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

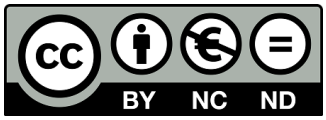
---

Javier Martín Pizarro

Joao Ricardo Pereira Valente

Leganés, Madrid

June, 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%.

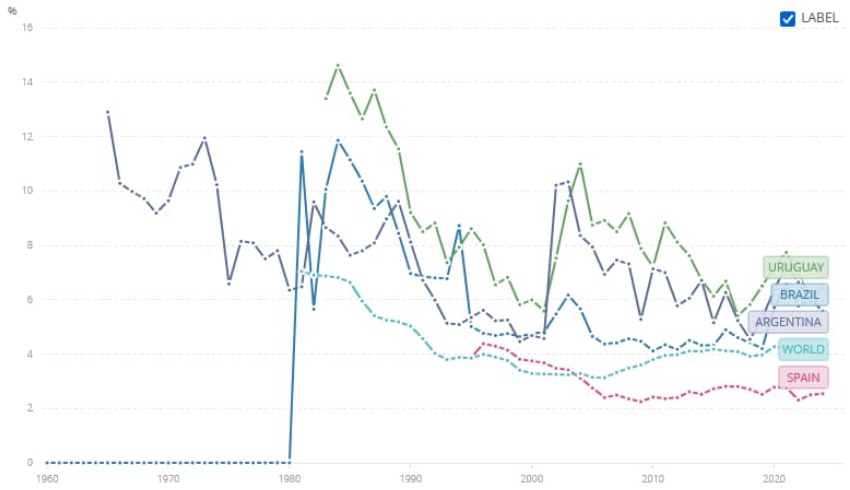


Fig. 1.1. Gross Domestic Product (GDP) share of agriculture, forestry, and fishing (in %) between the years 1960 and 2024. Source: *World Bank Data* [1].

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 have been (and still 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.

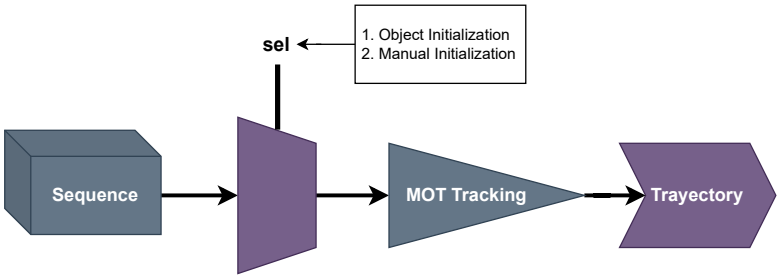


Fig. 1.4. Visual representation of the MOT problem. The multiplexer determines the approach: either Detection-based Tracking or Detection-Free Tracking.

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?

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

1.2.1. Asociation Algorithms

1.3. Regulatory Framework

1.4. Socio-economic impact

2. MODELS

2.1. 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.

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.

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].