uc3m | Universidad Carlos III de Madrid

Bachelor Degree in Computer Science & Engineering

2025-2026

Bachelor Thesis

Advancing Plant Phenotyping in Rangelands Through Drone-Derived Imagery and Video Data

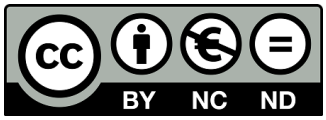
---

Javier Martín Pizarro

Joao Ricardo Pereira Valente

Leganés, Madrid

March, 2026



ABSTRACT

Site-specific weed control (SSWC) has been a problem since humans started to develop the field of agriculture. During the ages, existing methodologies did not change at all: they were based on manual detection and extraction. However, this changed with the appearance of modern computers.

In the last fifty years, there has been an extreme evolution in the detection of non-desired weeds. However, it was not until recently that newer computational methodologies based on Deep Learning and Artificial Neural Networks — more specifically Convolutional Neural Networks — were introduced into the field that the advances became trully efective.

This work addresses the development of a computer vision model capable of locating and counting instances of *Eryngium horridum*—commonly known as "cardilla"—a perennial, spiny weed species native to the grasslands of Uruguay, Argentina, and southern Brazil. By leveraging drone-derived imagery and video data, this thesis aims to contribute to precision agriculture and ecological monitoring, enhancing the sustainability and efficiency of rangeland management practices.

Keywords: Site-specific Weed Control, Object Counting, Object Segmentation, Computer Vision, Artificial Intelligence

DEDICATION

I would like to dedicate this work to the two fundamental pillars of my life:

To my family. You have been there for me no matter what. Even when I did not deserve it. You will always be the beacon I need when I must return to safe ports.

To my brothers. We are not blood-related, but you will always have a seat at my table. With you, I found loyalty even in the worst of the storms.

" We will all die and the universe will carry on without care. All that we have is that shout into the wind—how we live. How we go. And how we stand before we fall. "

— Pierce Brown, *Golden Son*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND MOTIVATION

The purpose of this chapter is to introduce the topic of the thesis, present the current state of the art — from both theoretical and practical perspectives — and outline the motivations behind it, as well as the regulatory framework and the associated economic impact.

1.1. Motivation

Agriculture plays a crucial role in the economy of the vast majorities of the countries in the world. Although in developed countries such as Spain it plays a less important role (near 2.3% in 2024[1]), in some developing countries such as Hispanic America it can raise up to 8%.

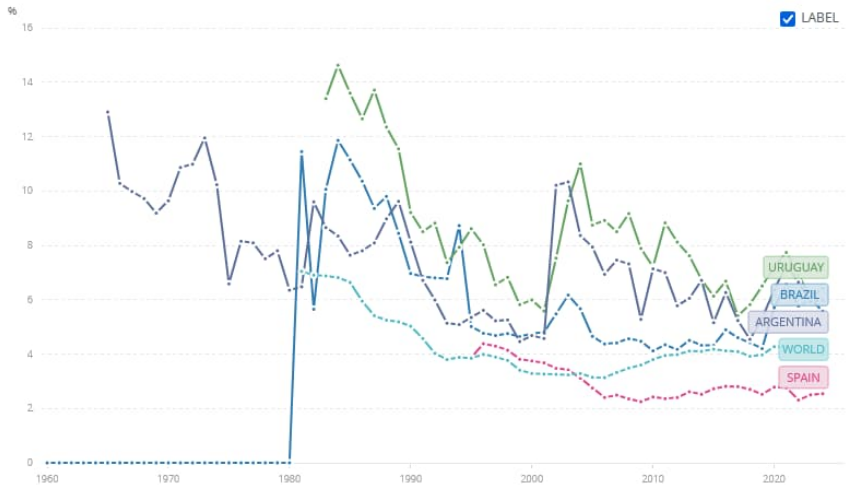


Fig. 1.1. Gross Domestic Product (GDP) share of agriculture, forestry, and fishing (in %) between the years 1960 and 2024. Source: *World Bank Data* [1].

Maintaining an adequate rhythm of production is vital not only for the economy, but for sustaining the quality of life of the population. Thus, innovating with new technologies in this field has always been necessary to supply the increasing demand.

Weeds have been consistently a problem; not only they reduced the quality and quantity of the crops, but detecting them was an extenuating job. Until the development of modern machinery, it was mainly done by hand — covering entire fields and removing them — or using agriculture techniques for avoiding their apparition — mainly grazing and crops rotation —, with mediocre results.

In the early years of the XX century, tractors were starting to be more and more common, reducing manual labour activities a lot. However, there was still the possibility

of developing weeds in the fields and not being able of estimating the total amount of them in the total land size.

Estimating the amount per hectare (or other desired unit of measure) is vital for understanding how weeds are influencing in the growth and quality of crops. Depending on the density of these unwanted plants, different quantities of herbicides can be used, reducing toxins and improving the condition of the batches.

Modern computation technologies were firstly used near the 70s: archaic solutions based on reflecting-based living ("green") plants with photoelectric diodes[2] were proposed. However, these methods were highly dependant on the ability of controlling the constantly in-change environment.

In the 80s, with the appearance of digital cameras, a new bunch of possibilities appeared. As the spectral-colour range cameras were more and more affordable, a totally new world for exploring this field was discovered.

The first predecessor of formal Convolutional Neural Networks — known as the neocognitron —, presented by Kunihiko Fukushima was a totally game-changer. It was a multilayer perceptron (MLP) able to extract features and predict handwritten numerals from "0" to "9"[3]. Fukushima also proposed several unsupervised training algorithms. Although they were revolutionary, after the proposal of back-propagation [4] (which is heavily used in computer vision currently) they fell into disuse.

After it, a spiral of hype and constant changes for these neural networks came into scene. In 1998, the LeNet-5[5] was the first neural network to include back-propagation end-to-end which was tested with the MNIST dataset.

In 2012, the neural network AlexNet was proposed. With nearly 22,000 categories (labels), this neural network was able to generalise a vast number of different objects with high precision. However, the most innovative thing was that it was **trained using Graphics Processing Units — GPUs —**[6], something that was never used in the field. This rapidly raised the level of training methodologies, reducing the time elapsed.

Only three years later, Microsoft engineers proposed a new method to lighten the weight of neural networks. After benchmarking and stating that **"the deeper network has higher training error, and thus test error"**[7]. This means that the more layers a network has (above a critical limit), the less precise it gets. They propose the Deep Residual Learning, based on the difference (error) between the expected value and the obtained one.

$$F(x) = H(x) - x$$

$$y = F(x) + x$$

Using this approach, the net eases the learning process compared to the standards of the moment. As a subsequent effect, they are able to reduce the weight of the networks up

to an 80% (compared with VGG nets).

It was not until 2012 that convolutional neural networks were mature enough in order to be applied in the agriculture field. However, there are important limitations when using these methodologies — which are by far the most effective—.

The main limitation is not computational or algorithmic, but the datasets used for training the nets. It is common to have a very limited dataset with very few instances of useful data to preprocess and work with. Restricted to the regions where they were obtained, it is complicated to make a dataset that is representative for an extensive region.

Nonetheless, the region is not the only factor, but also the seasons of the year. Generating a fine dataset that shows every phenotype of a given plant in a specific moment is expensive — economically and humanly —.

Thus, the aim of this work is not only to create a model able to segment and count instances of the cardilla, but to create a dataset competent enough to provide the sufficient information for future works about the field.

1.1.1. About the *Eryngium horridum*

Original from Hispanic America, the *Eryngium horridum* —also known as cardilla or caraguatá— is common to locate in the plain lands of Uruguay, southern Brazil and central-eastern Argentina.

The plant is a perennial forb with a highly distinctive morphology; it is a rosette with numerous spiny linear leaves that can reach up to 65 cm in length and 2 cm in width. Its inflorescence axis can grow as tall as 2 meters [8].

.

Fig. 1.2. *Eryngium Horridum*: Left image: bottom part of the plant. Upper right: flower. Bottom left: inflorescence

One peculiarity of this plant is its resilience in adverse conditions. After experiencing fires or frosts, it has been mentioned to see the floral part of the stem to grow quicker and bigger than before.

Although there are no common uses for this plant, it is known that it helps in the scarring process. However, here ends its applications. It is not harvested, at the contrary, it is heavily prosecuted because of the rapid growing through fields, destroying useful terrain for cropping in a few months. Not even the livestock desires to eat it, unless excessive hunger.

A quickly identification of the plant is needed to solve the issue before it ruins the field and the crops already planted.

1.1.2. Technical challenge

Although it is clearly obvious the distinctive shape of the plant, trying to observe them from an aerial perspective gets much more complicated. Depending on the height of the UAV (Unmanned Aerial Vehicle) and the quality and resolution of the camera, the difficulty can increase exponentially.

Fig. 1.3. Recorded aerial view from an unknown height. Sizes differ between each individual deppending on different factors, such as the dryness of the plant.

Estimating the possibles individuals per hectare by hand, even with the help of UAVs, is complex enough. In the last ten years, the trend of computer vision has also arrived to this field.

Computer vision, which a field of artificial intelligence — more specifically from machine learning —, is a discipline that allows computers to process images (frames) and to extract meaningfully data and make decisions.

This challenge is based on the Multiple Object Tracking (MOT); which aims for identifying an arbitrary number of n individuals with the maximum precision possible.

In the literature, there are currently two approaches for solving the MOT problem: **Detection-based Tracking (DBT)** and **Detection-Free Tracking (DFT)**. Whereas the DBT is used an external and automatic detector for localising the objects in the frame, the DFT needs some manual input at first glance in order to keep up with the trajectory of the item.

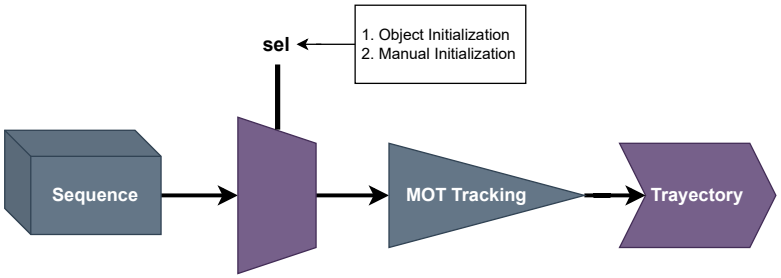


Fig. 1.4. Visual representation of the MOT problem. The multiplexer determines the approach: either Detection-based Tracking or Detection-Free Tracking.

These approaches, although similar, are not used in the same use cases. The applications differ, as well as the advantages and disadvantages depending on the scenaries where they are used.

Variables	**DBT**	**DFT**
Initialization	automatic, possibly imperfect	manual, perfect
# of objects	varying	fixed
Applications	specific type of objects	any types of objects
Advantages	ability to handle varying number of objects	free of object detector
Disadvantages	performance depends on object detection	manual initialization

TABLE 1.1. Comparison between DBT and DFT. Adapted from [9]

Whereas the DBT is made up from a pipeline: the **detector** (in charge of predicting the boxes of the items) and the **asociator** (in charge of asociating each box with a label or class) and then uses associations algorithms (see 1.2.1) for identifying the movement of an object, the DFT is much more complex. The net itself learns how to predict the trayectory of the individual, using transformers and tokens for maintining the individual identity of the model.

At first, DFT may seem like a reasonable option to choose. However, its complexity its not trivial. It is important to mention that usually is required to have big datasets for these types of models. In this work, the quantity and quality of the frames are not as large and precise as it would be prefereable. Thus, Detection-Based Tracking is a better option.

- Freedom for controlling the detector (YOLO, Faster R-CNN...)

- Atomic: is easier to modify local features.
- Different methods of association can be used for experimenting with them.

In summary, while both DBT and DFT approaches seems to be good enough results, the **Detection Based Tracking** method can give us more interpretability and control while developing the framework, specially with the data limitations of this study.

Therefore, the main focus of this study will be on the DBT pipeline, combining different classification models with associations algorithms. However, to evaluate the potential of more recent and capable technologies, the **T-Rex 2** model [10] — a representative of the DFT family — will be also tested as a comparative benchmark.

From this technical review, several questions arise:

1. How effectively can a detection-based model identify, localize and track the *Eryngium horridum*?
2. How do supervised models performed (DBT) compared to end-to-end transformers-based arquitectures such as T-Rex 2 (DFT)?
3. What are the limitations and generalization capabilities of these models when faced across different plant densities, flight altitudes, and observation perspectives?

The source code developed for this project is publicly accessible on GitHub, baptised as CARLA — Cardilla Recognition and Localization Analyzer [11].

1.2. State of Art: An introduction to Object Detection and Segmentation

1.2.1. Associations Algorithms

1.3. Regulatory Framework

1.3.1. European Regulatory Framework

The regulatory framework about these emerging technologies — such as AI and UAVs — are still being defined worlwide. In the European Union, several laws about the usage of AI and UAVs have been already defined for the correct usage of both technologies.

European Regulatory Framework of UAVs

Unmanned Aerial Vehicles regulatory framework started to develop in the early 2020s, with the approval of the Implementing Regulation (UE) 2019/947, which defines the rules and procedures for the usage of drones in european aerial space and the Implementing

Regulation (UE) 2018/1139, which integrated the drone normative with the European Union Aviation safety Agency (EASA).

One of the most important points of this regulation is the classification of the drones: *open*, *specific* or *certified*, which depends on several factors such as the danger of the mission and the capabilities of the vehicle. Depending on the category of the vehicle, different requirements may be needed (for the pilot and the plane) to fly in european space.

1. **Open**: *low danger operations*. Most are recreational or simple comercial flights. No previous authorization is required is the drone follows several requirements: weight does not surpass 25 kg, maximum flight altitude is 125 meters high and the vehicle will not approach any vital infraestructure (such as airports). The pilot must have constantly visual sight of the vehicle.

2. **Specific**: *medium danger operations*. In this category, flights required to have a previous authorization of the corresponding authorities (AESA in Spain). However, the limitations of these flights are less restricted than in the open category: night operations, increase size and weight of the vehicle and the posibility of flying in urban zones.

3. **Certified**: *high danger operations*. Compared to traditional aviation operations. This category includes vehicles that transports passengers or dangerous commodities or that used autonomous flights without any possible human interaction. Currently, there are very few operations that are currently in the need of this category; it is aimed for future projects.

In addition, the law in Spain (and thus in the entire European Union) requires that every drone pilot must be certified with a license, obtaining an unique identity number valid in the entire European Union.

This work aims to contribute to the open and specific categories. Here, individuals such as farmers or researchers may be in the need of using drones for recognising and identifying weeds in their fields.

However, these services can be also done through a company specialized in this. The pilots the company must be certified pilots as well. And because of the high amount of flights per week that these pilots may need to do, exists the **Light UAS Operator Certificate** (LUC) for frequent operators. These operators do not need to ask for previous authorization to AESA, reducing the bureocracy and speeding up the process of asking for authorization, which can last up to ten working days.

European Regulatory Framework of Artificial Intelligence

In 2024 the European Union adopted the first worldwide normative with respect to artificial intelligence: the reglament (UE) 2024/1689, also known as the Law of Artificial Intel-

ligence or AI Act. It is defined as a machine-based system designed to operate between different levels of autonomy.

The AI Act introduces a well-defined categorization of risk for AI systems. Depending on the risk category, some systems are forbidden, whereas others may require and extensive assessment before market entry.

- **Unacceptable risk**: applications such as social scoring systems and manipulative technologies are absolutely banned in the EU.
- **High risk**: applications based on risk evaluation (such as risk or infraestructure) must pass strict compliance reviews.
- **Limited risk**: limited applications such as recommender systems, image or video processing and more need to clarify to the user its usage and the usage of AI for generating the output. Data quality, transparency and fairness standards are required even for these types of applications
- **Minimal risk**: spam filters or or AI in video games fall in this category. They are not subjected to any regulatory framework.

The work proposed in this document clearly falls into the *limited risk* category. Thus, it is important to remark that transparency and fairness are vital. This is because the output of an AI model could cause confussion or influence decision-making capabilities of the user.

Several mininmal obligations are required to these systems for the sake of the protection of the user:

- User must always know that is using AI.
- Artificially generated contents must be labeled as "AI Generated" in order to avoid confusion or wiles.
- A transparent design means a less biased model.
- It is recommended to include ethics by design. This can be done by evaluating biases during training or applying *algorithmical justice*

The AI Act also mentions that the systems must be explainable (XAI). The goal of XAI is quite ambitious: surpassing the black box issue and integrating algorithms with human values, empowering and enhancing the individual in the decision-making process [12].

European Regulatory Framework of Personal Data and Privacy

In Europe, every use of personal data needs to follow the privacy policy established by the European Union; more specifically the GDPR — General Data Protection Regulation — which its effective date was 2018.

The GDPR imposes some fundamental principles about the usage of data, which can be used in AI proyects. These principles are based on transparency, loyalty and legality. Moreover, a limitation must be setted for the usage of the data in order to prevent a misuse of it.

Furthermore, the data must be stored following the principles of integrity and confidentiality up to an arbitrary number of years defined during the process of collection.

One relevant change that of the GDPR is that the population has rights about its own data: they have the right to access, change, remove or oppose to their usage in some applications. This means that every institution that automates processes using AI must be transparent about their usage and how it arrive to the conclusion given, which again bring us to XAI (eXplainable AI).

It is important to mention that the GDPR and AI Act were created to coexist: if an institution want to apply a service using AI, it must subjected to both the AI Act and the GDPR.

1.3.2. Regulatory Framework in Hispanic America

In contrast to the European Union's harmonized legal ecosystem, Hispanoamerica presents a more fragmented regulatory landscape regarding artificial intelligence, data privacy, and UAV (drone) operations. However, several countries have made considerable advances toward regulating these technologies, particularly in the areas of **data protection** and **civil aviation**.

Data Protection and Privacy Laws

Many Hispanoamerican countries have enacted specific laws to regulate the collection, processing, and use of personal data, largely inspired by international frameworks such as the EU's GDPR.

- **Brazil** enacted the *Lei Geral de Proteção de Dados* (LGPD - Law No. 13.709/2018), a comprehensive data protection law closely modeled after the GDPR, which applies to both public and private sector data handling. It includes principles such as consent, purpose limitation, data minimization, and accountability, and is enforced by the national authority ANPD.
- **Argentina** was a pioneer with its Data Protection Law 25.326, which regulates

personal data processing and guarantees data subject rights. The law also prohibits the dissemination of image or video data containing identifiable individuals without prior consent.

- **Chile**, **Mexico**, and **Uruguay** have similar legislative frameworks, each including provisions that require informed consent, secure data processing, and individuals' right to access or correct their data. In Uruguay, the Law 18.331 governs personal data protection and is overseen by the Unidad Reguladora y de Control de Datos Personales (URCDP).

These national frameworks establish that any AI or drone-based system capturing or processing images, biometrics, or location data must comply with consent and purpose principles. This significantly affects agricultural or surveillance applications involving UAVs or computer vision.

AI-Specific Legislative Developments

Unlike the EU's AI Act, Hispanoamerica does not yet have unified legislation targeting artificial intelligence. Nevertheless, several countries have launched national AI strategies and draft bills focused on ethical guidelines, transparency, and fostering innovation.

- **Brazil** released its *Estratégia Brasileira de IA* (2021), outlining policy objectives including responsible AI, scientific research, and AI in public services. A federal AI Bill was proposed in 2022 and is under parliamentary discussion.
- **Chile** introduced its National AI Policy in 2021, emphasizing development based on human rights, inclusion, and explainability.
- **Argentina**, **Colombia**, and **Uruguay** are also developing their national frameworks, supported by regional cooperation through organizations like CEPAL and the IADB.

Though still in progress, these initiatives aim to provide a regulatory basis for ethical AI deployment in agriculture, education, and digital public infrastructure.

UAV and Drone Legislation

Drone regulations in Hispanoamerica vary considerably by country. However, civil aviation authorities generally restrict operational parameters such as maximum altitude, speed, and flight over populated areas.

- In **Brazil**, the National Civil Aviation Agency (ANAC) regulates UAVs through RBAC-E94, which establishes over 70 provisions for certification, weight categories, and operation types. Drones used for commercial, agricultural, or research purposes require operator registration, flight authorization, and in some cases, insurance.

- **Argentina** maintains similar strict controls, requiring UAV registration, operator licenses, and compliance with maximum altitude (120 m) and speed (161 km/h) limits [13].
- In contrast, **Uruguay** has a more permissive regime for recreational drones: no formal permit is required for sport or leisure flights, provided that basic safety measures are followed and no sensitive areas are overflown.
- Countries such as **Colombia**, **Peru**, and **Mexico** follow hybrid approaches based on international aviation safety standards, including pilot certification and risk assessment for advanced operations (e.g., BVLOS, agricultural spraying).

Overall, while regulation is less standardized than in Europe, most Hispanoamerican countries are actively updating their frameworks to address UAV usage, particularly as drone-based AI systems become more prevalent in agriculture, infrastructure monitoring, and environmental surveillance.

1.4. Socio-economic impact

something

1.5. Planification and budget

2. DATASET

The dataset elaboration is usually the most expensive part of the project (in terms of human labour), specially when working with such a complicated problem. In this chapter, the methodologies for the data collection and the analysis of it are explained in detail, commenting the difficulties found when labelling the images and critically analizing the final result of it.

2.1. Data Collection Procedures

The collection of the entire dataset was done in the Treinta y Tres department of Uruguay (-33.253618º S, -54.503245º W in WGS84 (World Geodetic System 1984), 52 m a.s.l.). The team in charge was located at the "Palo a Pique" Research Station from the National Institute of Agricultural Research (in Spanish: Instituto Nacional de Investigación Agropecuaria).

Fig. 2.1. Dataset collection location. The image was obtained through Google Maps.

For the aerial recording of the environment, the usage of a quadcopter was necessary. The aerial platform was a quadcopter, DJI Mavic Air 2, equipped with an integrated RGB sensor, model FC3170, aperture of f/2.8, exposure time of 1/500 s, and focal distance of 4 mm. The flight was made on 21 October 2024, at noon with a clear sky. The flight was scheduled on a cross-path over 0.4 ha, at 10 m altitude (this can vary depending on the frame), with 75 % frontal and 70 % lateral overlap, and a camera angle of 75 ° (this varies in some videos in order to obtain different angles of view from the same point) (adapted from [8]).

In order to increase the posibilities of experiments for this works, videos were recorded in two modalities: **RGB** and **hyperspectral**. RGB stands for Red, Green and Blue, which combined in different measures different colors can be generated.

Fig. 2.2. A frame from the RGB dataset. In this case, weeds can be easily identified due to its size and color constract with respect to the soil.

On the other hand, hyperspectral is based on traditional imaging with a combination of spectroscopy to capture detailed information across a give range of wavelenghts. Spectroscopy is the study of how matter and electromagnetic radiation interact between other through absorption, emission or scattering.

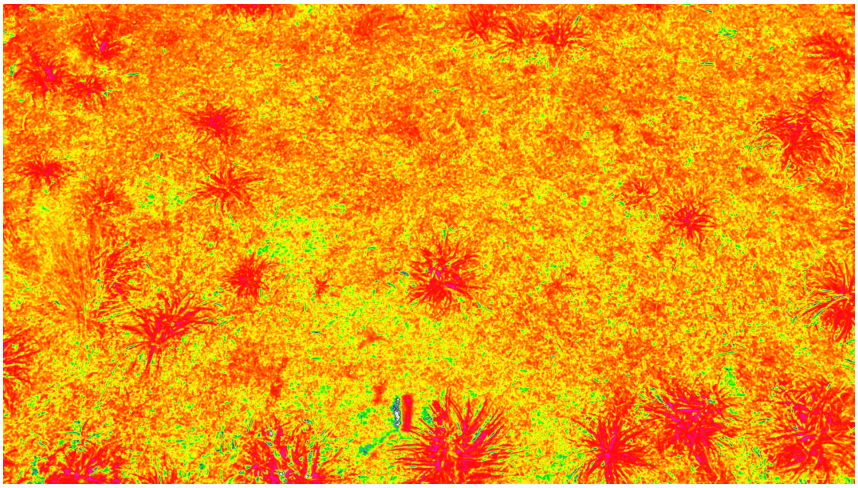


Fig. 2.3. The corresponding hyperspectral image for 2.2. Notice they slightly differ between each other.

As it is possible to see in the previous image 2.3, the specimens of the weed can be clearly distinguished from the soil and other minor vegetation, adding an extra layer of precision for its detection.

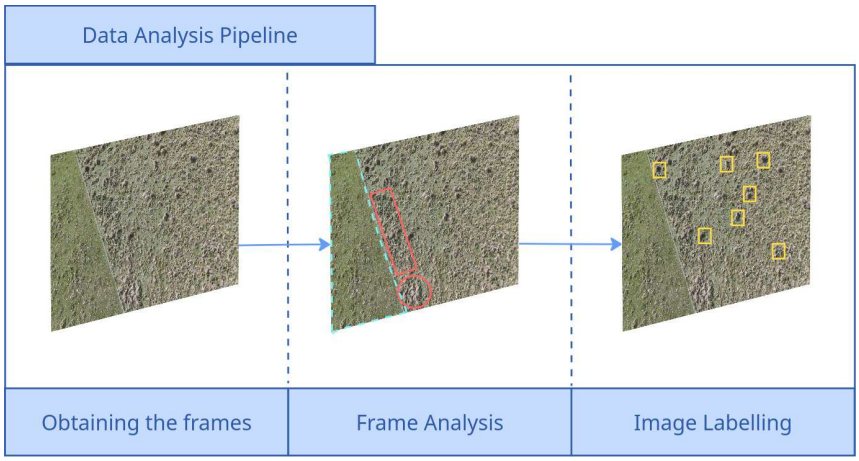
For the same flight, both modes were used. However, due to external factor the videos

are not exactly the same so the corresponding frames neither are. A script was executed to reduce these differences to the minimum.

The entire dataset frame collection has lasted, at least, one year — as it started on the October 22, 2024 and at the current date of writing this document (December 1, 2025) more data is currently being recorded for increasing the size of the dataset.

2.2. Data Analysis

This part embraces the entirety of a defined pipeline in order to incorporate the frames obtained from the videos into the dataset. The following diagram represents the pipeline used:



2.2.1. Transforming the videos into useful frames

Processing the videos was the first step for obtaining useful frames that could be used to fed future models. Thus, a script was coded to process each video, obtaining n frames depending on an offset t (in seconds) given as a parameter. The offset acts as a separator, only returning one frame each t seconds. This was done in order to obtain the highest amount possible of significant images tha may differ (slightly or very) between each other. Doing this, the dataset would have less images, but richer in quality.

All the images extracted have the same dimensions: 3,840 x 2,160 (width, height) with an average weight of 14 MB aproximately. Thus, the resolution and image quality is really high. This is really important when training the models, specially YOLO.

2.2.2. Image Analysis

After obtaining the frames, it is important to remove useless images or similar images, as they add noise to the training (redundant images), making it slower and can generate overfitting is the images are too similar.

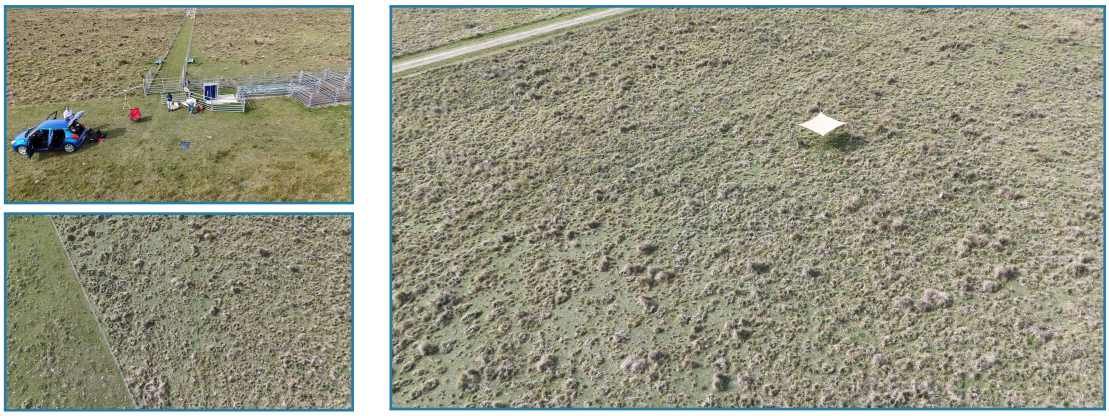


Fig. 2.4. A sample of some frames after processing the entirety of the videos. Notice the changes on the drone orientation with respect to the ground on images b (bottom left) and c (right).

For removing this images, it was taken into account the images previously inspected. If the frame was too near to the timestamp of another one, it was checked in order to see if the drone did moved enough in order to consider the new frame a significant one or the image should be removed. It is important to mention that not a lot of images were needed to be removed except in one video, were the drone was almost flying statically without moving too much. From this last video a total of 70 frames were generated, which only 37 of those were considered relevant.

After that, a deep analysis of the images is done in order to comprenhend the nature of the problem and the structure of the dataset.

Environment analysis

Because of the location where the dataset was recorded, the environment is not rich in diversity. However, the cardilla can be found in Brazil, Uruguay and Argentina mainly, which share similar conditions.

Although these arises problems that will be later discused in section 2.2.4, it is possible to assume a generalization for the region (for the moment).

Deepining more into the environment, it is possible to obtain different characteristics from the images studied:

- Except the cardilla, no other relevant type of plants are found out in the images. It is possible to assume that the density per hectare of other species is minimal.
- It is possible to find the cardilla in two possible scenarios:
  - **The density is not enough to form an almost indistinguible shape made of cardillas**, being possible to count and identify them with relatively ease.
  - **The density is high or really high**; the plants are too near between each other and is really complicated to identify them (as you can see in image 2.5).

- The ground is almost always covered by a thin layer of green, yellow or brown grass. The dryer the soil, the browner the grass is. The density of cardillas increases in that specific region when this fact is true. Cardillas were also found in soil with yellow grass, but the quantity was negligible.

- Depending on the angle of the camera is easier to identify some individuals. In some cases, the closer it forms a 90º angle with respect to the ground is easier to identify the individuals. However, when the density increases too much, the complexity of the problem escalates. Frames with an angle in the range of [50, 70] degrees are still required and are incredibly useful.

Fig. 2.5. Cardilla density examples: high density (left) and low-to-medium density (right). As the density increases, the soil gets darker, whereas when the density is lower it is possible to see green grass.

After this research, it is confirmed that different densities can be found of these type of weed, directly affecting the environment, avoiding the appereance of other species in the designated region. This is important to take into account, as no other possible plant species were identified during this process.

Furthermore, it is important to take into consideration the density of the weed in certain region, as it was previously mentioned. The human capacity of detecting the weeds is limited and that affects directly the labelling (see paragraph 2.2.3) used for the e cardilla is vividly green with respect to the other cardillas prmodel to be trained. Only perfect instances will be annotated to reduce overfitting or possible errors.

Plant analysis

An aerial analysis of the plant is also done in order to know how to correctly label instances of it. For a more precise analysis of the plant, please refer to the subsection 1.1.1, as it contains an extense explication about its main characteristics.

From an aerial perspective, the cardilla can be identified from two different angles: forming an angle in the range of [50, 70] degrees with respect to the ground or forming an angle close to 90 degres with respect to the ground. Both points of view are valid and give

important information about the individuals of the specific region, but limitations need to be considered.

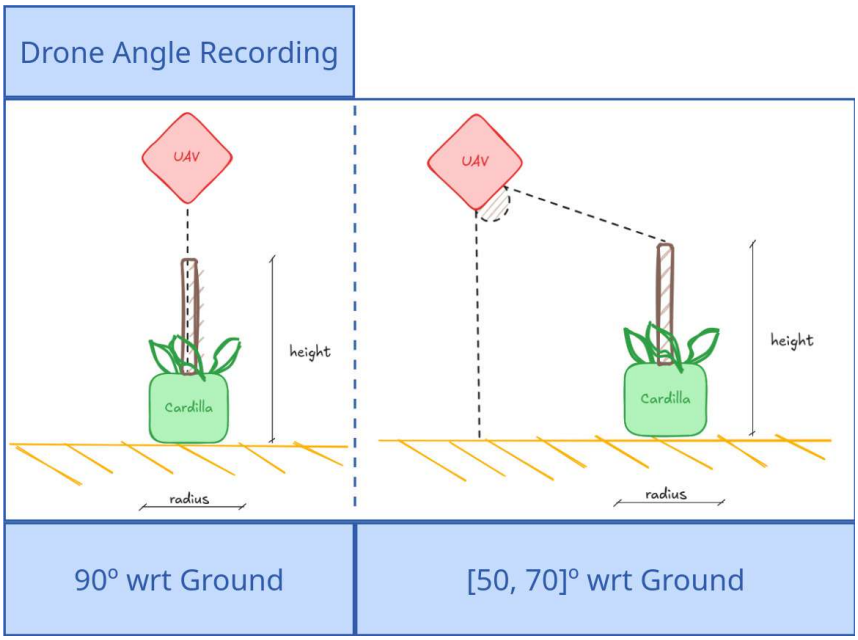


Fig. 2.6. A representation of the two possible options when recording on flight.

Depends on the ground and the context, but usually the weed has clear shape, circular with leaves rooting from the origin of it. Leaves are usually greener in the inner part (the closer to the center) whereas the parts near the end of the leaves are drier, being browner or grey. This is vital, as makes it easier to identify them with respect to the soil.

However, smaller individuals can be entirely green, making the previous method for identification invalid and hard to use. For those, it is important to consider the circular shape of the leaves and, if possible, the height of the plant.

With frames in the range of [50, 70] degrees with respect to the ground it is relatively easy to identify the biggest specimens as it stands out the individual. Notice that these weeds can be up to 2 meters tall if the stem is considered.

The biggest specimens are easy to identify: the stem is unchallenging to distinguish. Although its great size, the cardilla is vividly green with respect to the other cardillas previously mentioned with smaller size.

Fig. 2.7. Low-to-med density environment. The smaller cardillas are greener than the medium-sized ones. A taller, greener cardilla constrasts in the image on the left.

2.2.3. Image Labelling

For carrying out this task, the tool Roboflow has been used due to its versatility and capabilities for image processing and labelling. Roboflow is a SaaS (Software as a Service) that allows to manage, anotate and prepare dataset used in computer vision. Facilitates assignments such as labelling, generating augmented versions of the dataset and the exportation to different training formats such as YOLO or COCO.

For the image labelling, the usage of **bounding boxes** was crucial. The concept of bounding boxes can be defined as the following: a rectangular outline that completely encloses an object or an specific area, defined by coordinates. Those coordinates are usually inside a quadruple (if the shape is rectangular) with the form of:

$$\text{Bounding Box} = [x\_min, y\_min, x\_max, y\_max]$$

Where x corresponds with the coordinate (pixel) in the x-axis and y corresponds with the one in the y-axis.

As this problem only addresses only one type of weed, the cardilla, the usage of only one class was needed in order to label the entire dataset.

Labelling Criteria

After the extensive research commented in the section 2.2.2.

This is the most time-consuming part of the dataset creation. Identifying all possible samples from the dataset manually is complicated without the help of a team specialised in the topic. Knowing this, it is probable that the dataset labelling may contain some errors that may cause the appereance of false positives or reduced performance of our model.

2.2.4. Limitations of the dataset

At the current date of writing this document (December 3, 2025), the dataset is not perfect. In fact, it has several limitations that affect directly to the performance of the

current models:

- Limited location variety. Due to other limitations outside this scope, the recordings were done only in the field commented in the section 2.1. These won't allow to generalise properly to the model.
- Reduced sample size. The variety of the images is limited to the environment subject to study. Furthermore, the terrain may differ between other location (can be darker or lighter), directly affecting to the metrics of the model.

3. MODELS

3.1. T-Rex 2

- DFT
- Tiene una alta precisión para identificar las cardillas, pero no alcanza a capturar el tamaño total de dicha planta con una sola imagen como muestreo. Si se añadiesen más imágenes, se podría tener una mayor precisión.
- De nuevo, dependiendo de la imagen del muestreo el resultado cambia mucho. Cuando el muestreo es una imagen clara de las cardillas — en perpendicular con el suelo — la precisión mejora mucho para ese tipo de imágenes, pero no mejora (incluso empeora) en las imágenes que no son de ese tipo.
- De nuevo, por el tamaño de muestreo, no es capaz de generalizar correctamente si nos encontramos con unidades secas o no.
- En fotos con mucha maleza, la generalización no funciona como se esperaba (importante mencionar que ninguna foto de muestreo incluía este tipo de fotos).

A pesar de todo esto, tenemos un IoU relativamente alto para algunas imágenes. La media de IoU está en *0.697*

Se debería de probar con más tipos de imágenes de muestreo. No tengo tokens suficientes. -> T-Rex no permite esto. Por ello, no lo hace un buen modelo para este tipo de situaciones.

Hipótesis: el modelo es muy bueno, pero se debe de ejecutar con un dataset relativamente grande y bueno para que sea capaz de servir como un muestreo amplio para muchos casos genéricos.

3.2. YOLO

- Obtenemos muy buenas precisiones (60%): YOLOv8s + YOLOv8l + YOLOv11s + YOLOv11l (mejor precisión con YOLOv11, pero resultados muy aceptables con YOLOv8).
- Falla cuando hay densidad muy alta de cardilla, lo que reduce la precisión enormemente creando boxes demasiado grandes
- Limitado por el hardware actual. Modelos con confianza relativamente media-alta. Limitados por dataset.

3.2.1. Results

3.3. Modified YOLO: exploring the limits of object detection

3.4. YOLO + SAM2: improving accuracy through masks

BIBLIOGRAPHY

[1] W. Bank, *World development indicators: Agriculture, forestry, and fishing, value added (% of gdp)*, https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?end=2024&locations=BR-UY-AR-ES-1W&start=1960&view=chart, Accessed October 2025, 2024.

[2] G. R. Coleman et al., "Weed detection to weed recognition: Reviewing 50 years of research to identify constraints and opportunities for large-scale cropping systems," *Weed Technology*, vol. 36, no. 6, pp. 741–757, 2022. DOI: 10.1017/wet.2022.84.

[3] K. Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," *Neural Networks*, vol. 1, no. 2, pp. 119–130, 1988. DOI: https://doi.org/10.1016/0893-6080(88)90014-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0893608088900147.

[4] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for boltzmann machines," *Cognitive Science*, vol. 9, no. 1, pp. 147–169, 1985. DOI: https://doi.org/10.1016/S0364-0213(85)80012-4. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0364021385800124.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, Dec. 1998. DOI: 10.1109/5.726791.

[6] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems*, vol. 25, Jan. 2012. DOI: 10.1145/3065386.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," vol. 7, Dec. 2015.

[8] A. Quinones, J. V. Savian, A. Hirigoyen, and J. Valente, "Towards improved weed detection in native grasslands using high-resolution drone imagery and artificial intelligence," 2025.

[9] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: A literature review," *Artificial Intelligence*, vol. 293, p. 103 448, 2021. DOI: https://doi.org/10.1016/j.artint.2020.103448. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370220301958.

[10] Q. Jiang, F. Li, Z. Zeng, T. Ren, S. Liu, and L. Zhang, *T-rex2: Towards generic object detection via text-visual prompt synergy*, 2024. arXiv: 2403.14610 [cs.CV].

[11] J. M. Pizarro, *Carla*, https://github.com/jmartinpizarro/carla.

[12] S. Ali et al., "Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence," *Information Fusion*, vol. 99, p. 101 805, 2023. DOI: https://doi.org/10.1016/j.inffus.2023.101805. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253523001148.

[13] G. J. Sánchez-Zuluaga, L. Isaza-Giraldo, G. D. Zapata-Madrigal, R. García-Sierra, and J. E. Candelo-Becerra, "Unmanned aircraft systems: A latin american review and analysis from the colombian context," *Applied Sciences*, vol. 13, no. 3, 2023. DOI: 10.3390/app13031801. [Online]. Available: https://www.mdpi.com/2076-3417/13/3/1801.