INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Object Detection of Megalithic Dolmens in Google Satellite Imagery

Daniel André Barbosa Marçal

Master in Data Science,

Supervisor:
Professora Doutora Ana Maria de Almeida, Associate Professor,
ISCTE Instituto Universitário de Lisboa

Cossupervisor:
Professor Doutor João Pedro Oliveira, Assistant Professor,
ISCTE Instituto Universitário de Lisboa

October, 2023

TECNOLOGIAS
E ARQUITETURA

BUSINESS
SCHOOL

Departamento de Métodos Quantitativos para Gestão e Economia

Departamento de Ciências e Tecnologia da Informação

Object Detection of Megalithic Dolmens in Google Satellite Imagery

Daniel André Barbosa Marçal

Mestrado em Ciência de Dados,

Orientadora:
Professora Doutora Ana Maria de Almeida, Professora Associada,
ISCTE Instituto Universitário de Lisboa

Co-Orientador:
Professor Doutor João Pedro Oliveira, Professor Auxiliar,
ISCTE Instituto Universitário de Lisboa

Agradecimento

A dissertação de mestrado é sem dúvida um dos maiores desafios enfrentados no meu percurso académico. Apesar de ser um trabalho que cabe ao orientando produzir, nunca será possível a sua realização sem o acompanhamento e contributo de outras pessoas.

A professora Ana de Almeida, orientadora da dissertação, estou eternamente agradecido todo o apoio e disponibilidade que teve neste percurso, como também todo o seu contributo para o presente trabalho.

O professor João Oliveira, coorientador da tese, agradeço todo o apoio e orientação na escolha dos melhores métodos e abordagens para realizar esta dissertação.

Agradeço também a Ariele Câmara, uma investigadora da área da arqueologia por toda a informação que de outra forma eu sozinho não conseguiria obter e por toda a contribuição de uma profissional da área.

Estou eternamente grato à minha família, especialmente ao meu irmão, por todo o apoio incondicional a todos os níveis durante este ano, que infelizmente não se carateriza por ser um dos mais jubilosos da nossa vida. A todos um grande obrigado.

Resumo

A deteção de estruturas megalíticas tem uma grande importância, a nível humano ou arqueológico. Neste documento, será descrito o trabalho desenvolvido em prol do desafio de automatizar a deteção de dolmens, através de um caso de estudo destes monumentos na região do Alentejo, Portugal. Para a automatização, serão utilizados algoritmos de aprendizagem profunda, especialmente as arquiteturas YOLOv8 e FasterRCNN, conjugando a informação do terreno, de modo a refinar os resultados.

A motivação por detrás desta dissertação, surge da abundância de dolmens por descobrir, resultando na necessidade de sistemas de deteção eficientes. A metodologia envolve o pré processamento de imagens de modo a revelar elementos de interesse e a aplicar aprendizagem profunda, fazendo prospeção automática de dolmens numa região similar.

A descoberta principal do estudo foi a eficácia conseguida com a aproximação FasterRCNN na deteção de dolmens, no qual conseguimos atingir 93% de precisão média. Criando um sistema robusto, a ser utilizado como ferramenta de trabalho aos arqueólogos na identificação de potenciais áreas de prospeção de monumentos megalíticos.

Contudo, foram encontradas limitações, devido à falta de disponibilidade de imagens de qualidade na *Google Earth*, o que afeta, necessariamente, a precisão dos resultados. Trabalho futuro poderá focar em adquirir imagens de maior resolução, por forma a melhorar o desempenho do algoritmo.

Este estudo criou um sistema promissor para deteção automática de dolmens através da aprendizagem profunda. No entanto, será sempre necessário procurar a melhoria contínua, quer na capacidade do sistema, quer pela aquisição de dados de diferentes fontes de informação.

Abstract

The detection of dolmens, ancient megalithic structures, holds significant human and archaeological importance. We address the challenge of automating dolmen detection in the region of Alentejo, Portugal, by leveraging deep learning algorithms, specifically YOLOv8 and FasterRCNN, and exploring the potential influence of terrain information, including distance to water and rocky outcrops for refining the detection results.

The motivation behind this research stems from the abundance of undiscovered dolmens in the area, prompting the need for an efficient and accurate detection pipeline. The methodology involves preprocessing the images to enhance relevant features and applying the Deep Learning approach for dolmen detection.

The key finding of the study demonstrates the efficacy of FasterRCNN architectures in dolmen detection, which have achieved a confidence degree of 93% (average precision). These findings offer valuable insights and practical assistance to local archaeologists in identifying small megalithic structures in similar regions.

However, limitations were encountered, mostly due to the unavailability of high-quality images in *Google Earth* databases, thereby affecting the precision of the results. Future work should focus on acquiring more comprehensive and high-resolution image datasets to enhance the performance of the detection algorithm.

Our research provides a promising pipeline for automated dolmen detection using Deep Learning algorithms. However, we must emphasize the need for continued improvement and for the acquisition of additional data from different information sources to enhance the accuracy, efficiency, and generalization capacity of dolmen detection algorithms.

Index

Figures Index

Tables Index

List of Abbreviations and Acronyms

CNNs - Convolutional Neural Networks

DCN - Dilated Convolutional Network

DFL - Dynamic Feature Learning

DSR - Design Science Research

FN - False Negative

FP - False Positive

FPN - Feature Pyramid Network

GIS - Geographic Information System

IoU - Intersection over Union

MSE - Mean Squared Error

RPN - Region Proposal Network

TP - True Positive

YOLO - You Only Look Once

CHAPTER 1

Introduction

1.1. Context

Nowadays, the increasing usage of satellite imagery allows for a better understanding of potential archeological sites to help archeologists with their discoveries and studies. Archeology is the study of tangible remains from the past and the present [1], usually involving terrain prospection to discover human remains that enable the study of ancient societies. Computer vision makes it possible to obtain more features faster, by training suitable Deep Learning (DL) algorithms and applying adequate pre-processing to the images before using it for object detection. Therefore, the last decade has seen an increasing interest in remote-sensing technologies and methods for monitoring cultural heritage [2].

The objects that the work focuses on detecting are dolmens. These are megalithic monuments built with rocks in a table shape, formed by a central camera that can vary in dimensions, sometimes covered by vegetation. Its name came from Breton origins where “dól” means table and “men" means rock, so (rock table) [3].

In terms of structural and contextual characterization, Câmara describes the main differences in Portuguese dolmens from the rest of the dolmens in the world, based on their geometry, architecture, and spatial distribution, which turns them unique [3]. While Portugal is known to be one of the richer areas in terms of small megalithic monuments, it is believed that there are still many unknown monuments, particularly in the Alentejo region [4].

This work directs its focus on a significant research gap in terms of automated archeological prospection in the country where the study is taking place: Portugal. We believe that, combining terrain and expert information with DL shows significant benefits for remote identification, where the landscape and the object must be studied together and where some terrain features, such as water, rock outcrops, and vegetation, could indicate key areas with higher probability of occurrence of dolmen sites [5].

1.2. Motivation

There is an increase in the usage of satellite imagery within several areas of study, particularly in archaeology, which has led to a need to optimize the traditional manual methods used in

image prospection. Some investigations and case studies are taking place in various locations and involving various object types, as evidenced by multiple studies [6] – [16].

To this day, satellite sensors are still evolving, but researchers keep having data accessibility restrictions. Moreover, there are not that many studies in remote automation within the archeology field and the ones that exist demonstrate the need to automate their processes from end-to-end and enhance the difficulty in the extraction of useful knowledge to get the best remote identification zones, that is, with higher likelihood of finding (buried) monuments.

However, an important issue arises with object detection in an image, especially in dolmen detection that this work tries to tackle. Usually, object detection of small or (partially) buried monuments in a remote image ends with a high number of false positive detections. The reasons vary from the images having to many small objects, the dolmen being very similar to a rock outcrop, or, as previously mentioned, the object being buried or hidden in vegetation [5].

Therefore, our motivation is to create a new and effective approach in this area of study for dolmen detection decreasing the number of false positives and presenting a likelihood for each object detection. This work intends is built on a previous work on the same case study here used that employed pixel-based detection and applied Machine Learning methods based on the pixel detection of a dolmen [14]. The innovation concerns mainly the fact that we employ a different approach with focus on the dolmen object itself, contributing for a new methodological process for remote dolmen identification thus contributing to the area of archeological prospection on remote imagery.

1.3. Research Questions

This dissertation main goal is that of finding the best DL framework to further build a continuous learning pipeline capable of automatically recognizing dolmens in images. We use the Alentejo Portuguese region as a case study, gathering open-source images provided by *Google Earth*.

Our aims are (a) to understand the viability of employing DL architectures for the detection of dolmens in a mostly granitic region by nature, and (b) to investigate the use of expert domain information for detection improvement. Therefore, the results of the application of the DL model will be further analyzed using information from the terrain to understand if there are any landscape features that can be attached to critical areas for improving precision and, either decreasing the number of false positives or enabling the labelling of positives with an effective probability for the existence of a dolmen at that location.

To achieve this purpose, we decided to investigate the following research questions:

1) How important is it to enhance and preprocess the images before feeding the algorithm?
2) Which DL detection algorithm shows the best results in the detection of megalithic monuments, and which is the one presenting fewer false positive results?
3) Is the information available about the landscape features and the background of the object correlated with the existence of a dolmen?

1.4. Objectives

Taking into consideration the previous context and research questions, we can formulate the following research objectives. Firstly, to create the best possible system for dolmens detection in remote images. The images used were extracted from an open-data source and we have thrived to implement effective preprocessing and useful image augmentations in order to highlight the monuments, since these can be partially or almost totally covered by vegetation or other surrounding structures.

One other objective relates with the investigation of the role that the landscape and available expert domain information may have to refine the results from the detection model. In addition, we intend to find which would be the best terrain features to employ in this approach, in order to see if there is any correlation between the monuments and these features.

In the end, the main objective is to be able to contribute to help experts in their archeological prospection with some form of automation of their processes and help archeologists to detect monuments in a more efficient way by exploring a new automated methodology.

1.5. Contributions

This work makes significant contributions to the fields of Deep Learning as a tool in Archaeology by introducing novel approaches for the detection of small megalithic objects in the Portuguese Alentejo region in poorly contrastive images obtained from satellite imagery.

Through the utilization of advanced object detection techniques in the domain of DL, this research presents innovative methods that exhibit remarkable accuracy in identifying these small megalithic structures, thereby enhancing the efficiency of archaeological site identification. Moreover, this work extends its applied impact beyond algorithmic advancements, addressing practical challenges faced by local archaeologists.

By developing automated processes for the detection and localization of megalithic objects, the dissertation contributes to streamlining archaeological investigations. This automation not only expedites the detection process but also empowers archaeologists by providing them with powerful tools to aid in their analyses. In amalgamating cutting-edge technology with archaeological exploration, this thesis not only expands the horizon of DL applications but also facilitates and enriches archaeological research efforts in the culturally significant region of Alentejo, working as a proof-of-concept for an automated tool that experts can use in other regions and monuments.

The dataset used in this work is one other contribution. The data was gathered from *Google Earth* and was manually labeled. This dataset is available and public on *RoboFlow* [17].

Finally, it should be noticed that a conference paper describing the DL detection system and preliminary results has been submitted for peer review under the usual paper submission process in conference “2023 Symposium of Applied Computing”.

1.6. Research Methodology

The research methodology used has been the Design Science Research (DSR) [18], which is a methodology that aims to investigate, develop, and evaluate innovative artifacts, models, or systems to address specific real-world problems. It is commonly applied in fields such as information systems, computer science, and engineering [18]. DSR combines rigorous scientific investigation with the creation of practical solutions, bridging the gap between theory and practice.

DSR is divided into seven main steps next described.

(1) Problem identification and motivation: identify a specific problem or opportunity in a particular domain.

(2) Design and requirements definition: define the design requirements based on the identified problem and the objectives associated.

(3) Design and development: create a conceptual design for the artifact that addresses the identified problem. This step involves determining the architecture, components, interfaces, algorithms, or other relevant aspects of the artifact.

(4) Evaluation: assess the artifact's effectiveness, efficiency, and utility.

(5) Reflection and learning: reflect on the evaluation results and analyze the findings to gain insights into the artifact's strengths, weaknesses, and potential improvements. This step guides the refinement and enhancement of the artifact based on the obtained insights.

(6) Communication and dissemination: document and communicate the research process, outcomes, and artifacts to the relevant audience.

(7) Repeat iteratively in case it is needed: DSR may act as an iterative process, and steps 2 to 5/6 are usually repeated multiple times until a satisfactory artifact is developed. Each iteration builds upon the knowledge gained from previous iterations, leading to an enhanced understanding of the problem and the artifact's effectiveness.

In our case, chapter I and II illustrates the first step, where we identify the problem and opportunities in this domain, also the entire literature that is essential to get the state of art based on the methodologies applied.

Step two dates to early in this year, where it was defined the plan and design of the requirements based on our object detection problem, and also taking account the resources and datasets available.

Step three to five are referred to chapters III to V respectively, where the development and design of methodologies are described and determined, the results obtained and evaluated, and the main conclusions based on the evaluation results are made. Analyzing all the results and giving insights for future work.

1.7. Dissertation Structure

This document is divided into five chapters that describe the different phases of the work and give the organization necessary to address every step.

The first chapter introduces the dissertation, where it is given the context, motivations, objectives, and contributions, as well as the research questions and research method.

The second chapter addresses all the necessary literature review, to give context to the dissertation and to fundament some points made along the document. Based on existent scientific work.

The third chapter describes the entire proposed work for this thesis. Where it is addressed the methodologies and technologies used.

The fourth chapter is the implementation and results of the proposed work described in the previous chapter.

In the fifth chapter, is where the main conclusions about the results and methodologies are described, as well as future work recommendations.

CHAPTER 2

Literature Review

This chapter describes all the necessary literature review, to give context to the dissertation and to fundament some points made along the document. Based on existent scientific work, this chapter can be divided into six sections: the first contextualizes the archeology study field; the second addresses the monuments and its ontology; the third describes remote sensing techniques used nowadays and how it has evolved; the fourth presents the state of art in image enhancement techniques to data used in object detection; the fifth describes some of the object detection approaches in general; the sixth gives context about related work and studies in archelogy, by describing all the methodologies and their best takeaways and conclusions.

2.1. Archeological Context

Alentejo is the region selected to gather all the data for this dissertation. Having just 60 dolmens, consist of relatively small megalithic funerary monuments that can be visualized from satellite images. Some of these structures can date more than 4000 years, being essential for the social and cultural development during the Neolithic [3].

Dolmens are made of large stones composing geometrics chambers that can be circular, semi-circular, or quadrangular. Because they have survived the rise and fall of several civilizations, they may have been reused, buried, destroyed, or annexed to other structures and can be found in rural or urban areas [4]. The majority of these monuments it´s not in their original shape, most of them are destroyed, but their structure still can be visualized even those who are not complete anymore, or the existence of soil marks indicate their presence [4].

These monumental structures were not visible in their original form because they were covered by successive layers of earth and stone, known as barrows or tumulus; the presence of the tumulus may have vanished over the long period of time since its construction. The dolmens are scattered throughout the region, with most of them located near major riversides and rocky outcrops. They are commonly found in groups of up to a dozen, with relatively short distances between them. In Figure 1 is possible to see the structure of a dolmen [4][5].

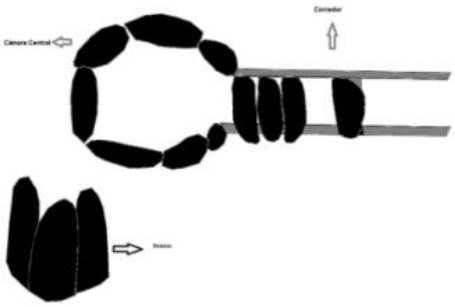


Figure 1 - Dolmen Structure (Source: [4]).

Given the natural environment's irregular placement of rocks, the geometric shape present in the chamber stands out from the surroundings, even though such regular forms are found in nature, being up to meters smaller than the diameter of the dolmens chamber [5].

2.2. Ontology in Archeology

Ontology, in the context of knowledge representation, refers to the formal specification of concepts, their properties, and the relationships between them within a particular domain. It provides a structured framework for organizing and categorizing knowledge, facilitating effective communication, and reasoning [19].

Ontologies are widely used in various domains, including artificial intelligence, information science, and the semantic web. By creating a shared understanding of a subject area, ontologies enable interoperability between different systems and support tasks such as data integration, information retrieval, and decision-making [20].

In archaeology, ontology plays a vital role in knowledge organization and data management. An archaeological ontology aims to represent the concepts, classifications, and relationships relevant to archaeological data, such as artifact types, excavation methods, and archaeological contexts [21]. It provides a standardized vocabulary and framework for describing and categorizing archaeological information, improving data consistency, and facilitating data sharing among researchers. Ontologies in archaeology enable more efficient data integration, support advanced querying and analysis, and enhance the interoperability of archaeological databases and information systems.

2.3. Remote Sensing

Remote sensing refers to the collection of information about an object or area without direct physical contact, typically using satellites, aircraft, or drones equipped with sensors [22]. Satellite imagery, a key application of remote sensing, provides valuable data about the Earth's

surface and atmosphere from space. This imagery offers numerous benefits across various domains, including environmental monitoring, agriculture, urban planning, and disaster management [23], providing a wide range of advantages. Firstly, it allows for large-scale coverage of areas that may be inaccessible or difficult to survey on the ground. This capability is particularly useful in remote or hazardous regions. Secondly, satellite imagery provides a historical record of the Earth's surface, enabling the analysis of changes over time.

This longitudinal perspective is crucial for understanding long-term trends, such as deforestation, urban growth, or climate change impacts [22]. Additionally, satellite images offer multi or hyper-spectral data, capturing information beyond what is visible to the human eye. By analyzing different wavelengths, researchers can infer properties such as vegetation health, land surface temperature, or oceanic phenomena, aiding in ecosystem monitoring and resource management [24].

Moreover, remote sensing using object detection techniques has gained prominence. Object detection algorithms automatically identify and delineate specific objects or features within satellite imagery [25]. This approach facilitates the extraction of valuable information at scale. Object detection enables the identification of objects such as buildings, roads, vehicles, and vegetation, providing crucial inputs for urban planning, infrastructure development, and disaster response [26].

2.4. Image Enhancement Approaches

Image enhancement techniques play a crucial role in improving object detection performance, especially in challenging environments. These techniques aim to enhance the quality, contrast, or clarity of images, making it easier for object detection algorithms to identify and localize objects of interest. Several studies highlight the importance of image enhancement techniques in object detection, emphasizing their ability to enhance image quality, improve visibility, and mitigate various challenges [27]-[31].

Image enhancement techniques were employed to improve object detection in challenging environments. The authors highlighted the significance of enhancing images to mitigate issues such as low contrast, noise, or uneven illumination, which can negatively impact the accuracy of object detection algorithms [27].

Aladem et al. emphasized the importance of color enhancement techniques in enhancing object detection performance [28]. By manipulating the color properties of images, such as saturation, brightness, or contrast, the authors demonstrated improved object detection results, especially in scenarios with complex backgrounds or lighting variations.

In the field of surveillance systems, the authors of [29] explored image enhancement techniques for better object detection. The authors emphasized the need for enhancing low-resolution or low-contrast surveillance images to enhance the visibility of objects and increase the accuracy of detection algorithms. Improving the contrast between objects and their backgrounds, showed significant improvements in the detection accuracy and robustness of object detection algorithms [30].

2.5. Object Detection Approaches

The use of satellite imagery these days allows for a better understanding about potential archeological sites to help archeologists in their discoveries and studies [32]. The issue lies in the fact that, given the cost and labor-intensive nature of traditional methods, archaeologists cannot effectively analyze these datasets [33].

Through computer vision, it's possible to obtain more features faster, by training suitable DL algorithms and applying adequate pre-processing to the imagery before using it. The last decade has seen an increasing interest in remote-sensing technologies and methods for monitoring cultural heritage [33].

Luo et al. made a review on a century of work in this topic, from 1917 to 2017, and discovered that remote sensing technology was not originally designed for archaeological purposes, but it has become an indispensable and powerful tool for archeologists and is being applied for various purposes. He also described that exists a lot of potential to use DL and cover many research gaps for the experts in the field [34].

Object detection is a computer vision task that involves identifying and localizing objects of interest within an image or video [35]. The goal is to draw bounding boxes around these objects and classify them into predefined categories. Object detection has numerous applications, including autonomous vehicles, surveillance systems, medical imaging, and more [36].

There are various techniques for object detection and one popular approach is the use of DL models, particularly convolutional neural networks (CNNs). CNN-based object detection methods have shown remarkable performance in recent years due to their ability to learn complex features from images. The classification technique is based on the obtained segments, and each segment is assigned to a class based on the target object characteristics, such as geometry and background relationships.

One of the widely used CNN-based object detection models is Faster R-CNN (Region-based Convolutional Neural Networks), which was proposed by [37] in their 2020 paper. Faster R-CNN improved upon earlier methods like R-CNN and Fast R-CNN by introducing a region proposal network (RPN) that efficiently generates region proposals, allowing the model to focus on potential object locations and refine its predictions.

2.6. Object Detection Methodologies

Some applications using the detection of archelogy objects have been successfully used. In archaeology, these applications have a significant disadvantage, where each application is created based on the specific features of a region and structural monumental typology, resulting in a non-generalizable solution for other applications in different geographical or cultural contexts [38]. This is due to the extremely high structural differences present in archaeological remains, which are caused by the culture and available materials in the area, resulting in highly divergent characteristics or environmental insertions [38][39].

Around the world, some authors used these methodologies to detect multiple archeological objects such as mounds, barrows, and Celtic fields [6] - [16]. Choosing Faster R-CNN proves to perform well on difficult objection detection tasks, and it is well suited for semi-automatic detection of cultural heritage [6] - [13]. It has been demonstrated that Faster R-CNN is convenient, in terms of the consumer's accuracy, for automated detection of cultural heritage objects in high resolution data. However, one may expect that the method must be improved in terms of the accuracy in order to limit the number of false positives when applied to large areas for detailed archaeological mapping. One such study worked in this precise area, using pixel-based detection and classifiers to obtain detection results, having 83% accuracy in their developed system and being able to eliminate 71% of false positive detections [14].

YOLO has been used in the detection of Portuguese burial mounds in Alto Minho, obtaining a 72.53% positive rate in their detections [16]. They also conclude that is a faster method than traditional CNN approaches that also mitigates the issue of false positive detection.

The pre-processing of the data is also a crucial step before training the algorithms. The use of image augmentation shows to be an essential technique for improving the performance and robustness of object detection models [12] [13]. Augmentation involves applying various transformations to the original training data to create new, slightly altered versions of the images. These transformed images are then used to train the object detection model, since they enable for a more robust model, that is, a with a better generalization capacity [13]. Some authors conclude that overall, this pre-processing technique increases data diversity, by helping the model to generalize better to unseen data and different real-world scenarios and improves model robustness: by exposing the model to augmented data during training, it becomes more resilient to various challenges, such as changes in lighting conditions, different object scales, rotations, and occlusions [6] [7].

Most of these studies found in this literature review used the average precision and F1-score as metrics to evaluate their algorithms [6] [7] [10] [11] [13], basing the results on their test and validation sets. Table 1 summarizes the few studies on object detection in archeology that have been here reviewed, gathering the state of art in this area of study.

Table 1 - Object Detection Studies in Archeology.

Reference	Country	Year	Method	Object of the detection
6	France	2018	CNN's	Mounds
7	Norway	2019	Faster RCNN	Mounds, Kilns
8	Norway	2021	Faster RCNN	Mounds
9	Netherlands	2019	Faster RCNN	Barrows, Celtic fields
10	USA	2021	Faster RCNN	Shell Rings
11	Brazil	2021	RetinaNet	Mounds
12	USA	2018	CNN's	Mounds
13	Pakistan	2023	Faster RCNN	Mounds
15	Spain	2019	CNN's	Mounds
16	Portugal	2023	YOLO	Mounds

A few conclusions we can have from this chapter are that Satellite imagery is a valuable tool for archaeologists, providing enhanced insights into potential archaeological sites. However, the conventional methods for analyzing such datasets are hindered by their high cost and labor-intensive nature.

The integration of computer vision techniques, particularly deep learning algorithms, with remote sensing technologies has revolutionized the field of cultural heritage monitoring. This advancement has not only accelerated the extraction of features from imagery but also addressed longstanding research gaps, showcasing the potential of deep learning in archaeological applications. The introduction of Faster R-CNN, with its innovative region proposal network, represents a significant stride forward in CNN-based object detection, offering improved localization and classification capabilities for archaeological studies.

Faster R-CNN and YOLO emerge as a promising choice for complex object detection tasks in megalithic monuments. While it demonstrates good performance in terms of consumer's accuracy, there is room for improvement, particularly in reducing false positive detections when applied over large areas for detailed archaeological mapping.

CHAPTER 3

Dolmens Detection System

This chapter describes the system proposed for the detection of dolmens in remote images of terrain. We start by introducing the necessary technologies, tools, and methodologies supported by related work. Firstly, we describe the system's pipeline scheme that sums up the work developed, divided by all the main steps that took place in this investigation. Beginning with the *Google Earth* images acquisition, describing why it is viable and how it is used; the techniques for image augmentation and enhancements proposed; the use of *RoboFlow* during the process of the image's preparation; the options of algorithms used and proposed for this work; the necessary evaluation for these algorithms.

Following up on the methodological design research process, this chapter encompasses step three where we create a conceptual design for the artifact that addresses the identified problem.

3.1. Pipeline Architecture

Considering all the important aspects of the proposed work, it is possible to build a pipeline that can be automated to obtain the desired results. From image gathering to dolmen detection in an image, the pipeline is presented in Figure 2 and will be briefly described over the next sections.

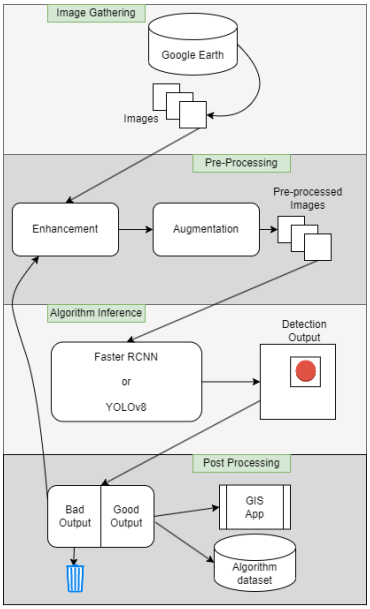


Figure 2 - Pipeline Architecture.

3.1.1. Data Gathering

This phase relates to the acquisition of all the image data that can be used to train and test the algorithm. This could be done by using *Google Earth* satellite imagery or other high-resolution sources and image databases. We procceded by extracting all the data from *Google Earth.* For future automation of the process, this dataset can be further used for testing images of interest or, since the images have been labelled, it can be used as a training dataset for new models.

3.1.2. Image Processing

This step will be described further down and where all the input images are going to be enhanced and augmented with techniques, in this case using *RoboFlow*.

3.1.3. Algorithm Inference

With the images ready and fed to train a model, this will receive the new inputs and load its weights and parameters to obtain a final output of the detection (or not) with a certain level of confidence and precision.

The outputs from this step can be stored and further analyzed or discarded, depending on the level of detection pretended and the result obtained.

3.1.4. Post Processing

The outputs now can be taken to a GIS application to be georeferenced in order to get the coordinates for the location of the positive hits resulting from the previous step.

Before taking this outputs to the application, an analysis of the output comes in place to decide if the image and its results are good enough (based on their score and level of confidence) to feed the algorithm with more of this new labelled data so that it can became more robust and overall better at detecting new objects. To this end, a bad detection optionally can be reprocessed from the image preprocessing step or discarded to not become part of the learning process of the algorithm.

3.2. *Google Earth* Imagery

In this study all the images were gathered from *Google Earth* Pro, which can be extracted and saved with 8k resolution, providing high quality images.

DL models have demonstrated better capabilities in detecting and localizing objects within images. However, the accuracy and reliability of these models heavily rely on the quality of the input data [40]. Thus, higher quality images play a crucial role in improving object detection performance in Deep Learning. When higher quality images are used for object detection, several benefits emerge. Firstly, high resolution images contain more detailed information, enabling more differences between background and objects [40].

Higher quality images tend to have less noise, which interferes with object detection. Noise reduction techniques, such as better sensor technology, improved image preprocessing, or higher signal-to-noise ratios, result in cleaner images with fewer distractions. Cleaner data allows the model to focus on relevant features, leading to improved object detection performance [41].

Additionally, higher quality images often have better color fidelity and dynamic range, which can benefit object detection in certain scenarios. For example, in applications where contrast or distinguishing objects based on color is essential, better image quality can provide more accurate information for the model to learn from [40][41].

Google Earth platform has become a valuable tool for researchers in a wide range of fields, namely in Archeology, enabling them to explore and analyze the Earth's surface in detail [42]-[44]. The authors of [44] utilized *Google Earth* imagery to identify potential archaeological sites in Egypt. The high-resolution imagery allowed the researcher to identify subtle variations in surface features, indicating buried archaeological structures, which were then confirmed through ground-truthing and archaeological investigations.

We can extract images from *Google Earth* picturing the Alentejo, Portugal area dating since 2006. After 2015, each year presents data with multiple timeframes within the year, with quality varying throughout the timeline. This does not necessarily mean better quality over the years, but rather depends on which satellite captured the images. For example, images from 2017 show better resolution than images from 2021 (see Figure 3 for comparison).

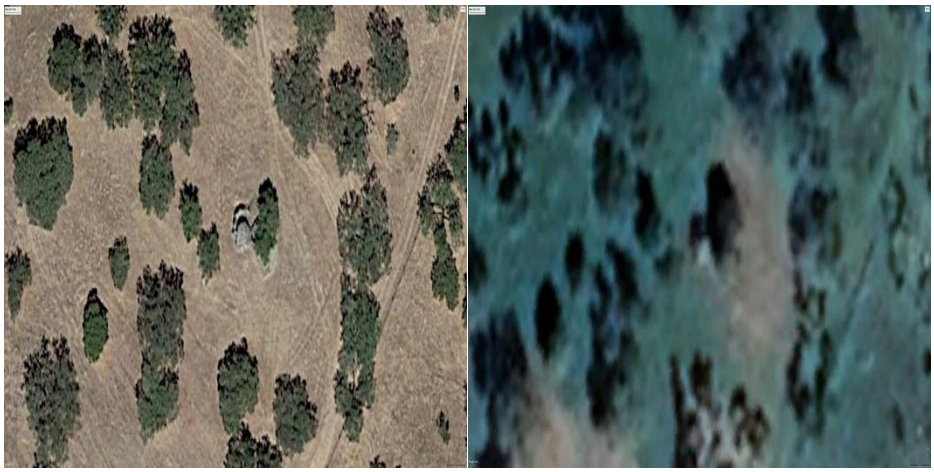


Figure 3 - Google Earth Images from 2017 (left) vs 2021 (right) for the same region.

3.3. Image Augmentation

Image augmentation, artificially increase the size and diversity of training datasets is used as a technique in computer vision that applies a variety of transformations to the original images. These transformations introduce variations in the data that serve to improve the generalization robustness and capacity of machine learning models towards new data. Image transformations employed, will be next described.

Rotation

Rotating an image by a certain angle can help create additional variations in the dataset. For example, rotating an image by 90 degrees, 180 degrees, or between those angles can be used to simulate different viewpoints [45].

Translation

Translating an image by shifting it horizontally or vertically can add diversity to the dataset. This transformation can simulate different object positions within the image [46].

Scaling

Scaling an image by changing its size and scale can account for variations in object sizes. It can simulate objects being closer or farther from the camera [47].

Flipping

Mirroring an image horizontally or vertically can create additional training samples that are horizontally or vertically flipped versions of the original images [48].

Noise Injection

Adding random noise to an image can enhance the model's robustness to improve generalization [46].

Image color augmentation is also a technique used to introduce variations in the color properties of images during the data augmentation process. It can be beneficial for training robust computer vision models that can handle variations in lighting conditions, color shifts, and other color-related factors. Some of the commonly used image color augmentation techniques and adjustments, will be described next.

Brightness

This technique involves changing the overall brightness of an image. By scaling the pixel values, you can make the image brighter or darker. It helps models become more robust to varying lighting conditions [49].

Contrast

Contrast adjustment alters the difference between the light and dark areas of an image. It can enhance or reduce the contrast to simulate different lighting environments and improve model generalization [50].

Saturation

Saturation refers to the intensity of colors in an image. Modifying the saturation level can make colors more vibrant or less saturated (depending on the range set), providing the model with variations in color appearance [49].

Hue

Hue adjustment involves shifting the colors in an image along color spectrums. It allows for modifications in the color tone, simulating different lighting conditions and different color variations [49].

Color Channel Shuffling

Color channel shuffling involves rearranging the color channels (e.g. RGB) of an image. This technique interferes with the original color distributions, promoting robustness to color channel variations [50].

Color Inversion

Color inversion flips the color values in an image, changing each pixel's hue and brightness values. It can be useful for training models to recognize objects in negative or inverted color representations [50].

Color augmentation techniques can be applied individually or in combination between themselves or even with other data augmentation methods. The choice of techniques always depending on the specific problem requirements of the detection task and the types of variations expected in the data.

3.4. Roboflow

Roboflow is a computer vision platform that provides tools and infrastructure to help developers and researchers create, train, and deploy computer vision models in a more efficient way. Offering features and services that simplify the process of creating, managing, and deploying computer vision models and algorithms. In order to take advantage from *Roboflow* there are some key aspects to be taken into account.

3.4.1. Data preparation and Annotation

It supports different data formats, from images to videos, additionally adding capabilities for data preprocessing, augmentation, and labeling. These tools help in preparing high-quality training datasets for computer vision models.

3.4.2. Data Versioning and Collaboration

Enables versioning and collaboration on datasets, allowing all team members to work on data and images annotation and management. With a version control features implemented, it is possible to track changes between different versions, and collaborate effectively on the dataset preparation process.

3.4.3. Model Training and Evaluation

By integrating with popular DL frameworks, such as TensorFlow and PyTorch, the model training is simplified on previous and existing versions of the annotated datasets. It provides infrastructure and automates the algorithm training, hyperparameter tuning, and model evaluation, making the training process more streamlined and efficient.

3.4.4. Model Deployment and Integration

Offering deployment options for object detection models, so the users can deploy the trained models.

It is also possible to integrate with popular deployment platforms and frameworks, simplifying the process of the model deployment into production environments.

3.4.5. Monitoring and Performance Tracking

It provides tools for monitoring and tracking model performance. After training it is possible to analyze the metrics and insights to obtain the performance of deployed models, including accuracy, precision, recall, and other relevant metrics. This facilitates users monitoring and improving the performance of their object detection models.

Overall, *Roboflow* aims to provide a comprehensive platform that covers all the stages of the object detection pipeline, from data preparation to model deployment. It aims to automate and accelerate the development and deployment of computer vision applications by simplifying complex tasks and offering infrastructure and tools to support the entire workflow.

3.5. Object Detection Algorithms

This sub-chapter serves as an introduction to the object detection algorithms discussed in this thesis, namely Faster R-CNN and YOLO. These algorithms represent two distinct approaches to object detection, each with its unique strengths and trade-offs. By providing an overview of these methods, it is possible to establish a foundation for the subsequent discussions on their architectures, training procedures, analyses, compare practical uses, and also setting the stage for the exploration of these algorithms, laying the groundwork for the subsequent chapters in this thesis.

3.5.1. Faster R-CNN

Faster R-CNN (Region-based Convolutional Neural Network) is a popular object detection framework that combines the advantages of DL and region-of-interest (RoI) proposal methods. It improves upon earlier region-based object detection methods, such as R-CNN and Fast R-CNN (Figure 4), by introducing a Region Proposal Network (RPN). The RPN shares convolutional layers with the detection network, enabling end-to-end training and significantly speeding up the process [51]. It generates region proposals by sliding a small network over the convolutional feature map. It predicts objects scores and refined bounding box coordinates for potential object locations called anchor boxes. These proposals are then passed to the subsequent detection network for object classification and bounding box regression [51]. A region of interest pooling layer is introduced to adaptively resize the fixed-size feature maps generated by the RPN to a fixed spatial extent. This allows the detection network to operate on region proposals of variable sizes and aspect ratios [52].

The training procedure of the Faster R-CNN consists of a two-stage process (see Figure 4). In the first stage, the RPN is trained to generate accurate region proposals. In the second stage, the region proposals are used to train a Fast R-CNN network, which performs object classification and bounding box regression.

Faster R-CNN has become a popular choice for object detection tasks due to its accuracy and real-time performance. It has been widely adopted in various applications, including autonomous driving, surveillance systems, and image analysis [51].

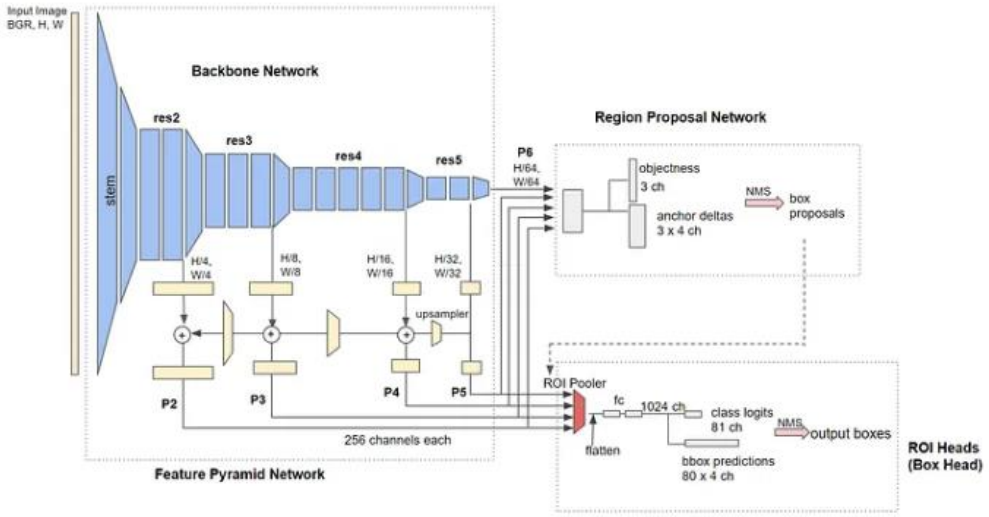


Figure 4 - Structure of a FastRCNN algorithm (Source: [35]).

The authors of [35] introduced Faster R-CNN and the bottleneck of the architecture, stating that Fast R-CNN is a selective search. This search accounts for a significant portion of the architecture's training time. It was replaced by the region proposal network in Faster R-CNN [53]. First and foremost, it passes the image into the backbone network in this network (Figure 5). A convolution feature map is generated by this backbone network. These feature maps are then fed into the network of region proposals. The region proposal network generates the anchors from a feature map. These anchors are then passed into the classification layer (which classifies whether there is an object or not) and the regression layer (which localizes the bounding box associated with an object) [54].

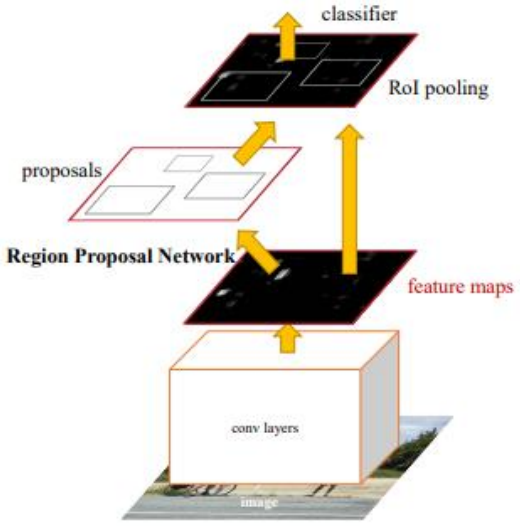


Figure 5 - Simple Structure of a FasterRCNN algorithm (Source: [35]).

3.5.2. YOLO

YOLO (You Only Look Once) is an object detection algorithm that achieves real-time object detection in images and videos with impressive accuracy [55]. YOLO approaches object detection as a regression problem, directly predicting bounding box coordinates and class probabilities in a single pass through the network. Unlike traditional object detection methods that involve multiple stages, YOLO provides a unified architecture that simultaneously predicts object locations and class labels [55].

The input preparation (image) is resized to a fixed size that the YOLO network expects. It is divided into a grid of cells, typically, with a size of, for example, 13x13 or 19x19. This technique employs a grid-Based approach, where the input image is divided into a grid and each grid cell is responsible for detecting objects. For each grid cell, YOLO predicts a fixed number of bounding boxes with associated confidence scores and class probabilities. The predictions are made based on the features extracted from the entire image using a convolutional neural network (CNN) that also predicts a fixed number of bounding boxes.

These bounding boxes are defined by their coordinates relative to the grid cell. For each bounding box, YOLO predicts the probability of containing an object and the associated class probabilities for different object categories [56].

In a next level, it uses anchor boxes of different shapes and sizes within each grid cell to improve the detection of objects with various aspect ratios and scales. These anchor boxes serve as reference templates and are used to predict the offsets for bounding box regression. In the end, YOLO assigns class probabilities to each bounding box, indicating the likelihood of containing different object categories. It employs softmax activation to produce class probabilities across all possible classes.

The network architecture employs a custom CNN architecture called Darknet, which was specifically designed to optimize the trade-off between accuracy and speed in real-time object detection.

This approach has been widely adopted for various applications, including object detection in images and videos, robotics, autonomous vehicles, and surveillance systems [55], due to its ability to provide real-time object detection on resource-constrained devices. Since the original YOLO release, several improved versions, from YOLOv2 to YOLOv8 (Figure 6), have been introduced, each bringing enhancements in terms of accuracy and speed [57]. Yolov8 medium was further used in this work.

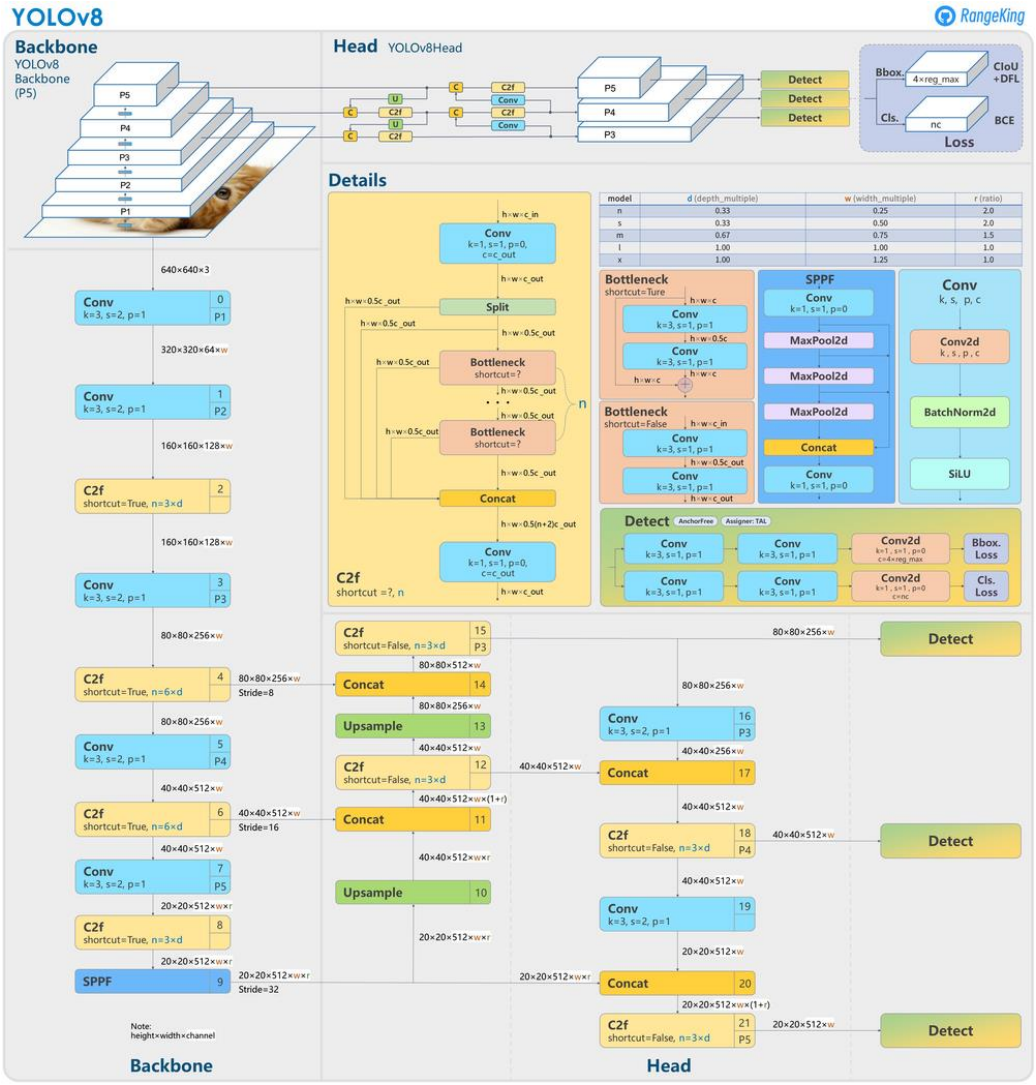


Figure 6 – Structure of YOLOV8 algorithm (Source: [58]).

The key advantages of YOLO are its efficiency and real-time performance. By eliminating the need for region proposal networks and performing detection in one pass, YOLO achieves impressive speed while maintaining competitive accuracy [56].

3.6. Algorithm Evaluation

When evaluating an object detection model, several metrics are commonly used to assess its performance [59]. The choice of metrics depends on the specific requirements and goals of the application. Some important metrics for evaluating object detection models [60] are: True Positive (TP) count, that is, counting the number of times that the model correctly identifies and locates an object within the image. The model's prediction of the object's presence is accurate and matches the actual object's location. The False Positive (FP) count, when the model incorrectly detects an object that isn't present in the image. It's a case of the model producing a positive detection when it shouldn't. The False Negative (FN) count of when the

model fails to detect an object that is present in the image. It's a case of the model missing a positive detection that should have been made. These counts are also used to compute the Average Precision (AP), a widely used metric for object detection evaluation. It calculates the precision-recall curve by varying the confidence threshold for object detection. AP summarizes the overall performance of the model by considering the area under the precision-recall curve. Other metrics are: Mean Average Precision (mAP) calculates the average AP across multiple object categories. The Intersection over Union (IoU) is the one that measures the overlap between the predicted and ground truth bounding boxes. It is defined as the ratio of the intersection area to the union area of the two boxes. A higher IoU indicates better object localization accuracy. Commonly used IoU thresholds include 0.5 (IoU=0.5), 0.75 (IoU=0.75), and 0.5 to 0.95 with a step of 0.05 (IoU=0.5:0.05:0.95) to evaluate different levels of object localization precision. On its own, Precision defines the ratio of the true positive detections to the total number of positive detections (both true positives and false positives), while Recall, known as the sensitivity or true positive rate, it measures the ratio of true positive detections to the total number of ground truth objects.

These metrics help assess the model's ability to correctly detect objects while minimizing the false positives, and F1-score is the mean of precision and recall, harmonized. It is a metric that provides a measure that balances both precision and recall. F1 score is useful when there is an imbalance between the number of positive and negative examples.In DL, box loss, classification loss, and dynamic feature learning loss are terms associated with object detection models, specifically those utilizing anchor-based methods. These losses are used to train the model and improve its accuracy in detecting objects within an image [61]. Box Loss or regression loss is applied in object detection where each object is typically represented by a bounding box, which consists of four values representing the coordinates of the box's top-left and bottom-right corners. It measures the discrepancy between the predicted box coordinates and the ground-truth box coordinates.

Common regression loss functions include smooth L1 loss and mean squared error (MSE). Classification Loss, along with predicting bounding box coordinates, is used in models that also need to classify the objects present in the image. It calculates the difference between the predicted class probabilities and the true class labels. Common loss functions for classification include softmax cross-entropy and binary cross-entropy. Dynamic Feature Learning Loss (DFL Loss), a component specific to the Region-based Convolutional Neural Network (Cascade R-CNN) framework, which is used for accurate object detection. It is applied after the initial detection stage and aims to refine the bounding box predictions. The DFL loss measures the discrepancy between the refined bounding box coordinates and the ground-truth box coordinates. This loss helps in iteratively improving the accuracy of object localization.

These loss functions are typically combined to form a multi-task loss (total loss), which is then minimized during the training process using techniques like backpropagation and gradient descent. The relative importance of each loss component can be adjusted through hyperparameters to achieve the desired balance between localization accuracy and classification performance in the object detection model [61].

CHAPTER 4

Implementations and Results

This chapter describes the implementation of the previous pipeline, step by step. First, we start by introducing the geographical area of the case study to give context to all the implementations and results obtained. At the end of the chapter, we find a discussion of all the results obtained after this implementation.

Following the design research methodology, we terminate its third step where implementations of the artifact are made. Next, we step into an evaluation that intends to assess the artifact's effectiveness, efficiency, and utility.

4.1. Case Study's geographical area

This study targets the region of Mora and Arraiolos, Alentejo, Portugal, because it is the one for which we have both expert knowledge and data to evaluate our proposal's success. The geographical area covers 185 km2 and its situated in the southern half of the country, in central Alentejo. According to [4], it is primarily composed of alkaline granites, granodiorites, tonalites, and trondhjemite. Other authors state that this territory contains the most extensive plateaus of Portugal, with local topography curves averaging an altitude of around 200m and little variation in declivity or relief, also containing three major hydrographic basins, one being in the Mora region.

4.2. Images and Data

After acquiring all the dolmens' locations, the next step was to get their satellite images. Using *Google Earth Pro* we got around 60 dolmens visible - from a software perspective - and took five images of each dolmen, making sure that the monument's position varies from image to image and therefore the background changes too as seen in Figure 7. In total, we collected a dataset of around 300 images. In order to test the algorithm on an image that doesn't contain any dolmen, two more images were collected. These images are similar in terms of background but, as far as we know, are void of monuments.

Figure 7 - Five images used of the same dolmen in different backgrounds.

Images saved from *Google Earth Pro* are in pan-sharpened (or simply, pan) format. Pan-sharpening is the process of combining a higher-resolution panchromatic (black and white) image with a lower-resolution multispectral (color) image to create a single high-resolution color image as depicted in Figure 8. This process improves the image's visual quality and details.

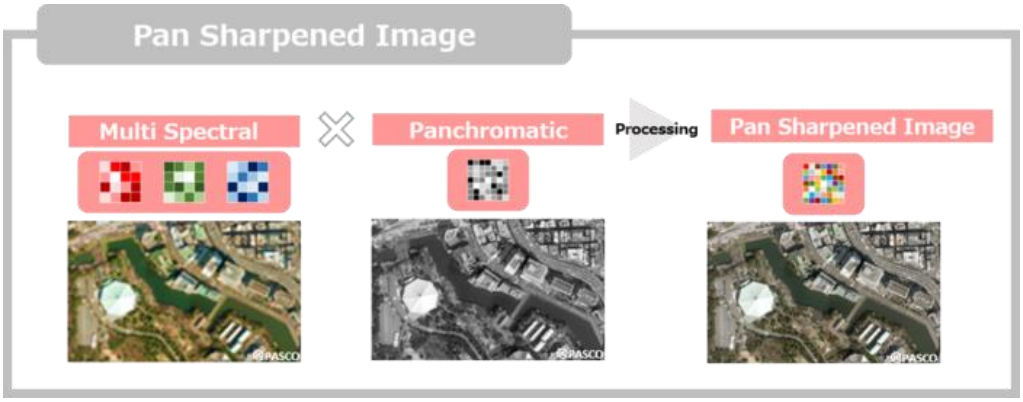


Figure 8 - Process of a Pan Sharpened Image (Source: [62]).

Google Earth Pro saved images are typically in pan-sharpened format, which provides a more detailed and visually appealing representation of the Earth's surface.

To train and test the algorithms the data had to be split. The test images set contains 17 images: 15 of only three distinct dolmens (*São Pedro da Gafanhoeira 1*, *Telhal 1*, *Anta de Prates 7*) that contribute with five images each. Moreover, we added the two previously mentioned images that are void of monuments. The remaining 285 images were used to train the algorithms. The following image in Figure 9 explains how the data was split.

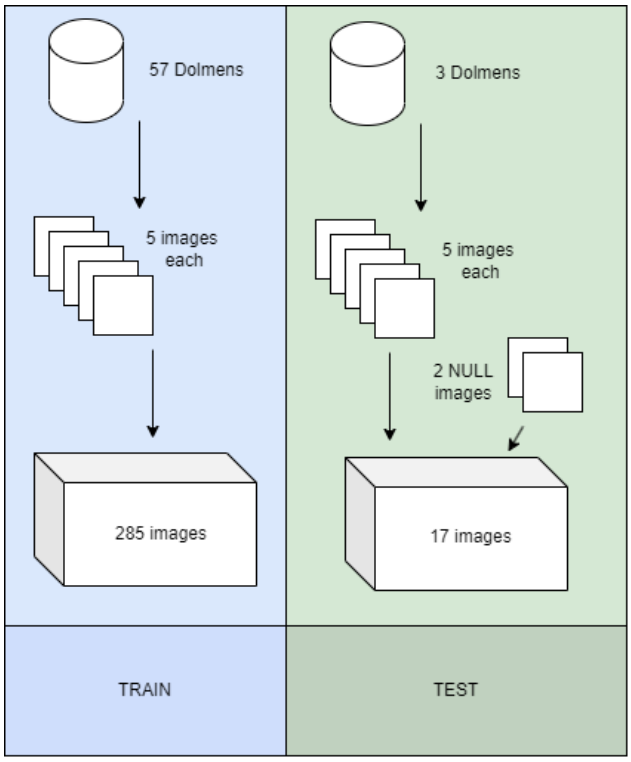


Figure 9 - Train and test data split of the images.

4.3. Image Pre-Processing

Before training the model, images had to be pre-processed using some of the image enhancement and augmentation techniques already referred to in previous chapters.

We begin by describing the enhancement techniques that have been used and afterwards we describe the augmentations that we have performed.

4.3.1. Image Enhancement

We decided to only use one type of image enhancement: adjustment of the image's contrast. *Roboflow* displays two pre-processing tools that use histogram and adaptive histogram equalizations of the contrast. It consists in an algorithm that uses histograms computed over different tiles of the image so that local details can be enhanced, even in regions that are darker or lighter than most of the image, like we can see in Figure 10. These features help particularly to enhance images where the vegetation index is higher.

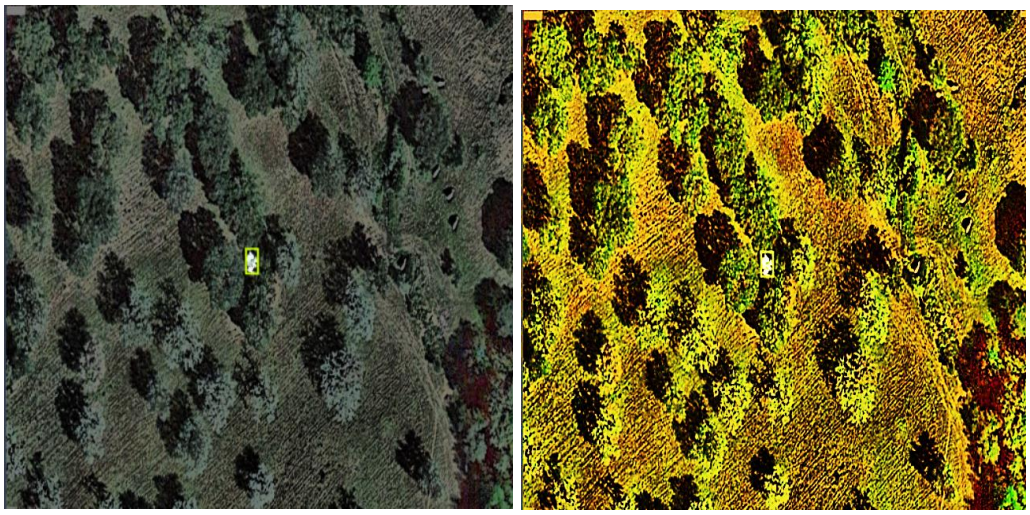


Figure 10 - Image before (on the left) and after enhancement (on the right).

4.3.2. Image Augmentation

As previously stated, augmentation is a crucial step before training any algorithm. After reviewing the state of the art, we could conclude that the best way to apply augmentation is to add random values of different types of augmentation until finding the best set of modifications in order to obtain better results.

Cropping and Rotation were two of the non-color-based augmentation used, where cropping could vary from 0% to 50% maximum zoom and rotation from -45º to 45º.

The color-based augmentations added were hue, saturation, brightness, and exposure, all in a range -50% to 50%, with a random spectrum of 100 points for each augmentation, where images could vary between those ranges.

Figure 11 shows two examples of finished preprocessing images after suffering both enhancement and augmentation.

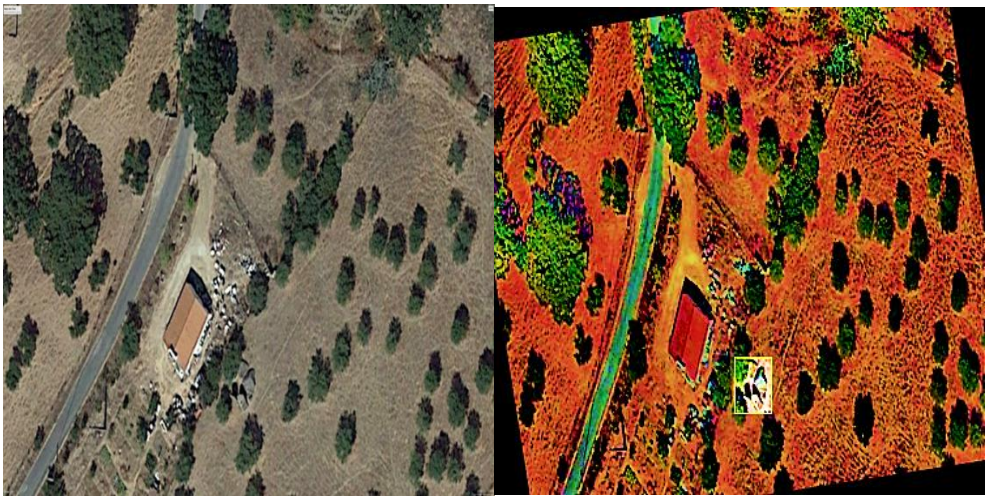


Figure 11 - Examples of image used in training after pre-processing: original (on the left) and pre-processed (on the right).

4.4. Algorithm Implementation

After pre-processing the images was time to develop and train the algorithms. As previously discussed in Chapter 2, the algorithms used were YOLO in its newest version (version 8) and FastRCNN using nine different architectures.

After augmentation, 855 images were used for training, where the 285 initial training images were multiplied by 3 using different augmentation techniques (crop and rotation).

Auto-tunning was used with both algorithms to obtain the best parameters and hyperparameters possible for the data used in training.

Using *Google Colab,* an algorithm takes approximately 40 minutes to train 5000 iterations to get the minimum total loss possible and to get closer to the established learning rate for each algorithm, overall consuming 10 GB of RAM and 8 GB of GPU memory.

Regarding the implementation of Faster-RCNN models, two types of backbone networks were used - ResNet-101 and ResNet-50 - along with three different network structures - Feature Pyramid Network (FPN), Dilated Convolutional Network (DCN), and Convolutional Network (CN). Additionally, two types of training schedules were used (1x and 3x) in the architectures.

4.5. Results

After all the iterations and training processes were concluded, the models were tested using the previously described test set. Table 2 presents the average precision and F1-score metrics that have been obtained for each of the trained models for the test set.

Table 2 – Chosen models test results (with the Top-3 best results highlighted).

Model	Average Precision	F1-Score
R_50_FPN_3x	0.67	0.51
R_50_FPN_1x	0.69	0.64
R_50_DC_3x	**0.70**	**0.65**
R_50_DC_1x	**0.93**	**0.78**
R_50_C_3x	0.61	0.57
R_50_C_1x	0.60	0.57
R_101_FPN_3x	0.71	0.64
R_101_DC_3x	0.74	0.71
R_101_C_3x	0.63	0.59
YOLOV8	**0.79**	**0.71**

As previously mentioned, the test set consisted in images from three different dolmens, São *Pedro da Gafanhoeira 1*, *Telhal 1*, and *Anta de Prates 7*, shown the images in Figure 12, respectively, where the dolmen can be seen in the middle of each image.

Figure 12 - Test Dolmens Images (with the dolmen in the center of each one).

For each set of images next presented (Figure 13-15), we show the models' results, with the "R_101_DC_3x" detection at the left (a), the "YOLOV8" detection in the middle (b), and the "R_50_DC_1x" result at right (c), since these are the three models that obtained a better average precision and F1-score among all the models tested. The detection results are presented next.

(a) (b) (c)

Figure 13 - Object detection results for São Pedro da Gafanhoeira 1.

In Figure 13 the "R_101_DC_3x" model, image (a), has a TP detection with 96% confidence degree but, however, has four FP detections whose confidence ranges range from 71% to 88%. The YOLOv8 model (b) presents a TP of 80% confidence degree and two FP detections with 64% and 53% each. The "R_50_DC_1x" model (c) detected a TP with 90% confidence degree and one FP with 85%.

(a) (b) (c)

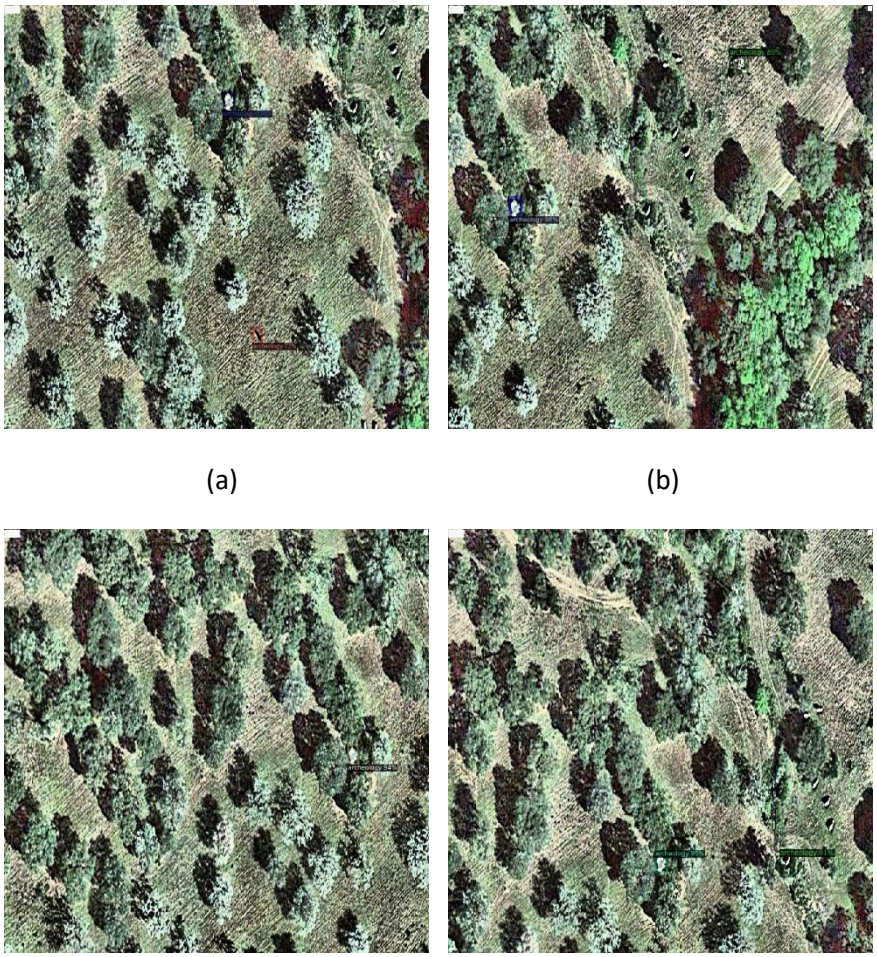
Figure 14 - Object detection results for Telhal 1.

In Figure 14 model “R_101_DC_3x” (a) and “R_50_DC_1x” (b) both detected a TP with 99% confidence degree while YOLOv8 (c) detected a TP of 71% confidence degree.

(a) (b) (c)

Figure 15 - Object detection results for Anta de Prates 7.

The Figure 15 shows that the model “R_101_DC_3x” (a) detected a TP with 95% confidence degree and a FP with 97%. The YOLOv8 model detected only a TP with 61% confidence degree. Model “R_101_DC_1x” (a) detected a TP with 91 % confidence degree and two false positives, one with 95% (same FP that (a) detected) and 76% confidence degree.



(a) (b)

(c) (d)

Figure 16 - Results from model R_50_DC_1x São Pedro da Gafanhoeira 1 (4 test images)

The image (a) on the top left has a TP with 96% and one FP with 77% confidence degree. The image (b) contains a TP with 96% confidence degree and one FP of 80%. Image (c) only has a TP of 94% and (d) has a 95% TP and a FP of 81% confidence degree.

It is also important to refer that the model R_50_DC_1x did not detected anything in the images that had no dolmens in it, which is a good indicator, and the best result possible, that being because any detection would be a FP detection. The other models, three had FP with an accuracy between 34% and 68%.

The remaining results from all the models can be visualized in Anex A.

4.6. Discussion of Results

The results of the object detection experiments reveal an intriguing trade-off between both Faster R-CNN and YOLO models. The Faster R-CNN models demonstrate a higher overall accuracy in detecting objects, indicating their proficiency in precise localization. However, the YOLO models exhibit a noteworthy characteristic: despite their lower accuracy, they present fewer FP detections.This suggests that the YOLO model excel in minimizing the instances where non-existent objects are mistakenly identified.

This trade-off has practical implications. While Faster R-CNN models are adept at precise object localization, the YOLO models may offer advantages in scenarios where minimizing FP is crucial. Although the number of FP is higher in FasterRCNN models, it's not a big number per image detected, and some are with a much lower level of confidence degree than the TP, indicating that we could reduce the number of final FP, considering a desired or suggested level of confidence degree.

The object detection results regard true positive detections across varying confidence levels. It becomes apparent that certain detections exhibit discrepancies in confidence scores, likely attributed to differences in background contexts. This phenomenon highlights the sensitivity of the detection models to environmental factors. Objects situated against distinct backgrounds may receive varying confidence ratings, underscoring the influence of contextual information on the model's detection process. This nuanced behavior implies the importance of considering background diversity in training datasets to enhance the robustness and adaptability of object detection models across diverse real-world scenarios. These findings emphasize the need for a comprehensive evaluation of detection performance that considers the characteristics of the surrounding environment.

FasterRCNN using a ResNet-50 backbone network and Dilated Convolutional Network for structure and a standard training schedule proved to obtain the best performance values and results for this test set, therefore it was chosen to be employed the in the final pipeline architecture.

During the training of this algorithm, it was possible to track a few metrics to detect any sign of over and underfitting during the epochs defined. The losses and accuracy metrics throughout the training can be analyzed from the graphs that can be found in Figure 16-21. The axis of these graphs correspond to the metric percentage versus the number of epochs (5000).

Since classification loss represents the difference between predicted class probabilities and the true class labels, ideally during training we want to minimize this metric along the epochs.

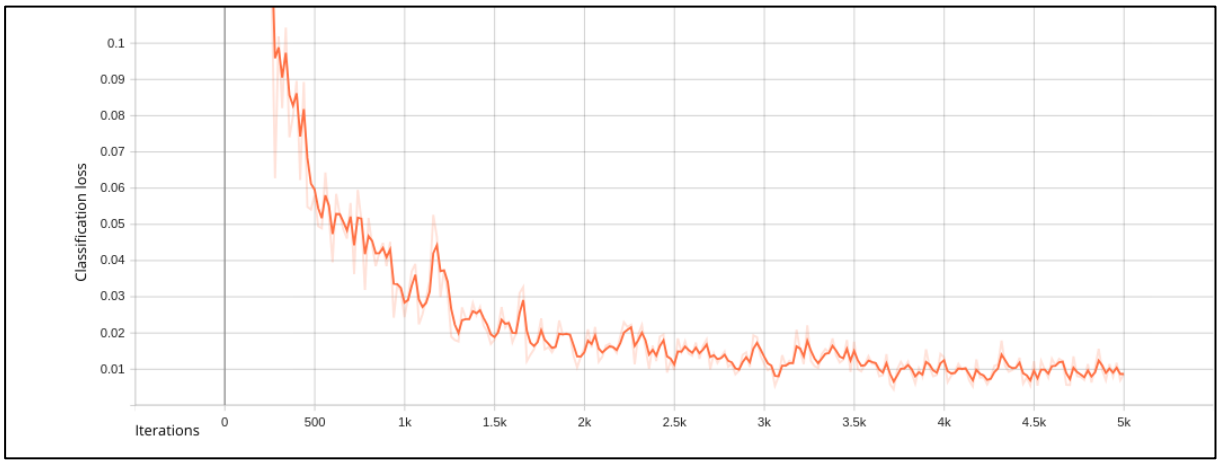


Figure 17 - Classification loss graph.

Clearly, *Figure 17* shows the classification loss decrease that appears to become more stable after 4000 epochs, converging towards an optimal solution. This suggests that the training process has effectively learned the underlying patterns in the data, enabling the model to make accurate predictions. The stabilization of the loss indicates a reduced sensitivity to minor fluctuations in the training data, which is indicative of a well-generalized model [63].

Box Loss measures the discrepancy between the predicted box coordinates and the ground-truth box coordinates, minimizing this value throughout the epochs is ideal for training [63].

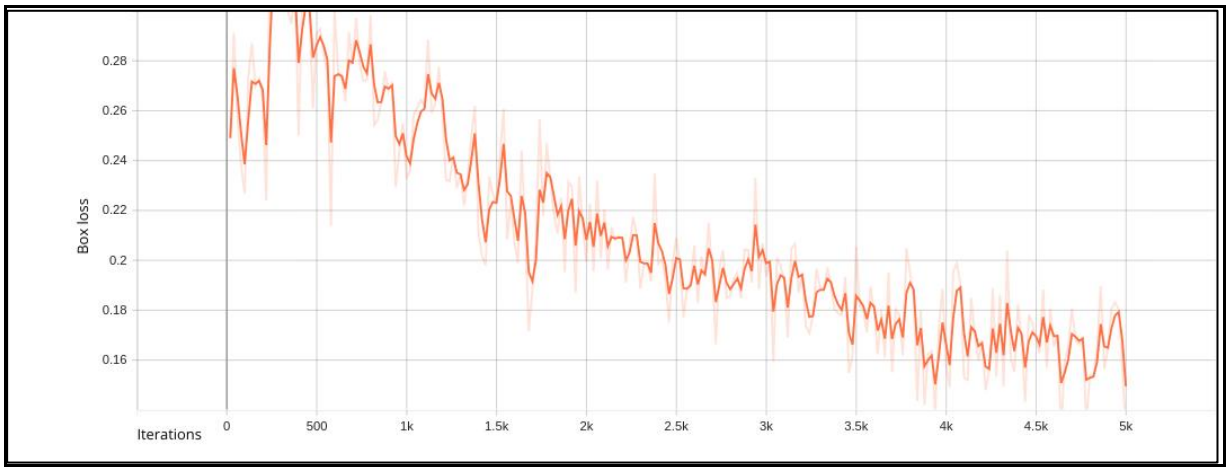


Figure 18 - Box Loss graph.

The box loss graph in *Figure 18* exhibits a generally favorable trend of minimization over the course of 5000 epochs. This indicates that the optimization process is effectively converging towards the desired objective. However, the presence of occasional upward and downward spikes suggests that there might be certain points in the training process where the model encounters local optima or fluctuations in the loss landscape. These spikes could be attributed to various factors, such as noisy or inconsistent data, learning rate adjustments, or the model's sensitivity to specific input patterns. It's crucial to monitor and investigate these spikes further, as they could potentially lead to suboptimal performance or indicate areas for improvement in the training process [64]. Overall, while the trend is positive, attention should be given to these spikes to ensure the model's robustness and stability.

Minimizing the Classification loss in the Region Proposal Network, the RPN learns to predict high objects scores for anchors that overlap significantly with ground-truth object bounding boxes and low scores for anchors that are far from any object. Ideally, the algorithm must focus on relevant regions likely to contain objects and ignore the regions likely to be background or irrelevant [63].

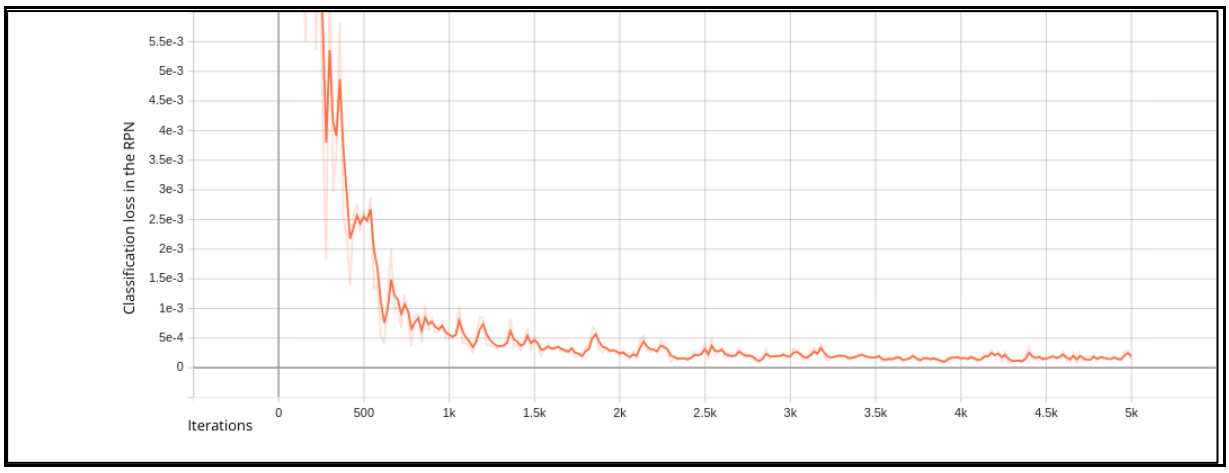


Figure 19 - Classification loss in the Region Proposal network graph.

Observing *Figure 19*, we can see that after around 2000 epochs the model starts to stabilize until the end of the training, suggesting that the model has reached a point of diminishing returns in terms of classification improvement. This stabilization is a positive sign, indicating that the RPN has likely converged to a satisfactory level of classification accuracy. However, it's important to note that further training beyond this point may not yield significant additional benefits [63] [64].

In *Figure 20*, another loss analyzed was the localization loss in the RPN where the minimization is essential because it ensures that the algorithm learns to accurately predict the correct bounding box coordinates for the positive region proposals. Better training results on the regression of the predicted bounding box coordinates for each positive anchor to align it with the ground-truth bounding box [63].

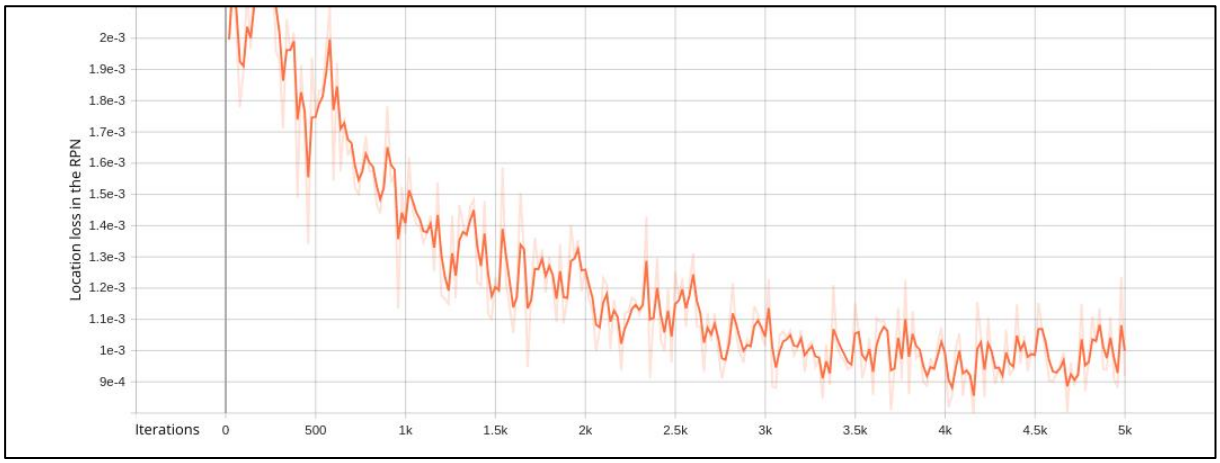


Figure 20 - Location loss in the Region Proposal Network graph.

However, the presence of numerous spikes suggests that there are instances where the model encounters challenges in precisely pinpointing object locations. These spikes may be attributed to various factors, such as complex object geometries, augmentations, or the image background. Despite the spikes, the overall decreasing trend indicates that the RPN is on the right path towards achieving accurate object localization [64].

The total loss (**Erro! A origem da referência não foi encontrada.**) can be calculated by summing the location and classification losses. During training, the algorithm aims to minimize the total loss, which means it seeks to minimize both the classification loss and the localization loss simultaneously [64].

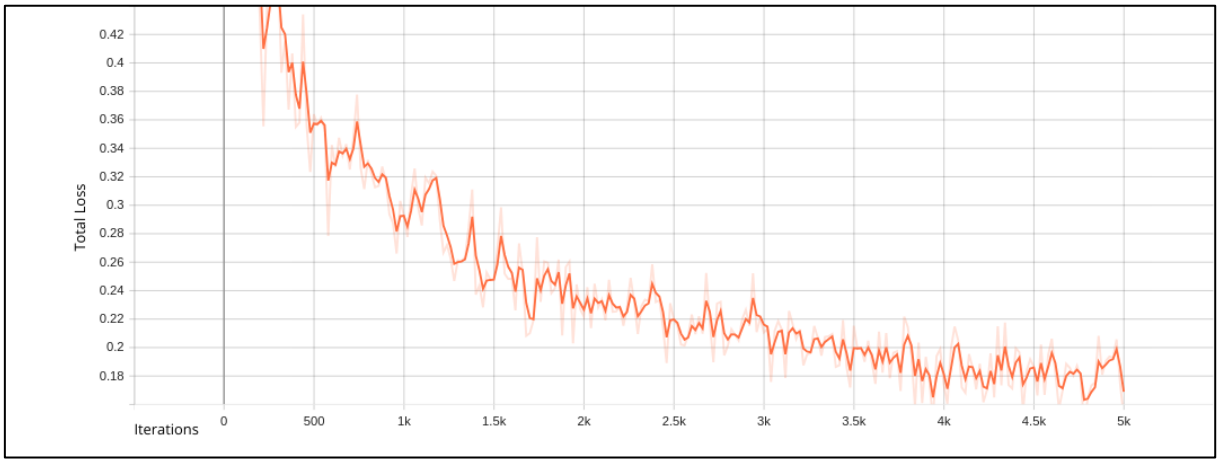


Figure 21 – Total Loss

The graph demonstrates a favorable trend of consistent reduction, signifying that the model is effectively converging towards the desired objective. This suggests that both the classification and localization components of the Region Proposal Network (RPN) are jointly improving in performance. The diminishing total loss indicates that the model is successfully learning to classify objects and refine their precise locations [63] [64].

For this algorithm, the classification accuracy (*Figure 22*) was also measured by epoch, where during training, the algorithm aims to adjust the object detections in order to classify with better precision the objects. The higher accuracy, the more the algorithm is adjusting to the training set [65].

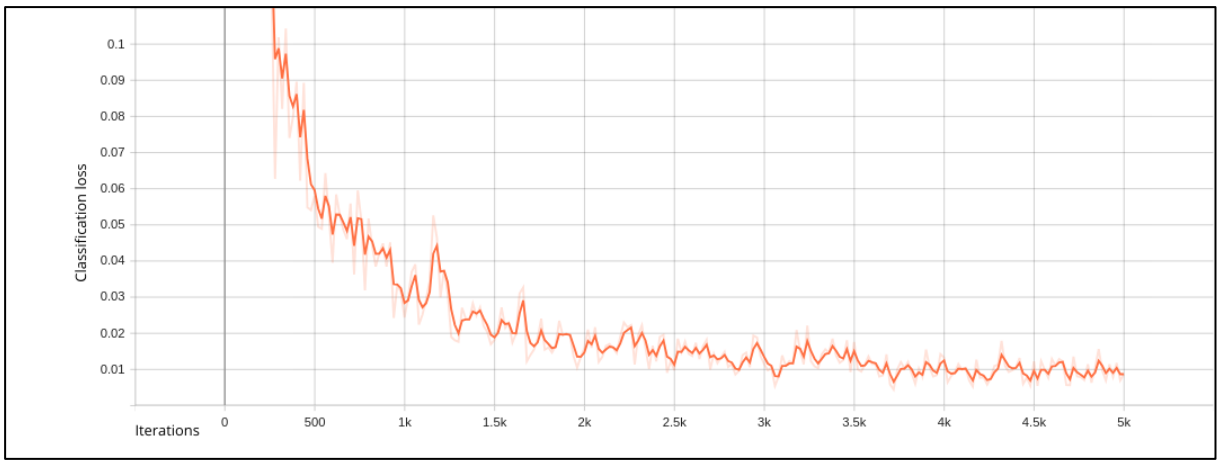


Figure 22 - Classification Accuracy graph.

The training classification accuracy graph depicts a highly positive trend, as it consistently approaches a value of 1. This indicates that the model's ability to correctly classify objects in the training set is steadily improving. The increasing accuracy signifies that the model is effectively learning the underlying patterns and features associated with the detection. This is a strong indication that the training process is successful and that the model is becoming increasingly proficient in its classification task. It's important to continue monitoring this metric to ensure that the model doesn't plateau prematurely, and to validate its performance on unseen data [65].

The successful training of the Faster R-CNN algorithm with minimal losses and good accuracy demonstrates proficiency in implementing object detection models. The faster R-CNN model is ready to be deployed to perform object detection in similar applications. Depending on the specific use case, fine-tuning the model on domain-specific data can be applied to further improve its performance in a specific context [64].

However, we stress that, despite achieving a lower confidence rate in true positive results, YOLOV8 stands out as the model with the fewest false positive detections. In terms of our overall objective of building a useful tool to help archeologists in their prospection work, a lower number of false positives is an important advantage.

Using only these images to test, it is not possible to fully demonstrate that there is a best model for future training sets because the results may vary from set to set and other algorithms could perform better in those other tests.

Finally, when compared with the pixel-based study previously performed in the context of the same case study ([14]), our overall accuracy results show an improvement although slight. Nevertheless, our approach was tested using a different set of images and trained in different samples.

CHAPTER 5

Conclusions

Entering now in the fifth step of the research method, it is time to reflect on the evaluation results and analyze the findings and potential improvements.

This dissertation investigated the possibility of developing an automatic object detection system that uses satellite imagery to detect Portuguese dolmens and related archeological monuments. For this investigation, Alentejo satellite images were used, namely in the areas of Mora and Pavia. Ten different algorithms were trained and tested in order to get the best model, that is, the one to detect with high accuracy and high confidence levels new objects in a set of dolmens previously identified by experts. A pipeline was built, from the data gathering to the final algorithm results, being open for further training and tunning.

This conclusion chapter is divided in two sections. main conclusions, where the objectives and research questions are addressed, and proposals for future work.

5.1. Main Conclusions

Looking back at the objectives and research questions set for this thesis, for the first question we can conclude that the augmentation and enhancement of the images is a crucial step to be taken for the algorithm training. Using the right amounts and techniques we could verify that it plays a significant role in improving the performance and robustness of the models. Techniques, such as rotation, flipping, scaling, and translation, help create variations of the original images, effectively increasing the dataset's size and diversity. These augmented images exposed the model to different object appearances, orientations, and scales, making it more adaptable to various situations. With these, the object detection model became more robust and less sensitive to small changes in the inputs, leading to better generalization of unseen data. From the training results, we could conclude that the augmentation acts as a form of regularization, preventing the model from memorizing specific instances in the training data. It helped the model to focus on learning important features and relationships between objects and backgrounds rather than memorizing individual training samples.

Regarding the second research question, it is not yet possible to select the best algorithm for object detection since it may vary with different test sets and new geographical locations. Although, for this investigation, the FasterRCNN using a ResNet-50 backbone network and Dilated Convolutional Network for structure, trained with a standard training schedule proved to show the best evaluation metric values for the test set used, the YOLO architecture showed the lower number of FP. Nevertheless, therefore FasterRCNN is our choice to be the in the final pipeline architecture for the current case study. In fact, false positive detection is a major issue in this area since in satellite imagery a lot of rock outcrops can be miss perceived as dolmens. YOLOV8 was the best model at avoiding false positives, despite the lower confidence rate on true positive detection. Now it is up to the archeologists and experts to choose what they feel better fits their job, balancing the pros and cons of each of the models. Despite having a few false positive detections, the FasterRCNN proves to be reliable, that being because the false positive detections have a lower confidence rate detection and accuracy compared to the true positive detections.

As usual, there were limitations for the work here presented. One such limitation is the small dataset used, despite the techniques used to multiply the images on the dataset resulting in a total of 855 images, for DL model training this is deemed too small a dataset. In the case of FasterRCNN, it is recommended to use around 5000 samples [35] and, in the case of YOLO, the recommendation is even higher: 10,000 instances per class [66].

5.2. Future Work and Research

Megalithic monuments exhibit different characteristics that preclude the application of automated approaches to new geographical areas. Through time, this topic is being further addressed and improved and new findings are always appearing. From our revision of the literature, we noted that such systems have mostly been developed and applied in foreign countries, leaving Portuguese regions and monuments with a significant lack of work and applications. Portuguese monumental heritage deserves further attention and investigation, especially because it presents unique characteristics worth been registered and studied.

Despite the good results here presented, it would be interesting to expand the system by adding new domain knowledge based on the terrain characteristics and environment capable of further refining areas presenting higher probability of monument’s presence.

Regarding the third research question and the last objective, that was to try to use the terrain information for further analysis of the object detection results or to complement the algorithms with the given information about the terrain to build up the confidence in the detection. Experts discovered that dolmens in Alentejo are usually located within a hundred meters up to one kilometer from lines of water, and less than one kilometer from rock outcrops. With this information is possible to create a sort of a hot zone area where dolmens are more likely located. We achieved to created buffers in order to understand where dolmens are more likely to be as presented in Figure 23.

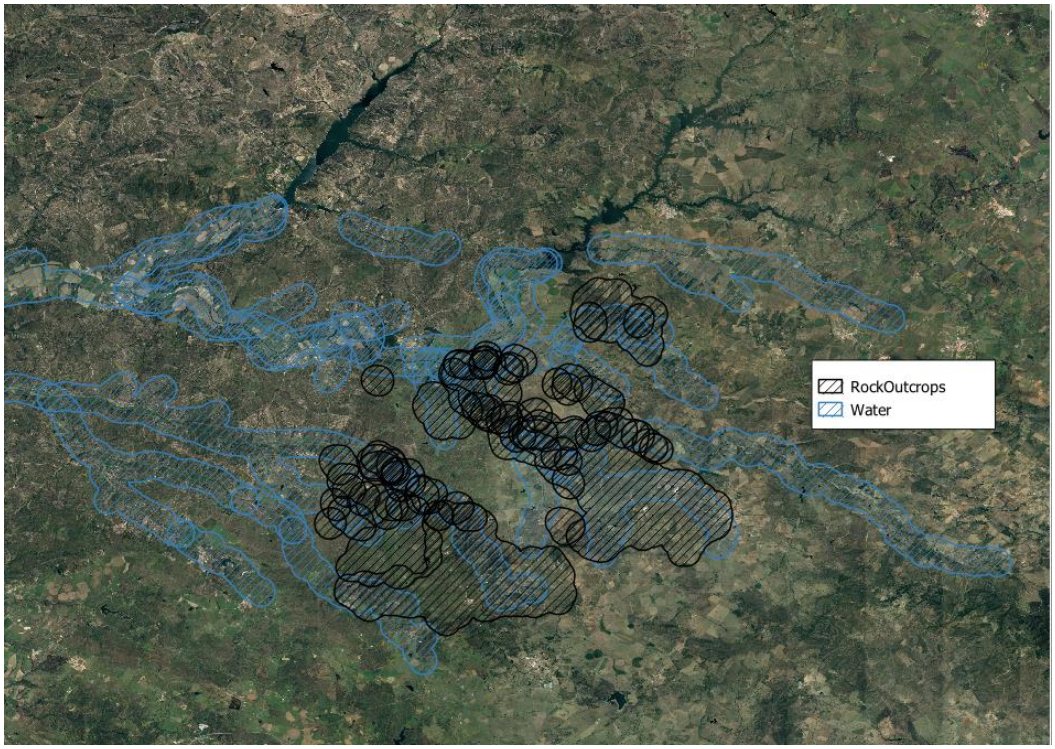


Figure 23 - Mora and Arraiolos regions with Rock outcrops and water buffered layers.

Future work on how to best hybridize this information with the models is expected to achieve better results, by filtering the information and being able to recalculate the accuracy based on location too.

References

[1] A. H. Jamil, F. Yakub, A. Azizan, S. A. Roslan, S. A. Zaki and S. A. Ahmad, "A Review on Deep Learning Application for Detection of Archaeological Structures," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 26, no. 1, pp. 7-15, 2022.

[2] M. Küçükdemirci, G. Landeschi, N. Dell'Unto and M. Ohlsson, "Mapping Archeological Signs from Airborne Lidar Data Using Deep Neural Networks: Primary Results," ArchéoSciences, no. 45, p. 291–293, 2021.

[3] A. Câmara, "A fotointerpretação como recurso de prospeção arqueológica. Chaves para a identificação e interpretação de monumentos megalíticos no Alentejo: aplicação nos concelhos de Mora e Arraiolos", 2017.

[4] A. Câmara, "Photointerpretation as a Tool to Support the Creation of an Ontology for Dolmens" Meeting of the Portuguese Association for Classification and Data Analysis, October 2020.

[5] L. R. a. T. B. A. Câmara, "Fotointerpretação e Sistemas de Informação Geográfica: Contributo para a Identificação de Dólmens em Portugal: O Caso de Mora e Arraiolos," Iber. Conf. Inf. Syst. Technol. Cist., 2017.

[6] Ø. Due, "Detection of cultural heritage in airborne laser scanning data using Faster R-CNN," TRIER, November 2019.

[7] L. H.-M. a. T. L. A. Guyot, "Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques," Remote Sensing, vol. 10, no. 2, p. 225, February 2018.

[8] Ø. D. Trier, J. H. Reksten and K. Løseth, "Automated mapping of cultural heritage in Norway from airborne lidar data using faster R-CNN," International Journal of Applied Earth Observation and Geoinformation, vol. 95, March 2021.

[9] W. B. V.-v. d. Vaart and K. Lambers, "Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands," Journal of Computer Applications in Archaeology, vol. 2, pp. 31-40, 2019.

[10] D. S. Davis, "Using deep learning to detect rare archaeological features: A case from coastal South Carolina, USA," Society of Exploration Geophysicists, 2021.

[11] J. Sales and J. M. Junior, "Retinanet Deep Learning-Based Approach to Detect Termite Mounds in Eucalyptus Forests," IEEE International Geoscience and Remote Sensing Symposium , July 2021.

[12] D. Davis, M. Sanger and C. Lipo, "Automated Mound Detection Using Lidar and Object-Based Image Analysis in Beaufort County, South Carolina," SSRN Electronic Journal, 2018.

[13] Berganzo-Besga Hector A. Orengo, A. Khan, M. Suárez-Moreno, J. Tomaney, R. C. Roberts and C. A. Petrie, "Curriculum learning-based strategy for low-density archaeological mound detection from historical maps in India and Pakistan," Scientific Reports, vol. 13, no. 1, 2023.

[14] D. Caçador, Automatic Recognition of Megalithic Objects in Areas of Interest in Satellite Imagery, October 2020.

[15] E. Cerrillo-Cuenca and P. Bueno-Ramírez, "Counting with the invisible record? The role of LiDAR in the interpretation of megalithic landscapes in south-western Iberia (Extremadura, Alentejo and Beira Baixa)," vol. 26, no. 3, pp. 251-264, May 2019.

[16] D. Canedo, J. Fonte, L. G. Seco, M. Vázquez, R. Dias and T. D. Pereiro, "Uncovering Archaeological Sites in Airborne LiDAR Data with Data-Centric Artificial Intelligence," IEEE Access, vol. 11, 2023.

[17] "ISCTE DetectionArq Dataset," Roboflow Universe, [Online]. Available: https://universe.roboflow.com/iscte/detectionarq/dataset/13. [Accessed 12 June 2023].

[18] S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact," MIS Quarterly, vol. 37, no. 2, pp. 337-355, 2013.

[19] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, vol. 5, no. 2, pp. 199-220, June 1993.

[20] K. Kintigh, "Grand Challenges for Archaeology," Cambridge University, vol. 79, no. 1, january 2017.

[21] J. Merchant, "Remote Sensing of the Environment: An Earth Resource Perspective," Cartography and Geographic Information Science, vol. 7, October 2000.

[22] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," International Journal of Remote Sensing, vol. 28, no. 5, p. 823–870, March 2007.

[23] G. Chen, "Object-based change detection," International Journal of Remote Sensing, August 2011.

[24] M. Wulder and Steven Franklin, "Remote sensing of forest environments: Concepts and case studies," January 2012.

[25] M. Shafaey, "Deep Learning for Satellite Image Classification," International Conference on Advanced Intelligent Systems and Informatics 2018, pp. 383-391, 2018.

[26] C. Tao, J. Qi, M. Guo, Q. Zhu and H. Li, "Self-Supervised Remote Sensing Feature Learning: Learning Paradigms, Challenges, and Future Works," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-26, 2023.

[27] H. Guo, "Dynamic Low-Light Image Enhancement for Object Detection via End-to-End Training," International Conference on Pattern Recognition, January 2021.

[28] M. Aladem, S. Baek and S. A. Rawashdeh, "Evaluation of Image Enhancement Techniques for Vision-Based Navigation under Low Illumination," Journal of Robotics, pp. 1-15, March 2019.

[29] M. Widyaningsih, T. K. Priyambodo, M. E. Wibowo and Muhammad, "Optimization Contrast Enhancement and Noise Reduction for Semantic Segmentation of Oil Palm Aerial Imagery," International Journal of Intelligent Engineering and Systems, vol. 16, no. 1, p. 597–609, February 2023.

[30] J. Guo, J. Ma, Á. F. García-Fernández, Y. Zhang and H. Liang, "A survey on image enhancement for Low-light images," Heliyon, vol. 9, no. 4, April 2023.

[31] M. Altaweel, A. Khelifi, Z. Li, A. Squitieri, T. Basmaji and M. Ghazal, "Automated Archaeological Feature Detection Using Deep Learning on Optical UAV Imagery: Preliminary Results," Remote Sensing, vol. 14, no. 3, p. 553, January 2022.

[32] C. M. Albrecht, "Learning and Recognizing Archeological Features from LiDAR Data," IEEE International Conference on Big Data (Big Data), 2019.

[33] A. Guyot, M. Lennon, T. Lorho and L. Hubert-Moy, "Combined Detection and Segmentation of Archeological Structures from LiDAR Data Using a Deep Learning Approach," Journal of Computer Applications in Archaeology, vol. 4, no. 1, February 2021.

[34] L. Luo, X. Wang, H. Guo, R. Lasaponara, X. Zong, N. Masini, G. W. P. Shi, H. Khatteli, F. Chen, S. Tariq, J. Shao, N. Bachagha, R. Yang and Y. Yao, "Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: A review of the century (1907–2017)," Remote Sensing of Environment, vol. 39, no. 6, June 2017.

[35] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, June 2017.

[36] D. Davis, "Object-based image analysis: a review of developments and future directions of automated feature detection in landscape archaeology," 2019.

[37] W. B. V.-v. d. Vaart, K. Lambers, W. Kowalczyk and Q. P. J. Bourgeois, "Combining Deep Learning and Location-Based Ranking for Large-Scale Archaeological Prospection of LiDAR Data from The Netherlands," ISPRS International Journal of Geo-Information, vol. 9, no. 5, May 2020.

[38] D. Davis, "Object-Based Image Analysis: A Review of Developments and Future Directions of Automated Feature Detection in Landscape Archaeology," SSRN Electronic Journal, 2018.

[39] Y. Tian, S. Tian, Y. Xu and Y. Zha, "Image object detection based on local feature and sparse representation," Journal of Computer Applications, October 2013.

[40] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, M. S. Bernstein and F.-F. Li, "Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations," International Journal of Computer Vision, vol. 123, no. 32–73, February 2017.

[41] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," University of Freiburg, May 2015.

[42] L. Lin, Z. Hao, C. Post and E. A. Mikhailova, "Monitoring Land Cover Change on a Rapidly Urbanizing Island Using Google Earth Engine," Applied Sciences, vol. 10, no. 20, October 2020.

[43] N. D. Burgess, B. Bahane and T. Clairs, "Getting ready for REDD+ in Tanzania: a case study of progress and challenges," Oryx, vol. 44, no. 3, July 2017.

[44] P. Sarah, "Satellite remote sensing for archaeology," Choice Reviews Online, vol. 47, no. 4, December 2009.

[45] J. Im, J. Y. Louhi, Kasahara, H. MaruYama, H. Asama and A. Yamashita, "Advanced Random Mix Augmentation: Data Augmentation Method for Improving Performance of Image Classification in Deep Learning Using Unduplicated Image Processing Combinations," Journal of the Japan Society for Precision Engineering, vol. 89, no. 1, p. 105–112, 2023.

[46] H. Furukawa, "Deep Learning for Target Classification from SAR Imagery: Data Augmentation and Translation Invariance," University of Bristol, August 2017.

[47] V. Dumoulin, "A Learned Representation for Artistic Style," ICLR, July 2022.

[48] C. Lee, "Deeply Supervised Nets," Cornell University, September 2014.

[49] D. Pathak, "Context Encoders: Feature Learning by Inpainting," IEEE, November 2016.

[50] "When to Use Contrast as a Preprocessing Step," Roboflow Blog, [Online]. Available: https://blog.roboflow.com/when-to-use-contrast-as-a-preprocessing-step. [Accessed 29 July 2023].

[51] M. Srikar, "A Real Time Object Detection in Integral Part of Computer Vision using Novel Image Classification of Faster R-CNN Algorithm over Fast R-CNN Algorithm," Journal of Pharmaceutical Negative Results, vol. 13, 2022.

[52] R. Girshick, "Fast R-CNN," IEEE, December 2015.

[53] J. Kim and J. Cho, "RGDiNet: Efficient Onboard Object Detection with Faster R-CNN for Air-to-Ground Surveillance," Sensors, vol. 21, no. 5, March 2021.

[54] E. Ayman, "Depth Based Region Proposal: Towards Robust Two-Stage Real-Time Object DetectionSSRN Electronic Journal," SSRN Electronic Journal, June 2022.

[55] J. Redmon, "You Only Look Once: Unified, Real-Time Object Detection," IEEE, 2016.

[56] C. Zheng, "Stack-YOLO: A Friendly-Hardware Real-Time Object Detection Algorithm," IEEE Access, vol. 11, 2023.

[57] A. I, "Real Time Object Detection Using YoloReal Time Object Detection Using Yolo," International Journal for Research in Applied Science and Engineering Technology, vol. 9, no. 11, November 2021.

[58] "What's New in YOLOv8?," Roboflow Blog, [Online]. Available: https://blog.roboflow.com/whats-new-in-yolov8. [Accessed 2023 June 12].

[59] C. Tong, X. Yang, Q. Huang and F. Qian, "NGIoU Loss: Generalized Intersection over Union Loss Based on a New Bounding Box Regression," Applied Sciences, vol. 12, no. 24, December 2022.

[60] M. Tan, R. Pang and Q. Le, "EfficientDet: Scalable and Efficient Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[61] X. Ren, F. Luo and S. Li, "Mixed geometric loss for bounding box regression in object detection," Journal of Electronic Imaging, vol. 29, no. 5, September 2020.

[62] A.-W. 3D. [Online]. Available: https://alos-pasco.com/en/solutions/detail/post-254.html. [Accessed 2023 June 13].

[63] I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio, "Deep Learning," MIT press Cambridge, vol. 1, 2016.

[64] S. Ruder, "An overview of gradient descent optimization algorithms," 2016.

[65] C. Zhang, S. Bengio, M. Hardt, B. Recht and O. Vinyals, "Understanding deep learning requires rethinking generalization," 2016.

[66] L. Arie, "The practical guide for Object Detection with YOLOv5 algorithm," Medium, 2023.

Anex A

This Anex refers to all the other seven FasterRCNN detection results in the images used for the test set, five from each of the three dolmens.

Object Detection Results from the FasterRCNN algorithms:

Faster_rcnn_R_50_FPN_3x

São Pedro da Gafanhoeira 1

Telhal 1

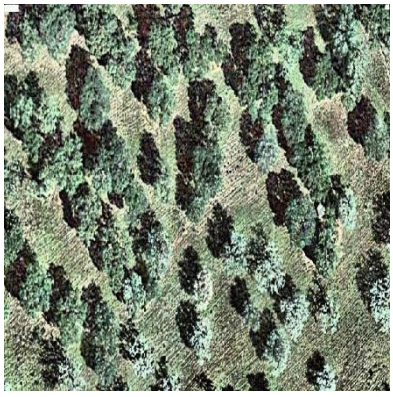

Anta de Prates 7

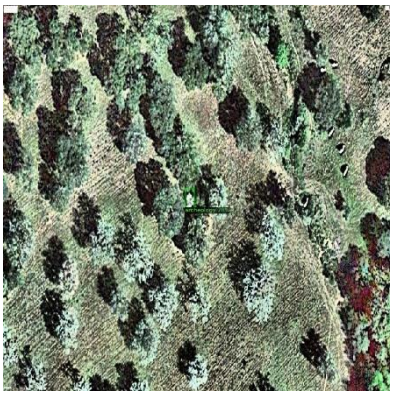




Faster_rcnn_R_50_FPN_1x

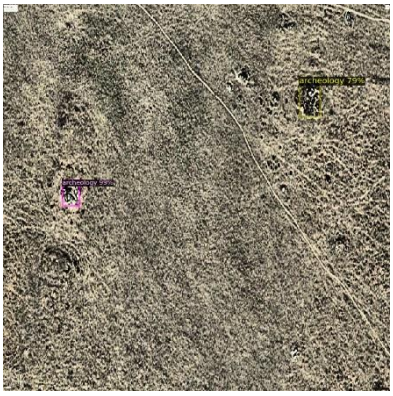
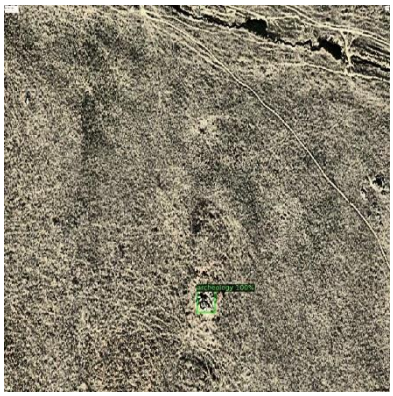
São Pedro da Gafanhoeira 1



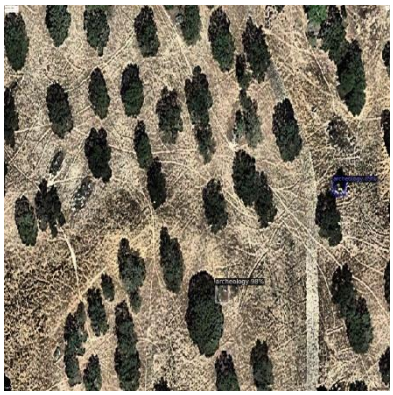
Telhal 1




Anta de Prates 7





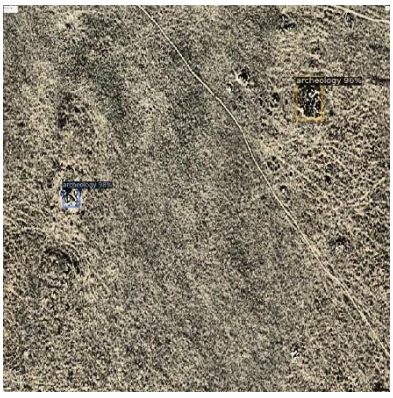

Faster_rcnn_R_50_DC5_3x

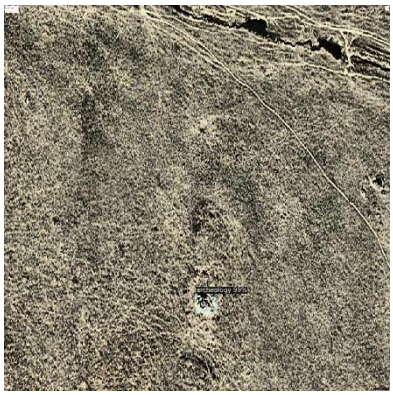
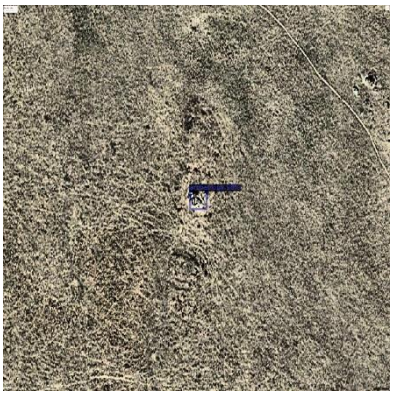
São Pedro da Gafanhoeira 1





Telhal 1

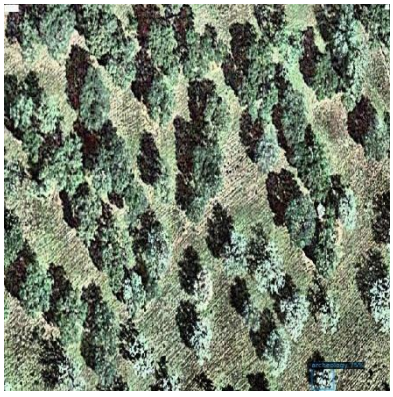
Anta de Prates 7

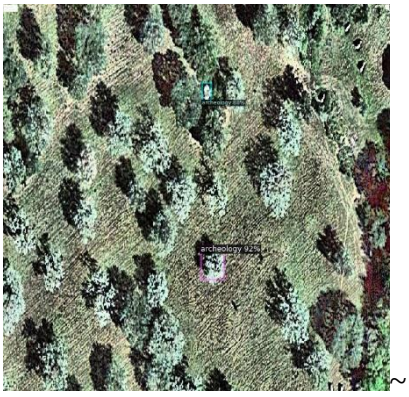




Faster_rcnn_R_50_C_3x

São Pedro da Gafanhoeira 1






Telhal 1



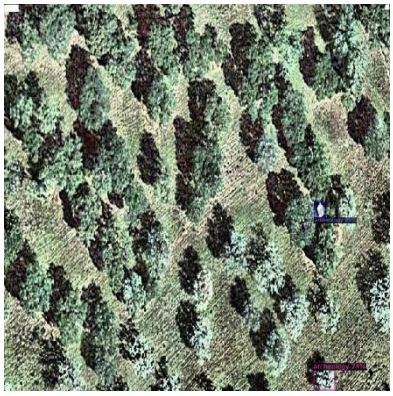


Anta de Prates 7



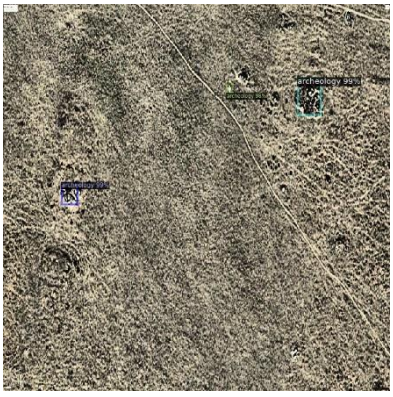

Faster_rcnn_R_101_FPN_3x

São Pedro da Gafanhoeira 1





Telhal 1



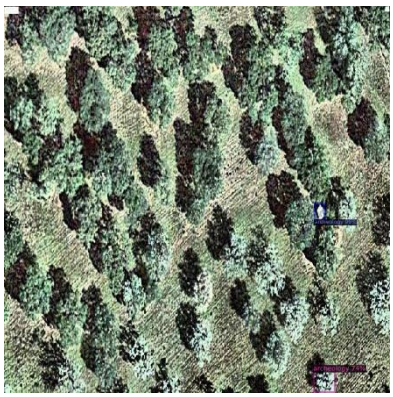
Anta de Prates 7

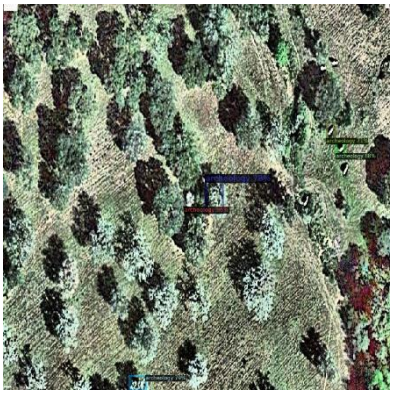




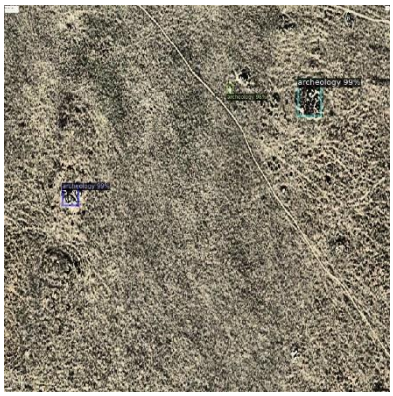
Faster_rcnn_R_101_DC_3x

São Pedro da Gafanhoeira 1



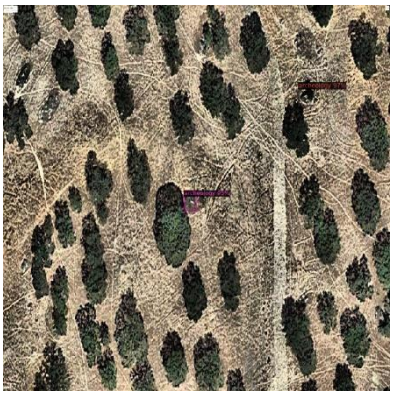
Telhal 1





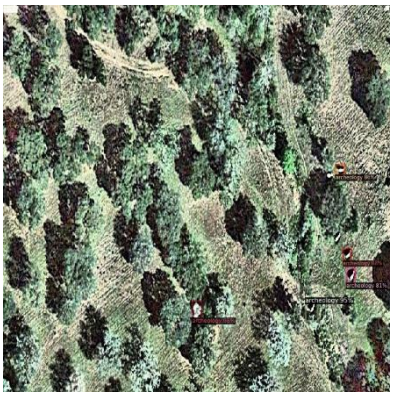
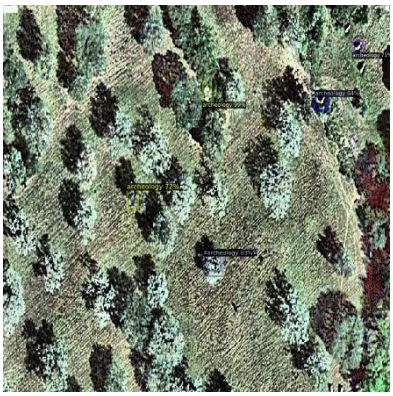
Anta de Prates 7



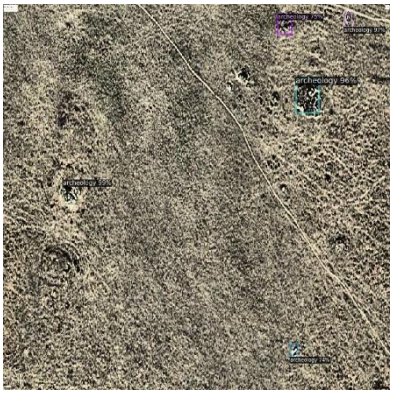



Faster_rcnn_R_101_C4_3x

São Pedro da Gafanhoeira 1

Telhal 1



Anta de Prates 7