Recall**	**F1-score**	**Optimizer**	**Precis.**	**Recall**	**F1-score**
EfficientNet	Adadelta	0.14	0.14	0.11	Adadelta	0.28	0.21	0.19
EfficientNet	Adagrad	0.74	0.68	0.69	Adagrad	0.80	0.78	0.78
EfficientNet	Adam	0.03	0.07	0.04	Adam	0.01	0.07	0.02
NASNetMobile	Adadelta	0.19	0.17	0.14	Adadelta	0.28	0.26	0.23
NASNetMobile	Adagrad	0.32	0.19	0.13	Adagrad	0.25	0.22	0.18
NASNetMobile	Adam	0.16	0.15	0.10	Adam	0.01	0.07	0.02
Xception	Adadelta	0.29	0.23	0.21	Adadelta	0.55	0.52	0.51
Xception	Adagrad	0.83	0.75	0.77	Adagrad	0.86	0.85	0.85
Xception	Adam	0.14	0.18	0.14	Adam	0.48	0.40	0.37

In most scenarios for both, balanced and unbalanced, datasets unsatisfactory outcomes were achieved from the nine experiment configurations (with F-scores mostly below 0.3). However, two good outcomes were obtained with Xception and EfficientNet when employing the Adagrad optimizer. Adagrad is better suitable for sparse data, which is the case for some vegetation classes in this study. It outperforms the other two optimizers (Adadelta and

Adam), which have a fixed learning rate momentum. In this study, combining a fixed-sized CNN (i.e., Xception) and an adaptive learning rate optimizer (i.e., Adagrad) produced the best F1-score results.

The learning curve for the best configuration (Xception and Adagrad) is shown in Figure 7 and confirms that the choice of 1000 epochs for training was enough. It could have even been reduced because, after 40 epochs, the validation loss seems to decrease, indicating a possible start of overfitting. EfficientNet and Xception reveal a good performance identifying the plant classes, which is indicated by the high values of the diagonal of the confusion matrix (Figure 8). The success of the DL method varied depending on the class. The highest errors were given for the Carix riparia species that has been mistakenly classified as Other in 14% of the test set samples with both architectures; EfficientNet and Xception. EfficientNet presented a lower correct classification rate (64%) than Xception (70%) for this class. The brownish color, which also appears in some of the training samples from the class Others, may have confused both DL algorithms. NASNetMobile could not learn to classify most of the classes, shown by the producer's accuracy value of the confusion matrix being only 0.78 for the dead plants class.

Solely a small number of samples (i.e., 318 superpixels per class) was needed in this study to train the networks. Manual labeling is one of the costliest aspects of a supervised image classification workflow. Thus, the need for a small number of samples is an auspicious advancement for improved mapping of complex vegetation systems (Figure 9).

4 |DISCUSSION

Vegetation mapping is an important technical undertaking for managing natural resources. Traditional approaches (e.g., field surveys or collateral and supplementary data analysis) are time-consuming, data-lagging, and often prohibitively expensive. Remote sensing in combination with machine learning technologies can provide a potential practical and cost-effective way to study changes in vegetation cover. Results of this study utilizing machine learning methods to UAV images achieved F1-scores that are comparable to previous studies of plant segmentation (Torres et al. (2020); Elkind et al. (2019); Martins et al. (2021)). Some studies implemented additional information such as digital surface models (DSMs) or multi-spectral data to classify species to obtain similar or better F1-scores (Husson et al.(2016); Chabot et al.(2018); Durgan et al.(2020); Benjamin et al.(2021); Schulze-Brüninghoff et al.(2021)) However, retrieving DSMs might be difficult in kettle holes, e.g., under windy conditions due to moving vegetation (Pätzig et al.(2020)) Moreover, the application of multi-spectral cameras makes the approach more expensive. Furthermore, these studies relied on extensive, hand-crafted feature selection (e.g., spectral and textural variables and band indices), which were then used with random forest or support vector

machine classifiers, whereas CNN's allow for end-to-end learning. Another study by (Bhatnagar et al.(2020)) , also using CNNs achieved F-score up to 90%, however classifying vegetation communities (grouped plant species).

In order to begin vegetation preservation and restoration projects, it is important to first determine the existing status of vegetation cover, using UAV in combination with DL methods we have a highly efficient kettle hole mapping. Nonetheless, remote sensing of plant species detection is still complex. The potentially high changes of plant communities in a small area due to high abiotic gradients within kettle holes and meta-community processes within kettle holes and beyond demand punctual high resolution spatial data for a good discretization of the environment. Furthermore, in some classification cases an information richer dataset, such as multi-spectral or hyper-spectral datasets, still needs to be applied (Rossi et al. (2021)). Also, even trained specialists may produce mistakes caused by poor image quality, influences of shadow or a pixel being covered by several species (mixed pixels).

When using the results of vegetation mapping from remote sensing imagery, further aspects need to be considered (Rapp et al. (2005)): how well does the chosen classification system represent actual vegetation community composition, how effective do remote sensing images capture distinguishing features of each mapping unit and how well can these mapping units be delineated by photo-interpreters. Thus, we must evaluate the applicability of each chosen machine learning method for each specific task. To better depict plant community compositions, a well-fit vegetation classification system should be carefully established according to the study's purpose. There are no superior image classifiers that can be used in all applications equally. Applying or developing new classifiers fit for certain applications is a challenging task needing further research.

Kettle holes in Germany have received legislative protection, but existing conservation methods are insufficient in terms of preserving potential habitat functions, which are dependent on the specific environmental conditions (Berger et al. (2011), Pätzig et al. (2012)). A significant achievement of this study is correctly classifying plant species with different plant forms at the same time with more than 85% F-1 score. This is a pioneering and forward-looking methodological requirement for efficient characterization of many kettle holes and thus a better understanding of the dynamics, the functions and services of kettle hole ecosystems. In a next step, creating spatio-temporal land-cover change maps of these regions, allied with information about the current agricultural practices, can help to better understand how different agricultural strategies influence these ecosystems. This can support the provision of specific protection measures for each kettle hole and can also build the base for transferring experiences to other regions.

The advance of this research is the implementation of an unsupervised machine learning algorithm SLIC that is capable of detect and segment features of color, shape and texture, and them feed this feature information into the Deep learning model, this process makes the training and production phase less complex and resource demanding because instead of working with all the pixels of the original images (5472 x 3648 pixels) we are working with a predetermined number of grouped pixels in the case of this work 4000 for each image. This also makes the labeling process easier by reducing the need of manual labeling every feature that we want to detect in the image, for that, we inputted some labelled features to the superpixel algorithm segmented them and created the feature patches dataset for the machine learning training, performed the training and then transferred the learning for the rest of the images.

5 |CONCLUSION

DL with CNNs is especially suited for information extraction from UAV imagery and it can become essential for the assessment of the challenging ecosystem of kettle holes. These ecosystems greatly benefit from high resolution monitoring of their diversity and change over time. We demonstrated that it is possible to train a suitable network for classifying different plant species and dead plants in kettle hole systems. Three different networks and three different optimizers were tested to automatically classify plant species which are typical in this landscape, and which were captured by UAV images. Best results (f1-score = 0.86) were achieved with the CNN Xception and the optimizer Adagrad. A balanced dataset was necessary for successful training. However, solely a small number of samples (i.e., 318 superpixels per class) was needed. A workflow was introduced to sample the training data efficiently from patchily annotated images.

This study provides a method to reduce the current gap between the sciences of machine learning and remote sensing and ecology, including conservation. Machine learning is a viable tool to access the health of ecosystems and should be increasingly implemented in long-term monitoring schedules to facilitate evidence-based conservation programs. In the next step, the segmented images, whose exterior and interior camera geometry are known, will be projected to an orthomosaic to improve the quantification of plant species cover. Furthermore, the classification needs to be extended to different periods of the year to capture and quantify seasonal changes and better assess the dynamics of the kettle hole ecosystem.

Acknowledgements

This research has been partly funded by the Leibniz Centre for Agricultural Landscape Research (ZALF) through the integrated priority project SWBTrans: “Smart Use of Heterogeneities of Agricultural Landscapes". We would like to thank Dorith Henning for labelling the images. We would like to thank the Graduate Program of Environmental Technologies of the Federal University of Mato Grosso do Sul (UFMS) that supported the doctoral dissertation of the first author. We would like to thank NVIDIA Corporation for the donation of the GPU used in this research.

Author’s contributions

JMJ, HP and AE drafted the ideas and designed methodology; MP collected the data; JM, HP and DS analysed the data; JM, MP and AE led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Conflict of interest

No potential conflict of interest was reported by the authors.

Data sharing

For the reproduction of the results the codes used can be accessed via the link: https://github.com/Jose-Augusto-C-M/KettleHole_SLIC_CNN.

6 References

Abu Bakar, M. Z., Samad, R., Pebrianti, D., Mustafa, M. and Abdullah, N. R. H. (2015) Finger application using K-Curvature method and Kinect sensor in real-time. *2nd International Symposium on Technology Management and Emerging Technologies, ISTMET 2015 - Proceeding*, 218–222.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. and Süsstrunk, S. (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, 2274–2281.

Anderson, K. and Gaston, K. (2013) Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, **11**, 138–146.

Azpiroz, A. B., Isacch, J. P., Dias, R. A., Di Giacomo, A. S., Fontana, C. S. and Palarea, C. M. (2012) Ecology and conservation of grassland birds in southeastern south america: a review. *Journal of Field Ornithology*, **83**, 217–246.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L. and Bareth, G. (2015) Combining uav-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, **39**, 79–87.

Bera, S. and Shrivastava, V. K. (2020) Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. *International Journal of Remote Sensing*, **41**, 2664–2683. URL: https://doi.org/10.1080/01431161.2019.1694725.

Berger, G., Pfeffer, H. and Kalettka, T. (2011) Amphibienschutz in kleingewässerreichen ackerbaugebieten. *Natur & Text, Rangsdorf*.

Calik, R. C. and Demirci, M. (2018) Cifar-10 image classification with convolutional neural networks for embedded systems. *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, 1–2.

Casado, M. R., Gonzalez, R. B., Kriechbaumer, T. and Veal, A. (2015) Automated identification of river hydromorphological features using uav high resolution aerial imagery. *Sensors*, **15**, 27969–27989.

Cen, H., Wan, L., Zhu, J., Li, Y., Li, X., Zhu, Y., Weng, H., Wu, W., Yin, W., Xu, C. et al. (2019) Dynamic monitoring of biomass of rice under different nitrogen treatments using a lightweight uav with dual image-frame snapshot cameras. *Plant Methods*, **15**, 1–16.

Chabot, D., Dillon, C., Shemrock, A., Weissflog, N. and Sager, E. P. (2018) An object-based image analysis workflow for monitoring shallow-water aquatic vegetation in multispectral drone imagery. *ISPRS International Journal of Geo-Information*, **7**, 294.

Chollet, F. (2017) Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258.

Costa, C. S., Tetila, E. C., Astolfi, G., Sant'Ana, D. A., Pache, M. C. B., Gonçalves, A. B., Zanoni, V. A. G., Nucci, H. H. P., Diemer, O. and Pistori, H. (2019) A computer vision system for oocyte counting using images captured by smartphone. *Aquacultural Engineering*, **87**, 102017.

Cruzan, M. B., Weinstein, B. G., Grasty, M. R., Kohrn, B. F., Hendrickson, E. C., Arredondo, T. M. and Thompson, P. G. (2016) Small unmanned aerial vehicles (micro-uavs, drones) in plant ecology. *Applications in plant sciences*, **4**, 1600041.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009) Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Dobson, A. P., Bradshaw, A. and Baker, A. á. (1997) Hopes for the future: restoration ecology and conservation biology. *Science*, **277**, 515–522.

Dos Santos Ferreira, A., Freitas, D. M., da Silva, G. G., Pistori, H. and Folhes, M. T. (2017) Weed detection in soybean crops using convnets. *Computers and Electronics in Agriculture*, **143**, 314–324.

Duchi, J., Hazan, E. and Singer, Y. (2011) Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, **12**, 2121–2159.

Dutta, A. and Zisserman, A. (2019) The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19. New York, NY, USA: ACM. URL: https://doi.org/10.1145/3343031.3350535.

Elkind, K., Sankey, T. T., Munson, S. M. and Aslan, C. E. (2019) Invasive buffelgrass detection using high-resolution satellite and uav imagery on google earth engine. *Remote Sensing in Ecology and Conservation*, **5**, 318–331.

Feichtinger, H. G. and Zimmermann, G. (1998) A banach space of test functions for gabor analysis. In *Gabor analysis and algorithms*, 123–170. Springer.

Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D. and Knapp, A. (2003) The importance of land-use legacies to ecology and conservation. *BioScience*, **53**, 77–88.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Hu, M.-K. (1962) Visual pattern recognition by moment invariants. *IRE transactions on information theory*, **8**, 179–187.

Kalettka, T. and Rudat, C. (2006) Hydrogeomorphic types of glacially created kettle holes in north-east germany. *Limnologica*, **36**, 54–64.

Kingma, D. P. and Ba, J. (2014) Adam: A method for stochastic optimization. URL: http://arxiv.org/abs/1412.6980. Cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

Lozada-Gobilard, S., Stang, S., Pirhofer-Walzl, K., Kalettka, T., Heinken, T., Schröder, B., Eccard, J. and Joshi, J. (2019) Envi-

ronmental filtering predicts plant-community trait distribution and diversity: Kettle holes as models of meta-community systems. *Ecology and evolution*, **9**, 1898–1910.

Martins, J. A. C., Nogueira, K., Osco, L. P., Gomes, F. D. G., Furuya, D. E. G., Gonçalves, W. N., Sant'Ana, D. A., Ramos, A. P. M., Liesenberg, V., dos Santos, J. A. et al. (2021) Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. *Remote Sensing*, **13**,3054.

Pätzig, M. and Düker, E. (2021) Dynamic of dominant plant communities in kettle holes (northeast germany) during a five-year period of extreme weather conditions. *Water*, **13**, 688.

Pätzig, M., Kalettka, T., Glemnitz, M. and Berger, G. (2012) What governs macrophyte species richness in kettle hole types? a case study from northeast germany. *Limnologica*, **42**, 340–354.

Rapp, J., Wang, D., Capen, D., Thompson, E. and Lautzenheiser, T. (2005) Evaluating error in using the national vegetation classification system for ecological community mapping in northern new england, usa. *Natural Areas Journal*, **25**, 46–54.

Raunkiaer, C. et al. (1934) The life forms of plants and statistical plant geography; being the collected papers of c. raunkiaer. *The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer*.

Rossi, C., Kneubühler, M., Schütz, M., Schaepman, M. E., Haller, R. M. and Risch, A. C. (2021) Spatial resolution, spectral metrics and biomass are key aspects in estimating plant species richness from spectral diversity in species-rich grasslands. *Remote Sensing in Ecology and Conservation*.

Shen, D., Wu, G. and Suk, H.-I. (2017) Deep learning in medical image analysis. *Annual review of biomedical engineering*, **19**, 221–248.

Soh, L.-K. and Tsatsoulis, C. (1999) Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 780–795. URL: http://ieeexplore.ieee.org/document/752194/.

Sotille, M. E., Bremer, U. F., Vieira, G., Velho, L. F., Petsch, C. and Simões, J. C. (2020) Evaluation of uav and satellite-derived ndvi to map maritime antarctic vegetation. *Applied Geography*, **125**, 102322.

Swain, M. J. and Ballard, D. H. (1991) Color indexing. *International journal of computer vision*, **7**, 11–32.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2015) Rethinking the inception architecture for computer vision. *CoRR*, **abs/1512.00567**. URL: http://arxiv.org/abs/1512.00567.

Tan, M. and Le, Q. (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning* (eds. K. Chaudhuri and R. Salakhutdinov), vol. 97 of *Proceedings of Machine Learning Research*, 6105–6114. PMLR. URL: https://proceedings.mlr.press/v97/tan19a.html.

Torres, D. L., Feitosa, R. Q., Happ, P. N., La Rosa, L. E. C., Junior, J. M., Martins, J., Bressan, P. O., Gonçalves, W. N. and Liesenberg, V. (2020) Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. *Sensors (Switzerland)*, 563.

Torres-Sánchez, J., Pena, J. M., de Castro, A. I. and López-Granados, F. (2014) Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from uav. *Computers and Electronics in Agriculture*, **103**, 104–113.

Triggs, B., McLauchlan, P. F., Hartley, R. I. and Fitzgibbon, A. W. (1999) Bundle adjustment—a modern synthesis. 298–372.

Turner, D., Lucieer, A., Malenovskỳ, Z., King, D. and Robinson, S. A. (2018) Assessment of antarctic moss health from multi-sensor uas imagery with random forest modelling. *International journal of applied earth observation and geoinformation*, **68**, 168–179.

Vahrson, W. and Frielinghaus, M. (1998) Soil erosion through farming in a young moraine landscape in ne germany. *Beitrage fur Forstwirtschaft und Landschaftsokologie (Germany)*, 109–114.

Van Iersel, W., Straatsma, M., Addink, E. and Middelkoop, H. (2018) Monitoring height and greenness of non-woody floodplain vegetation with uav time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, **141**, 112–123.

Van Klaveren, E. P., Michels, J. P. J. and Schouten, J. A. (1999) THE ORIENTATIONAL AND STRUCTURAL PROPERTIES OF N 2 and N 2 –AR SOLIDS AT HIGH PRESSURE. *International Journal of Modern Physics C*, **10**, 445–453. URL: https://www.worldscientific.com/doi/abs/10.1142/S0129183199000334.

Vasić, F., Paul, C., Strauss, V. and Helming, K. (2020) Ecosystem services of kettle holes in agricultural landscapes. *Agronomy*, **10**, 1326.

Wilson, A., Kasy, M. and Mackey, L. (2020) Approximate cross-validation: Guarantees for model assessment and selection. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (eds. S. Chiappa and R. Calandra), vol. 108 of *Proceedings of Machine Learning Research*, 4530–4540. PMLR. URL: http://proceedings.mlr.press/v108/wilson20a.html.

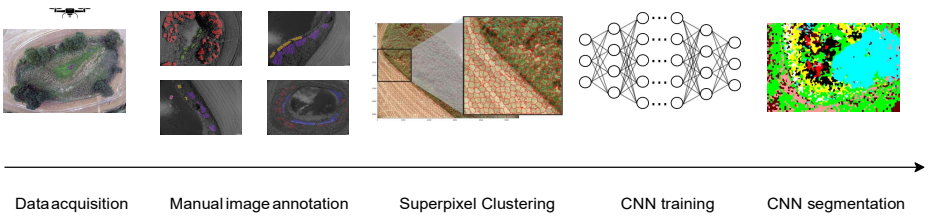
Wong, T.-T. and Yeh, P.-Y. (2019) Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, **32**, 1586–1594.

Zeiler, M. D. (2012) Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

Zoph, B. and Le, Q. V. (2016) Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*.

Zoph, B., Vasudevan, V., Shlens, J. and Le, Q. V. (2018) Learning transferable architectures for scalable image recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8697–8710.

F I G U R E 1 Example of a Kettle hole in the AgroScapeLab Quillow region.



F I G U R E 2 Illustration of the workflow proposed in the research.

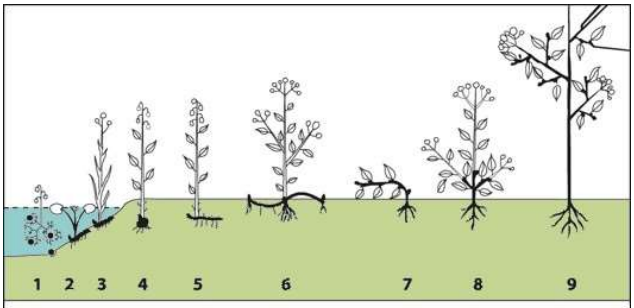
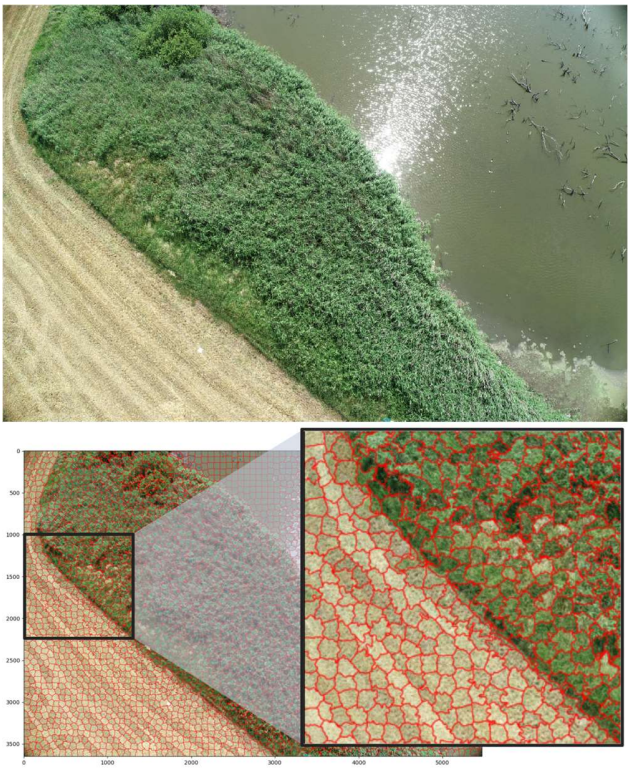


Diagram of the most important life forms based on the classification by Raunkiaer (1934)

1 + 2	Hydrophytes	(Hydr.)	water plants
3	Helophytes	(Helo.)	winter buds under water flowering plants above water
4 + 5	Crytophytes or geophytes	(Geof.)	winter buds below ground
6	Hemicryptophytes	(Hemi.)	winter buds above or just below ground
7 + 8	Chamaephytes	(Cham.)	winter buds up to 50 cm above ground
9	Phanerophytes	(Phan.)	winter buds at least 50 cm above ground (i.e. trees, shrubs and lianes)

F I G U R E 3 Plant life forms of which some were classified in this study as described by (Raunkiaer et al. (1934).



F I G U R E 4 Example of an image made of superpixels after applying the SLIC method with a K value of 4000. Each superpixel has an average approximated area of 83 cm²

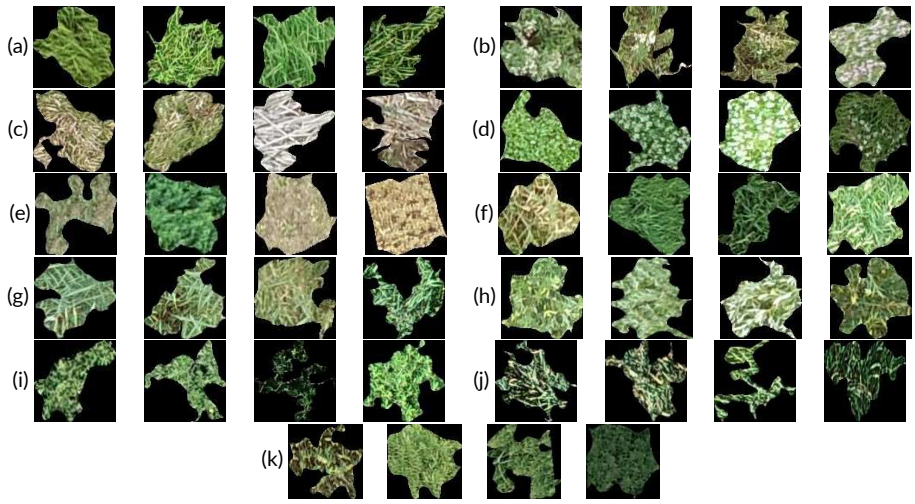


FIGURE 5 Superpixel examples (generated based on the features color, texture, shape and gradient) for eleven different classes: (a) *Carex riparia*, (b) *Cirsium arvense*, (c) Dead plants, (d)*Oenanthe aquatica*, (e) Others, (f) *Phalaris arundinacea*, (g) *Phragmites australis*, (h) *Salix alba*, (i) *Salix cinerea*, (j) *Typha latifolia*, (k) *Urtica dioica*. Each superpixel has a size of about 9.2 cm x 9.2 cm.

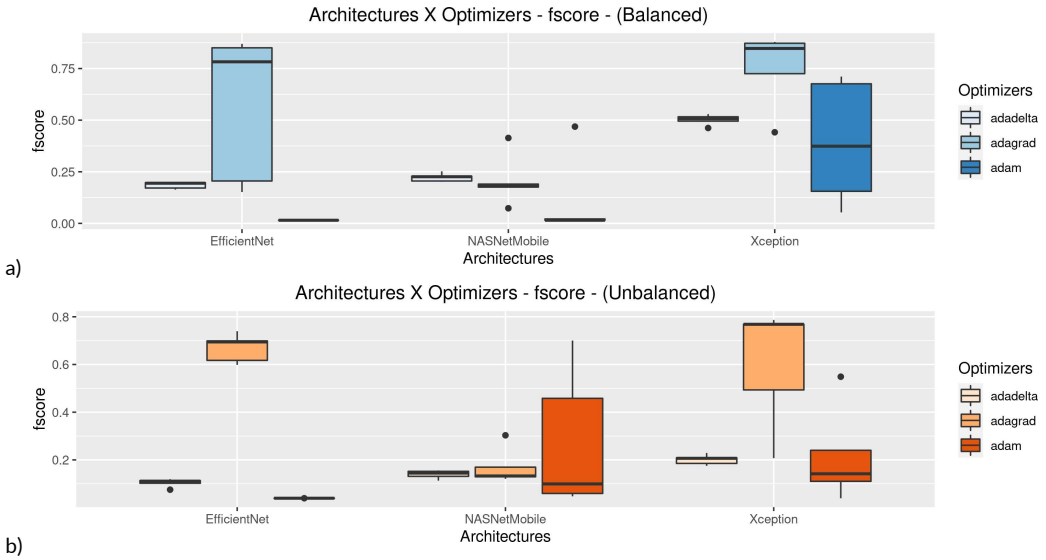
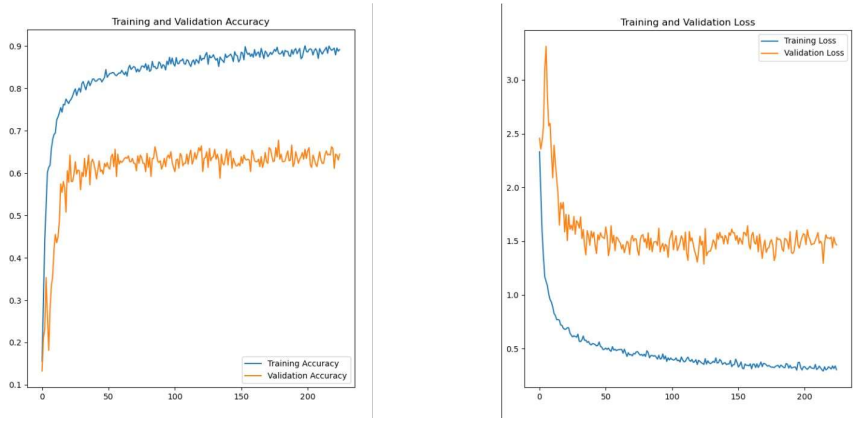


FIGURE 6 Boxplots for precision, recall and F1-score for all tested configurations using the a) Balanced and b) Unbalanced dataset.



F I G U R E 7 Learning curves for the Xception architecture over the training and validation set using accuracy and loss. the y value represents the error and the x axis represents the number of epochs.

Confusion Matrix for EfficientNet (Balanced)

Predicted \ Measured	*Carex riparia*	*Cirsium arvense*	*Dead Plants*	*Oenanthe aquatica*	*Others*	*Phragmites australis*	*Phalaris arundinacea*	*Salix alba*	*Salix cinerea*	*Typha latifolia*	*Urtica dioica*
Urtica dioica	0.08	0.00	0.00	0.02	0.05	0.06	0.00	0.00	0.06	0.00	0.85
Typha latifolia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.00
Salix cinerea	0.02	0.03	0.00	0.02	0.05	0.00	0.02	0.02	0.91	0.03	0.11
Salix alba	0.02	0.03	0.00	0.00	0.03	0.05	0.00	0.97	0.00	0.00	0.00
Phragmites australis	0.05	0.00	0.00	0.00	0.00	0.05	0.95	0.00	0.02	0.02	0.02
Phalaris arundinacea	0.08	0.00	0.00	0.00	0.00	0.83	0.03	0.00	0.02	0.00	0.00
Others	0.14	0.06	0.09	0.02	0.83	0.00	0.00	0.02	0.00	0.02	0.02
Oenanthe aquatica	0.00	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Dead Plants	0.00	0.00	0.86	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Cirsium arvense	0.00	0.88	0.03	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00
Carex riparia	0.64	0.00	0.02	0.05	0.00	0.03	0.00	0.00	0.00	0.00	0.00

Confusion Matrix for NASNetMobile (Balanced)

Predicted \ Measured	*Carex riparia*	*Cirsium arvense*	*Dead Plants*	*Oenanthe aquatica*	*Others*	*Phragmites australis*	*Phalaris arundinacea*	*Salix alba*	*Salix cinerea*	*Typha latifolia*	*Urtica dioica*
Urtica dioica	0.08	0.02	0.02	0.11	0.11	0.08	0.10	0.08	0.11	0.00	0.24
Typha latifolia	0.00	0.05	0.00	0.00	0.00	0.00	0.05	0.00	0.02	0.17	0.02
Salix cinerea	0.02	0.06	0.00	0.03	0.02	0.08	0.02	0.03	0.30	0.11	0.10
Salix alba	0.02	0.03	0.00	0.02	0.00	0.00	0.03	0.03	0.00	0.02	0.00
Phragmites australis	0.06	0.10	0.00	0.10	0.05	0.24	0.46	0.00	0.10	0.21	0.05
Phalaris arundinacea	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Others	0.60	0.32	0.21	0.29	0.60	0.33	0.16	0.52	0.21	0.13	0.41
Oenanthe aquatica	0.05	0.10	0.00	0.38	0.03	0.06	0.03	0.02	0.14	0.02	0.08
Dead Plants	0.13	0.11	0.78	0.02	0.14	0.11	0.11	0.22	0.03	0.05	0.05
Cirsium arvense	0.05	0.22	0.00	0.06	0.05	0.06	0.03	0.10	0.10	0.30	0.06
Carex riparia	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00

Confusion Matrix for Xception (Balanced)

Predicted \ Measured	*Carex riparia*	*Cirsium arvense*	*Dead Plants*	*Oenanthe aquatica*	*Others*	*Phragmites australis*	*Phalaris arundinacea*	*Salix alba*	*Salix cinerea*	*Typha latifolia*	*Urtica dioica*
Urtica dioica	0.03	0.05	0.00	0.02	0.02	0.08	0.00	0.00	0.05	0.00	0.88
Typha latifolia	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.03	0.95	0.00
Salix cinerea	0.02	0.00	0.00	0.03	0.02	0.00	0.02	0.02	0.86	0.05	0.09
Salix alba	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.02
Phragmites australis	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.05	0.00	0.02
Phalaris arundinacea	0.09	0.00	0.00	0.00	0.02	0.83	0.00	0.00	0.00	0.00	0.00
Others	0.14	0.03	0.08	0.00	0.88	0.02	0.00	0.00	0.02	0.00	0.00
Oenanthe aquatica	0.02	0.02	0.02	0.89	0.00	0.00	0.00	0.03	0.00	0.00	0.00
Dead Plants	0.00	0.00	0.86	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Cirsium arvense	0.00	0.89	0.05	0.05	0.03	0.00	0.00	0.02	0.00	0.00	0.00
Carex riparia	0.70	0.00	0.00	0.00	0.02	0.06	0.00	0.00	0.00	0.00	0.00

F I G U R E 8 Confusion matrices for the combination of architecture and optimizer that presented the best F1-score results using the balanced dataset, as presented in Table 5. The values correspond to the percentage of a total of objects and were normalized by columns and summed up to one except for rounding effects.

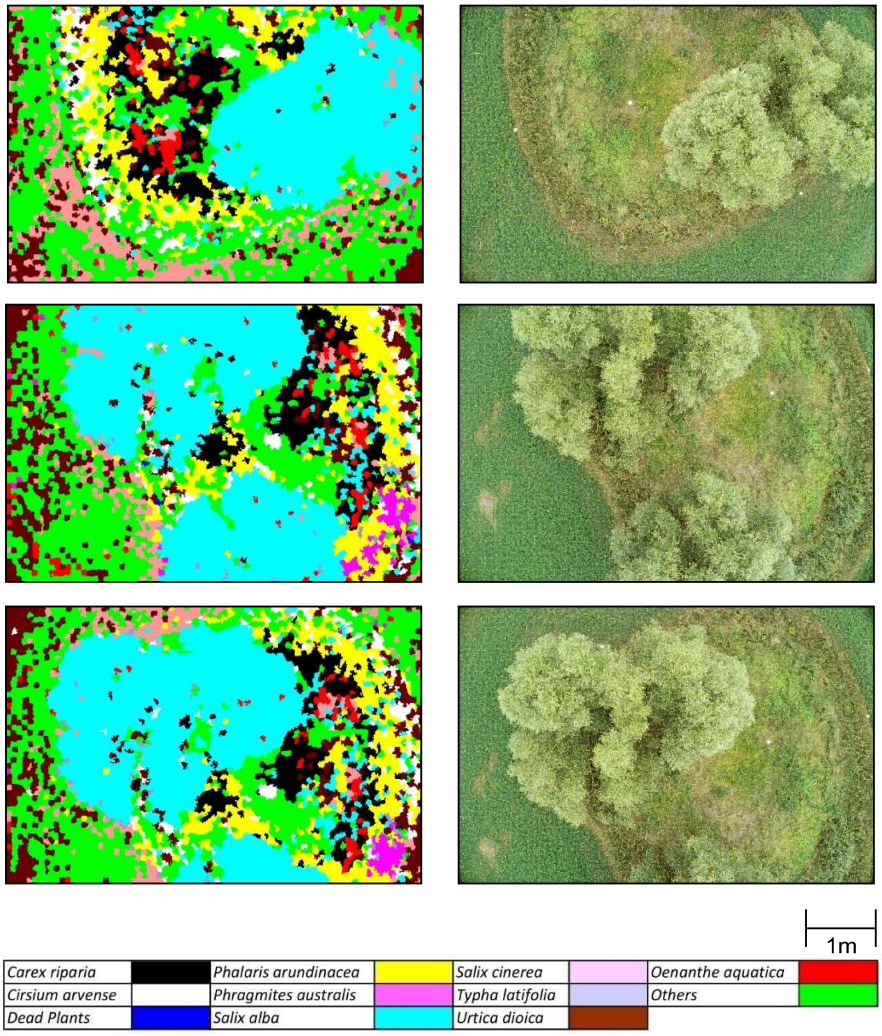
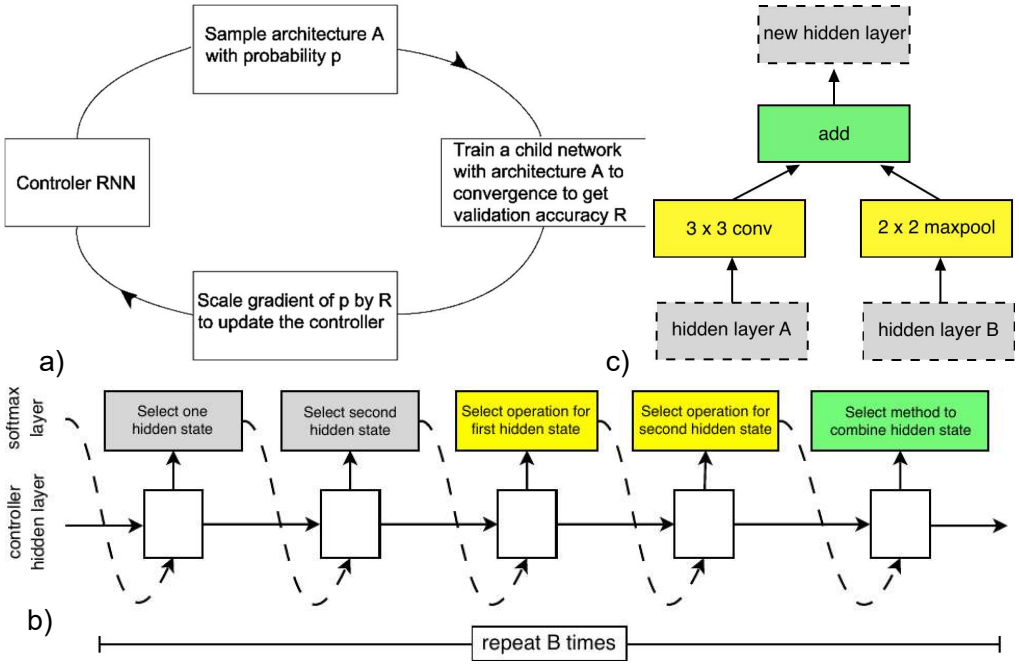


FIGURE 9 Resulting map of identified species using the trained CNN with the best testing result, i.e. Xception optimized with Adagrad. The maps are presented using a SLIC image as a base. Therefore, we have this fragmented output. However, the trained network can also be applied to raw images. Please, note that the provided scale is an approximation as the images have not be georeferenced, yet. Furthermore, the results are not verified with true data and hence the actual vegetation map may differ.

7 A | APPENDIX

This section will provide artistic representations with short descriptions of the CNNs, and the plant classification used in this research. For more details of each architecture, we advise the reader to search for the papers. Often, the CNN architectures include several building blocks, such as con- volution, pooling, dense and fully connected layers. The CNNs used in this research are advances of CNN structures in regard of the search for better performance in terms of metrics and processing capabilities.

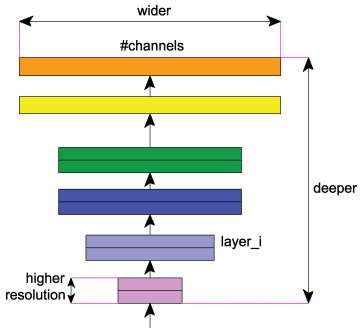


A1 a) Optimal neural network architecture search with NASNet, proposed by (Zoph and Le (2016)) and used in (Zoph et al. (2018)). b) NasNet controller model architecture for recursively constructing one block of a convolutional cell. Each block requires the selection of five discrete parameters, each corresponding to the output of a softmax layer. An example for a block construction is shown in c). A convolutional cell contains B blocks. B may vary depending on the complexity of the experiment and the operator of the network. This figure is adapted from the research that proposes this network (Zoph et al. (2018)).)

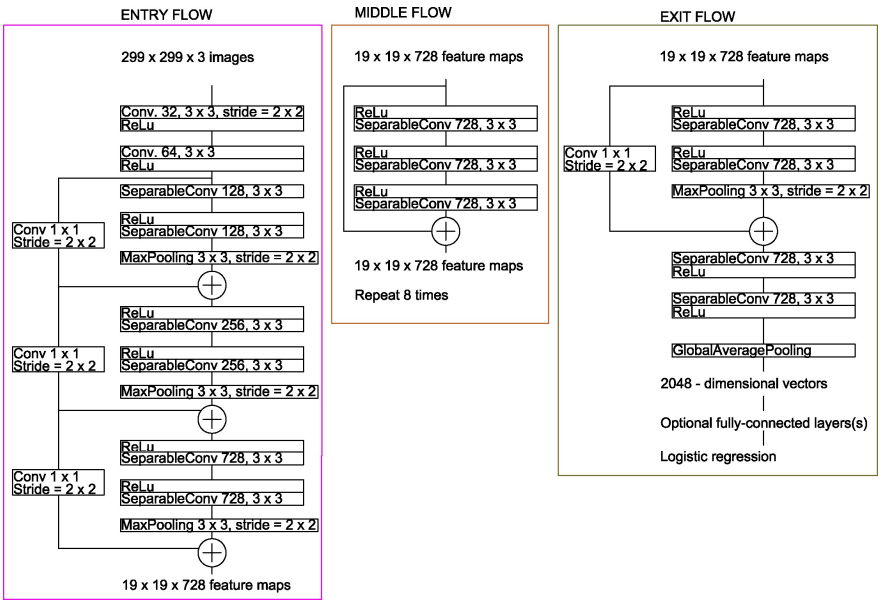
a)

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

b)



A2 a) EfficientNet-B0 architecture as proposed in (Tan and Le (2019)). b) The main building block of the EfficientNet is mobile inverted bottleneck MBConv (Tan and Le (2019)), that can change the size of the network channels to adapt to the task and to the available hardware resources.



A3 Xception architecture as proposed in (Chollet (2017)). All convolution and separable convolution layers are followed by batch normalization (not included in the diagram). All separable convolution layers use a depth multiplier of 1 (no depth expansion) (Chollet (2017).)

A4 Description of the characteristics of the Kettle holes. KH_No = kettle hole number according to numbering within the SWBTrans project (ZALF). HGMT. = hydrogeomorphological type: BS-S = big and shallow storage type, BS-SO = big and shallow shore overflow type, SS-S = small and shallow storage type, SS-SO= small and shallow shore overflow type. KA = mean area of the kettle hole between 2016 and 2020, measured up to the tilled area of the adjacent field in square meter, google earth measurement. KA_C= size class of the mean area of the kettle hole in hectare: 1 = very small (≤0.03 ha), 2 = small (≤0.10 ha), 3 = medium (≤0.32 ha), 4 = large (≤1.0 ha), 5 = very large (>1 ha). SS_C = class of shore slope in percent: 1 = flat (≤10%), 2 = oblique (≤20%), 3 = very oblique (≤30%), 4 = steep (≤40%). LP = long-term pond permanence class for the period 2011-2017. Perm = permanent; S.per = semi-permanent; Epis = Episodic; Peri = Periodic; LSOT = long-term shore overflow tendency class for the period 2011-2017. P.O = Partly overflowing; N.S.O = No Shore Overflow; F.O = Fully Overflowing. Note: due to a prolonged drought since 2018 the long-term pond permanence shifted for many kettle holes, e.g., all examined permanent kettle holes were not any longer permanent.

KH_No	HGMT.	KA	KA_C	SS_C	LP	LSOT
1	SS-S	2271	3	4	Perm.	N.S.O.
2	BS-S	5800	3	3	Epis.	N.S.O.
3	SS-S	685	2	4	Peri.	N.S.O.
4	BS-SO	3224	4	2	S.per.	F.O.
5	BS-S	7771	4	3	Epis.	N.S.O.
7	SS-S	1878	3	3	Peri.	N.S.O.
8	BS-SO	12034	4	3	Perm.	F.O.
9	SS-S	2910	3	3	S.per.	N.S.O.
10	SS-SO	789	2	2	Epis.	P.O.
12	SS-S	842	2	4	Peri.	N.S.O.
13	SS-S	1017	2	4	Peri.	N.S.O.
14	SS-S	379	2	4	Epis.	N.S.O.
15	SS-S	1560	3	4	Epis.	N.S.O.
16	SS-SO	1014	2	3	S.per.	F.O.
17	SS-SO	1660	3	2	S.per.	P.O.
18	SS-S	3852	4	3	S.per.	N.S.O.
20	BS-SO	4644	4	2	Peri.	P.O.
23	SS-SO	2288	3	2	Perm.	F.O.
25	SS-S	4594	3	3	Peri.	N.S.O.

CHAPTER 5

Paper 3: Deep learning and vision transformers applied to vegetation mapping for the region of Brazilian Pantanal

Manuscript being developed.

Deep learning and vision transformers applied to vegetation mapping for the region of Brazilian Pantanal

José Augusto Correa Martins[1], Maximilian Jaderson de Melo[2], Wesley Nunes Gonçalves[2], José Marcato Junior[1]

Abstract

Pantanal is the largest continuous wetland in the world, but its biodiversity is currently endangered by wildfires and the advance of the local agro-industry. The information available for the area regarding location and the extent of deforestation practices today is generated with datasets that have spatial resolutions ranging from 30 m up to 1 km, meaning that small forest clearing often is not detected using those methods. For better monitoring, we want to propose a method that will improve these assessments and create more accurate information that will assist in environmental actions. Using PlanetScope imagery with 4.77m spatial resolution and state-of-the-art, deep learning (DL) segmentation methods in the form of convolutional neural networks (CNN) and Transformed-based networks. These techniques are highly capable of extracting information from remote sensing imagery. Here we combine these DL methods and high-resolution planet imagery to segment forested areas in the Brazilian Pantanal wetland. We compared the performances of multiple DL-based networks. For the transformer method, we used Segformer and for CNN networks we used DeepLabV3+ and OCRNet. considering RGB and near-infrared within a large dataset of 722 image patches (4096×4096 pixels). We later verified the generalization capability of the model by dividing the dataset into three equal-sized parts. As a result, the transformer method SegFormer presented an average of each batch of the divided result of F1-score 96.34% $\pm$0.99 and IoU of 92.95% $\pm$1.83 for the validation set of the dataset and provided the best results between the analyzed networks.

For the update of the paper we also will provide a link for the finished paper in the repository `https://github.com/Jose-Augusto-C-M?tab=repositories`, since at the time of the thesis presentation this paper will not yet be published.

Keywords:

1. Introduction

A wetland consists of an ecosystem that depends on constant or recurrent flooding or saturation, shallow or close to the surface of the substrate, that

[1]FAENG - Universidade Federal de Mato Grosso do Sul, Brazil
[2]FACOM - Universidade Federal de Mato Grosso do Sul, Brazil

is, flooding. And due to this characteristic, over time, nature adapts, and an ecosystem develops. Therefore, another characteristic must be the presence of physical, chemical, and biological characteristics that reflect these episodes of flooding or recurrent and sustained saturation. Another common diagnostic characteristic of wetlands is hydric soils, and hydrophilic vegetation [1]. Until recently, worldwide policies were intended to encourage or subsidize the conversion of wetlands to drained or filled lands that could be used for agriculture, urbanization, or other purposes not compatible with the existence of wetlands [2]. Wetland conversion is a global phenomenon at a global scale [3]. It is estimated that freshwater wetlands have lost nearly half of their area since 1900 [3]. The reasons for that are many; we can cite: water withdrawal, infrastructure development (dams, dikes, levees, diversions), land conversion, over-harvesting, land exploitation, the introduction of exotic species, pollution [4]. Between the 1780s and 1980s, wetlands in the contiguous US decreased by 53%, with losses in Ohio and California reaching 90% and 91%, respectively. [5] Over 80% of Canada's wetlands, which are mostly in the southern quarter, have been converted to agricultural or urban uses; the former being the primary reason for 85% of the country's wetland losses [6]. So we can observe that the conversion to agricultural and arable land primarily causes the wetland land conversion. It is predicted that by 1985, 56–65% of the accessible wetland in Europe and North America, 27% in Asia, 6% in South America, and 2% in Africa had been drained [4]. And more recently, it has been estimated that at least 33% of the world's wetlands, including 2.64 million km² of open water and 4.58 million km² of non-water wetlands, had been destroyed as of 2009 [7]. These estimations are based on meta-analysis [8, 9] of wetland data bibliography and case studies. Significant discrepancies exist in information on wetlands, and estimates of the global wetlands remaining area range from 1.53 to 14.86 million km2 [7].

Suppose we want to have an objective mapping in different scales of the elements that are the components of the landscape of a big and rich environment such as the wetlands. For that task, we may need to perform a semantic segmentation computer vision task. It gives pixel-wise category predictions rather than predictions for the whole image [10]. The semantic segmentation task can be interpreted as a kind of thematic mapping and is largely used to extract data as demanded from maps and any type of digital tensors. The chosen wetland for this research is the Pantanal, one of the largest wetlands on the planet.

The objective is to map the vegetation cover of the Pantanal. The justification for that is quantifying tree canopy cover change for a dynamic system like Pantanal. By doing this, we can identify the impact of events and estimate the effects of these events on the conservation of the region. Deforestation is one of the listed sustainable development goals (SGD), being the goal of number 15 [11]. And we can take the permission to say that the loss of forest cover and the changes in the configuration of terrestrial life greatly impact almost all of the objectives.

We want to target the problem of quantifying the trees with purely remote sensing as data source and process it with deep learning (DL) techniques [12, 13].

Remote sensing is the science and skill of gathering data about an object via a device that is placed remotely from the source of information [14]. For this case, the overall objective of computer vision procedures is to "automatically categorize all pixels in an image into land cover classes or themes [14]". Two more valuable statements: Spectral patterns, that is, pixels with comparable spectral combination classes of reflectance and emissivity, are combined to represent specific types of surface characteristics [14]. Spatial pattern recognition means classifying image pixels according to how they are positioned with respect to other pixels [14].

The introduction of DL in more recent years has sparked an interest in neural networks. Since 2014, the remote-sensing community has focused on DL since it has shown significant success in various image analysis tasks, such as object recognition and land use and land cover (LULC) classification [15–18], data fusion [19], segmentation [20], change detection [21] and registration [22]. The object of quantifying an area of trees canopy cover of a dynamic system such as Pantanal is made possible with remote sensing and deep learning tools in hand. For this objective, we want to map the vegetation resources of the Brazilian Pantanal wetland with deep learning methods that use Convolutional Neural Network (CNN) architectures [23], and transformer-based architectures [13].

2. Materials and Methods

2.1. Study Area

The study area is located in Brazilian Pantanal. This region is one of the largest wetlands on the planet. Pantanal is situated in the geographic heart of South America and drains the Cerrado of central Brazil it covers 361,666 km² estimated for the Upper Paraguay Basin [24]. Of that, 140,000 km² is in Brazilian territory, or 38.21% of the basin's total area [25]. The study area map is in Figure 1.

Pantanal is a vibrant and diverse ecosystem with immeasurable value in terms of habitats and species of fauna and flora due to the wide variety of soil types and flood regimes present in the biome. As an example of the wealth of natural biodiversity present in the Pantanal, we have that it is the richest wetland in bird species in the world, with 463 cataloged and with new studies carried out over time with different sampling techniques and bigger numbers obtained. [24]. The wetland consists of an ecosystem dependent on constant or recurrent flooding or soil saturation, shallow or close to the surface of the substrate. Moreover, due to this flooding characteristic, nature adapts over time, and a flooding-dependent ecosystem develops. Therefore, another characteristic of wetland zones must be the presence of physical, chemical, and biological structures that reflect these episodes of flooding or recurrent and sustained saturation over time. Other common diagnostic characteristics of wetlands are hydric soils and hydrophilic vegetation [1]. On Figure 2, we can see some examples of scenic environments that we can find on Pantanal.

Wildfires are currently the most significant concerning aspect of the Pantanal's conservation [27, 28]. Every year, the Pantanal experiences 7512 fire spots starters

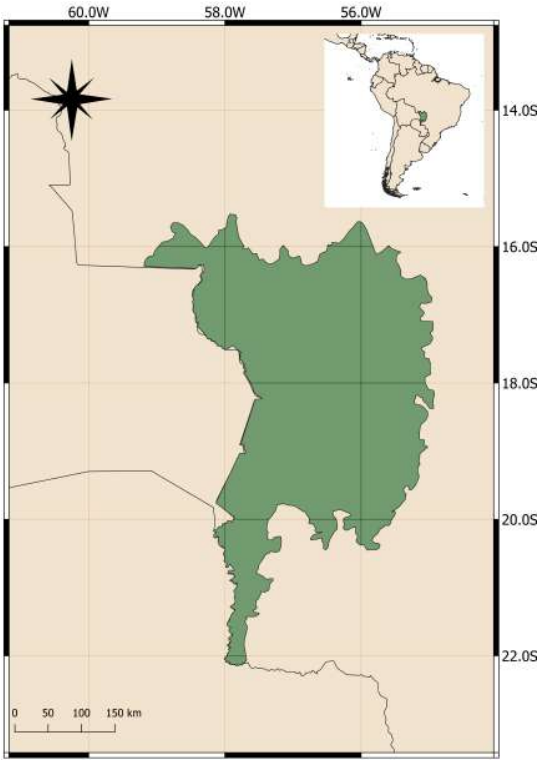


Figure 1: Map of the study area, in detach the Brazilian area of Pantanal wetland.

on average [29]. One of the worst years in recent memory, 2020, saw fires in the Pantanal spread to 43 % of the area, killing around 17 million vertebrates [27, 28, 30, 31]. Additionally, over the past twenty years, the Pantanal has demonstrated a propensity to see an increase in burned regions [32]. Another major concern for the conservation of this region is advance of the agro-industries. Pantanal is a region increasingly threatened by large development programs; these programs alter the region's flood condition and severely change the type of land use, increase deforestation and promote the spread of all kinds of environmental contaminants [24, 33–36].

2.2. Dataset

In this study, Planet and NICFI Basemaps for Tropical Forest Monitoring - Tropical Americas [37] satellite imagery was used. The dataset covers the whole Brazilian Pantanal Biome with a spatial resolution of 4.77 meters and blue, green, red, and near-infrared bands. NICFI mosaics are composed of monthly and semester collections. For this study case images from July 2022 were used.

The dataset was downloaded using the portal `https://code.earthengine.google.com/`, and its download functions using a python API `https://geemap.`

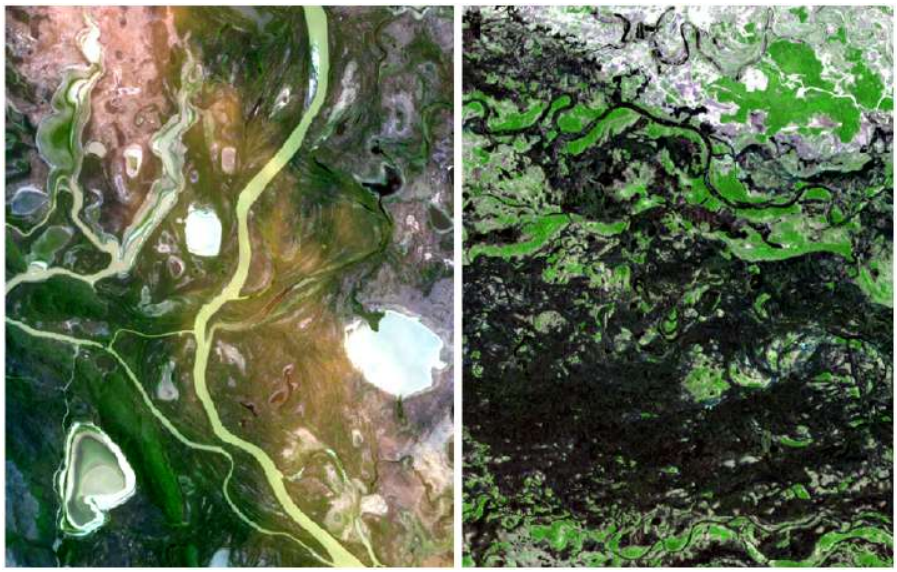


Figure 2: These figures were created using data from the google earth engine visualization tool [26], accessed via the link: `https://code.earthengine.google.com/`, the scale of the figures is 1:100000.

`org/` developed by [38]. For this, we created a script to download a batch of what represents the images that are inside our study area for the period determined, which was the first image of the dataset of 2022. The file format for the download was .tiff and also can be .npy, but the training procedure was executed using the .npy files because they are a simpler format than the format tiff.

To gather the reference data (i.e. vegetated areas), manual labeling was performed by specialists with the assistance of the Geographical Information System (GIS) open-source software QGIS 3.22. Within the Pantanal. After the download of the dataset we created annotations for each image, called masks or ground truth, which was inputted to the network in a .png format, and the raw image patch was inputted as mentioned before in a .npy format for each corresponding mask. It is important to note that some of the annotations are provided blank.

2.3. Deep learning methods

To segment and map the vegetated areas, we used state-of-the-art semantic segmentation networks. We compared recent ViT-based methods, such as SegFormer [10] with known CNN-based methods, like OCRNet [39] and DeepLabV3+ [40]. In general, segmentation methods take an image as input and return a pixel-wise classification. In our case, the result of each method is an image where the value of each pixel is categorized in a class lying between background or a burned area. Traditional DL methods use convolution, pooling, and fully

connected layers such as DeepLabV3+ and OCRNet. As stated, Transformers have been used as a replacement for convolution layers to take advantage of both global and local attention in the image. As traditional CNN methods are commonly explored in remote sensing, we did not describe them in detail. Below, we only describe the focused Transformer-based methods: SegFormer.

[41] developed a deep CNN model named DeepLabv3+. This model is an improvement of the previous DeepLab versions. This versions uses and encoder which takes advantages of several blocks designed to capture features from various scales at once, such as atrous convolutions, depthwise separable convolutions, Atrous Spatial Pyramid Pooling (ASPP) and a simple but effective decoder capable of refining the segmentation result map especially in object boundaries. A modified Xception version was used as the network backbone.

[39] proposed a Object Contextual Representation (OCR) scheme for semantic segmentation. The method consists in to augment the representation of one pixel given the context of the object representation in which that pixel is contained. The method intuition is simple yet effective. First, a coarse representation is obtained from a conventional DCNN. In sequence, the object regions are estimated. Lastly, the object contextual representations are computed and augmented.

SegFormer [10] is an efficient semantic segmentation method that combines a Transformer encoder and multilayer perceptron decoder. The structure of Segformer is encoder-decoder. Through hierarchically organized Transformers, multi-scale features are obtained from the image in the encoder. Convolutional layers, as opposed to the conventional Transformer, are used to implement the position encoder on the encoder because they perform better at various image resolutions. The multi-scale features are combined in the decoder to represent both local and global data. Finally, the input image is segmented using the combined features. Despite having a straightforward decoder, SegFormer outperformed other methods in conventional computer vision benchmark datasets.

2.4. Experimental Design

We split the areas into patches of size 256×256 pixels without overlap due to the input dimension limitations of DL methods. A total of 722 patches were obtained from the images. To train and evaluate the models, the dataset was divided into training, validation, and testing sets with approximately 70%, 15%, and 15%, respectively. The models were evaluated considering three repetitions of the splits. In each iteration, the validation and test sets were chosen so there were no repeated samples between the current and previous splits. To ensure reliable results, the mean and standard deviation were estimated concerning all evaluated metrics. The models were trained using the open-source semantic segmentation toolbox MMSegmentation [42]. The state-of-the-art Deeplabv3+, OCRNet and SegFormer architectures were applied. The Deeplab v3+ and OCRNet models used ResNet 101 [43] model as the backbone. SegFormer is not compatible with convolution-based backbones and uses its own transformer-based backbone. Mit-b5 `https://huggingface.co/nvidia/mit-b5` was used as SegFormer backbone. All models were trained with $40k$ iterations using the

AdamW optimizer algorithm with a learning rate of 6×10^{-5} and default weight decay of 10^{-2}. The learning rate scheduler makes the model head learning rate ten times bigger than the backbone learning rate. The batch size was set to 8 for all models, except for SegFormer, which was set to 2 due to GPU memory constraints.

The training and inference were performed inside a google co-laboratory environment where we can access GPU on demand, also we have 32GB of RAM memory available. The GPU used on the occasion of the project was a NVIDIA T4 `https://www.nvidia.com/en-us/data-center/tesla-t4/`. The codes used in this research will be available in `https://github.com/Jose-Augusto-C-M/Pantanal_benchmark_Segmentation.git`.

2.5. Evaluation metrics

The performance of the models is evaluated using the metric F1-score (F_1) (Equation 3), pixel accuracy (Equation 1), and the Intersection over Union (IoU) (Equation 2), as they are currently used to assess semantic segmentation experiments [44].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$IoU = \frac{|GT \cap Prediction|}{|GT \cup Prediction|} \tag{2}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

The F1-score metric is calculated based on the weighted average of Precision and Recall, where an F1-score reaches its best value at one and the worst score at 0. The precision metric is defined as the number of True Positives (TP) divided by the number of true positives (TP) plus the number of False Positives (FP). The Recall metric is defined as the number of true positives (TP) over the number of true positives (TP) plus the number of False Negatives (FN). The IoU, also known as the Jaccard Index, is the ratio between the intersection and the union between the ground truth (GT) and the prediction masks.

3. Results

3.1. Quantitative evaluation of Image Segmentation Methods

The results for the pixel accuracy, F1-score and IoU metrics are presented in Tables 1 and 2. We report metrics separately for background (BG) and vegetated area (Tree) pixels for a complete analysis of the results, as the occurrence of vegetated area pixels tends to be lower than the overall background. As we can see, SegFormer excelled for the metrics for the two classes, background and

vegetated areas, for the validation sets and in some cases, it was outperformed by the Deeplabv3+, but with also a big standard deviation for the Deeplabv3+.

Considering the IoU of the burned area, the Segformer obtained 92.14 against 90.75 for the DPT, the second-best method. This evidences the robustness of Transformers against convolutional layers, as both methods are based on this recent advance. Considering pixel accuracy, the best segmentations were from SegFormer, FCN, and DPT with metrics above 96%. For the F-score, the methods presented similar results for SegFormer, DPT, OCRNet, and FCN.

Table 1: Semantic segmentation results in test sets.

	Accuracy		F1-score		IoU	
	BG	Tree	BG	Tree	BG	Tree
Deeplab v3+	$95,89 \pm 1,77$	$95,29 \pm 1,54$	$96,53 \pm 1,95$	$96,09 \pm 2,19$	$92,92 \pm 4,34$	$92,54 \pm 4,02$
OCRNet	$95,57 \pm 0,06$	$93,23 \pm 0,96$	$94,53 \pm 0,44$	$94,29 \pm 0,61$	$89,63 \pm 0,79$	$89,20 \pm 1,08$
SegFormer	$96,25 \pm 1,16$	$94,71 \pm 1,88$	$95,61 \pm 0,42$	$95,44 \pm 0,74$	$91,59 \pm 0,78$	$91,28 \pm 1,35$

Table 2: Semantic segmentation results in validation sets.

	Accuracy		F1-score		IoU	
	BG	Tree	BG	Tree	BG	Tree
Deeplab v3+	$95,79 \pm 0,03$	$94,54 \pm 1,39$	$95,34 \pm 0,68$	$95 \pm 0,65$	$91,11 \pm 1,24$	$90,48 \pm 1,17$
OCRNet	$96,11 \pm 1,15$	$94,77 \pm 1,52$	$95,64 \pm 1,36$	$95,27 \pm 1,31$	$91,67 \pm 2,48$	$90,98 \pm 2,43$
SegFormer	$96,71 \pm 0,71$	$96,25 \pm 1,36$	$96,61 \pm 1,01$	$96,34 \pm 0,99$	$93,45 \pm 1,90$	$92,95 \pm 1,83$

3.2. Qualitative evaluation of Image Segmentation Methods

The last experiment was carried out to determine the robustness and generalizability of the model built in the earlier steps using the SegFormer network. To segment the vegetated areas in Brazilian Pantanal forest zones, we used the SegFormer, which was exclusively trained on Pantanal photos.

4. Discussion

Vegetation mapping is an important study area in remote sensing and earth observation. It is vital in various fields, including conservation, land management, agriculture, and climate change research. Vegetation mapping can provide information about the distribution and type of vegetation in a given area, which is used to make informed decisions about land use, conservation efforts, and sustainable resource management. One important use of vegetation mapping is in conservation and land management. Vegetation mapping can have the finality of identifying and mapping areas of high biodiversity and tracking changes in vegetation cover over time. This information is used to develop conservation strategies and monitor the effectiveness of conservation efforts. Another important use of vegetation mapping is in agriculture. By mapping the distribution and type of vegetation in a given area, farmers can better understand

Figure 3: These figures represent examples of the result of the inference model of the Segformer architecture for the whole Pantanal region.

the resources available to them and make informed decisions about crop selection, irrigation, and fertilizer use. Vegetation mapping can identify areas suitable for crop cultivation, which can help improve food security in regions where arable land is limited. Climate change research is another area where vegetation mapping is essential. Vegetation plays a critical role in the global carbon cycle, and vegetation mapping can provide an understanding of how vegetation covers changes and species may affect the global climate. Vegetation mapping can also be used to monitor how different land uses, such as deforestation or reforestation, may affect the global carbon cycle. Overall, vegetation mapping is a critical tool for understanding and managing the earth's resources, and it is vital for conservation, land management, agriculture, and climate change research.

Transformer networks are a type of neural network architecture that was introduced in the paper "Attention Is All You Need" by Google researchers in 2017 [45]. The architecture is based on the idea of self-attention, where the network can weigh the importance of different parts of the input when making a prediction. This allows the network to effectively process sequential data, such as natural language processing (NLP), where the order of the words is important. The transformer network uses a multi-head self-attention mechanism to process the input, which allows it to capture dependencies between different parts of the input in a more sophisticated way than traditional recurrent neural networks (RNNs) or convolutional neural networks (CNNs). Additionally, the transformer network uses a feed-forward neural network as the core of the model, which allows it to learn a more complex function. The transformer network architecture is highly effective for various natural languages processing tasks, such as machine translation and language understanding. It has become the de facto standard for these tasks. The transformer-based models are also used in computer vision and other domains. Many popular pre-trained transformer-based models such as GPT-2, BERT, T5, RoBERTa, etc have been introduced by researchers and are being used widely in the NLP-related tasks. Because of the success in NLP fields transformers networks have been tested in various remote sensing applications, including vegetation mapping. In this context, the goal is typically to use satellite or aerial imagery to map the distribution and type of vegetation in a given area.

Segmenting vegetated areas in the largest wetland ecosystem on the planet is an important procedure that environmental and governmental institutions can use in decision-making tasks. As regarded previously, current information for the affected areas mapped in this study is produced by an online platform called Mapbiomas [46] `https://mapbiomas.org/`. The method presented in this paper uses a finer spatial resolution dataset (4.77m) as presented in section 2.2 and new methods of deep learning that are presented in section 2.3. Creating a more accurate vegetation map for the region. And with that, we demonstrated that combining deep learning methods and remote sensing imagery, such as the PlanetScope with RGB + NIR spectral bands is suitable to map these areas in the Pantanal. As this method prove feasible to return high-detailed maps, it also demonstrates the potential of using daily data provided by planet [37]. Increasing the frequency of monitoring for deforestation being useful for environmental planning in both controlling the current damages and restoring the destroyed areas.

It should be noted that mapping vegetated areas in wetlands is also a difficult task for humans mainly because of the amount of humid, water bodies and other sources of noise inside the environment [47]. Making the labeling process a complex task. When considering only RGB + NIR information, some of these regions tend to confuse manual labeling processes because it is difficult to distinguish between agricultural areas or open grass fields and tree canopy cover. Regardless, the DL methods tested were quite capable of dealing with the wetland's natural characteristics. Quantitatively Tables 1 and 2. Returned similar results, and visually the images presented in Figure 3 present a good prediction of the forested areas inside Pantanal. The model could differentiate

both natural water bodies and agricultural regions of bare soil, burned areas, and humid soils.

Further studies should consider the combination of preliminary segmentation methods and DL networks, evaluating the impact of, for example, weakly-supervised methods and how well the methods are capable of improving the original segmentation. Another important piece of information to be evaluated is an analysis of multi-temporal imagery segmentation. Daily monitoring of vegetation is important to detect and direct actions to stop it as soon as possible, minimizing the damage. Lastly, domain adaptation techniques to deal with multiple sensor data and few-shot and sparse labeling investigations may be useful in novel approaches to improving the current method's generalization. These processes are considered state-of-the-art approaches [48–51] in computational vision tasks, and remote sensing imagery may greatly benefit from its integration with current ViT or CNN-based methods to investigate forested areas. Regardless, the current method proved satisfactory performance over difficult analysis situations. It indicates that visible to near-infrared regions and high-spatial detailed imagery is suitable for mapping forested areas in the wetlands.

5. Conclusion

We investigated the capabilities of deep learning methods, in specific Transformer-based networks, in mapping forested areas in the Brazilian Pantanal wetland. The results demonstrated that the networks based on vision transformers resulted in better accuracy, F1-score, and IoU than traditional CNNs architectures such as Deeplabv3+, with a better and more regular result. We have used a method of dividing the dataset into three random sets of images that had the same number of images and together composed the entirety of the data. For the validation sets of analysis, the architecture SegFormer returned an F1 score of 95.44% with a deviation of 0.74% and an IoU of 91.28 with a deviation of 1.35% between the sets. The experimental results and the division of the dataset indicate that DL models trained with a small fraction of the dataset can be generalized to other areas inside Pantanal. Furthermore, we conclude that Transformer-based networks are fit to deal with vegetated areas inside Pantanal using high-spatial-resolution imagery. Future studies should bring on vision transformer architectures to perform said task.

References

[1] E. Carp, Directory of wetlands of international importance in the western Palearctic, Iucn, 1980.

[2] N. R. Council, et al., Wetlands: Characteristics and boundaries, National Academies Press, 1995.

[3] S. v. Asselen, P. H. Verburg, J. E. Vermaat, J. H. Janse, Drivers of wetland conversion: a global meta-analysis, PloS one 8 (11) (2013) e81292.

[4] M. E. Assessment, Synthesis report, Island, Washington, DC.

[5] T. E. Dahl, Wetlands losses in the United States, 1780's to 1980's, US Department of the Interior, Fish and Wildlife Service, 1990.

[6] P. Douaud, T. A. Radenbaugh, Changing Prairie Landscapes, Vol. 32, University of Regina Press, 2000.

[7] S. Hu, Z. Niu, Y. Chen, L. Li, H. Zhang, Global wetlands: Potential distribution, wetland loss, and status, Science of the total environment 586 (2017) 319–327.

[8] L. V. Hedges, Meta-analysis, Journal of Educational Statistics 17 (4) (1992) 279–296.

[9] R. Rosenthal, M. R. DiMatteo, Meta-analysis: Recent developments in quantitative methods for literature reviews, Annual review of psychology 52 (1) (2001) 59–82.

[10] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, Segformer: Simple and efficient design for semantic segmentation with transformers (2021). doi:10.48550/ARXIV.2105.15203.
URL https://arxiv.org/abs/2105.15203

[11] J. D. Sachs, From millennium development goals to sustainable development goals, The lancet 379 (2012) 2206–2211.

[12] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436–444.

[13] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.

[14] T. Lillesand, R. W. Kiefer, J. Chipman, Remote sensing and image interpretation, John Wiley & Sons, 2015.

[15] Y. Chen, Z. Lin, X. Zhao, G. Wang, Y. Gu, Deep learning-based classification of hyperspectral data, IEEE Journal of Selected topics in applied earth observations and remote sensing 7 (6) (2014) 2094–2107.

[16] Q. Zou, L. Ni, T. Zhang, Q. Wang, Deep learning based feature selection for remote sensing scene classification, IEEE Geoscience and Remote Sensing Letters 12 (11) (2015) 2321–2325.

[17] D. Marmanis, M. Datcu, T. Esch, U. Stilla, Deep learning earth observation classification using imagenet pretrained networks, IEEE Geoscience and Remote Sensing Letters 13 (1) (2015) 105–109.

[18] R. Ghali, M. A. Akhloufi, W. S. Mseddi, Deep learning and transformer approaches for UAV-based wildfire detection and segmentation, Sensors 22 (5) (2022) 1977. doi:10.3390/s22051977.
URL https://doi.org/10.3390/s22051977

[19] J. Gao, P. Li, Z. Chen, J. Zhang, A survey on deep learning for multimodal data fusion, Neural Computation 32 (5) (2020) 829–864.

[20] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, D. Terzopoulos, Image segmentation using deep learning: A survey, IEEE transactions on pattern analysis and machine intelligence.

[21] L. Khelifi, M. Mignotte, Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis, Ieee Access 8 (2020) 126385–126400.

[22] Z. Zhang, Y. Dai, J. Sun, Deep learning based point cloud registration: an overview, Virtual Reality & Intelligent Hardware 2 (3) (2020) 222–246.

[23] S. Albawi, T. A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, in: 2017 international conference on engineering and technology (ICET), Ieee, 2017, pp. 1–6.

[24] M. B. HARRIS, W. TOMAS, G. MOURAO, C. J. D. SILVA, E. GUIMARAES, F. SONODA, E. FACHIM, Safeguarding the pantanal wetlands: Threats and conservation initiatives, Conservation Biology 19 (2005) 714–720. doi:10.1111/j.1523-1739.2005.00708.x.

[25] J. Silva, M. d. M. Abdon, Delimitacao do pantanal brasileiro e suas subregioes, Pesquisa Agropecuaria Brasileira, Brasilia, v. 33, p. 1703-11.

[26] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, R. Moore, Google earth engine: Planetary-scale geospatial analysis for everyone, Remote sensing of Environment 202 (2017) 18–27.

[27] L. C. Garcia, J. K. Szabo, F. de Oliveira Roque, A. d. M. M. Pereira, C. N. da Cunha, G. A. Damasceno-Júnior, R. G. Morato, W. M. Tomas, R. Libonati, D. B. Ribeiro, Record-breaking wildfires in the world's largest continuous tropical wetland: integrative fire management is urgently needed for both biodiversity and humans, Journal of environmental management 293 (2021) 112870.

[28] R. Libonati, C. C. DaCamara, L. F. Peres, L. A. Sander de Carvalho, L. C. Garcia, Rescue brazil's burning pantanal wetlands, Nature Publishing Group 588.
URL https://media.nature.com/original/magazine-assets/d41586\-020-03464-1/d41586-020-03464-1.pdf

[29] J. F. de Oliveira-Junior, P. E. Teodoro, C. A. da Silva Junior, F. H. R. Baio, R. Gava, G. F. Capristo-Silva, G. de Gois, W. L. F. Correia Filho, M. Lima, D. de Barros Santiago, et al., Fire foci related to rainfall and biomes of the state of mato grosso do sul, brazil, Agricultural and Forest Meteorology 282 (2020) 107861.

[30] W. M. Tomas, C. N. Berlinck, R. M. Chiaravalloti, G. P. Faggioni, C. Strüssmann, R. Libonati, C. R. Abrahão, G. do Valle Alvarenga, A. E. de Faria Bacellar, F. R. de Queiroz Batista, et al., Distance sampling surveys reveal 17 million vertebrates directly killed by the 2020's wildfires in the pantanal, brazil, Scientific reports 11 (1) (2021) 1–8.

[31] R. Libonati, J. L. Geirinhas, P. S. Silva, A. Russo, J. A. Rodrigues, L. B. Belém, J. Nogueira, F. O. Roque, C. C. DaCamara, A. M. Nunes, et al., Assessing the role of compound drought and heatwave events on unprecedented 2020 wildfires in the pantanal, Environmental Research Letters 17 (1) (2022) 015005.

[32] D. B. Correa, E. Alcântara, R. Libonati, K. G. Massi, E. Park, Increased burned area in the pantanal over the past two decades, Science of The Total Environment 835 (2022) 155386.

[33] C. J. Alho, L. M. Vieira, Fish and wildlife resources in the pantanal wetlands of brazil and potential disturbances from the release of environmental contaminants, Environmental Toxicology and Chemistry: An International Journal 16 (1) (1997) 71–74.

[34] A. F. Seidl, J. d. S. V. de Silva, A. S. Moraes, Cattle ranching and deforestation in the brazilian pantanal, Ecological Economics 36 (3) (2001) 413–425.

[35] P. M. Fearnside, et al., Can pasture intensification discourage deforestation in the amazon and pantanal regions of brazil, Deforestation and land use in the Amazon (2002) 283–364.

[36] J. Silva, M. d. M. ABDON, S. M. A. da SILVA, J. A. de MORAES, Evolution of deforestation in the brazilian pantanal and surroundings in the timeframe 1976-2008.

[37] P. Team, Planet team, planet application program interface: In space for life on earth, https://api.planet.com, 2017 (2017).

[38] Q. Wu, geemap: A python package for interactive mapping with google earth engine, Journal of Open Source Software 5 (2020) 2305.

[39] Y. Yuan, X. Chen, X. Chen, J. Wang, Segmentation transformer: Object-contextual representations for semantic segmentation`doi:10.48550/ARXIV.1909.11065`.
URL `https://arxiv.org/abs/1909.11065`

[40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation (2018). `doi:10.48550/ARXIV.1802.02611`.
URL `https://arxiv.org/abs/1802.02611`

[41] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

[42] M. Contributors, MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark, https://github.com/open-mmlab/mmsegmentation (2020).

[43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition (2015). doi:10.48550/ARXIV.1512.03385.
URL https://arxiv.org/abs/1512.03385

[44] X. Yuan, J. Shi, L. Gu, A review of deep learning methods for semantic segmentation of remote sensing imagery, Expert Systems with Applications 169 (2021) 114417.

[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, CoRR abs/1706.03762. arXiv:1706.03762.
URL http://arxiv.org/abs/1706.03762

[46] C. M. Souza, J. Z. Shimbo, M. R. Rosa, L. L. Parente, A. A. Alencar, B. F. T. Rudorff, H. Hasenack, M. Matsumoto, L. G. Ferreira, P. W. M. Souza-Filho, S. W. de Oliveira, W. F. Rocha, A. V. Fonseca, C. B. Marques, C. G. Diniz, D. Costa, D. Monteiro, E. R. Rosa, E. Vélez-Martin, E. J. Weber, F. E. B. Lenti, F. F. Paternost, F. G. C. Pareyn, J. V. Siqueira, J. L. Viera, L. C. F. Neto, M. M. Saraiva, M. H. Sales, M. P. G. Salgado, R. Vasconcelos, S. Galano, V. V. Mesquita, T. Azevedo, Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine, Remote Sensing 12 (2020) 2735. doi:10.3390/rs12172735.

[47] L. Higa, J. M. Junior, T. Rodrigues, P. Zamboni, R. Silva, L. Almeida, V. Liesenberg, F. Roque, R. Libonati, W. N. Gonçalves, J. Silva, Active fire mapping on brazilian pantanal based on deep learning and CBERS 04a imagery, Remote Sensing 14 (3) (2022) 688. doi:10.3390/rs14030688.
URL https://doi.org/10.3390/rs14030688

[48] D. Tuia, C. Persello, L. Bruzzone, Domain adaptation for the classification of remote sensing data: An overview of recent advances, IEEE Geoscience and Remote Sensing Magazine 4 (2016) 41–57. doi:10.1109/MGRS.2016.2548504.

[49] H. Wang, E. Cimen, N. Singh, E. Buckler, Deep learning for plant genomics and crop improvement, Current Opinion in Plant Biology 54 (2020) 34–41. doi:10.1016/J.PBI.2019.12.010.

[50] Y. F. Li, L. Z. Guo, Z. H. Zhou, Towards safe weakly supervised learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (2021) 334–346. doi:10.1109/TPAMI.2019.2922396.

[51] A. Farahani, S. Voghoei, K. Rasheed, H. R. Arabnia, A brief review of domain adaptation (2021) 877–894doi:10.1007/978-3-030-71704-9_65/COVER.
URL https://link.springer.com/chapter/10.1007/978-3-030-71704-9_65

CHAPTER 6

Discussions and Conclusions

6.1 Remote sensing, deep learning and vegetation mapping

By combining remote sensing and deep learning, vegetation mapping can be performed at different scales with different sensors with high accuracy. Remote sensing data provides a synoptic view of the Earth's surface, while deep learning algorithms can extract complex patterns and relationships within this data to produce detailed vegetation maps. These maps can be used to monitor and understand vegetation distribution, composition, and dynamics over time and to support a wide range of applications, such as land-use planning, forest management, and monitoring of vegetation health (COLOMINA; MOLINA, 2014; WHITE et al., 2016; SHEN, 2018). One example of remote sensing and deep learning for vegetation mapping is the use of satellite imagery and deep neural networks to map forests and other vegetation types as presented in Chapter 5. In this application, satellite imagery is fed into different deep neural networks, which uses complex algorithms to analyze the raw data and produce a vegetation map. The resulting map can be used to monitor and understand the distribution, composition, and health of forests and other kinds of vegetation, providing valuable information for land-use planning, forest management, and conservation efforts. Remote sensing and deep learning have the potential to revolutionize vegetation mapping, providing a cost-effective and efficient way to monitor and understand the Earth's vegetation. Combining these technologies makes it possible to generate detailed and accurate vegetation maps at a large scale, which can support a wide range of applications.

In the papers presented in this research we developed techniques of quantifying vegetation inside different environments and scales of environments. In Chapters 3, 4 and 5. are presented a workflows that contains a supervised and semi-supervised methods for the creation of datasets and computer models that perform image segmentation of vegetation inside a diversity of environments, and be applied to a great diversity of uses. Segmentation and the development of computer models related to vegetation inside cities and forests have a wide range of applications as presented in (BARÓ et al., 2014), some of then we can describe:

- Urban Forest Management: Computer models can be used to understand the distribution, composition, and dynamics of urban forests, providing valuable information for urban forest management and planning. For example, by using satellite imagery and machine learning algorithms, urban forest canopies can be segmented and analyzed to understand their contribution to the urban environment.
- Urban Green Space Planning: Segmentation and computer models can be used to understand the distribution and distribution of urban green spaces, such as parks and gardens, and how these spaces contribute to the urban environment. This information can be used to inform green space planning, design, and management, ensuring that urban green spaces are optimized for their desired functions and benefits.
- Biodiversity Assessment: Computer models can be used to understand the composition and structure of vegetation in forests and other natural areas, providing valuable information for biodiversity assessment and conservation planning. For example, by using remote sensing data and machine learning algorithms, species distribution and abundance can be estimated, allowing for biodiversity assessment in a given area.
- Forest Health Monitoring: Computer models can be used to monitor the health and productivity of forests, providing valuable information for forest management and conservation. For example, by using remote sensing data and machine learning algorithms, changes in vegetation structure and productivity can be detected, allowing for early detection of threats such as disease, pest infestations, and deforestation.

6.2 Thesis contributions

This chapter will explain the scientific and technical implications for the academic community and society of the research findings. The doctoral dissertation included a thorough investigation of remote sensing and deep learning data processing and demonstrated applications in the fields of urban vegetation mapping and wetland vegetation mapping, which have possibilities for improvement and proposals of novel techniques such

as the ones presented in this doctoral dissertation. This is because computer vision and digital image processing techniques are indispensable in contexts where it is necessary to obtain information quickly, updated, and automatically on digital images. Although remote sensing images are easily available and in large quantities, their use presents a real difficulty in processing by having a large amount of undesirable noise and imperfections that impair the analysis for feature extraction. Automated vegetation detection or any other information from remote sensing imagery remains challenging.

At the start of the doctoral dissertation research, we were faced with a demand for smart cities-related applications of segmenting trees inside an urban area with high-resolution images (10cm GSD). This demand resulted in the article of chapter 3, in which we developed a method that can be described as a supervised learning method. We picked random aerial orthoimages, annotated them, and then trained five state-of-the-art Convolutional Neural Networks to identify trees inside the urban area. We created a method for urban managers to measure and quantify the urban landscape vegetation.

In chapter 4, we have a workflow containing a semi-supervised method for creating the dataset to perform the training. The use of labeled data for the training of algorithms is the most usual way; however, it is considered a laborious task. The computer vision community, over the years, developed ways of diminishing and mitigating this task or even, in some cases excluding it. We created a more diverse and rich dataset for deep learning algorithms. Furthermore, since we were working with many different plant species, this increase of information for training was fundamental to accurately segmenting the plant species area. We generated results to support the measurement and quantification of the plant health of a region.

In chapter 5, we have a workflow containing supervised method. We used planet satellite data provided via a project Planet and NICFI Basemaps for Tropical Forest Monitoring (PLANET, 2017). The satellite imagery was collected by Planet's Dove and SkySat satellites, which are able to capture images of the earth's surface with a high level of resolution and accuracy. The article uses Planet and NICFI images and combination of deep learning architectures in the form of convolutional neural networks (CNN) and transformer based neural network to create detailed distribution vegetation maps of the Brazilian Pantanal. These maps can be used to support a wide range of activities related to tropical forest monitoring, such as carbon stock assessments, monitoring of deforestation and degradation, and land use planning. The authors would also like to highlight the importance of these detailed maps in the context of REDD+ (Reducing Emissions from Deforestation and Forest Degradation) programs, which aim to create financial value for the carbon stored in forests, offering incentives for developing countries to reduce emissions from forested lands and invest in low-carbon paths. Overall, the article presents a promising approach for wetland forest monitoring using high-resolution satellite imagery

and deep learning. It highlights the potential for cost-effective and accurate monitoring of wetlands and the importance of this kind of data for management and conservation.

This document presents three papers with very different manners of obtaining and processing the data, to get to the finality of teaching a mechanical machine to identify a pattern. This research contribution has provided new and innovative approaches to vegetation mapping, allowing for a more comprehensive understanding of the distribution, composition, and dynamics of vegetation over time. The results of this research can be used to inform a wide range of decision-making processes, from urban green space planning to conservation efforts. By providing a cost-effective and efficient way to monitor and understand vegetation.

6.3 Future directions

The use of remote sensing and deep learning for vegetation mapping is a rapidly evolving field with many exciting opportunities for future research. Some potential future directions for this field may include:

- Integration of Technologies: Integrating remote sensing and deep learning with other technologies, will continue to provide new and innovative ways to monitor and understand vegetation. And with the increased spatial and temporal Resolutions will make possible the development of new and improved remote sensing platforms. Providing a more comprehensive understanding of vegetation dynamics over time and support various applications, such as monitoring vegetation health and productivity. Generating information to support a wide range of decision-making processes.
- For vegetation monitoring we also have the necessity of increasing automation and scalability. The development of automated and scalable methods for vegetation mapping with remote sensing and deep learning will enable the mapping of vegetation at a larger scales with increased efficiency. This will provide valuable information for a wide range of applications, such as land-use planning and deforestation monitoring.

6.4 Conclusion

In conclusion, using remote sensing and deep learning for vegetation mapping is a rapidly evolving field with many exciting opportunities for future research. By developing and improving these methods, it is possible to generate more detailed and accurate vegetation maps and support a wide range of decision-making processes related to environmental monitoring and management.

My doctorate formation has greatly improved my skills as a scientific researcher and writer. During my doctorate, I had the opportunity to participate in many research articles presented in the Annex section. For my formation as a doctor, the participation of my Laboratory structure, partnerships, and colleagues was fundamental. During this period, I had a taste of what is the possibility of making science and being able to contribute to a better future for the planet.

Bibliography

ABADI, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>. Citado na página 27.

AGREEMENT, P. United nations. *United Nations Treaty Collect*, p. 1–27, 2015. Citado na página 26.

ALBUS, J. S. 4d/rcs: a reference model architecture for intelligent unmanned ground vehicles. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Unmanned Ground Vehicle Technology IV*. [S.l.], 2002. v. 4715, p. 303–310. Citado na página 15.

ALONZO, M. et al. Mapping urban forest structure and function using hyperspectral imagery and lidar data. *Urban Forestry Urban Greening*, v. 17, p. 135–147, 2016. ISSN 1618-8667. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1618866715300194>. Citado na página 29.

ARANTES, B. L. et al. Urban forest and per capita income in the mega-city of sao paulo, brazil: A spatial pattern analysis. *Cities*, v. 111, p. 103099, 2021. ISSN 0264-2751. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0264275120314475>. Citado na página 29.

ASKNE, J.; NORDIUS, H. Estimation of tropospheric delay for microwaves from surface weather data. *Radio Science*, AGU, v. 22, n. 03, p. 379–386, 1987. Citado na página 15.

BAN, K.-m. Sustainable development goals. United Nations, 2016. Citado na página 14.

BARÓ, F. et al. Contribution of ecosystem services to air quality and climate change mitigation policies: The case of urban forests in Barcelona, Spain. *Ambio*, 2014. ISSN 00447447. Citado 2 vezes nas páginas 29 and 98.

BARTHOLOME, E.; BELWARD, A. S. Glc2000: a new approach to global land cover mapping from earth observation data. *International Journal of Remote Sensing*, Taylor & Francis, v. 26, n. 9, p. 1959–1977, 2005. Citado na página 15.

BLUM, J. *Urban forests: Ecosystem services and management*. [S.l.]: CRC Press, 2017. Citado na página 21.

BOLAND, D. H. *Trophic classification of lakes using Landsat-1 (ERTS-1) multispectral scanner data*. [S.l.]: US Environmental Protection Agency, Office of Research and Development . . . , 1976. Citado na página 15.

CAESAR, H. et al. nuscenes: A multimodal dataset for autonomous driving. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2020. p. 11621–11631. Citado na página 15.

CALDERÓN-LOOR, M.; HADJIKAKOU, M.; BRYAN, B. A. High-resolution wall-to-wall land-cover mapping and land change assessment for australia from 1985 to 2015. *Remote Sensing of Environment*, Elsevier, v. 252, p. 112148, 2021. Citado na página 15.

CHANG, M. *Forest hydrology: an introduction to water and forests*. [S.l.]: CRC press, 2006. Citado na página 13.

CHEN, W. Y.; WANG, D. T. Urban forest development in china: Natural endowment or socioeconomic product. *Cities*, v. 35, p. 62–68, 2013. ISSN 0264-2751. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0264275113000954>. Citado na página 29.

CHOI, D.; GAO, Z.; JIANG, W. Attention to global warming. *The Review of Financial Studies*, Oxford University Press, v. 33, n. 3, p. 1112–1145, 2020. Citado na página 20.

CIHLAR, J. Land cover mapping of large areas from satellites: status and research priorities. *International journal of remote sensing*, Taylor & Francis, v. 21, n. 6-7, p. 1093–1114, 2000. Citado na página 15.

CLIFFORD, P. Markov random fields in statistics. *Disorder in physical systems: A volume in honour of John M. Hammersley*, p. 19–32, 1990. Citado na página 27.

CLINE, W. R. et al. Economics of global warming, the. *Peterson Institute Press: All Books*, Peterson Institute for International Economics, 1992. Citado na página 20.

COLOMINA, I.; MOLINA, P. *Unmanned aerial systems for photogrammetry and remote sensing: A review*. 2014. Citado na página 97.

CRACKNELL, A. P. *Advanced very high resolution radiometer AVHRR*. [S.l.]: Crc Press, 1997. Citado na página 15.

CRACKNELL, A. P.; VAROTSOS, C. A. *Editorial and cover: Fifty years after the first artificial satellite: from sputnik 1 to envisat*. [S.l.]: Taylor & Francis, 2007. Citado na página 15.

DAVIDSON, E. A. et al. The amazon basin in transition. *Nature*, Nature Publishing Group, v. 481, n. 7381, p. 321–328, 2012. Citado na página 20.

DELWORTH, T. L.; KNUTSON, T. R. Simulation of early 20th century global warming. *Science*, American Association for the Advancement of Science, v. 287, n. 5461, p. 2246–2250, 2000. Citado na página 20.

D'ODORICO, P. et al. The global food-energy-water nexus. *Reviews of geophysics*, Wiley Online Library, v. 56, n. 3, p. 456–531, 2018. Citado na página 14.

DONCHYTS, G. et al. Planetary-scale surface water detection from space. In: *AGU Fall Meeting Abstracts*. [S.l.: s.n.], 2017. v. 2017, p. IN44A–01. Citado 3 vezes nas páginas 10, 22, and 24.

ECONOMIC, U. D. of; AFFAIRS, P. D. S. *World Population Prospects 2019: Highlights*. [S.l.]: UN-United Nations, 2019. Citado 3 vezes nas páginas 20, 21, and 29.

ESCH, T. et al. Breaking new ground in mapping human settlements from space–the global urban footprint. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 134, p. 30–42, 2017. Citado 2 vezes nas páginas 13 and 20.

FENG, Q.; LIU, J.; GONG, J. UAV Remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sensing*, 2015. ISSN 20724292. Citado na página 29.

FERNANDES, T.; OLIVEIRA, E. Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption. *Journal of Business Research*, Elsevier, v. 122, p. 180–191, 2021. Citado na página 15.

FERREIRA, H. S.; CÂMARA, G. Current status and recent developments in brazilian remote sensing law. *J. Space L.*, HeinOnline, v. 34, p. 11, 2008. Citado na página 15.

FRIEDL, M. A. et al. Global land cover mapping from modis: algorithms and early results. *Remote sensing of Environment*, Elsevier, v. 83, n. 1-2, p. 287–302, 2002. Citado na página 15.

GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. [S.l.]: O'Reilly Media, 2019. Citado 2 vezes nas páginas 19 and 27.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado na página 27.

GORELICK, N. et al. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, Elsevier, v. 202, p. 18–27, 2017. Citado 2 vezes nas páginas 13 and 25.

GRIMMOND, S. U. Urbanization and global environmental change: local effects of urban warming. *Geographical Journal*, Wiley Online Library, v. 173, n. 1, p. 83–88, 2007. Citado na página 20.

HANSEN, M. C. et al. High-resolution global maps of 21st-century forest cover change. *science*, American Association for the Advancement of Science, v. 342, n. 6160, p. 850–853, 2013. Citado 3 vezes nas páginas 16, 22, and 26.

HENRICH, V. et al. *IDB-https://www. indexdatabase. de/, Entwicklung einer Datenbank für Fernerkundungsindizes. AK Fernerkundung*. [S.l.]: Bochum, 2012. Citado na página 27.

HUANG, X. et al. The apolloscape dataset for autonomous driving. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2018. p. 954–960. Citado na página 15.

JENNINGS V., L. L.; YUN, J. Advancing sustainability through urban green space: Cultural ecosystem services, equity, and social determinants of health. *International journal of environmental research and public health*, v. 13, n. 2, 2016. ISSN 1660-4601. Disponível em: <https://www.mdpi.com/2072-4292/13/4/767>. Citado na página 29.

JIM, C.; CHEN, W. Y. Ecosystem services and valuation of urban forests in china. *Cities*, v. 26, n. 4, p. 187–194, 2009. ISSN 0264-2751. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0264275109000456>. Citado na página 29.

JR, J. R. et al. *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*. [S.l.], 1974. Citado na página 26.

KARDAN, O. et al. Neighborhood greenspace and health in a large urban center. *Scientific Reports*, 2015. ISSN 20452322. Citado na página 29.

KATTENBORN, T.; EICHEL, J.; FASSNACHT, F. E. Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution uav imagery. *Scientific reports*, Nature Publishing Group, v. 9, n. 1, p. 1–9, 2019. Citado na página 27.

KERR, R. A. Global warming is changing the world. *Science*, American Association for the Advancement of Science, v. 316, n. 5822, p. 188–190, 2007. Citado na página 20.

KILLOUGH, B. Overview of the open data cube initiative. In: IEEE. *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium*. [S.l.], 2018. p. 8629–8632. Citado na página 25.

KIM, D.-H. et al. Global, landsat-based forest-cover change from 1990 to 2000. *Remote sensing of environment*, Elsevier, v. 155, p. 178–193, 2014. Citado na página 26.

La Rosa, D.; WIESMANN, D. Land cover and impervious surface extraction using parametric and non-parametric algorithms from the open-source software R: An application to sustainable urban planning in Sicily. *GIScience and Remote Sensing*, 2013. ISSN 15481603. Citado na página 29.

LALIBERTE, A. S. et al. Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (uav) imagery for rangeland monitoring. *Photogrammetric Engineering & Remote Sensing*, American Society for Photogrammetry and Remote Sensing, v. 76, n. 6, p. 661–672, 2010. Citado na página 22.

LECUN, Y.; BENGIO, Y. et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, v. 3361, n. 10, p. 1995, 1995. Citado 2 vezes nas páginas 20 and 27.

LILLESAND, T.; KIEFER, R. W.; CHIPMAN, J. *Remote sensing and image interpretation*. [S.l.]: John Wiley & Sons, 2015. Citado 3 vezes nas páginas 13, 19, and 20.

LIM, S.-H.; BAE, T.-S. Estimation of gnss zenith tropospheric wet delay using deep learning. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, v. 39, n. 1, p. 23–28, 2021. Citado na página 15.

MARTINS, J. A. C. et al. Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. *Remote Sensing*, v. 13, n. 16, 2021. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/13/16/3054>. Citado 3 vezes nas páginas 10, 15, and 19.

MCHUGH, N. et al. Modelling short-rotation coppice and tree planting for urban carbon management - a citywide analysis. *Journal of Applied Ecology*, 2015. ISSN 13652664. Citado na página 29.

MCLEAN, G.; OSEI-FRIMPONG, K. Hey alexa ... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, v. 99, p. 28–37, 2019. ISSN 0747-5632. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0747563219301840>. Citado na página 15.

MELLO, M. P. et al. A simplified bayesian network to map soybean plantations. In: IEEE. *2010 IEEE International Geoscience and Remote Sensing Symposium*. [S.l.], 2010. p. 351–354. Citado na página 27.

Memarian Sorkhabi, O.; ASGARI, J.; AMIRI-SIMKOOEI, A. Monitoring of caspian sea-level changes using deep learning-based 3d reconstruction of grace signal. *Measurement*, v. 174, p. 109004, 2021. ISSN 0263-2241. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0263224121000403>. Citado na página 15.

NORDHAUS, W. D.; BOYER, J. *Warming the world: economic models of global warming*. [S.l.]: MIT press, 2003. Citado na página 20.

PACKER, J.; REEVES, J. Romancing the drone: Military desire and anthropophobia from sage to swarm. *Canadian Journal of Communication*, v. 38, n. 3, 2013. Citado na página 28.

PASZKE, A. et al. Automatic differentiation in pytorch. 2017. Citado na página 27.

PLANET. *Planet NICFI Basemaps for Tropical Forest Monitoring - Tropical Americas, https://api.planet.com*. 2017. Citado 2 vezes nas páginas 25 and 99.

PÖRTNER, H.-O. et al. *Climate change 2022: Impacts, adaptation and vulnerability*. [S.l.]: IPCC Geneva, Switzerland:, 2022. Citado na página 14.

SALLAB, A. E. et al. Deep reinforcement learning framework for autonomous driving. *Electronic Imaging*, Society for Imaging Science and Technology, v. 2017, n. 19, p. 70–76, 2017. Citado na página 15.

SENTINEL. 2022. Accessed: 2022-05-10. Disponível em: <https://www.sentinel-hub.com/>. Citado na página 25.

SEPAL. *System for Earth Observation Data Access Processing and Analysis for Land Monitoring SEPAL UNSPIDER Knowledge Portal*. 2022. Accessed: 2022-05-10. Disponível em: <https://www.un-spider.org/links-and-resources/gis-rs-software/system-earth-observation-data-access-processing-and-analysis>. Citado na página 25.

SEXTON, J. O. et al. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of modis vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, Taylor & Francis, v. 6, n. 5, p. 427–448, 2013. Citado na página 26.

SHEN, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, Wiley Online Library, v. 54, n. 11, p. 8558–8593, 2018. Citado na página 97.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. p. 1–14, 2014. ISSN 15352900. Disponível em: <http://arxiv.org/abs/1409.1556>. Citado na página 27.

SONG, X.-P. et al. Global land change from 1982 to 2016. *Nature*, Nature Publishing Group, v. 560, n. 7720, p. 639–643, 2018. Citado na página 26.

VANCUTSEM, C. et al. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Science Advances*, American Association for the Advancement of Science, v. 7, n. 10, p. eabe1603, 2021. Citado na página 26.

VASWANI, A. et al. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017. Citado 3 vezes nas páginas 20, 27, and 28.

VITOUSEK, P. M. Beyond global warming: ecology and global change. *Ecology*, Wiley Online Library, v. 75, n. 7, p. 1861–1876, 1994. Citado na página 20.

WALLACH, H. M. Conditional random fields: An introduction. *Technical Reports (CIS)*, p. 22, 2004. Citado na página 27.

WANG, F.-M. et al. New vegetation index and its application in estimating leaf area index of rice. *Rice Science*, Elsevier, v. 14, n. 3, p. 195–203, 2007. Citado na página 27.

WENG, Q. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, Elsevier, v. 117, p. 34–49, 2012. Citado 2 vezes nas páginas 15 and 21.

WHITE, J. C. et al. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Canadian Journal of Remote Sensing*, v. 42, n. 5, p. 619–641, 2016. ISSN 17127971. Citado na página 97.

WULDER, M. A. et al. Land cover 2.0. *International Journal of Remote Sensing*, Taylor & Francis, v. 39, n. 12, p. 4254–4284, 2018. Citado na página 15.

ZHU, Z. et al. Benefits of the free and open landsat data policy. *Remote Sensing of Environment*, Elsevier, v. 224, p. 382–385, 2019. Citado na página 15.

Annex

ANNEX A

Technical and Scientific Production

- ARTICLES PUBLISHED IN JOURNALS:
  - Gonçalves, D. N., Junior, J. M., Carrilho, A. C., Acosta, P. R., Ramos, A. P. M., Gomes, F. D. G., **Martins, J.**, Libonati, R. (2023). Transformers for mapping burned areas in Brazilian Pantanal and Amazon with PlanetScope imagery. International Journal of Applied Earth Observation and Geoinformation, 116, 103151.
  - Araújo Carvalho, M., Junior, J. M., **Martins, J.**, Zamboni, P., Costa, C. S., Siqueira, H. L., Gonçalves, W. N. (2022). A deep learning-based mobile application for tree species mapping in RGB images. International Journal of Applied Earth Observation and Geoinformation, 114, 103045.
  - **Martins, J.**, Marcato Junior, J., Pätzig, M., Sant'Ana, D. A., Pistori, H., Liesenberg, V., Eltner, A. (2022). Identifying plant species in kettle holes using UAV images and deep learning techniques. Remote Sensing in Ecology and Conservation.
  - Sant'Ana, D. A., Pache, M. C. B., **Martins, J.**, Astolfi, G., Soares, W. P., de Melo, S. L. N., Pistori, H. (2022). Computer vision system for superpixel classification and segmentation of sheep. Ecological Informatics, 101551.
  - Bressan, P. O., Junior, J. M., **Martins, J.**, de Melo, M. J., Gonçalves, D. N., Freitas, D. M., ... Gonçalves, W. N. (2022). Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. International Journal of Applied Earth Observation and Geoinformation, 108, 102690.
  - **Martins, J.**, Menezes, G., Gonçalves, W., Sant'Ana, D. A., Osco, L. P., Liesenberg, V, Junior, J. M. (2021). Machine learning and SLIC for Tree Canopies segmentation in urban areas. Ecological Informatics, 66, 101465.
  - Sant'Ana, D. A., Pache, M. C. B., **Martins, J.**, Soares, W. P., de Melo, S. L. N., Garcia, V., Pistori, H. (2021). Weighing live sheep using computer vision techniques and regression machine learning. Machine Learning with Applications, 100076.
  - **Martins, J.**, Nogueira, K., Osco, L. P., Gomes, F. D. G., Furuya, D. E. G., Gonçalves, W. N., Junior, J. M. (2021). Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. Remote Sensing, 13(16), 3054.
  - Lobo Torres, D., Queiroz Feitosa, R., Nigri Happ, P., Elena Cue La Rosa, L., Marcato Junior, **Martins, J.**, Liesenberg, V. (2020). Applying fully convolutional architectures for semantic segmentation of a single tree species in

urban environment on high resolution UAV optical imagery. Sensors, 20(2), 563.

- ARTICLES PUBLISHED IN CONFERENCES:
  - **Martins, J.**, Nogueira, K., Zamboni, P., de Oliveira, P. T. S., Gonçalves, W. N., dos Santos, J. A., Marcato, J. (2021, July). Segmentation of Tree Canopies in Urban Environments Using Dilated Convolutional Neural Network. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 6932-6935). IEEE.
  - P. A. P. Zamboni, J. Marcato, G. T. Miyoshi, J. de Andrade Silva, **J. Martins** and W. N. Gonçalves, "Assessment of CNN-Based Methods for Single Tree Detection on High-Resolution RGB Images in Urban Areas," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 590-593, doi: 10.1109/IGARSS47720.2021.9553092.
  - **J. Martins**, D. A. Sant'Ana, J. M. Junior, H. Pistori and W. N. Gonçalves, "Aerial Image Segmentation In Urban Environment For Vegetation Monitoring," 2020 IEEE Latin American GRSS ISPRS Remote Sensing Conference (LAGIRS), 2020, pp. 375-379, doi: 10.1109/LAGIRS48042.2020.9165618.
  - **Martins, J.**, Junior, J. M., Menezes, G., Pistori, H., Sant, D., Gonçalves, W. (2019, July). Image segmentation and classification with SLIC superpixel and convolutional neural network in forest context. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 6543-6546). IEEE.
  - **Martins, J. A. C.**, Pessi, D. D., Paranhos, A. C. Python e Google Earth Engine no Monitoramento de Mudanças de Cobertura do Solo no Pantanal.