POLITECNICO DI TORINO

Master's Degree in Data Science and Engineering

Master's Degree Thesis

Enhancing Weed Detection and Segmentation with Advanced Deep Learning Algorithms leveraging RGB-NIR Imaging

Supervisors

Prof. Renato FERRERO

Dr. Nicola DILILLO

Candidate

Bahareh BEHROUZI

July 2024

Abstract

This study explores the potential of improving weed identification accuracy and reducing herbicide usage through the use of computer vision analysis, specifically incorporating imaging datasets that include both RGB and near-infrared (NIR) channels. The thesis investigates the application of deep learning models, namely YOLO and U-Net, to improve the detection of weeds and their semantic segmentation in precision agricultural contexts. It utilizes two meticulously annotated datasets: ACRE for object detection and Sunflower for semantic segmentation. Additionally, the study considers the effect of varying NDVI thresholds on segmentation performance, finding that fine-tuning these thresholds can be key to enhancing outcomes. The customized U-Net model deployed for segmentation offers promising results, outpacing several leading models in effectiveness, particularly with integrating NIR data. Combining RGB images with NIR channels has shown a notable boost in the deep learning models' ability to differentiate between weeds and crops—an essential step forward for precision agriculture. The findings reveal marked progress in segmentation accuracy, with the UNet-ResNet50 model incorporating RGB+NIR data achieving significant benchmarks: a mean Intersection over Union (IoU) score of 0.88, a Crop IoU of 0.93, and a Weed IoU of 0.71, respectively with an improvement of 2.7%, 4%, and 3%.

Conversely, YOLO models have demonstrated solid capabilities in object detection tasks utilizing only RGB data. YOLOv8 achieves a mean average precision (mAP) of 0.48, which is competitive with the state-of-the-art models in object detection. While surpassing some traditional methods, these encouraging outcomes highlight the promising role that multispectral imaging could play in transforming precision agriculture. This research suggests a path toward an era where the merging of sophisticated imaging technologies fundamentally changes precision agriculture. In summary, the study presents evidence that combining NIR data with RGB channels can significantly boost the precision and effectiveness of weed segmentation systems, contributing to the development of more eco-friendly agricultural practices.

Acknowledgements

I am profoundly grateful to Prof. Renato Ferrero for his invaluable guidance, insightful feedback, and continuous support throughout the course of this research. His knowledge and commitment have played a crucial role in the development of this thesis.
I am also profoundly thankful to Dr. Nicola Dilillo for his mentorship, technical advice, and encouragement. His contributions have been essential to the completion of this work, and his commitment to academic excellence has been a constant source of inspiration.
Lastly, I wish to extend my heartfelt thanks to my mother for her unwavering support and belief in me. Her love and encouragement have been my driving force, providing me with the strength and motivation to pursue and achieve my academic goals.

Table of Contents

List of Tables

List of Figures

Acronyms

AI
Artificial intelligence

IoT
Internet of Things

RGB
red, green, and blue

GPS
Global Positioning System

NIR
near-infrared

NDVI
Normalized Difference Vegetation Index

DL
Deep Learning

ML
Machine Learning

CNN
Convolutional Neural Networks

YOLO
You Only Look Once

IoU

Intersection over Union

NN

Neural Networks

SVM

Support Vector Machines

KNN

K-nearest Neighbor

Local Binary Pattern

Local Binary Pattern

OBIA

Object-Based Image Analysis

UAV

Unmanned Aerial Vehicles

GLSM

Gray-level Co-occurrence Matrix

ANN

Artificial Neural Networks

FCN

Fully Convolutional Networks

DNN

Deep Neural Networks

mAP

Mean Average Precision

TP

True Positive

FP

Flase Postive

TN

True Negative

FN

False Negative

AP

Average Precision

Chapter 1

Introduction

1.1 Smart Farming

The primary objective of smart farming is to fundamentally transform traditional agriculture through the deployment of cutting-edge technologies such as artificial intelligence (AI), Big Data, and the Internet of Things (IoT). These technologies serve to enhance the quality and quantity of agricultural output by enabling more efficient use of resources and reducing environmental impacts, thereby securing a stable food supply, promoting sustainable farming methods, and ensuring robust agricultural production [1]. At the heart of smart farming lies precision agriculture, which employs these sophisticated technologies and meticulous data analysis to optimize crop yields, curtail waste, and amplify productivity. This approach is marked by strategic decision-making and the meticulous integration of leading-edge technologies, enhancing the precision of agricultural operations. For instance, AI empowers farmers to make informed decisions by providing detailed insights into crop health, soil conditions, and climatic variables. Furthermore, computer vision technology equips machines with the ability to 'see' and 'interpret' the farm environment, facilitating early detection of potential issues and ongoing monitoring of crop development[1]. These technological tools collaborate seamlessly to enable farmers to achieve greater outputs with fewer inputs, thereby reducing costs and enhancing both the quantity and quality of the produce. The integration of AI and IoT in smart farming establishes a data-driven decision-making framework that significantly augments the capacity of farmers to anticipate and address various agricultural challenges effectively. Advanced sensing and imaging technologies play a crucial role in detecting subtle changes in plant health, which allows for timely interventions that can preempt crop diseases or pest infestations, thereby mitigating potential losses. This forward-thinking strategy not only preserves resources but also minimizes dependence on chemical

treatments, supporting more sustainable agricultural practices[1].

1.2 Challenges in Weed management

Weeds pose significant threats to agricultural productivity as they compete directly with crops for space and essential resources such as water, soil, nutrients, and sunlight and according to Wang[2], an average of 34% of crop production is lost due to this problem. Effective weed management is crucial for maintaining optimal crop yields and quality. Additionally, the detection of weeds becomes challenging due to their irregular distribution throughout the field and their tendency to blend in with crop plants. Traditional methods to control weed infestation have included manual weeding and the use of chemical herbicides. Although manual weeding is effective, it requires a significant amount of labor and is economically unsustainable on a large scale [3]. On the other hand, although chemical herbicides are more feasible for extensive farming operations, they carry environmental risks. These risks include potential pollution of ecosystems and the development of herbicide-resistant weed strains, which can further complicate weed control efforts. Instead of the conventional approach of broadly applying herbicides over entire cultivated fields, Precision Agriculture enhances farming practices through the automatic localization and classification of crops and weeds. This advanced method significantly boosts both the efficiency and sustainability of agricultural management. By precisely targeting herbicides only to the areas where they are needed, or even by mechanically removing weeds without any chemical usage, which decreases the most drawbacks of manual weeding[4]. However, mechanized weeding also presents its challenges, including high energy consumption, potential risk of crop damage, intricate system designs, and possibly substantial maintenance costs for the equipment[2]. it is also notable that the constant mutation of weeds and regulatory removal of some key herbicides have significantly reduced the availability of effective selective herbicides for many crops. Moreover, the lengthy process required for agribusinesses to discover and market new alternatives is further constraining the supply of such options[5]. Moreover, Weed management is complex due to the variability in weed behavior across different conditions and seasons. For example, while the tall, fast-growing fat hen weed poses a threat to nearby crops, its late-summer seedlings are smaller and less harmful. Additionally, some weeds coexist with crops initially but later compete vigorously for resources. Identifying the precise moment when weeds begin impacting crop yield is challenging. Given these challenges, there is a growing need for innovative weed management strategies that are both efficient and environmentally friendly.

1.3 Imaging Technologies

Imaging, whether through RGB or multispectral cameras, is increasingly utilized in agriculture for detecting and segmenting weeds from crops. These imaging technologies are central to innovative weed management techniques that are both efficient and environmentally friendly. RGB imaging captures light in red, green, and blue bands, mirroring what the human eye sees. it offers a snapshot of field conditions by showing the number and maturity of plants. This information aids ground-level scouting, helping to identify and locate areas of concern by integrating the data with Global Positioning System(GPS) coordinates for precise navigation to problem spots[6]. However, RGB cameras may not always distinguish between crops and weeds effectively, especially under variable lighting conditions or when the plants exhibit similar colors. Unlike RGB cameras, multispectral imaging captures light across multiple wavelengths beyond the visible spectrum, including near-infrared (NIR).To be more precise, "The RGB spectral bands are in the visible range (400-700 nm), whereas the NIR spectral band is beyond the visible range (700-1100 nm)"[7][8]. The NIR band, positioned in the electromagnetic spectrum between the visible and mid-infrared bands, spans a broad wavelength range and delivers clear image information, even under low-light conditions[8]. To be more specific, While RGB sensors provide a visible representation of the field, facilitating the counting of plants, multispectral sensors are designed to capture the way light reflects off plant leaves. This approach emphasizes different spectral properties, offering insights into plant health that are not visible to the naked eye. It also utilizes indices like the NDVI to offer further detailed analysis[6]. Hence, an image captured using RGB-NIR technology displays a diverse range of features. Additionally, the use of the NIR band in multispectral cameras enhances weed detection capabilities. The NIR band's ability to detect subtle differences in how crops and weeds reflect infrared light enables more precise identification of these plants. This increased precision in identifying weeds allows for more targeted herbicide application, potentially reducing overall herbicide use and minimizing environmental impact. By accurately distinguishing weeds from crops, NIR technology ensures that herbicides are applied only where necessary, thus supporting more sustainable agricultural practices[9]. A comparison between RGB and multispectral images is made in Fig1, Fig2.

Figure 1: An RGB image of a farm, showing the visible spectrum that provides clear views of plant coverage and field conditions. This image serves as a straightforward example of RGB imaging in agricultural applications[6].

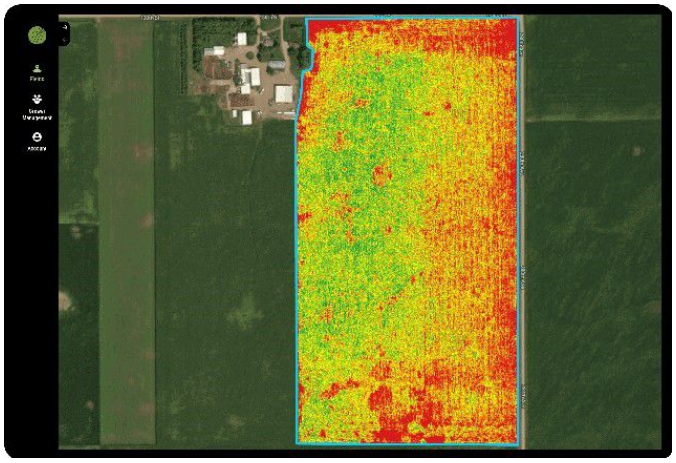

Figure 2: A multispectral Normalized Difference Vegetation Index(NDVI) image highlighting the health of vegetation by assessing the density and vitality of the crops. NDVI images like this are crucial for precision agriculture, offering detailed insights into plant health not visible in RGB spectra[6].

1.4 Applications of Multispectral Sensors

Multispectral sensors provide a range of insights that are crucial for effective agricultural management. They allow for the evaluation of how crops react to

essential inputs, which informs future application decisions. These sensors also track the crop canopy to monitor growth rates and assess plant maturity throughout the growing season. In terms of plant health, multispectral imaging helps spot areas of stress, aiding in the development of targeted mitigation strategies. Additionally, they offer an understanding of soil health by analyzing how effectively plants absorb nutrients from the soil. Moreover, multispectral sensors can forecast potential drought conditions, enabling more precise management of water resources and the formulation of strategic irrigation plans. Integrating remote sensing with these capabilities enhances the ability to observe and manage agricultural fields remotely, leading to more informed and precise farming practices[6].

1.5 Role of Deep Learning

Deep learning(DL), a specialized branch of machine learning(ML), leverages multi-layered neural networks to analyze diverse data types effectively. This approach is particularly advantageous in image analysis, where it excels at autonomously identifying complex patterns and features within extensive datasets. Such capabilities render DL an invaluable asset for critical agricultural tasks like object recognition, classification, and anomaly detection. At the core of DL's utility in image analysis are Convolutional Neural Networks (CNNs). These networks are tailored to extract features from images automatically and efficiently—a key factor in precise weed detection. Unlike conventional image processing methods that depend on manual feature selection and adjustment, CNNs adapt to recognize vital features directly from provided training data. This adaptation simplifies the analysis process and enhances its reliability, making it especially suited to the dynamic and diverse conditions encountered in agriculture. In the realm of precision agriculture, DL frameworks such as You Only Look Once(YOLO) and U-Net have transformed weed detection and management. YOLO excels in real-time object detection, swiftly identifying and categorizing various objects, including weeds, across different agricultural fields. U-Net, initially developed for biomedical image segmentation, is adept at intricately segmenting complex images, thus effectively differentiating between weeds and crops at a granular level [10]. The incorporation of these advanced models into farming practices facilitates more precise and efficient weed management strategies. By enhancing weed detection accuracy, deep learning not only furthers sustainable agriculture initiatives—optimizing resource use and maximizing crop yields—but also supports the overarching goals of modern, sustainable agriculture. Thus, deep learning not only boosts the efficacy of weed detection and segmentation systems but also contributes significantly to the broader objectives of contemporary,

sustainable agriculture by promoting smarter, data-driven farming decisions. These advancements mark a pivotal evolution in agricultural practices, bringing them into closer alignment with the principles of precision agriculture.

1.6 Research Problem and Objectives

Despite these technological advances, significant challenges remain in effectively managing weed populations without incurring high environmental costs or labor demands. The primary research problem this thesis addresses is the need for an integrated approach that enhances the accuracy and efficiency of weed detection and segmentation, reducing reliance on chemical herbicides while ensuring crop health and yield. To comprehensively address this problem, this study will embark on several focused objectives: 1. Evaluating the effectiveness of convolutional neural networks (CNNs) in differentiating weeds from crops using multispectral imaging, specifically comparing the performance of the RGB dataset against the RGB+NIR dataset: Implementing YOLO algorithms for object detection training on standard RGB dataset and comparing results with models trained on combined RGB+NIR dataset using the UNet architecture for precise image segmentation 2. Investigating the impact of integrating additional spectral data with standard RGB channels on the accuracy of weed segmentation: Exploring several combinations of spectral data integration, including RGB+NIR (near-infrared), G+NIR (green and near-infrared), and G+NIR+NDVI (green, near-infrared, and Normalized Difference Vegetation Index). The effectiveness of these combinations will be evaluated based on their Intersection over Union (IoU) metrics to determine which configuration most significantly enhances weed detection and segmentation accuracy. This analysis will help identify the optimal spectral channel combination for improving the precision of automated weed management systems.

1.7 Scope and Limitations

Scope of the Study: The study focuses on optimizing convolutional neural network architectures, utilizing UNet for RGB-NIR datasets to enhance the precision of weed identification in agricultural fields while employing YOLO exclusively on RGB datasets. The analysis is based on existing datasets that have been annotated for crop and weed detection and segmentation to train and test the proposed models. Limitations of the Study: 1. Data Availability and Quality: The effectiveness of deep learning models in agricultural applications is heavily dependent on the quantity and quality of the training data. One significant limitation is the availability of comprehensive, well-annotated datasets specifically for weed detection in agriculture. The scarcity of labeled weed data can constrain

the training process, potentially affecting the generalizability and accuracy of the models. 2. Visual Characteristics of Crops and Weeds: Differentiating between crops and weeds using image data is challenging due to their similar visual characteristics. Traditional image segmentation techniques that rely on preprocessing and pixel-based classification may not adequately address the subtle differences between these plant types, necessitating more sophisticated DL approaches. 3. Environmental and Operational Variabilities: The variability in environmental conditions such as lighting, weather, and stress on plants can significantly impact image quality and the consistency of image data. Variations in illumination at different times of the day or changes in weather conditions during data collection can introduce inconsistencies that complicate the segmentation and detection tasks. Potential errors such as mislabeled crop/weed samples or shifts in crop/weed boundaries due to growth or external factors also pose challenges, affecting the training accuracy and model reliability.
In response to the critical dependency of deep learning models on high-quality, well-annotated training data, our research utilized two key datasets: ACRE Dataset for Object Detection: We employed the ACRE dataset, a well-annotated resource known for its comprehensive coverage of agricultural scenes, specifically designed for object detection tasks. This dataset has been instrumental in training our models to accurately detect and differentiate between various objects within agricultural fields, including weeds and crops. Sunflower Dataset for Semantic Segmentation: For segmentation tasks, particularly focusing on precise delineation of plant boundaries, we utilized the Sunflower dataset. This dataset provides detailed annotations necessary for training models like U-Net, which require pixel-level accuracy for effective image segmentation. Furthermore, we used transfer learning leveraging pre-trained models on similar tasks which allowed us to utilize learned features that can be adapted to the specific requirements of weed detection along with segmentation, thereby reducing the dependency on extensive labeled weed data.

1.8 Research Questions and Hypothesis

1.What advancements have been made in the application of deep learning for weed detection and segmentation in agriculture over the past decade? This question seeks to explore the progression and technological evolution in the use of deep learning models for identifying and managing weed species across various agricultural settings. We answered this Question in Chapter2.
2. How effective are different deep learning models, such as YOLO and UNet, when applied to RGB and multispectral (RGB-NIR) imaging datasets for weed detection along with segmentation? This inquiry focuses on evaluating the

performance of specific deep learning architectures in processing and analyzing different types of imaging data to enhance the accuracy and efficiency of weed detection. This question is answered in Chapter4.
3. Can an existing Convolutional Neural Network, trained using an RGB dataset, be effectively utilized to segment weeds from a multispectral image? We hypothesize that the neural network pre-trained on RGB data will exhibit transferability to multispectral images for weed segmentation. The inherent features learned from the RGB dataset, combined with the additional spectral information available in multispectral imagery, are expected to enhance the network's performance in accurately delineating weeds from the background. This question is answered in Chapter 3.
4. What challenges and limitations are currently faced in the practical application of deep learning for weed detection, and how can these be addressed to improve future models? This question aims to identify key obstacles such as dataset limitations, computational demands, and model adaptability, proposing pathways for research and development to overcome these challenges. This question is answered in Chapter2.

1.9 Organization of the Thesis

Chapter 1: Introduction

The introduction of this thesis outlines the advancements and challenges in smart farming and precision agriculture, explores the role of imaging technologies and deep learning, and defines the research objectives and significance within the context of enhancing sustainable agricultural practices

Chapter 2: Literature Overview

The literature overview chapter provides a comprehensive review of existing research on ML and imaging techniques in agriculture. It discusses traditional methods, advancements in deep learning applications such as YOLO and U-Net, and their optimization to enhance the precision of weed identification in agricultural fields.

Chapter 3: Methods

Chapter 3 of the thesis details the methods, describing the datasets, the specific configurations of the convolutional neural network architectures used, and the experimental setups for both the RGB and RGB-NIR datasets to systematically evaluate the performance of YOLO and U-Net in weed detection and segmentation.

Chapter 4: Results and discussions

This chapter presents the experimental results, analyzing the effectiveness of the YOLO and UNet models in weed detection and segmentation across different datasets, and discusses the impact of integrating NIR data on the accuracy and reliability of weed identification. Also, this chapter offers a discussion of the findings, interpreting the implications of the experimental results in comparison with state-of-the-art methods

Chapter 5: Conclusion

Chapter 5 represents the Conclusion by assessing the contributions of the study to the field of precision agriculture, along with recommendations for future research and potential improvements in weed management strategies.

Chapter 2

Literature Review

2.1 Traditional Machine Learning Weed Detection Methods

In the realm of weed identification research, a variety of traditional ML methods, grounded in image processing techniques, have been utilized. These methods include support vector machines (SVM) [11],decision tree[12]-based random forest algorithm[13],and K-nearest neighbor (KNN) classifiers[14]. These techniques primarily focus on extracting features such as color, texture, shape, and spectrum from weed images through intricate manual processes. For example, Le et al.[15] successfully differentiated between corn and certain weeds using SVM along with Local Binary Pattern (LBP) texture features. Chen et al.[16] created a multi-feature method for locating weeds in soybean fields based on shape and color attributes, and Zhu et al.[17] devised a classification approach for five different weed types in agricultural settings focusing on shape and texture. Zhang et al.[18] performed a comparative study on the grayscale distribution of weeds in RGB, HSV, and HIS color spaces at the pea seedling stage, developing a method for weed segmentation and extraction based on R-B color difference features. Additionally, some researchers have enhanced identification accuracy using plant height[19] or positional data[20], though these techniques can be affected by vibration and other uncontrollable movements in field applications. Despite these advances, challenges persist, particularly in distinguishing similar weed species when image extraction is not comprehensive or features are obscured. In the foundational stages of this field, researchers widely adopted these ML algorithms in conjunction with image features to efficiently detect weeds. These traditional methods are beneficial due to their low requirement for large data samples and short training periods; they also place minimal demands on graphics processing units. Consequently, they provide a cost-effective solution for incorporation into

agricultural machinery, enabling reliable plant identification and weed detection via image processing. This effective use of machine vision technology, which continuously evolves to refine the process of extracting elementary weed features for subsequent classification, underscores the successful integration of smart technologies in agricultural applications. Initially, the differentiation between crops and weeds is achieved by assessing the image's textural, shape, color, and spectral properties. Additionally, certain studies have concentrated on utilizing a single feature for plant identification, resulting in low accuracy and limited reliability. To address the challenges posed by complex field environments and the limitations of using single features for plant recognition, researchers have developed methods that combine multiple features to enhance accuracy and stability. For example, He et al.[21]integrated various recognition information including plant leaf shape, fractal dimension, and texture, leveraging the strengths of SVM for small sample classifications and the benefits of Dempster–Shafer evidence theory in managing incomplete and uncertain information. This approach of multi-feature decision fusion has demonstrated improved stability and higher recognition accuracy compared to single-feature methods. Similarly, Sabzi et al[22] developed a machine vision prototype that utilizes video processing and meta-heuristic classifiers, incorporating features from the Gray-level Co-occurrence Matrix (GLCM), color, texture, invariant moments, and shape to identify and classify a large dataset of potato and weed species with high accuracy online. Deng et al.[23] integrated color, shape, and texture features into a 101-dimensional feature set to address the poor recognition accuracy of single features in rice fields. Although these innovations mark significant advances in image-based plant recognition and weed detection technologies, they predominantly focus on leaf identification rather than precise in-field crop or weed detection. Furthermore, there remains a scarcity of studies that tackle the identification and localization of plants and weeds against complex, practical field backgrounds, indicating that further research is essential to enhance weed detection and identification in actual farmland settings. table1 summarizes various studies on the use of traditional ML techniques for identifying or classifying plant leaves. These methods are effective in controlled settings with specific plant species and backgrounds but fall short in rapid, large-scale image processing in natural environments. The application of drone imagery for large-scale vegetation classification and weed detection is gaining traction. Object-Based Image Analysis (OBIA) is progressively replacing older pixel-based classification methods. One of the main challenges with OBIA is determining the optimal parameter settings. To address this, Torres-Sánchez et al.[24] utilized Unmanned Aerial Vehicles images(UAV) to develop an automatic thresholding algorithm that operates within the OBIA framework, facilitating unsupervised classification of various herbaceous row crops. UAV offer the advantage of less restriction by field conditions, enabling broad-scale monitoring of weed

populations. They also provide high-resolution images, which are crucial for detecting low densities of weeds and allow for flexible timing of image capture, thereby enhancing the potential for extensive applications in high-input agricultural environments. Table 1 presents the accuracy and associated problems of various traditional ML methods used for plant and leaf identification.

Table 1: Overview of Traditional Machine Learning Techniques and Their Challenges

Purpose	**Accuracy**	**Problems**
Utilizing HOG features alongside SVM for grape leaf identification	83.50%	Single-feature detection suffers from low stability and accuracy.
Improved detection of plant leaves using enhanced LBP	79.35%	Single-feature detection suffers from low stability and accuracy.
Investigating the impact of SVM and ANN on sugar beet and weed detection	93.33%	Inadequate feature selection analysis.
Utilizing Gabor wavelet and gradient field distribution for weed categorization	93.75%	Inadequate feature selection analysis.
Categorizing fresh tea with improved LBP and GLCM	94.80%	Only identifies one type of leaf.

2.2 Conventional Methods and Their Pros and Cons for Common Weed Detection

The traditional methods for weed detection through image processing, which are a focus of discussion in this section of the thesis, rely on differentiating plant leaves from weeds based on distinct image features. This discussion critically examines the advantages and drawbacks of utilizing four key features—texture, shape, spectrum, and color—in the detection and recognition of weeds.

2.2.1 Texture Features

Texture features are key identifiers in image classification, reflecting the spatial relationships between pixels. Commonly found in plant leaves, these features vary based on vein patterns and surface textures, aiding in the differentiation of crops

from weeds. These texture methods fall into four main types: statistical, structural, model-based, and transform-based. Commonly used descriptors include (GLCM)[25] LBP, which are effective in analyzing microstructures and maintaining consistency across image rotations and translations. These methods, however, face challenges under complex agricultural conditions, such as overlapping plants and varying weed densities. Advanced applications of these descriptors are crucial for enhancing precision in detecting subtle differences between plant types, thereby improving both the efficiency and accuracy of automated weed management systems.

2.2.2 Shape Features

In the field of image analysis for weed detection, shape features are critically important. They describe the geometrical attributes of objects within an image, including aspects like perimeter, area, major and minor axis lengths, and more intricate measures such as eccentricity, compactness, and solidity. These features help in distinguishing between different plant species by their morphological traits. Additionally, sophisticated algorithms such as Hu moment invariants[26] are employed to ensure these features are invariant to rotation, scaling, and translation, enhancing their utility across varied imaging conditions. However, the effectiveness of shape features alone is limited when plant leaves are distorted by environmental factors such as disease, insect damage, or mechanical impacts. Furthermore, field conditions often present challenges like overlapping leaves and occlusion, complicating weed identification. To overcome these challenges, combining shape features with other descriptors is essential to enhance the accuracy and robustness of the detection process.

2.2.3 Spectral Features

Spectral features are essential in distinguishing plants based on their leaf colors and are particularly effective when the spectral reflectance between weeds and crops is distinct. These features are less sensitive to partial occlusions and are efficient in computation. Researchers have harnessed visible light, near-infrared spectra (Vis–NIR)[27], multispectral/hyperspectral imaging, and fluorescence[28] to identify various plant species. Additionally, capturing a multispectral image is influenced by daily climatic conditions, which alter plant reflectivity based on light absorption. Techniques that leverage spectral indices have been successful in differentiating between crop species by analyzing chlorophyll and carotenoid content. While spectral sensors provide detailed data, their ability to distinguish between plant species in early growth stages is limited due to similar reflective characteristics. Despite promising results in identifying weeds using specific

spectral bands, the effectiveness is often compromised by environmental factors like moisture, disease, and growth stages. Hence, combining spectral data with other features like shape and texture is advisable to improve identification accuracy in diverse field conditions[29].

2.2.4 Color Features

The accuracy of color-based plant detection relies heavily on the specific plant species and the contrast in their colors. This method excels in resisting changes in scale, size, and orientation, and is often utilized to separate plants from backgrounds by capitalizing on these color differences. Notable research by Hamuda et al[30] and Tang et al[31] has detailed the strengths and limitations of color index-based segmentation approaches. Despite the robustness of color properties, they are not infallible and are influenced by various environmental factors like lighting, which can obscure the color distinctions necessary for effective weed and crop differentiation. Consequently, many researchers opt to transition from the RGB color spectrum to other color spaces such as HIS, HSV, Lab, and YCrCb, aiming to improve the accuracy of segmentation and analysis. However, color remains a variable feature in plant identification, susceptible to alterations by plant health, seasonal changes, or inconsistent lighting conditions, potentially complicating the discrimination of weeds from crops under real-world conditions. Table 2 Table 2 offers an examination of the strengths and weaknesses associated with four prevalent image features utilized in weed detection.

Table 2: Comparison of the Advantages and Disadvantages of Four Common Features Used in Weed Detection

Features	**Advantages**	**Disadvantages**
Texture	Offers high precision and adaptability, with a strong resilience against disruptions	GLCM processing is slow and not suitable for real-time applications.
Shape	Operates independently of any geometrical transformations and is robust against noise interference	Prone to deformation from external factors such as disease or physical damage, and may struggle with overlapping shapes.
Color	Remains consistent despite changes in scale, size, and orientation, providing vital differentiation	Struggles with similar colors in different plants, affected by environmental conditions.
Spectral	Effective even with partial blockages	Performance varies with plant growth stages and environmental conditions, less stable.

2.2.5 Multi-Feature Fusion

The challenge of detecting weeds using a single image feature is compounded by the similarity between weeds and crops. This has led researchers to explore multi-feature fusion methods, aiming to enhance accuracy and stability in weed detection, particularly in non-ideal field conditions. Multi-feature fusion in weed detection is an advanced approach that integrates different types of data derived from images to enhance the accuracy of identifying weeds among crops. This technique combines several types of image features—shape, texture, color, and spectral—each providing unique insights that contribute to a more robust classification system. For instance, shape features to focus on the geometry of the plants, texture features analyze the surface characteristics, color features capture the visible hues, and spectral features assess information from various light spectrums. An example of successful multi-feature fusion is the study by He et al[21], who explored the integration of these features for identifying crop and weed species. Their system improved detection accuracy significantly by leveraging the complementary strengths of each feature type, demonstrating how multi-feature fusion can overcome the limitations of using single features alone, particularly in complex agricultural environments. While multi-feature fusion has advanced weed detection, further research is needed to address persistent issues in experiment

accuracy and stability under varying field conditions.

2.2.6 Classifier

SVMs and ANNs are pivotal in the classification of crops and weeds[9], benefiting from their ability to handle nonlinear, high-dimensional patterns and small sample sizes. SVMs are especially effective in avoiding nonlocal minima issues, while ANNs excel in learning from untrained data due to their robust learning capabilities. Besides these, methods like KNN and random forests are frequently mentioned in research, alongside naive Bayesian[32] and AdaBoost algorithms[33] for their strong classification performances. Recent advancements have seen researchers applying these classifiers in varied contexts. For instance, Jeon et al[34] utilized ANNs for weed detection in outdoor settings, adapting to challenging lighting conditions. Chen et al[35] combined KNN with advanced algorithms to enhance weed classification, achieving high accuracy. Similarly, Rumpf et al[36] used sequential SVM models to differentiate between types of weeds and crops, illustrating the adaptability of SVMs to complex agricultural scenarios. Moreover, the integration of multiple classifiers is becoming commonplace to leverage the strengths of various approaches. Bakhshipour and Jafari[11] compared the effectiveness of SVMs and ANNs in detecting weeds in sugar beet fields, finding SVMs generally more accurate. The combined use of different classifiers not only enhances the precision but also broadens the applicability in diverse field conditions. However, despite the successes, challenges such as recognizing multiple weed types in complex environments persist. Continued research is crucial for refining these technologies to ensure they meet the demands of modern agriculture, indicating the importance of ongoing innovation in multi-feature fusion and classifier optimization.

2.3 Weed Detection and Identification Methods Based on Deep Learning

The significant advancements and widespread availability of image-capturing devices have simplified the process of capturing images. Concurrently, the reduction in computer hardware costs and the enhancement in GPU computing power have facilitated the application of deep learning in agriculture. Deep learning techniques have shown impressive results in weed detection and classification[9]. Although traditional ML methods are straightforward and have seen numerous enhancements, they are often validated in controlled environments with low-density images, posing challenges such as occlusion, clustering, and varying lighting conditions in natural settings. DL distinguishes itself with its

network feature structure, enabling more efficient feature extraction compared to manual methods. It achieves higher-level features by aggregating local features from lower levels and integrating them at higher levels, allowing for diverse task-specific features. In weed detection, DL exploits spatial and semantic feature variations to enhance the accuracy of identifying and detecting weeds and crops. Commonly employed deep learning networks in this domain include CNNs and Fully Convolutional Networks (FCNs). Additionally, semi-supervised and unsupervised methods have been developed to mitigate labeling costs. DL algorithms frequently outperform traditional algorithms in classification tasks, though they require extensive datasets for training. The difficulty of collecting comprehensive crop and weed images highlights a drawback of DL methods in weed identification. One of the key advantages of deep learning methods such as CNNs and FCNs is their automatic feature extraction capability, which surpasses the efficiency of manually defined features. The higher accuracy of deep learning models is attributed to their complexity and ability to tackle more intricate problems. Despite the outstanding performance of DL algorithms, challenges in data collection and environmental variability remain significant hurdles. In summary, while DL offers enhanced accuracy and automation in feature extraction for weed detection, the dependency on large datasets and environmental complexities continue to pose challenges that need addressing for more effective application in agricultural practices [3].

2.4 Weed Detection and Identification Methods Based on CNNs

CNNs have become increasingly popular in weed detection, thanks to their exceptional performance in classification and identification tasks. Various studies, such as those by Yu et al[37] have demonstrated the effectiveness of deep CNNs in this field. For instance, Potena et al[38] used two different CNNs to process RGB and NIR images, achieving rapid and accurate crop and weed identification. A lightweight CNN was employed for fast and robust vegetation segmentation, followed by a deeper CNN to classify the segmented pixels between crops and weeds. Beeharry and Bassoo[39] evaluated UAV-based weed detection algorithms, showing AlexNet's accuracy exceeded 99%, compared to 48% for ANN on the same dataset. Ramirez et al[40] compared their aerial image weed segmentation model with SegNet and U-Net, finding that balanced data and better spatial semantic information improved accuracy. You et al[41] introduced a semantic segmentation method for weed detection based on deep neural networks (DNNs), enhancing segmentation accuracy through four additional components, making it effective for weeds of various shapes in complex environments. These methods do

not rely on image preprocessing and can autonomously extract useful features from images, outperforming traditional ML methods with manually designed features. CNN frameworks like AlexNet[42], ResNet[43], VGG[44], GoogleNet[45], U-Net, MobileNets, and DenseNet[46] are widely used in weed detection, standing out from conventional index-based methods. For example, Chechliński et al[46] measured four different plants in diverse growing places and light conditions, using a custom framework combining U-Net, MobileNets, DenseNet, and ResNet, achieving impressive results. CNNs also excel in plant disease detection with high precision. However, DL methods have drawbacks, including longer training times, higher complexity, greater computational costs, and the need for large training datasets. This requirement for extensive training samples poses a challenge in acquiring sufficient crop and weed images for weed detection, necessitating improvements to build effective models with limited datasets. The application of CNNs in weed segmentation has shown significant promise. They effectively separate crops from weeds in complex environments, leveraging their deep feature extraction capabilities. By learning spatial and semantic features, CNNs enhance the accuracy of weed segmentation tasks, making them a valuable tool in precision agriculture. In summary, CNNs are increasingly used for weed detection and segmentation due to their high accuracy and ability to automate feature extraction, addressing many challenges faced by traditional ML methods. While they have some limitations, ongoing research aims to optimize their performance even with smaller datasets. To provide a comprehensive comparison of DL methods, Table 3 summarizes the key features, advantages, and challenges of five prominent CNN-based architectures. This table highlights the primary characteristics of AlexNet, VGGNet, ResNet, GoogleNet (Inception), and U-Net, showcasing their contributions to the field of DL and their respective areas of application.

Table 3: Comparison of Typical CNN-Based DL Methods

Method	Features	Advantages	Challenges
AlexNet	Convolutional layers, max pooling, ReLU activation, fully connected layers	High accuracy, efficient for large-scale images, reduced overfitting with dropout	Computationally expensive, large number of parameters
VGGNet	Deep convolutional layers, small receptive fields, max pooling, fully connected layers	Improved accuracy with deeper layers, simple and uniform architecture	High memory consumption, slow to train due to depth
ResNet	Residual blocks, skip connections, batch normalization	Mitigates vanishing gradient problem, allows training of very deep networks	Complex architecture, increased computational cost
GoogleNet (Inception)	Inception modules, multiple filter sizes, auxiliary classifiers	Efficient computation with fewer parameters, high accuracy	Complex architecture, requires careful tuning of hyperparameters
U-Net	Encoder-decoder structure, skip connections, upsampling layers	Effective for image segmentation, works well with small datasets	High computational cost, requires large memory

2.5 Object Detection

Object detection is a fundamental technology in precision agriculture, enabling the identification and classification of various objects, such as crops and weeds, within images. Leveraging DL algorithms, object detection analyzes visual data to make informed decisions, enhancing farming efficiency and sustainability. object detection in an image involves several steps. First, an image is passed through a neural network, which processes it to extract features. This network, often a CNN, scans the image in a grid-like fashion, creating feature maps that highlight important aspects such as edges, textures, and patterns. These feature maps are then analyzed to predict bounding boxes around objects, classifying each box with a confidence score. The final output includes the locations and labels of the detected objects within the image. Popular models used for object detection include YOLO, Faster R-CNN, SSD (Single Shot MultiBox Detector), and Mask

R-CNN. YOLO models, particularly YOLOv5 and YOLOv8, are favored for their high speed and accuracy, making them suitable for real-time weed detection and crop monitoring. These models significantly improve farming efficiency by enabling precise weed identification and targeted interventions, ultimately supporting sustainable agricultural practices.

2.5.1 YOLO: A Paradigm Shift in Object Detection

YOLO is a state-of-the-art object detection algorithm that has gained widespread popularity due to its speed and accuracy. Unlike traditional object detection systems that apply a classifier to different regions of an image, YOLO frames object detection as a single regression problem, a single regression task, directly from image pixels to bounding box coordinates and class probabilities. This approach significantly speeds up the detection process while maintaining high accuracy.

2.5.2 YOLOv5 in Weed Detection

YOLOv5, developed by Ultralytics[47], is an enhanced version of the original YOLO algorithm. It incorporates several improvements that make it highly effective for agricultural applications, including weed detection. YOLOv5's architecture is optimized for real-time object detection, making it suitable for integration with UAVs and ground robots used in precision agriculture. Key features of YOLOv5 include: Speed and Efficiency: YOLOv5 can process images at a high frame rate, enabling real-time weed detection and monitoring. Accuracy: The model's ability to accurately detect and classify weeds amidst crops is due to its advanced feature extraction and learning capabilities. Flexibility: YOLOv5 can be trained on a variety of datasets, making it adaptable to different agricultural environments and weed species. Recent studies have highlighted the effectiveness of YOLOv5 in agricultural applications. For instance, a study by Xie et al[48] demonstrated that YOLOv5 could accurately detect weeds in maize fields with a high degree of precision and recall, outperforming traditional ML methods. The model's robustness to varying lighting conditions and its ability to detect weeds at different growth stages were key factors in its success. The real-time detection capabilities of YOLOv5 enabled efficient monitoring and management of weeds, reducing the reliance on manual scouting.

2.5.3 YOLOv8: The Next Generation

YOLOv8, also developed by Ultralytics[49], builds upon the strengths of its predecessors with further enhancements in speed, accuracy, and efficiency.

YOLOv8 introduces new architectural improvements and training techniques that make it even more powerful for object detection tasks, including weed detection. Key improvements in YOLOv8 include: Enhanced Backbone Network: YOLOv8 features a more efficient backbone network that improves feature extraction, leading to better detection accuracy. Advanced Data Augmentation: Techniques such as mosaic augmentation and cutout regularization are employed to improve the model's robustness to various image conditions. Optimized Training Process: YOLOv8 uses a more efficient training process that reduces overfitting and improves generalization across different datasets. Research Using YOLOv8 for Weed Detection Emerging research has begun to explore the potential of YOLOv8 in agricultural settings. Preliminary studies indicate that YOLOv8 can achieve even higher accuracy rates compared to YOLOv5. For instance, Zhang et al[50] utilized YOLOv8 for weed detection in rice fields and reported an accuracy improvement of 3-5% over YOLOv5, particularly in detecting weeds under challenging conditions such as dense foliage and varying sunlight.

2.5.4 Challenges of YOLO Models in Weed Detection

While YOLO models are effective for real-time object detection, they face significant challenges in weed detection. One major issue is their difficulty in detecting small, overlapping objects, which is common in dense vegetation. YOLO's grid-based detection can miss fine details and precise boundaries, leading to inaccurate classifications and missed detections. This limitation is particularly problematic in agriculture, where distinguishing between closely situated weeds and crops is critical. Moreover, YOLO models struggle with high variability in weed appearances due to different growth stages, lighting conditions, and occlusions by other plants. These challenges necessitate the use of image segmentation models like U-Net or Mask R-CNN, which provide pixel-level accuracy essential for precise differentiation. Segmentation models excel in tasks requiring detailed analysis, offering improved performance by addressing the precision challenges posed by YOLO models in complex agricultural settings. Therefore, while YOLO is advantageous for speed, image segmentation models are crucial for accuracy in weed detection.

2.6 Image Segmentation

Image segmentation is a key image processing task that groups pixels into coherent regions based on similarities. Unlike general image classification, which assigns a single label to an entire image, semantic segmentation classifies each pixel, making the process more detailed and exhaustive. Traditional methods of segmentation include techniques like thresholding, region-based methods, edge detection,

watershed, and clustering. These methods rely heavily on local pixel differences and gradients to define segments. However, the advent of DL has revolutionized image segmentation, allowing for more precise results. Neural network models such as SegNet [51] and U-Net[52] have shown superior performance by leveraging DL techniques. These models learn to recognize complex patterns and features within images, making them highly effective for tasks requiring detailed pixel-level classification. A comparison between object detection and different types of image segmentation is made in Figure3.

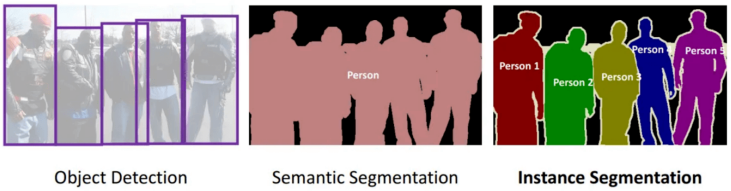

Figure 3: Comparison of Object Detection, Semantic Segmentation, and Instance Segmentation. Object detection identifies and locates objects within an image using bounding boxes. Semantic segmentation classifies each pixel into a category without distinguishing instances. Instance segmentation combines the strengths of both methods by identifying and delineating individual instances within the image

2.6.1 Weed Segmentation

Image segmentation is vital for vision-based tasks, including weed detection in precision agriculture. Segmentation of imagery data is crucial before further processing, such as quantifying weed presence and determining pesticide requirements. One of the main challenges in crop and weed segmentation is distinguishing their visual characteristics. Conventional segmentation techniques generally involve two stages: pre-processing and pixel-based classification. Pre-processing often includes image enhancement to reduce noise and illumination effects. In pixel-based segmentation, several traditional methods are used, including color index-based segmentation, threshold-based segmentation, and learning-based segmentation [30]. These methods often employ classifiers such as Random Forest and SVMs to categorize pixels as belonging to either weeds or crops. DL approaches, such as CNNs and DNNs [41], have been increasingly used to enhance the performance of traditional classifiers. These models are capable of learning complex patterns and features, which significantly improves segmentation accuracy. An illustration of Conventional Segmentation Methods is shown in Fig4.

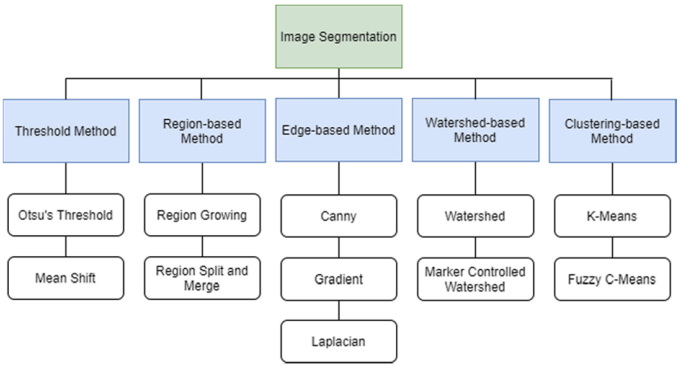

Figure 4: Conventional Segmentation Methods

Moreover, multi-modal approaches using RGB and hyperspectral Cameras have demonstrated promising outcomes in managing variations in weed and crop appearances, further improving segmentation accuracy. By integrating data from multiple sources, these approaches can better distinguish between different plant species under varying conditions, making them highly effective for precision agriculture.

2.6.2 U-Net-based segmentation

U-Net is a DL model designed specifically for image segmentation, introduced by Ronneberger, Fischer, and Brox in 2015[10]. It has become one of the most popular models in the field due to its ability to perform well with limited training data. The architecture of U-Net consists of two main parts: downsampling and upsampling. During downsampling, pooling layers reduce the image resolution while maintaining the number of channels, which helps in abstracting information and reducing complexity. The upsampling part then uses convolutional layers to increase the resolution, producing more precise outputs. The U-Net model has been widely applied across various image segmentation tasks, including medical imaging, remote sensing, and computer vision. It has been shown to outperform other DL models such as SegNet [51], FCNs, and ResNet in several benchmarks. In the context of weed detection, U-Net has been employed effectively. For instance, Asim et al [53] utilized U-Net alongside vegetation indices to detect weeds. Moreover, U-Net has been used for background removal by segmenting leaves and soil, facilitating tasks such as disease detection in cassava leaves[54] and herbal leaf classification [55]. This model's ability to handle such varied tasks underscores its versatility and robustness in image segmentation. An illustration of U-Net architecture is shown in Fig.5.

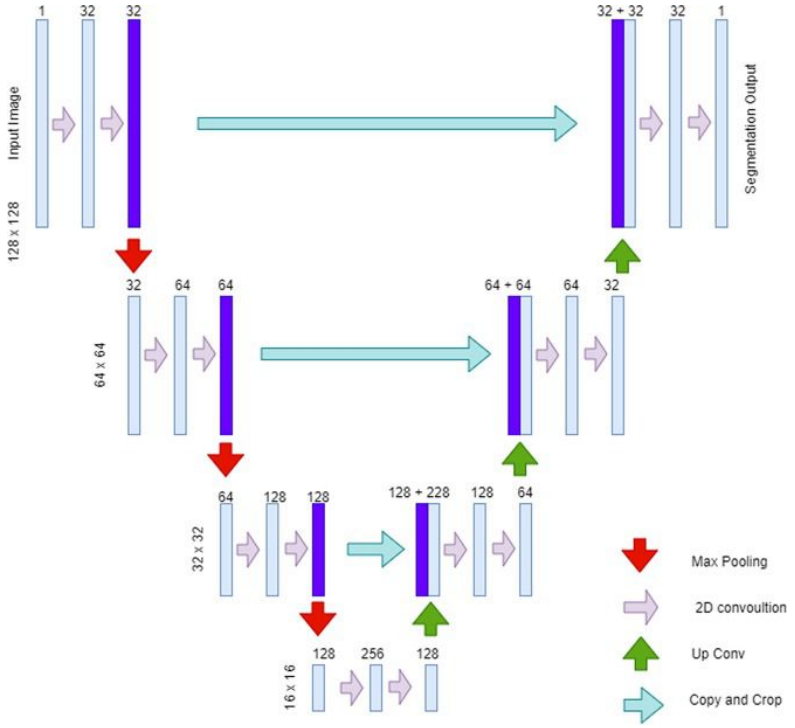

Figure 5: U-NET Architecture

Chapter 3

Methods

3.1 The ACRE Dataset

"The data collection for the ACRE dataset took place from June 8th to June 10th, 2022, during the first ACRE Field Campaign at the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) site in Montoldre, France. The experimental plots included two types of crops: maize (Zea mays) and beans (Phaseolus vulgaris). Moreover, four types of weeds were planted: ryegrass (Lolium perenne), mustard (Sinapis arvensis), matricaria (Matricaria chamomilla), and lamb's quarter (Chenopodium album). Maize was sown in two-row plots with a spacing of 0.75 meters between rows, while beans were sown in three-row plots with a 0.375-meter spacing between rows. The planting density for maize was 0.14 meters and for beans 0.07 meters between plants within the same row. These plots, each approximately 40 meters in length, were sown on May 19th, 2022. By the time of data collection, the bean plants had grown to a height of 0.08 to 0.1 meters, and the maize plants were between 0.1 and 0.15 meters tall. Data was gathered using a four-wheel skid-steering robot equipped with a Basler acA2000-50gc RGB camera, mounted perpendicularly to the ground. The camera, with a resolution of 2046 x 1080 pixels, captured images as the robot, teleoperated at an average speed of 0.2 meters per second, moved through the fields. Data collection occurred at different times of the day to capture various lighting conditions, including diffused sunlight using a white tissue and direct sunlight. A total of ten acquisition batches were performed, resulting in 1000 RGB images stored in ten folders, each representing a batch. Every image comes with an XML file containing instance segmentation annotations. The ACRE dataset is valuable for the development and enhancement of weed detection algorithms, applicable in tasks such as object detection, semantic segmentation, and instance segmentation"[56].

3.1.1 Annotation

"The annotation process for the ACRE dataset was supervised by the French Laboratoire national de métrologie et d'essais (LNE) to avoid potential machine learning biases. Manual annotations were carried out by subcontractors, ensuring accuracy and thoroughness. Each annotation in the dataset includes several critical elements: Polygons that precisely define the boundaries of each plant, referred to as "clipping" in the annotation XML files. If possible, the coordinates of the center of each plant where the stem meets the ground. Labels indicating the type of plant (crop, weed, or unknown species). Labels with the plant's name, either selected from a predetermined list or entered manually. A high-quality annotation should use as many vertices as needed to accurately outline the plant's shape, with a minimum of three vertices per polygon. Annotators follow specific rules for ambiguous cases, such as foliage gaps, plant overlaps, plants at the image edge, "natural" weeds (different from seeded varieties), blurred plants, and plants outside the field area. Annotators receive training from an expert using LNE-DIANNE software and practice on different image types. An example annotation is illustrated in Figure 6. After training, a "test batch" of 10% of the images is annotated and undergoes quality control by LNE to identify issues and provide guidance for improvement. Additionally, 10% of the images are annotated twice to ensure inter-annotator consistency, and a manual random sample quality control is conducted. Subcontractors report on the task and any encountered difficulties, noting image characteristics that caused ambiguity, such as high plant density, sunny weather leading to overexposure and harsh shadows, and motion blur from tall maize plants[56]." An illustrative example of an annotation is shown in Figure6.

Figure 6: The screenshot of the annotations displays crops outlined in green and weeds highlighted in yellow. Each plant's center is marked with a cross symbol.

3.1.2 Data Processing and Splitting

The ACRE dataset has been efficiently processed using Roboflow[57], a robust platform designed to streamline the preparation and management of image datasets for machine learning projects. Roboflow enhances the ACRE dataset by providing comprehensive tools for augmentation and preprocessing, which are crucial for training deep learning models. In the case of the ACRE dataset, Roboflow facilitated several key preprocessing steps. These included auto-orienting images to ensure consistent alignment and resizing the images to a uniform dimension of 640x640 pixels. Standardizing the input size is essential for optimizing model performance and ensuring uniformity during the training process. The dataset was methodically split into three distinct sets:700 images for training, 200 images for validation, and 100 images for testing were used. This strategic division ensures that the models have sample data for learning, validating, and testing, allowing for accurate evaluation and adjustment of the model's performance. By leveraging the capabilities of Roboflow, researchers can ensure high-quality preprocessing of the ACRE dataset, which is pivotal for developing robust and effective weed detection models. This preparation supports the training of deep learning models that can accurately identify and distinguish between crops and weeds, thereby advancing precision agriculture.

3.1.3 Model Architecture and Hyperparameter Configuration

All experiments were conducted using YOLO models developed by Ultralytics. Our primary goal was to identify the optimal configuration of hyperparameters and model size for weed detection on the ACRE dataset. To achieve this, we trained models on the ACRE dataset using various splits for training and validation. We did not exclude any images due to annotation errors or low quality; instead, we utilized the entire dataset as provided. We employed different YOLO versions for our experiments, specifically YOLOv5 and YOLOv8. For YOLOv5, we used the yolov5s and yolov5m models, while for YOLOv8, we experimented with yolov8s, yolov8m, and yolov8n models. The dataset was divided into training and validation sets to ensure robust model evaluation, with performance showcased on the ACRE "test_dev" split. To test the impact of various hyperparameters, we experimented with image sizes of 640x640, 1024x1024, and 1600x1600 pixels, and batch sizes of 16, 32, and 64. Notably, we observed that the choice of model size had minimal impact on the resulting mAP95 (mean Average Precision at an IoU=[0.50:0.95]) and mAP50 (mean Average Precision at an IoU=0.50). Throughout the training process, we closely monitored losses and metrics to prevent overfitting. We found that training for 50 epochs was sufficient for our dataset/batch combination. This careful monitoring and adjustment ensured that the models were effectively trained without overfitting, which is crucial for achieving high accuracy in weed detection tasks. In summary, the experiments conducted with YOLOv5 and YOLOv8 on the ACRE dataset helped us determine the most effective configurations for weed detection, taking into account various model sizes, image resolutions, and batch sizes. This process is essential for optimizing the performance of deep learning models in practical agricultural applications.

3.2 The Sunflower Dataset

For the segmentation part of the research, we utilized a publicly available dataset known as the Sunflower dataset Fawakherji et al[58]. Examples of scenes taken from the sunflower farm on different days are shown in Fig7. This dataset was specifically chosen for its relevance in crop and weed segmentation tasks. The dataset was gathered with a 4-channel (RGB + NIR) JAI AD-13 camera mounted on an agricultural robot in Jesi, Italy. The dataset comprises three subsets, Totaling 500 scene images, which were captured at various times and on different days, these images represent three distinct stages: the emergence stage, a subsequent growth stage, and the final stage for applying chemical treatments. From these, we used 318 images from the first two stages for our experiments. The

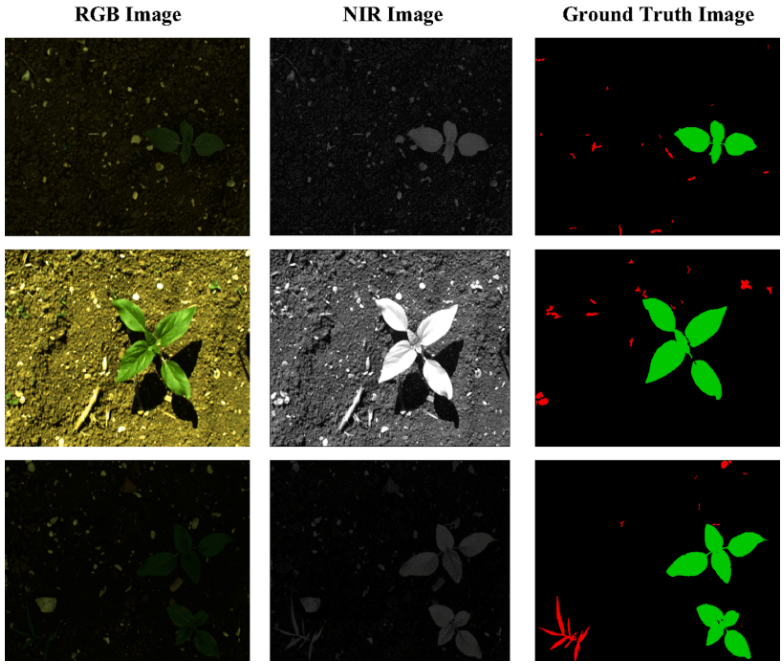

Figure 7: "An example of scenes from the field, from top to bottom: the emergence state, a subsequent growth state, and the last state for applying chemical treatments"[3]

U-Net model was trained using 254 images and tested on 64 images using the hold-out method to evaluate segmentation performance. This approach involves splitting the dataset into two parts: one for training the model and the other for testing its performance. By doing so, we ensured that the model's effectiveness was evaluated on unseen data, providing a more robust measure of its segmentation capabilities and helping to prevent overfitting. To ensure that images used for training were not used for testing, we employed the `random split` method from PyTorch, which guarantees a random split and that no image in the training set is included in the test set. The hold-out method is particularly advantageous as it helps prevent overfitting, ensuring that the model generalizes well to new, unseen images. This method enabled us to assess the U-Net model's effectiveness in accurately segmenting crops and weeds across various growth stages, demonstrating its potential application in precision agriculture. The ground truth images in this dataset are annotated with specific colors: black for soil, green for crops, and red for weeds.

3.2.1 Challenges of dataset

The field environment introduces several challenges in creating datasets and obtaining image samples for the Sunflower dataset. After planting the crops, they may experience various stresses and weather conditions that are not easily controllable. Additionally, due to data collection occurring on different days and at different times, the lighting or illumination conditions can vary significantly across scene images, as demonstrated in this dataset. This variability in lighting can complicate the training process for DL models. There are also potential issues such as mislabelled crop or weed samples and changes in crop or weed boundaries within the same samples. Mislabelled samples can significantly impact the performance of a deep learning model, leading to incorrect results. For instance, if a crop sample is mistakenly labeled as soil, the model might learn incorrect associations and produce erroneous outputs when applied to new data. Similarly, if a mislabelled sample is used in the test set, it would lead to an inaccurate evaluation of the model's performance because the mask IoU calculation would consider the incorrect label as the ground truth. The soil background in the field is often not purely soil; it may include aggregates, rocks, dead leaves, and other elements. For example, a dead leaf labeled as a weed in the ground truth image could be misclassified as soil by the model. This misclassification may contribute to a lower IoU for the weed class, which is already a minority class in the dataset. Furthermore, the pixel values at the boundaries of crops and weeds can change across consecutive scene images of the same sample. This variability can affect segmentation results because the model might be trained or tested on different labels for the same pixel values. This issue is likely due to the natural field environment and the resolution limitations of the image sensors used. Ensuring the accuracy and consistency of labels in both the training and test data is essential. Accurate labelling enables the model to learn correct relationships and provides a reliable evaluation of its performance, thereby improving its effectiveness in real-world applications.

3.2.2 Image Pre-processing

In the preprocessing step of this thesis, two significant techniques were applied to enhance the quality and utility of the dataset: NDVI and the Gaussian Bilateral Filter on NIR images. These methods are essential for improving the accuracy and reliability of the models used for weed detection and segmentation.NDVI is a widely used remote sensing index that measures vegetation health by comparing the red and NIR light reflected by vegetation. The formula for NDVI is:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{3.1}$$

This index ranges from -1 to 1, where higher values indicate healthier and denser vegetation. NDVI is particularly useful in agricultural applications for monitoring crop health, detecting plant stress, and differentiating between crops and weeds. By incorporating NDVI into the preprocessing pipeline, the model benefits from enhanced contrast between vegetation and non-vegetation areas, improving its ability to accurately identify and segment crops and weeds[59]. The Gaussian Bilateral Filter is a non-linear, edge-preserving, and noise-reducing smoothing filter[60]. It combines the advantages of Gaussian smoothing and edge-preserving properties by taking into account both the spatial closeness and the intensity difference of neighboring pixels. The bilateral filter works as follows:

$$\text{filtered}(x) = \frac{1}{W_p} \sum_{y \in \Omega} I(y) \cdot \exp\left(-\frac{\|x-y\|^2}{2\sigma_s^2}\right) \cdot \exp\left(-\frac{\|I(x)-I(y)\|^2}{2\sigma_r^2}\right) \tag{3.2}$$

where: $I(y)$ is the intensity value at pixel y, , Ω is the neighborhood of pixel x, ,σ_s is the spatial standard deviation, , σ_r is the range standard deviation, , W_p is the normalization factor, defined as:

$$W_p = \sum_{y \in \Omega} \exp\left(-\frac{\|x-y\|^2}{2\sigma_s^2}\right) \cdot \exp\left(-\frac{\|I(x)-I(y)\|^2}{2\sigma_r^2}\right) \tag{3.3}$$

In this study, the Gaussian Bilateral Filter was applied to NIR images to reduce

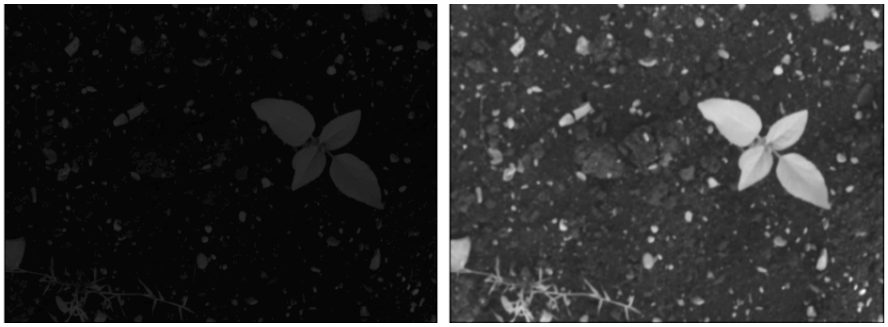

Figure 8: NIR image (left) and the result of the bilateral filter image (right)

noise while preserving important edges, which is crucial for accurate segmentation tasks. By smoothing out noise and maintaining the integrity of edges, the bilateral filter enhances the clarity and quality of the images, leading to better performance of the deep learning models. A NIR image and its corresponding Filtered-NIR are shown in Fig 8. Furthermore,in agricultural datasets, such as the sunflower dataset, certain data augmentation techniques may not always be appropriate. Vertical flips, for example, are not realistic because crops and weeds typically grow upwards due to gravity, and it is rare to observe them upside down naturally.

Horizontal flips might be more acceptable as they can represent a change in the camera's perspective relative to the rows of crops. Rotations can also be useful if the camera capturing the images might be tilted or if the scenes might be observed from different angles. However, after conducting experiments with various data augmentation techniques, it was found that they did not improve the results and, in some cases, led to lower performance. As a result, no data augmentation techniques were used in this part. This decision was made to ensure that the training data remained representative of real-world conditions, thus improving the model's generalization and performance. Moreover, the function transforms.Normalize(mean, std) is a common preprocessing step used in image processing with deep learning models, particularly when using the PyTorch library. we used this function to normalize the image pixel values to have a specific mean and standard deviation, which helps the neural network to learn more effectively. Normalizing the images helps the model to converge faster during training as it standardizes the input, ensuring that the input features have a similar range and distribution. This is a crucial step in the preprocessing pipeline for training deep learning models.

3.2.3 Proposed inputs and Model Architecture

In this work, three different input combinations for the UNet-ResNet50 (UNet with ResNet50 backbone for enhanced image segmentation) model were proposed to segment weed, crop, and soil areas in agricultural fields: RGB+NIR(4-channel input), G+NIR+NDVI(3-channel input), and G+NIR(2-channel input). Following figures show examples of the different channel combinations used. The combined image of RGB and NIR channels (9) provides a broad spectrum of information. This combination integrates the standard RGB channels with the NIR channel. The addition of the NIR channel enhances the ability to distinguish between crops and weeds because NIR can capture details not visible in the RGB spectrum. Vegetation typically reflects more NIR light compared to soil, which helps in identifying plant health and differentiating between different types of plants. For the 4-channel configuration (RGB+NIR), the first convolutional layer of the UNet-ResNet50 model was modified to accept four input channels. This modification allowed the model to combine the detailed color information from the RGB channels with the additional spectral information provided by the NIR channel. The combination of Green, NIR, and NDVI channels (10) offers enhanced contrast and feature detection. This combination uses the Green channel, the NIR channel, and the NDVI. NDVI highlights areas with healthy vegetation, as it measures the contrast between red and near-infrared reflectance of vegetation. The use of the Green channel in this combination is particularly effective as it has been shown to significantly influence weed detection. This combination offers enhanced

contrast and highlights the vegetation, aiding in precise segmentation. The 3-channel configuration (G+NIR+NDVI) involved adjusting the first convolutional layer to accept three input channels, allowing the network to utilize crucial spectral information. while the Green and NIR channels (11) focus on specific spectral bands that are significant for weed detection. The Green channel is sensitive to vegetation, and when combined with the NIR channel, it helps in highlighting the plants more distinctly against the soil background. This combination is simpler yet effective for certain segmentation tasks, leveraging the strong response of vegetation in these spectral bands. The Green channel (G) was

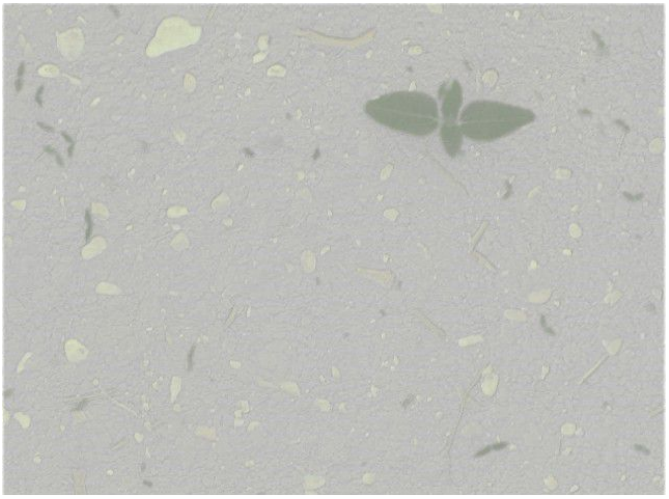

Figure 9: Combination of RGB and NIR channels enhancing the visibility of crops and weeds.

included in these combinations because, as indicated by recent state-of-the-art research[27], the Green channel has a more significant impact on weed detection than the Blue or Red channels. The original image size of 964x1296 was maintained for all models to preserve the spatial resolution essential for precise segmentation tasks. Each model also applied a Gaussian Bilateral Filter on the NIR images to enhance image quality by reducing noise while preserving edges. The number of epochs for training was set to eight due to memory constraints and the observation that additional epochs did not result in significant improvements. The hyperparameters used in these experiments are detailed in Table 4. In order to answer to the Research Question 3 presented in chapter1 we utilized the U-Net model with a ResNet-50 backbone. In this context, the idea of transfer learning is fundamental. By leveraging a pre-trained ResNet-50 model on an RGB dataset, we aimed to transfer the learned features to the multispectral domain. The ResNet-50 model, pre-trained on large-scale RGB image datasets, captures low-level to high-level features that are essential for effective image segmentation.

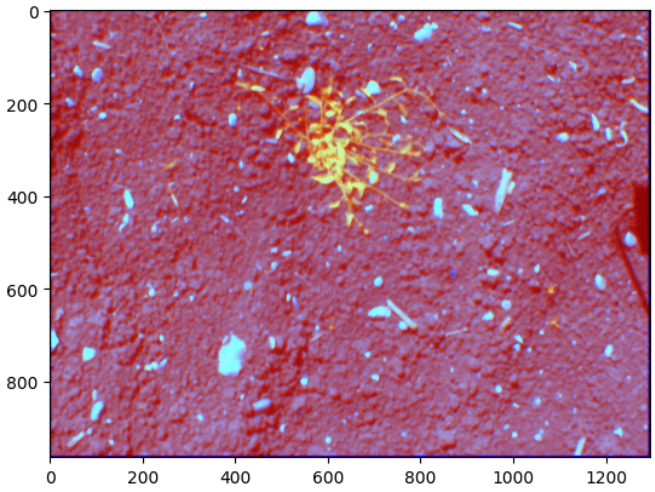

Figure 10: Combination of Green, NIR, and NDVI channels emphasizing the crop and weeds through high NDVI values.

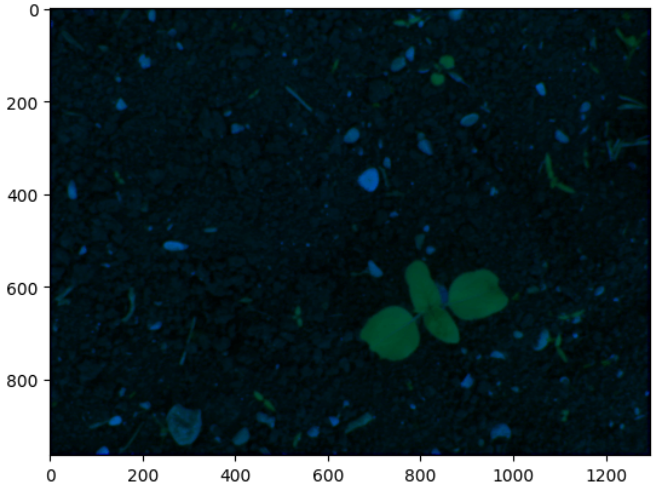

Figure 11: Combination of Green and NIR channels highlighting the crop and weeds.

These features include edges, textures, and complex patterns that are crucial for distinguishing between crops and weeds. When the pre-trained model is fine-tuned on multispectral images, the additional spectral information, particularly from NIR channel, provides complementary data that can significantly improve segmentation accuracy. For the ResNet-50 backbone and RGB-NIR channels, the parameters for the additional NIR channel in the first convolutional layer are not

Table 4: Hyperparameters for the U-Net model

Hyperparameters	**Value**
Epochs	8
Batch Size	2–4
Loss Function	Combined Loss (Dice Loss and Cross-Entropy)
Activation	Softmax
Optimizer	Adam
Evaluation Metric	IoU Score

present in the pre-trained model because it was trained on RGB images. To handle this, the first convolutional layer is modified to accept four input channels (RGB+NIR). The weights for this new convolutional layer are initialized randomly since pre-trained weights are only available for the original three RGB channels. During training, the entire network, including this new layer, is fine-tuned on the multispectral dataset to learn the appropriate features for all four channels. The ResNet-50 architecture is known for its depth and ability to capture intricate features through its residual learning framework. This architecture, when integrated into the U-Net framework, combines the robust feature extraction capabilities of ResNet-50 with the precise segmentation abilities of U-Net. Additionally, a custom U-Net architecture was implemented from scratch, with specific layers designed to optimize performance for this task. The custom architecture includes double convolution layers for downsampling, max pooling layers, and upsampling layers, which combine features from different levels of the network to produce high-resolution segmented images. Moreover,in this thesis, the U-Net architecture was selected due to its exceptional suitability for image segmentation tasks, especially in scenarios with high class imbalance, such as weed detection. U-Net is a CNN designed to accurately identify both the overall shapes and fine textures of objects within images. One of its key advantages is the incorporation of skip connections between the down-sampling and up-sampling paths, which allows the network to utilize both local and global information for decision-making. This feature significantly improves the accuracy of the model, particularly when the positive class (weeds) is underrepresented in the training data. Furthermore, U-Net's design is computationally efficient, an essential attribute for real-time applications in weed detection. The network's relatively small number of parameters ensures that it can perform effectively without requiring extensive computational resources. These characteristics make U-Net an ideal choice for our segmentation tasks, as it balances accuracy with efficiency. To optimize the training process, a combination of Dice Loss and CrossEntropyLoss was used as the loss function. Dice Loss is particularly effective for segmentation

tasks as it directly optimizes the overlap between the predicted and true masks, which is crucial for accurate segmentation. CrossEntropyLoss, on the other hand, is widely used for multi-class classification problems and helps to differentiate between the classes effectively. Combining these two loss functions leverages their strengths, improving both the accuracy and robustness of the segmentation model. For models incorporating NDVI, a custom NDVI thresholding function was implemented. This function calculates the NDVI index using the NIR and Red channels, scales the values to a range of 0-255, and applies a threshold to generate a binary mask. To implement these models, an RGB+NIR camera is necessary to capture the required channels. Although this might seem like a drawback, NIR cameras have become widely available and affordable. NIR images provide crucial information about the target area, which enhances segmentation accuracy. As demonstrated by [5], NIR cameras are valuable for segmentation tasks. BoniRob, an agricultural field robot with sensors and a JAI AD-130GE camera, utilizes an RGB+NIR camera, highlighting its practical applications in agricultural robotics.

Chapter 4

Results and Discussions

4.1 Evaluation Metrics

Evaluation metrics play a crucial role in understanding the performance of machine learning models, particularly in tasks like object detection and image segmentation. The appropriate choice of metrics provides insights into the strengths and weaknesses of the models and guides improvements. This section covers key evaluation metrics, including "Precision, Recall, mAP, and IoU."

4.1.1 Evaluation metrics for Object Detection Models

Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is a measure of the accuracy of the positive predictions made by the model. Mathematically, it is expressed in the formula 4.1.

$$Precision = \frac{TP}{TP + FP} \tag{4.1}$$

where TP is the number of true positives(number of correctly identified objects), and FP is the number of false positives(number of incorrectly identified objects). In agricultural settings, high precision in weed and crop detection ensures that the majority of identified weeds are indeed weeds, minimizing the risk of mistakenly identifying crops as weeds. This is vital because inaccurate identification can lead to unnecessary application of herbicides, which not only harms the crops but also wastes resources. Accurate detection is essential for efficient resource use and maintaining crop health, directly impacting agricultural productivity and profitability. Precision plays a significant role in reducing the misuse of herbicides. In precision agriculture, herbicides are applied based on the detection of weeds. High precision ensures that herbicides are applied only where necessary, thus preventing damage to crops and reducing chemical usage. Misidentifying crops as

weeds (false positives) leads to the unnecessary application of herbicides, harming crops and wasting resources. Therefore, precision is key to achieving sustainable agricultural practices. Moreover, precision in weed detection enhances crop yield by ensuring that weed removal processes are targeted and effective. Accurate identification and removal of weeds allow crops to grow without competition, promoting better growth and higher yields. This is particularly important in densely planted fields where competition for resources is intense. Achieving high precision in weed and crop detection, however, is challenging due to the visual similarities between crops and weeds, especially in their early growth stages. This similarity can make it difficult to distinguish between them. Additionally, varying outdoor conditions, such as different lighting and shadows, can affect the appearance of plants, making accurate detection even more complex. Overlapping plants and occlusions further complicate the task, requiring sophisticated models capable of handling these complexities. The quality of the training dataset and the accuracy of annotations significantly impact the model's precision. Properly labeled data is crucial for the model to learn the differences between crops and weeds accurately. High-quality annotations ensure that the model receives accurate information during training, leading to better performance. Recall (also known as Sensitivity or True Positive Rate) is the ratio of correctly predicted positive observations to all observations in the actual class. It indicates the model's ability to capture all relevant cases. The formula for Recall is:

$$Recall = \frac{TP}{TP + FN} \tag{4.2}$$

where TP is the number of true positives(number of correctly identified objects), and FN is the number of false negatives(the number of actual positive instances that were missed by the model). In agricultural object detection, high recall is particularly important because it ensures that most, if not all, of the crops and weeds present in the images are correctly identified. This is vital for several reasons. Firstly, effective weed management relies on identifying all weeds; missing any can lead to inadequate weed control, allowing the weeds to continue competing with crops for essential resources like nutrients, water, and sunlight, which can negatively impact crop yields and the overall health of the agricultural field. Secondly, accurate recall ensures that all crops are detected, which is essential for precise agricultural practices such as targeted fertilization, irrigation, and harvesting. Missing a crop could result in parts of the field not receiving the necessary attention, leading to suboptimal farming outcomes. Thirdly, in scenarios where the detection system informs the application of herbicides or other treatments, high recall ensures that all necessary areas are treated, preventing patches of weeds from being missed and subsequently proliferating.
mAP is a widely used metric for evaluating object detection models. It provides a

comprehensive performance measure by combining precision and recall across different classes and IoU thresholds. The calculation of mAP involves several steps: The first step in computing mAP is to determine the Average Precision (AP) for each class. The AP for each class is determined by calculating the area under the precision-recall curve. This curve plots precision against recall for different confidence thresholds. Once the AP is calculated for each class, the mAP is determined by taking the mean of these AP values across all classes. This provides a single scalar value that summarizes the model's performance across different object categories. There are different variations of mAP based on the IoU thresholds used: mAP@0.5 (mAP50): This metric calculates the average precision at a single IoU threshold of 0.5. It indicates how well the predicted bounding boxes overlap with the ground truth boxes by at least 50%. mAP@0.5:0.95 (mAP50-95): This metric calculates the average precision over multiple IoU thresholds ranging from 0.5 to 0.95 in steps of 0.05. It provides a more robust evaluation by considering the model's performance across a range of IoU thresholds. In the context of the ACRE dataset, mAP is used to evaluate the performance of YOLOv5 and YOLOv8 models in detecting weeds and crops. This dataset comprises various images with annotated bounding boxes for weeds and crops, making it ideal for object detection evaluation. By calculating mAP, we can understand how well these models perform in accurately identifying and localizing weeds and crops under different IoU thresholds.
The confusion matrix is a tool that provides a more detailed breakdown of the model's performance by displaying the counts of TP, true negative (TN), false positive(FP), and FN predictions. It helps to visualize the performance of the model and is instrumental in understanding the balance between precision and recall. The confusion matrix is structured as follows:

Table 5: Confusion Matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

- **True Positives (TP)**: The number of correctly predicted positive instances.
- **True Negatives (TN)**: The number of correctly predicted negative instances.
- **False Positives (FP):** The number of incorrectly predicted positive instances.
- **False Negatives (FN):** The number of incorrectly predicted negative instances.

Precision and Recall are often used together to provide a more comprehensive evaluation of a model's performance. They can be combined into a single metric known as the F1 Score, and it is calculated as the harmonic mean of Precision and Recall. The F1 Score provides a balance between Precision and Recall, especially useful when dealing with imbalanced datasets, where one class (e.g., weeds) is more prevalent than the other (e.g., crops). The formula for the F1 Score is:

$$F1\,\text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.3}$$

4.1.2 Evaluation Metrics for Semantic Segmentation Models

The performance of a segmentation model can be gauged by evaluating the quality of the segmentation results, which involves comparing how accurately the shapes and positions of the segmentation outputs align with the true labels or target masks. The most commonly used evaluation metric for this purpose is IoU. IoU is a robust metric that quantifies the overlap between the predicted segmentation and the ground truth. IoU is calculated as follows:

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}} = \frac{|A \cap B|}{|A \cup B|} \tag{4.4}$$

where A represents the predicted segmentation and B represents the ground truth segmentation. The intersection ($|A \cap B|$) is the number of pixels where the predicted and true segmentations overlap, while the union ($|A \cup B|$) is the total number of pixels present in both the predicted and true segmentations combined. Using IoU, we can comprehensively evaluate how well the model performs in segmenting different classes within the dataset. High IoU values indicate better performance, signifying that the model's predictions closely match the actual segmentation masks. In the context of the Sunflower dataset, IoU is particularly useful due to the presence of multiple segmentation challenges such as varying lighting conditions, overlapping plants, and different growth stages.

4.2 Exprimental Results

4.2.1 Results for ACRE Dataset

Results for Different Hyper-parameter Settings of the YOLOv5 Model

The table shows the performance of the YOLOv5s model on the ACRE dataset. It includes metrics for overall performance, as well as specific results for detecting

Table 6: Performance of a single model over the test_dev split of the ACRE dataset (model size: yolov5s - small).

Class	Images	Instances	Precision (P)	Recall (R)	F1 Score	mAP50	mAP50-95
All	100	11472	0.774	0.682	0.725	0.721	0.425
Crop	100	1205	0.85	0.807	0.828	0.845	0.577
Weed	100	10267	0.699	0.557	0.619	0.598	0.273
Epochs	50						
Batch Size	16						
Image Size	640						
Model Size	yolov5s						

crops and weeds. Additionally, the table lists the model's training parameters, including the number of epochs, batch size, and image size used during training.

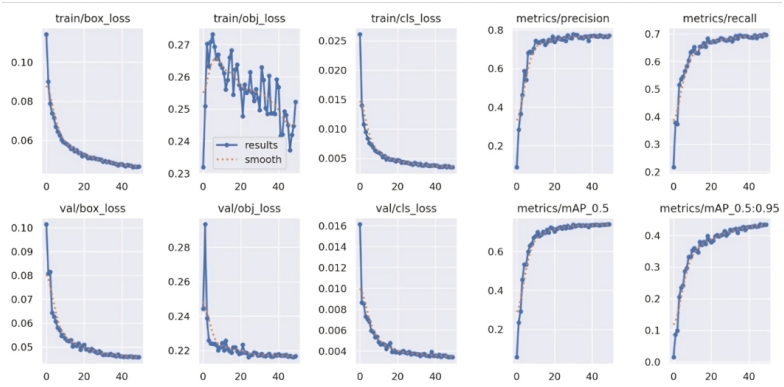

Figure 12: Training and Validation Loss and Metrics Plots

- These plots display the training and validation loss for the bounding box (box), objectness (obj), and class (cls) losses, as well as the metrics for precision, recall, mAP@0.5, and mAP@0.5:0.95 during the training of the YOLOv5s model over 50 epochs. These plots help in understanding the model's learning behavior and convergence during the training process. Components in the Plots: Train/Validation Box Loss: Measures the accuracy of the predicted bounding box coordinates compared to the ground truth. Train/Validation Objectness Loss: Evaluating how well the model predicts whether an object exists in a given box. Train/Validation Class Loss: Assesses the accuracy of the predicted class probabilities for each object. The training and validation losses for the bounding box (box), objectness (obj), and class(cls) show a steady decrease over the 50 epochs, indicating that the model is learning effectively and converging well. Lower values in the validation loss compared to the training loss suggest that the model is not

overfitting and is generalizing well on the validation data.

- The following confusion matrix illustrates the performance of the YOLOv5s model in predicting crops, weeds, and background classes. It shows the true positive, false positive, and false negative rates for each class, providing insight into the model's accuracy and misclassification rates.

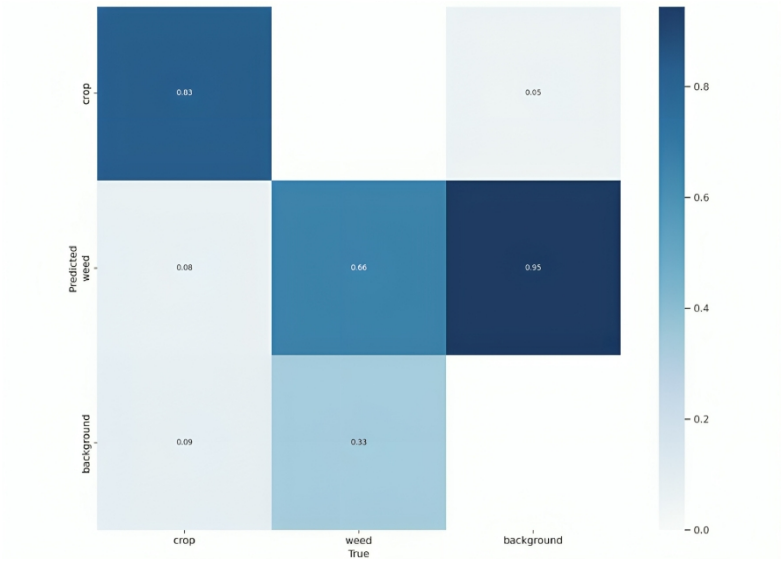

Figure 13: Confusion Matrix for Crop and Weed Detection

Table 7: Confusion Matrix Interpretation for Crop, Weed, and Background Classes

Class	Metric	Percentage	Description
Crop	True Positives (TP)	83%	Correctly identified crops
Crop	False Positives (FP)	5%	Incorrectly identified as crops
Crop	False Negatives (FN)	17%	Actual crops not identified
Weed	True Positives (TP)	66%	Correctly identified weeds
Weed	False Positives (FP)	8%	Incorrectly identified as weeds
Weed	False Negatives (FN)	33%	Actual weeds not identified

Table 8: Performance of a single model (yolov5s) over the test_dev split of the ACRE dataset.

Class	Images	Instances	Precision (P)	Recall (R)	F1 Score	mAP50	mAP50-95
All	100	11472	0.759	0.703	0.730	0.737	0.433
Crop	100	1205	0.829	0.831	0.830	0.863	0.587
Weed	100	10267	0.688	0.576	0.627	0.611	0.279
Epochs	50						
Batch Size	32						
Image Size	640						
Model Size	yolov5s						

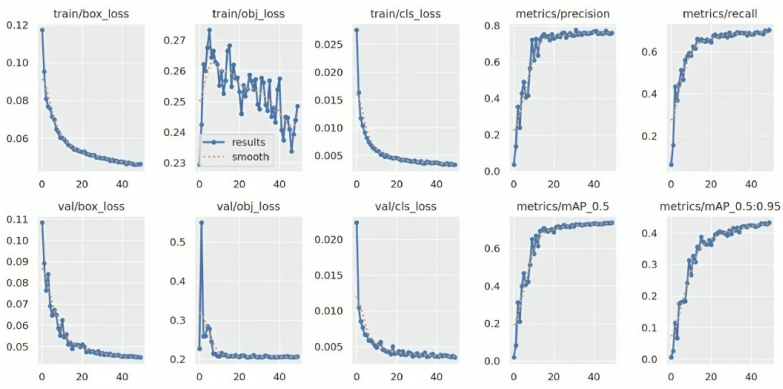

Figure 14: Training and Validation Loss and Metrics Plots

Table 9: Confusion Matrix Interpretation for Crop and Weed Classes

Class	Metric	Percentage	Description
Crop	True Positives (TP)	85%	Correctly identified crops
Crop	False Positives (FP)	6%	Incorrectly identified as crops
Crop	False Negatives (FN)	15%	Actual crops not identified
Weed	True Positives (TP)	68%	Correctly identified weeds
Weed	False Positives (FP)	39%	Incorrectly identified as weeds
Weed	False Negatives (FN)	32%	Actual weeds not identified

Table 10: Performance of a single model over the test_dev split of the ACRE dataset (model size: yolov5s - small).

Class	Images	Instances	Precision (P)	Recall (R)	F1 Score	mAP50	mAP50-95
All	100	11472	0.745	0.53	0.62	0.575	0.336
Crop	100	1205	0.822	0.581	0.682	0.647	0.433
Weed	100	10267	0.668	0.479	0.56	0.503	0.24
Epochs	50						
Batch Size	16						
Image Size	1024						
Model Size	yolov5s						

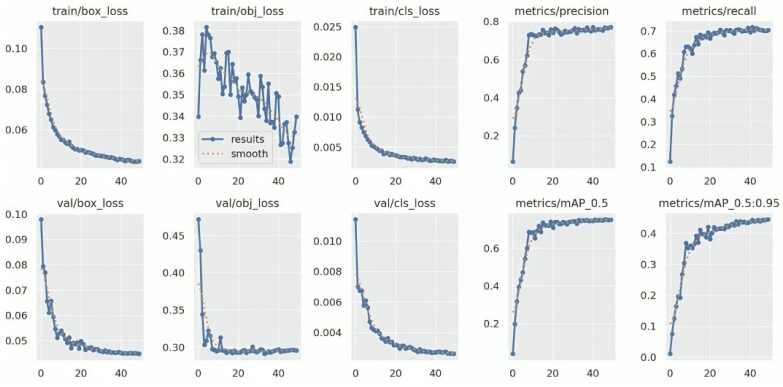

Figure 15: Training and Validation Loss and Metrics Plots

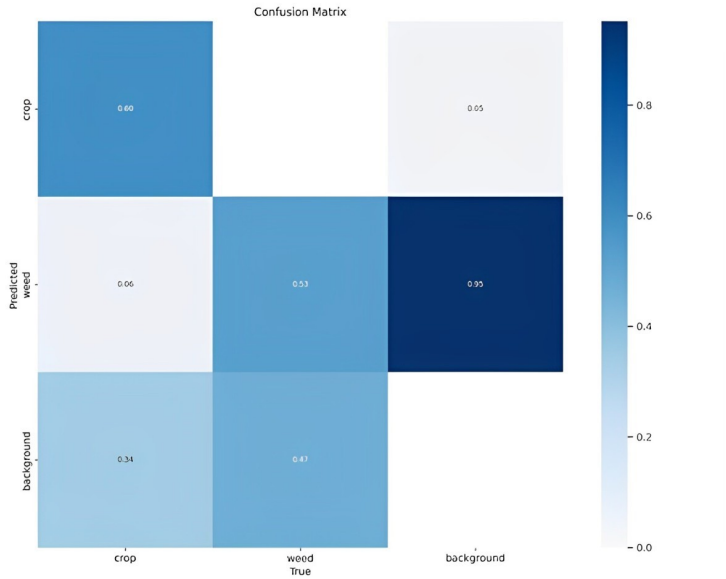

Figure 16: Confusion Matrix for Crop and Weed Detection

Table 11: Confusion Matrix Interpretation for Crop, Weed, and Background Classes

Class	**True Positives (TP)**	**False Positives (FP)**	**False Negatives (FN)**
Crop	60%	5%	40%
Weed	53%	6%	47%

Table 12: Performance of a single model over the test_dev split of the ACRE dataset (model size: yolov5m - medium).

Class	Images	Instances	Precision (P)	Recall (R)	F1 Score	mAP50	mAP50-95
All	100	11472	0.774	0.654	0.709	0.691	0.423
Crop	100	1205	0.855	0.752	0.800	0.789	0.553
Weed	100	10267	0.694	0.556	0.618	0.592	0.294
Epochs	50						
Batch Size	16						
Image Size	1024						
Model Size	yolov5m						

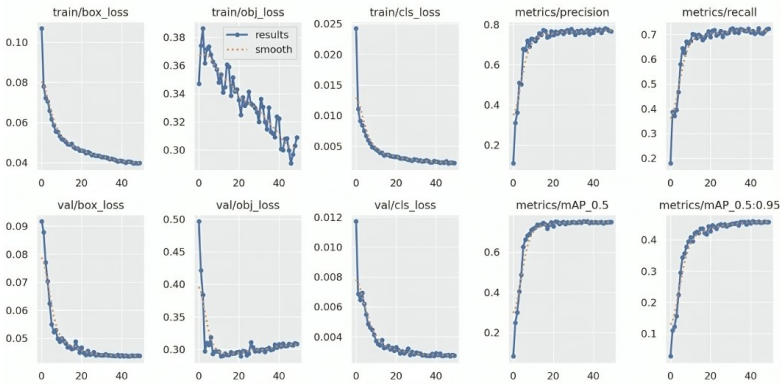

Figure 17: Training and Validation Loss and Metrics Plots

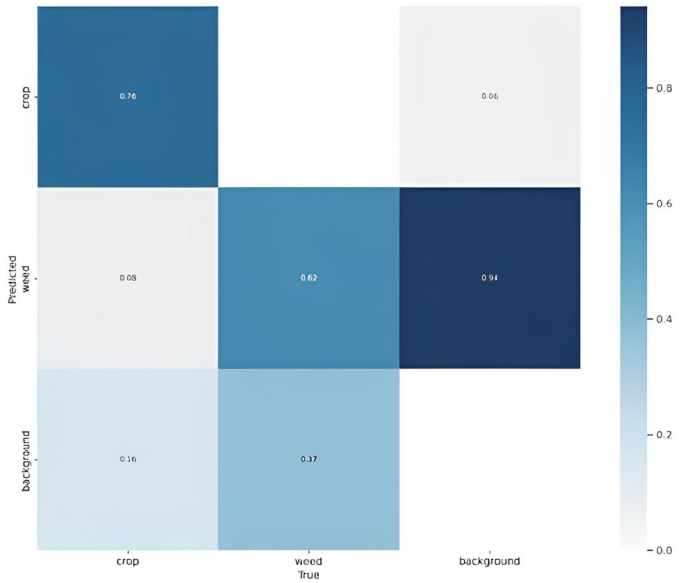

Figure 18: Confusion Matrix for Crop and Weed Detection

In this section, we interpreted the results from the confusion matrix for crop and weed classification. The table below provides an overview of TP, FP, and FN for both crop and weed classes.

Table 13: Confusion Matrix Interpretation for Crop, Weed, and Background Classes

Class	**True Positives (TP)**	**False Positives (FP)**	**False Negatives (FN)**
Crop	76%	6%	24%
Weed	62%	8%	37%

Table 14: YOLOv5 Model Performance Insights

Model	**Image Size**	**Batch Size**	**Performance**	**Insight**
YOLOv5s	640	16	Balanced performance with decent precision, recall, and F1 scores across all classes.	Lower batch size with moderate image size ensures the model learns effectively without overfitting.
YOLOv5s	640	32	Slightly decreases precision and recall but improves mAP, especially for crops.	Larger batch sizes provide better gradient estimates, enhancing mAP but slightly lowering precision and recall.
YOLOv5s	1024	16	Larger image size reduces recall and mAP, indicating increased complexity and noise.	Higher image resolution introduces more details, which might overwhelm the model, leading to poorer performance.
YOLOv5m	1024	16	Improves precision, recall, and mAP metrics, showing a significant performance boost.	A more complex model better handles larger image sizes and more details, resulting in improved detection accuracy.

Table 14 summarizes the effects of varying image sizes and batch sizes on model performance, including precision, recall, and mean Average Precision (mAP), along with insights into how these hyper-parameters influence the learning process and detection accuracy. Impact of Image Size: An image size of 640 was generally more effective for the smaller YOLOv5s model, leading to higher F1 scores and mAP values. This suggests that a moderate image resolution is sufficient for

accurate object detection without incurring the computational overhead associated with larger image sizes. Increasing the image size to 1024 for the YOLOv5s model resulted in a decrease in both recall and F1 Score for weed detection, indicating potential overfitting or increased difficulty in processing larger images efficiently within the given epoch limit. Batch Size Considerations:
A batch size of 32 demonstrated improved performance metrics for the YOLOv5s model compared to a batch size of 16. This highlights the advantage of larger batch sizes in achieving better gradient estimates during training, thereby enhancing model performance. For the YOLOv5m model, a batch size of 16 provided a balance between computational efficiency and model accuracy, particularly for crop detection.

Results for Different Hyper-parameter Settings of the YOLOv8 Model

The table blow presents the performance metrics of a YOLOv8n (nano) model evaluated on the ACRE dataset. The evaluation considered various hyperparameters, including an image size of 640, a batch size of 16, and 50 epochs. The metrics provided include Precision (P), Recall (R), F1 Score, mAP50, and mAP50-95 for three classes: All(which averages the performance across all instances), Crop, and Weed.

Table 15: Performance of a single model over the test_dev split of the ACRE dataset (model size: YOLOv8n - nano).

Class	**Images**	**Instances**	**Precision (P)**	**Recall (R)**	**F1 Score**	**mAP50**	**mAP50-95**
All	100	11472	0.75	0.685	0.715	0.732	0.447
Crop	100	1205	0.799	0.831	0.815	0.866	0.606
Weed	100	10267	0.701	0.539	0.61	0.599	0.287
Epochs	50						
Batch Size	16						
Image Size	640						
Model Size	YOLOv8n						

The YOLOv8n model with an image size of 640 and batch size of 16 demonstrates balanced performance with good precision and recall for both crops and weeds. The F1 score and mAP metrics are reasonably high, indicating effective detection capabilities for this lightweight model.

Table 16: Performance of the YOLOv8s model over the test_dev split of the ACRE dataset (model size: YOLOv8s - small).

Class	**Images**	**Instances**	**Precision (P)**	**Recall (R)**	**F1 Score**	**mAP50**	**mAP50-95**
All	100	11472	0.762	0.717	0.739	0.753	0.469
Crop	100	1205	0.814	0.852	0.833	0.874	0.627
Weed	100	10267	0.71	0.581	0.639	0.632	0.311
Epochs	50						
Batch Size	16						
Image Size	640						
Model Size	YOLOv8s						

The YOLOv8s model with an image size of 640 and batch size of 16 shows a consistent performance improvement over the YOLOv8n model. This model demonstrates better precision, recall,mAP50,mAP50-95 and F1 scores across all classes.

Table 17: Performance of YOLOv8m model over the test_dev split of the ACRE dataset (model size: YOLOv8m -medium).

Class	**Images**	**Instances**	**Precision (P)**	**Recall (R)**	**F1 Score**	**mAP50**	**mAP50-95**
All	100	11472	0.772	0.711	0.740	0.761	0.477
Crop	100	1205	0.815	0.878	0.844	0.848	0.631
Weed	100	10267	0.73	0.573	0.642	0.645	0.322
Epochs	50						
Batch Size	16						
Image Size	640						
Model Size	YOLOv8m						

The YOLOv8m model demonstrates superior performance, particularly in recall and mAP scores for both crops and weeds. When comparing the models, it is evident that the YOLOv8m model provides the best overall performance across all evaluation metrics. It shows higher precision, recall, F1 score, mAP50, and mAP50-95 values for both crops and weeds compared to the YOLOv8n and YOLOv8s models. For crop detection, the YOLOv8m model achieves the highest F1 score and mAP50, indicating its robustness in accurately identifying crops. For weed detection, although the recall is slightly lower for the YOLOv8m model compared to the YOLOv8s model, it compensates with higher precision and mAP50 scores.

Table 18: Performance of YOLOv8m model over the test_dev split of the ACRE dataset (model size: YOLOv8m -medium).

Class	**Images**	**Instances**	**Precision (P)**	**Recall (R)**	**F1 Score**	**mAP50**	**mAP50-95**
All	100	11472	0.777	0.716	0.745	0.762	0.478
Crop	100	1205	0.833	0.847	0.840	0.886	0.640
Weed	100	10267	0.721	0.584	0.645	0.638	0.316
Epochs	50						
Batch Size	32						
Image Size	640						
Model Size	YOLOv8m						

The YOLOv8m model with an image size of 640 and batch size of 32 shows an improvement in recall and mAP metrics compared to its batch size 16 counterpart. While the differences may be slight, the model with a batch size of 32 shows slightly better recall and mAP metrics overall.

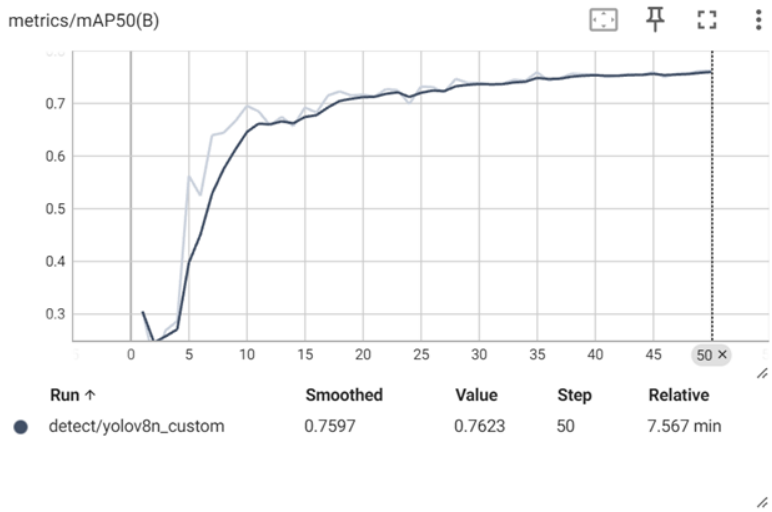

Figure 19: mAP50 Metric for All Class (which averages the performance across all instances.) YOLOv8m Model,Batch Size = 32, Image Size = 640

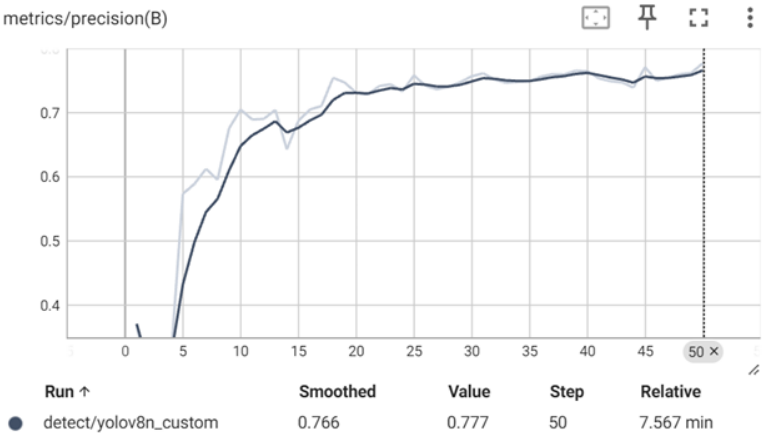

Figure 20: Precision Metric for All Class (which averages the performance across all instances). YOLOv8m Model,Batch Size = 32, Image Size = 640

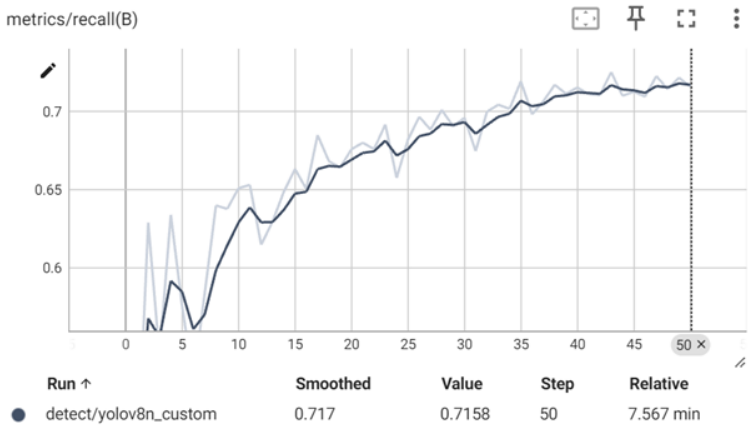

Figure 21: Recall Metric for All Class (which averages the performance across all instances). YOLOv8m Model,Batch Size = 32, Image Size = 640

Table 19: Performance of a single model over the test_dev split of the ACRE dataset (model size: YOLOv8m -medium).

Class	**Images**	**Instances**	**Precision (P)**	**Recall (R)**	**F1 Score**	**mAP50**	**mAP50-95**
All	100	11472	0.781	0.714	0.746	0.766	0.480
Crop	100	1205	0.832	0.853	0.842	0.888	0.639
Weed	100	10267	0.729	0.576	0.644	0.643	0.321
Epochs	50						
Batch Size	64						
Image Size	640						
Model Size	YOLOv8m						

The YOLOv8m model with an image size of 640 and batch size of 64 shows consistent performance with slight variations in metrics compared to the batch size 32 variant. Effect of Batch Size on YOLOv8m: Increasing the batch size from 16 to 32 improves recall and mAP metrics for both crops and weeds. However, further increasing the batch size to 64 results in minor variations, with a slight decrease in recall for weeds but consistent precision and F1 scores.

Table 20: Performance of a single model over the test_dev split of the ACRE dataset (model size: YOLOv8s - small).

Class	**Images**	**Instances**	**Precision (P)**	**Recall (R)**	**F1 Score**	**mAP50**	**mAP50-95**
All	100	11472	0.746	0.718	0.732	0.752	0.473
Crop	100	1205	0.803	0.832	0.817	0.859	0.612
Weed	100	10267	0.688	0.604	0.643	0.645	0.334
Epochs	50						
Batch Size	16						
Image Size	1600						
Model Size	YOLOv8s						

Increasing the image size to 1600 improves the model's ability to capture finer details, leading to better overall performance.

Table 21: Performance of a single model over the test_dev split of the ACRE dataset (model size: YOLOv8s - small).

Class	**Images**	**Instances**	**Precision (P)**	**Recall (R)**	**F1 Score**	**mAP50**	**mAP50-95**
All	100	11472	0.747	0.653	0.697	0.691	0.443
Crop	100	1205	0.789	0.768	0.778	0.797	0.583
Weed	100	10267	0.704	0.539	0.610	0.586	0.303
Epochs	50						
Batch Size	16						
Image Size	1024						
Model Size	YOLOv8s						

The performance metrics slightly decline compared to the image size of 1600, but it remains robust and effective.

Table 22: YOLOv8 Model Performance Insights

Model	**Image Size**	**Batch Size**	**Performance**	**Insight**
YOLOv8m	-	16 vs. 32	Consistent performance with slight improvements in recall and mAP metrics as batch size increases.	Indicates that the YOLOv8m model is relatively stable and can handle larger batch sizes without significant performance degradation.
YOLOv8m	-	32 vs. 64	Marginal increase in performance metrics when moving from a batch size of 32 to 64.	Suggests that the model's performance is optimized around these batch sizes, providing flexibility in training configurations.
YOLOv8s	1600 vs. 640	-	Larger image size significantly enhances the model's ability to detect finer details, improving performance metrics, particularly for crop detection.	Higher precision, recall, F1 score, and mAP metrics with larger image size.
YOLOv8s	1024 vs. 640	-	Slightly lower performance metrics than with an image size of 1600 but maintains robust detection capabilities.	Indicates that even a moderate increase in image size can enhance detection performance.

Table 22 compares YOLOv8 model configurations on the ACRE dataset, detailing the impact of varying image sizes and batch sizes on performance metrics such as precision, recall, F1 score, and mean Average Precision (mAP). The table provides insights into the stability and efficiency of the models under different training conditions, highlighting how changes in hyper-parameters affect detection accuracy and model learning. The YOLOv8m model with an image size of 640 and batch size of 32 provides the best overall performance for weed detection, with the highest F1 score and competitive mAP metrics. This configuration demonstrates the optimal balance between precision and recall for the specific task of weed detection in the ACRE dataset.

Improvements Using YOLOv8 Over YOLOv5: The YOLOv8m model achieves higher mAP50 and mAP50-95 scores compared to the YOLOv5s model, indicating improved detection capabilities and more accurate bounding box predictions.

Table 23: Comparison of YOLOv8 model results with YOLOv5 model results. This table summarizes the performance metrics, including mAP50, mAP50-95, and F1 score, highlighting the improvements in detection capabilities and architectural advancements in YOLOv8.

Metric	YOLOv8m	YOLOv5s	Analysis
mAP50	0.638	0.611	The YOLOv8m model shows a noticeable improvement in mAP50 compared to the YOLOv5s model, indicating better precision and recall at an IoU threshold of 0.50.
mAP50-95	0.316	0.279	The YOLOv8m model also outperforms the YOLOv5s model in mAP50-95, which averages precision over multiple IoU thresholds from 0.50 to 0.95. This suggests better overall performance across a range of IoU values.
F1 Score	0.645	0.627	The higher F1 score of the YOLOv8m model indicates a better balance between precision and recall for weed detection compared to the YOLOv5s model.

Additionally, the improved F1 score of YOLOv8m demonstrates a better trade-off between precision and recall, leading to more reliable detection of weeds without a significant increase in false positives or false negatives. These advancements are likely attributed to YOLOv8's enhancements in neural network architecture, including improvements in backbone networks, neck designs, and detection heads, which result in more efficient feature extraction and better overall performance.

4.2.2 Comparison of YOLOv5 and YOLOv8 Results with State-of-the-Art Results

To evaluate the performance of our models, we compared our results with state-of-the-art benchmarks in object detection tasks. The tables below summarize the best results obtained from YOLOv5[52] and YOLOv8[56] models in the existing literature.

YOLOv5 vs. State-of-the-Art

Table 24: Performance Comparison of YOLOv5 with State-of-the-Art Methods

Networks	**Precision (%)**	**Recall (%)**	**F1 Score**	**mAP(%)**
YOLOv3	89.0	99.0	0.88	91.80
YOLOv5	88.0	99.0	0.89	92.40
Faster R-CNN	65.9	98.0	0.78	92.15

While the precision and recall are slightly lower in the YOLOv5 ACRE results compared to the state-of-the-art benchmarks, the model still demonstrates a robust F1 score and mAP, indicating its effectiveness in the specific application of weed detection in agricultural datasets.

Table 25: This table presents the performance metrics for state-of-the-art YOLOv8 models on the ACRE dataset. The results include model size, image size, batch size, epochs, mAP95, precision (P), recall (R), and F1 score for both crop and weed detection. The average (avg) values for each metric are also provided, highlighting the differences in detection performance across various configurations

# imgs	model size	img size	batch	epochs	mAP95	key	P	R	F1
2479	SMALL	1024	16	50	0.4596	crop	0.79	0.85	0.82
						weed	0.69	0.68	0.69
						avg	0.7	0.7	0.7
2479	SMALL	1600	16	50	0.4765	crop	0.776	0.88	0.825
						weed	0.647	0.742	0.691
						avg	0.661	0.756	0.705
2479	MED	640	32	50	0.4561	crop	0.802	0.903	0.85
						weed	0.594	0.67	0.682
						avg	0.706	0.694	0.699
2479	MED	640	64	50	0.4581	crop	0.779	0.883	0.828
						weed	0.664	0.684	0.674
						avg	0.676	0.705	0.69

Table 26: Comparison of YOLOv8 Small Model Performance Metrics with Thesis Model. The table summarizes the precision, recall, F1 score, and mAP50-95 for both weed and crop detection, highlighting the differences in performance between the state-of-the-art YOLOv8 Small model and the Thesis YOLOv8s model under identical configurations (Image size: 1024, Batch size: 16).

Metric	**State-of-the-Art YOLOv8 Small**			**Thesis YOLOv8s**		
	Weed	Crop	Avg	Weed	Crop	All
Precision (P)	0.69	0.79	0.70	0.704	0.789	0.747
Recall (R)	0.68	0.85	0.70	0.539	0.768	0.635
F1 Score	0.69	0.82	0.70	0.610	0.778	0.697
mAP50-95	0.4596			0.303	0.583	0.443

Weed Class Performance: The thesis model achieves a higher precision of 0.704 compared to the state-of-the-art model's precision of 0.69. This indicates that the thesis model is more accurate in predicting true positive weed instances. However, the state-of-the-art model has a higher recall of 0.68, while the thesis model's recall is 0.539. This suggests that the state-of-the-art model is better at capturing all actual weed instances. The F1 score for the state-of-the-art model is 0.69, which is higher than the thesis model's F1 score of 0.610. This reflects a better balance between precision and recall in the state-of-the-art model. In terms of

mAP50-95, the state-of-the-art model's overall mAP95 is 0.4596, indicating superior performance compared to the thesis model's mAP50-95 for weeds at 0.303.
Crop Class Performance: Both models have similar precision for crops, with the thesis model at 0.789 and the state-of-the-art model at 0.79. The state-of-the-art model achieves a higher recall of 0.85 compared to the thesis model's 0.768. This suggests that the state-of-the-art model better captures true positive crop instances. The state-of-the-art model also has a higher F1 score of 0.82 compared to the thesis model's 0.778, indicating a better balance of precision and recall. The thesis model, however, achieves a higher mAP50-95 for crops at 0.583 compared to the state-of-the-art model's overall mAP95 of 0.4596. This indicates strong performance in crop detection by the thesis model.
Overall Performance (All Classes): When considering all classes, the thesis model has a higher overall precision of 0.747 compared to the state-of-the-art model's average precision of 0.70. However, the state-of-the-art model has a higher overall recall of 0.77 compared to the thesis model's recall of 0.635. The state-of-the-art model achieves a higher overall F1 score of 0.70, whereas the thesis model's F1 score is 0.697. In terms of mAP50-95, the thesis model has an overall score of 0.443, which is slightly lower than the state-of-the-art model's overall mAP95 of 0.4596.

Table 27: Comparison of YOLOv8 Small Model Performance Metrics with Thesis Model. The table summarizes the precision, recall, F1 score, and mAP50-95 for both weed and crop detection, highlighting the differences in performance between the state-of-the-art YOLOv8 Small model and the Thesis YOLOv8s model under identical configurations (Image size: 1600, Batch size: 16).

Metric	**State-of-the-Art YOLOv8 Small**			**Thesis YOLOv8s**		
	Weed	Crop	Avg	Weed	Crop	All
Precision (P)	0.647	0.776	0.661	0.688	0.803	0.746
Recall (R)	0.742	0.88	0.756	0.604	0.832	0.718
F1 Score	0.691	0.825	0.705	0.645	0.817	0.732
mAP50-95	0.4765			0.334	0.61	0.473

Weed Class Performance: The thesis model has higher precision (0.688) compared to the state-of-the-art model (0.647), indicating better accuracy in predicting true positive weed instances. However, the state-of-the-art model has higher recall (0.742 vs. 0.604), better capturing all actual weed instances. The state-of-the-art model also has a higher F1 score (0.691 vs. 0.645), showing a better balance between precision and recall.
Crop Class Performance: The thesis model achieves higher precision (0.803) than the state-of-the-art model (0.776), but the state-of-the-art model has higher

recall (0.88 vs. 0.832), suggesting it captures true positive crop instances better. The state-of-the-art model also has a slightly higher F1 score (0.825 vs. 0.817).
Overall Performance (All Classes): The thesis model demonstrates higher overall precision (0.746 vs. 0.661) but lower recall (0.718 vs. 0.756) compared to the state-of-the-art model. The thesis model has a higher F1 score (0.732 vs. 0.705), but slightly lower mAP50-95 (0.473 vs. 0.4765).

Table 28: Comparison of YOLOv8 Medium Model Performance Metrics with Thesis Model. The table summarizes the precision, recall, F1 score, and mAP50-95 for both weed and crop detection, highlighting the differences in performance between the state-of-the-art YOLOv8 Medium model and the Thesis YOLOv8m model under identical configurations (Image size: 640, Batch size: 32).

Metric	**State-of-the-Art YOLOv8 Medium**			**Thesis YOLOv8m**		
	Weed	Crop	Avg	Weed	Crop	All
Precision (P)	0.694	0.802	0.706	0.721	0.833	0.777
Recall (R)	0.691	0.903	0.694	0.584	0.847	0.716
F1 Score	0.682	0.85	0.699	0.645	0.84	0.745
mAP50-95	0.4561			0.316	0.64	0.478

Weed Class Performance: The thesis model achieves a higher precision of 0.721 compared to the state-of-the-art model's precision of 0.694. This indicates better accuracy in predicting true positive weed instances. However, the state-of-the-art model has a higher recall of 0.691, while the thesis model's recall is 0.584. This suggests that the state-of-the-art model is better at capturing all actual weed instances. The F1 score for the state-of-the-art model is 0.682, which is higher than the thesis model's F1 score of 0.645, reflecting a better balance between precision and recall in the state-of-the-art model.
Crop Class Performance: For crop detection, the thesis model achieves a higher precision of 0.833 compared to the state-of-the-art model's 0.802. The state-of-the-art model has a higher recall of 0.903 compared to 0.847 in the thesis model, suggesting better capture of true positive crop instances. The state-of-the-art model has a higher F1 score of 0.85 compared to the thesis model's 0.84.
Overall Performance (All Classes): The thesis model achieves a higher overall precision of 0.777 compared to the state-of-the-art model's average precision of 0.706. The thesis model has a higher overall recall of 0.716 compared to the state-of-the-art model's recall of 0.694. The state-of-the-art model achieves a lower overall F1 score of 0.699, whereas the thesis model's F1 score is 0.745. In terms of mAP50-95, the thesis model has an overall score of 0.478, which is slightly higher than the state-of-the-art model's overall mAP95 of 0.4561.

Table 29: Comparison of YOLOv8 Medium Model Performance Metrics with Thesis Model. The table summarizes the precision, recall, F1 score, and mAP50-95 for both weed and crop detection, highlighting the differences in performance between the state-of-the-art YOLOv8 Medium model and the Thesis YOLOv8m model under identical configurations (Image size: 640, Batch size: 64).

Metric	**State-of-the-Art YOLOv8 Medium**			**Thesis YOLOv8m**		
	Weed	Crop	Avg	Weed	Crop	All
Precision (P)	0.664	0.779	0.676	0.729	0.832	0.781
Recall (R)	0.684	0.853	0.705	0.576	0.853	0.714
F1 Score	0.674	0.828	0.69	0.644	0.842	0.746
mAP50-95	0.4581			0.321	0.639	0.48

Weed Class Performance: The thesis model achieves a higher precision of 0.729 compared to the state-of-the-art model's precision of 0.664. This indicates better accuracy in predicting true positive weed instances. However, the state-of-the-art model has a higher recall of 0.684, while the thesis model's recall is 0.576. This suggests that the state-of-the-art model is better at capturing all actual weed instances. The F1 score for the state-of-the-art model is 0.674, which is higher than the thesis model's F1 score of 0.644, reflecting a better balance between precision and recall in the state-of-the-art model.
Crop Class Performance: For crop detection, the thesis model achieves a higher precision of 0.832 compared to the state-of-the-art model's 0.779. The thesis model also has a similar recall of 0.853, suggesting equivalent capture of true positive crop instances. The thesis model has a higher F1 score of 0.842 compared to the state-of-the-art model's 0.828.
Overall Performance (All Classes): The thesis model achieves a higher overall precision of 0.781 compared to the state-of-the-art model's average precision of 0.676. The state-of-the-art model has a lower overall recall of 0.705 compared to the thesis model's recall of 0.714. The state-of-the-art model achieves a lower overall F1 score of 0.69, whereas the thesis model's F1 score is 0.746. In terms of mAP50-95, the thesis model has an overall score of 0.480, which is about 2% higher than the state-of-the-art model's overall mAP95 of 0.4581.

Figure 22 demonstrates the model's performance in identifying both categories. In this particular image, the model successfully distinguishes crops from weeds, with crops being labeled as 'crop' and weeds as 'weed.' Confidence scores are displayed for each detection, indicating the model's certainty. For instance, crops are detected with high confidence scores of 0.93 and 0.91, showcasing the model's effectiveness in crop identification. The presence of multiple weed detections with varying confidence levels also highlights the model's capability to handle a diverse and cluttered field environment, ensuring accurate weed identification. This performance is crucial for precision agriculture applications, where accurate detection and classification can significantly enhance weed management practices and crop yield predictions.

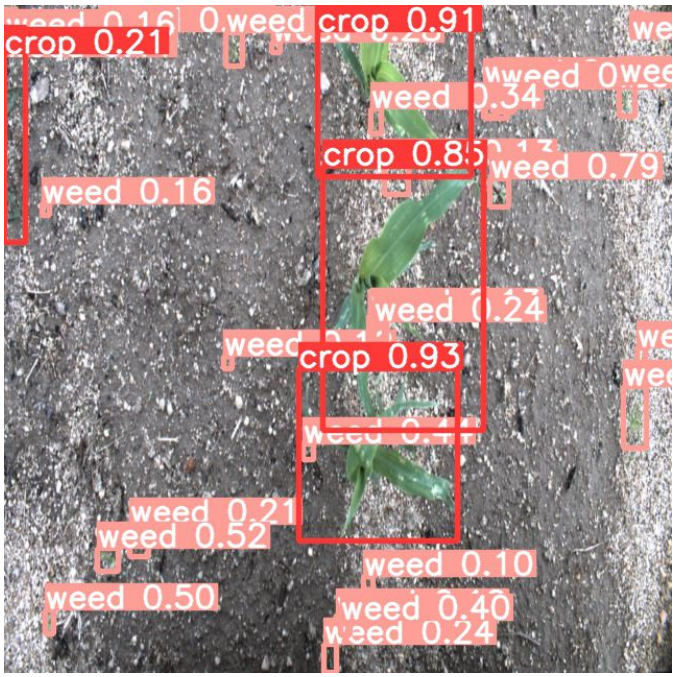

Figure 22: Detection and classification of crops and weeds using the YOLOv8 model. This image illustrates the model's ability to accurately detect and classify crops (labeled as 'crop') and weeds (labeled as 'weed') with associated confidence scores. Notable crop detections include labels with confidence scores of 0.93 and 0.91.

4.2.3 Results for Sunflower Dataset

Table 30: Results for Different Configurations of RGB+NIR Model for Sunflower Dataset

Combination of Channels	Num Epochs	Image Size	Batch Size	Filter on NIR	Model	NDVI Threshold	Soil IOU	Crop IOU	Weed IOU	Mean IOU
RGB+NIR	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	-	0.99	0.93	0.71	0.88
RGB+NIR	8	964x1296	2	Gaussian bilateral filter	UNET-RESNRT50	-	0.99	0.86	0.67	0.84
RGB+NIR	8	964x1296	4	Gaussian bilateral filter	Custom UNET	-	0.99	0.90	0.69	0.86
RGB+NIR	8	964x1296	2	Gaussian bilateral filter	Custom UNET	-	0.99	0.86	0.60	0.82

IoU metric for the soil class is consistently high (0.99) across all configurations, indicating excellent accuracy in identifying soil. Crop IOU metric shows variability. The best performance is observed with the UNET-RESNRT50 model using a batch size of 4 (0.93), while the lowest performance is with the Custom UNET model using a batch size of 2 (0.86). Weed IOU metric exhibits the most variability. The best performance is observed with the UNET-RESNRT50 model using a batch size of 4 (0.71), while the lowest performance is with the Custom UNET model using a batch size of 2 (0.60). The average IoU across all classes (soil, crop, and weed) is highest with the UNET-RESNRT50 model using a batch size of 4 (0.88) and lowest with the Custom UNET model using a batch size of 2 (0.82).
In summary, the UNET-RESNRT50 model generally outperforms the Custom UNET in terms of Mean IOU, particularly with a batch size of 4. The batch size appears to influence performance, with a larger batch size yielding better results for the UNET-RESNRT50 model. The Weed IOU is the lowest among the classes, indicating that distinguishing weeds from other classes is more challenging. This analysis suggests that for this specific task, the UNET-RESNRT50 model with a larger batch size performs better overall.

Table 31: Results for Different Configurations of G+NIR+NDVI Model for Sunflower Dataset

Combination of Channels	Num Epochs	Image Size	Batch Size	Filter on NIR	Model	NDVI Threshold	Soil IOU	Crop IOU	Weed IOU	Mean IOU
G + NIR + NDVI	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	0.5	0.99	0.86	0.61	0.82
G + NIR + NDVI	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	0.45	0.97	0.53	0.25	0.58
G + NIR + NDVI	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	0.4	0.98	0.36	0.36	0.57
G + NIR + NDVI	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	0.35	0.99	0.79	0.38	0.72
G + NIR + NDVI	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	0.3	0.98	0.40	0.23	0.53
G + NIR + NDVI	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	0.6	0.99	0.75	0.41	0.72

The NDVI threshold significantly impacts the segmentation performance for crops and weeds, while the soil IoU remains relatively stable across different thresholds. This is likely because the soil class is distinct and consistently identifiable regardless of the NDVI value, whereas the differentiation between crops and weeds is more nuanced and sensitive to the chosen threshold.

- Optimal Threshold: 0.5

An NDVI threshold of 0.5 yields the highest mean IoU (0.82). This threshold effectively balances the distinction between vegetation (crops and weeds) and non-vegetation areas (soil). At this threshold, the segmentation algorithm can accurately identify the boundaries between different classes, resulting in the best overall performance.

- Lower Thresholds (0.45, 0.4, 0.3):

These thresholds result in lower mean IoUs, particularly affecting the crop and weed IoUs. Lower NDVI thresholds can lead to misclassification because they may not effectively separate vegetation from non-vegetation. Specifically:

- 0.45: Decreases crop and weed IoUs significantly, indicating that some areas of vegetation are not being correctly identified as crops or weeds.

- 0.4: Further decreases in performance suggest that this threshold might be too low to distinguish crops and weeds accurately.

- 0.3: The lowest performance, indicating substantial misclassification. At this threshold, many vegetation pixels might be misclassified as non-vegetation, leading to poor segmentation of crops and weeds.

- Higher Threshold: 0.6

This threshold provides a relatively high mean IoU (0.72), similar to the threshold of 0.35. A higher NDVI threshold ensures that only the areas with the highest vegetation density are classified as crops or weeds. While this improves the accuracy of identifying very dense vegetation, it can lead to the exclusion of less dense areas, resulting in a balanced but slightly lower mean IoU compared to the optimal threshold.

Table 32: Results for Different Configurations of G+NIR Model for Sunflower Dataset

Combination of Channels	**Num Epochs**	**Image Size**	**Batch Size**	**Filter on NIR**	**Model**	**NDVI Threshold**	**Soil IOU**	**Crop IOU**	**Weed IOU**	**Mean IOU**
G+NIR	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	-	0.99	0.63	0.42	0.68
G+NIR	8	964x1296	4	Gaussian bilateral filter	UNET-RESNRT50	-	0.99	0.64	0.41	0.68
G+NIR	8	964x1296	2	Gaussian bilateral filter	UNET-RESNRT50	-	0.99	0.67	0.47	0.71
G+NIR	8	964x1296	2	Gaussian bilateral filter	UNET-RESNRT50	-	0.99	0.70	0.41	0.70

The Soil IOU remains consistently high (0.99) across all configurations, indicating excellent accuracy in identifying soil. This suggests that the model is very effective at distinguishing soil from other classes. The Crop IOU varies between 0.63 and 0.70. The best performance is seen with a batch size of 2 in Configuration 4 (0.70), suggesting that a smaller batch size may improve the model's ability to identify crops accurately. The Weed IOU ranges from 0.41 to 0.47. The best performance is observed with a batch size of 2 in Configuration 3 (0.47). This indicates that the model struggles more with distinguishing weeds compared to soil and crops, but a smaller batch size may slightly improve weed detection. The Mean IOU ranges from 0.68 to 0.71. The highest Mean IOU is achieved with a batch size of 2 in Configuration 3 (0.71), indicating that this configuration provides the best overall segmentation performance across all classes. The analysis suggests that the UNET-RESNRT50 model performs consistently well in identifying soil across all configurations. A smaller batch size (2) tends to improve the segmentation performance for crops and weeds, leading to higher Mean IOU values. The distinction between crops and weeds is more challenging for the model compared to soil, but performance can be optimized by adjusting the batch size.

Table 33: Comparison of RGB+NIR, G+NIR+NDVI, and G+NIR Results Using Different Configurations. The table summarizes the Weed IOU, Crop IOU, and Mean IOU for each configuration, highlighting the range of values observed across different experimental setups

Configuration	**Weed IOU**	**Crop IOU**	**Mean IOU**
RGB+NIR	0.60 to 0.71	0.86 to 0.93	0.82 to 0.88
G+NIR+NDVI	0.23 to 0.61	0.36 to 0.86	0.53 to 0.82
G+NIR	0.41 to 0.47	0.63 to 0.70	0.68 to 0.71

For the RGB+NIR configuration, the rich information from RGB and NIR channels aids differentiation, with RGB capturing color differences and NIR enhancing vegetation identification. Larger batch sizes support stable learning, making this the best and most consistent configuration, and the preferred choice.
The G+NIR+NDVI configuration's effectiveness varies with the NDVI threshold. Proper tuning improves detection, but this adds an optimization parameter. This configuration has less RGB information, relying more on indices, resulting in good performance with proper tuning but variable outcomes.
The G+NIR configuration uses simpler data with only Green and NIR channels, which lacks full RGB differentiation but maintains high soil IOU. Smaller batch sizes (2) yield better results, though the overall performance is stable but lower compared to the other configurations, making it less effective without additional channels.

Table 34: Best Results: ACRE vs. Sunflower Datasets

Task	Metric	Model (Data)	Value
Weed Detection and Segmentation	F1 Score	YOLOv5s (RGB)	0.627
	mAP50	YOLOv5s (RGB)	0.611
	mAP50-95	YOLOv5s (RGB)	0.279
	F1 Score	YOLOv8m (RGB)	0.645
	mAP50	YOLOv8m (RGB)	0.648
	mAP50-95	YOLOv8m (RGB)	0.321
	Weed IoU	UNet-ResNet50 (RGB+NIR)	0.71
Crop Detection and Segmentation	F1 Score	YOLOv5s (RGB)	0.830
	mAP50	YOLOv5s (RGB)	0.863
	mAP50-95	YOLOv5s (RGB)	0.587
	F1 Score	YOLOv8m (RGB)	0.842
	mAP50	YOLOv8m (RGB)	0.888
	mAP50-95	YOLOv8m (RGB)	0.64
	Crop IoU	UNet-ResNet50 (RGB+NIR)	0.93
Overall	F1 Score	YOLOv5s (RGB)	0.73
	mAP50	YOLOv5s (RGB)	0.737
	mAP50-95	YOLOv5s (RGB)	0.433
	F1 Score	YOLOv8m (RGB)	0.746
	mAP50	YOLOv8m (RGB)	0.766
	mAP50-95	YOLOv8m (RGB)	0.48
	Mean IoU	UNet-ResNet50 (RGB+NIR)	0.88

The best model for weed segmentation is UNet-ResNet50 (RGB+NIR), which benefits from the inclusion of NIR data. This additional spectral information highlights vegetation characteristics that are not visible in the RGB spectrum, enhancing weed segmentation capabilities. The pixel-wise segmentation approach of the UNet-ResNet50 model leverages this extra spectral data effectively, resulting in superior performance compared to YOLO models using only RGB data.
The combination of RGB and NIR channels provides a more comprehensive set of features, allowing the model to better differentiate crops from weeds and soil. This results in improved overall segmentation performance. The UNet model effectively utilizes the richer data provided by the RGB+NIR combination, leading to higher IoU scores for crops.

Examples of results for UNet-ResNet50 (Input: 964*1296 RGB+NIR with batch size 4)

Figure 23 displays the ground truth segmentation for a given agricultural field image. In this visualization, green regions represent crops, while red regions denote weeds and black regions represent soil. The ground truth image serves as a benchmark for evaluating the accuracy of the U-Net model's predictions. Figure 24 illustrates the segmentation results produced by the U-Net model. The model's

predictions are compared against the ground truth to compute Intersection over Union (IoU) metrics. These results highlight the effectiveness of integrating RGB and NIR channels in improving segmentation accuracy. The high IoU values across different classes indicate the U-Net model's potential for practical applications in precision agriculture, particularly in automated weed management and crop monitoring systems.

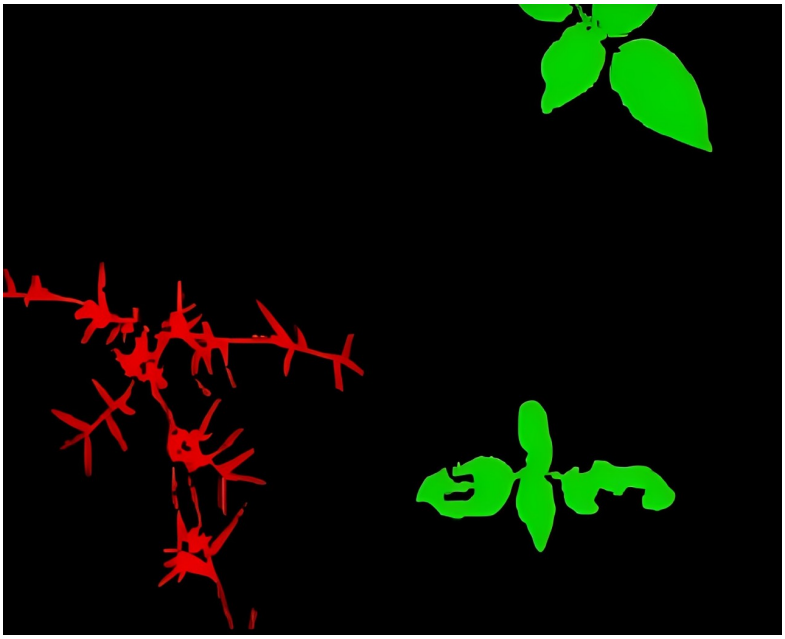

Figure 23: Ground Truth

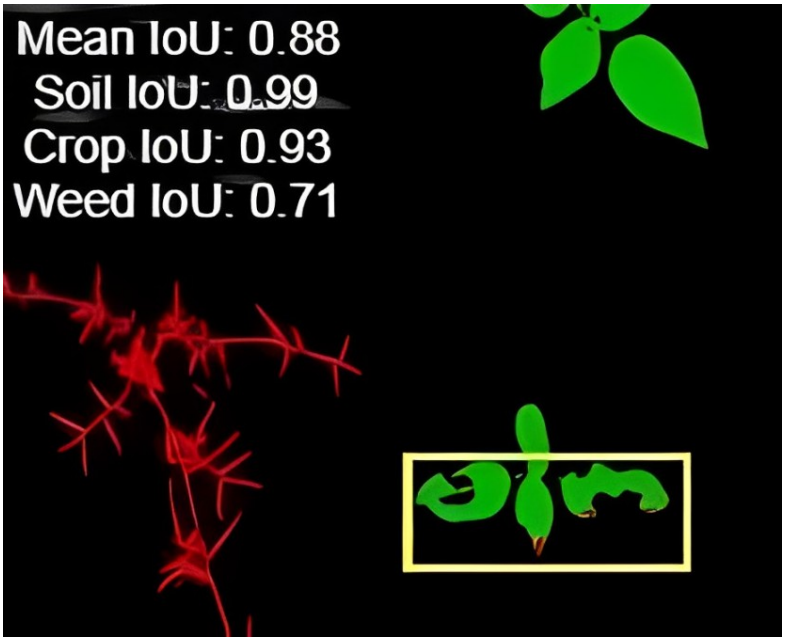

Figure 24: U-Net Prediction

4.2.4 Comparison of U-Net Segmentation Results with State-of-the-Art Results

Summary of Results from the State-of-the-Art Papers

Table 35: State-of-the-art Results from the Papers

Model / Input	Crop IoU	Weed IoU	Soil IoU	Mean IoU
UNet (704 × 704)[61]	0.90	0.76	0.86	0.84
UNet (512 × 512)[61]	0.88	0.75	0.83	0.82
UNet[58][61]	0.68	0.40	0.99	0.69
Bonnet[58][61]	0.88	0.69	0.99	0.86
UNet-ResNet[58][61]	0.70	0.48	0.99	0.72
RGB + NIR[3] (Image size: 512 × 512, Batch size: 8)	0.890	0.681	0.986	0.853
G + NIR + NDVI[3] (Image size: 512 × 512, Batch size: 8)	0.896	0.729	0.990	0.871
RGB + NIR [3](Image size: 704 × 704, Batch size: 4)	0.894	0.658	0.979	0.844
G + NIR + NDVI[3] (Image size: 704 × 704, Batch size: 4)	0.905	0.744	0.982	0.877

Table 36: Comparison of Custom U-Net (RGB+NIR) with State-of-the-Art U-Net Models

Model / Input	**Crop IoU**	**Weed IoU**	**Soil IoU**	**Mean IoU**
UNet (704 × 704)[61]	0.90	0.76	0.86	0.84
UNet (512 × 512)[61]	0.88	0.75	0.83	0.82
UNet[58][61]	0.68	0.40	0.99	0.69
Custom UNet (RGB+NIR), Thesis Results	0.93	0.69	0.99	0.86
Custom UNet (RGB+NIR), Thesis Results	0.86	0.60	0.99	0.82

The custom UNet (RGB+NIR) model shows a Crop IoU of 0.93, which is higher than the state-of-the-art UNet (704 × 704) at 0.90, representing an improvement of 3%, and significantly higher than the other two UNet configurations at 0.88 and 0.68, representing improvements of 5% and 25% respectively. In terms of Weed IoU, the custom UNet achieves 0.69, which is lower than the state-of-the-art UNet (704 × 704) at 0.76, representing a decrease of 7%, and UNet (512 × 512) at 0.75, representing a decrease of 6%, but significantly higher than the UNet[58][61] at 0.40, representing an improvement of 29%. The Soil IoU of the custom UNet (0.99) surpasses all the state-of-the-art models, indicating exceptional performance in soil detection, with improvements of 13%, 16% over the UNet (704 × 704), UNet (512 × 512) respectively. The Mean IoU of the custom UNet is 0.86, which is slightly higher than the state-of-the-art UNet (704 × 704) at 0.84 and UNet (512 × 512) at 0.82, representing improvements of 2% and 4% respectively, and much higher than the UNet[58] at 0.69, representing an improvement of 17%.

Table 37: Comparison of UNET-RESNET50 with State-of-the-Art Models

Model / Input	**Crop IoU**	**Weed IoU**	**Soil IoU**	**Mean IoU**
UNet-ResNet[58][61]	0.70	0.48	0.99	0.72
RGB + NIR[3] (Image size: 512 × 512, Batch size: 8)	0.890	0.681	0.986	0.853
RGB + NIR [3](Image size: 704 × 704, Batch size: 4)	0.894	0.658	0.979	0.844
UNET-RESNET50 (RGB+NIR), Thesis Results	0.93	0.71	0.99	0.88
UNET-RESNET50 (RGB+NIR), Thesis Results	0.86	0.69	0.99	0.84

The UNET-RESNET50 (RGB+NIR) achieves a Crop IoU of 0.93, which is significantly higher than the UNet-ResNet[58][61] at 0.70, representing an improvement of 23%, and higher than the other two configurations at around 0.894, representing an improvement of 3.6%. In terms of Weed IoU, the UNET-RESNET50 achieves 0.71, which is notably higher than the UNet-ResNet at 0.48, representing an improvement of 23%, and also higher than the other

configurations at around 0.68, representing an improvement of 3%. The Soil IoU of 0.99 for UNET-RESNET50 is consistently high across all configurations, indicating its strength in soil detection. The Mean IoU of 0.88 for UNET-RESNET50 is higher than all the compared models, demonstrating its superior overall segmentation performance, with an improvement of 2.7% over the highest other model. These results highlight the potential of integrating NIR with RGB channels and optimizing model architectures such as UNET-RESNET50 to achieve superior segmentation performance. The improvements in Weed IoU are particularly significant, as effective weed detection is crucial for precision agriculture. Further optimization and tuning of hyperparameters, as well as exploring additional spectral bands, could lead to even greater advancements in segmentation accuracy and overall model performance.

Comparison of UNet-RESNET50 (G+NIR+NDVI) with State-of-the-Art Models

Table 38: State-of-the-Art G+NIR+NDVI Models

Model / Input	**Crop IoU**	**Weed IoU**	**Soil IoU**	**Mean IoU**
G + NIR + NDVI[3] (Image size: 512 × 512, Batch size: 8)	0.896	0.729	0.990	0.871
G + NIR + NDVI[3] (Image size: 704 × 704, Batch size: 4)	0.905	0.744	0.982	0.877

Table 39: Thesis G+NIR+NDVI Results

Model	Batch Size	NDVI Threshold	Crop IoU	Weed IoU	Soil IoU	Mean IoU
UNET-RESNET50	4	0.50	0.86	0.61	0.99	0.82
UNET-RESNRT50	4	0.45	0.53	0.25	0.97	0.58
UNET-RESNET50	4	0.40	0.36	0.36	0.98	0.57
UNET-RESNET50	4	0.35	0.79	0.38	0.99	0.72
UNET-RESNET50	4	0.30	0.40	0.23	0.98	0.53
UNET-RESNET50	4	0.60	0.75	0.41	0.99	0.72

When comparing the state-of-the-art G+NIR+NDVI models to the thesis results, it is evident that the state-of-the-art models achieve higher metrics across the board. The state-of-the-art model with an image size of 512×512 and a batch size of 8 achieves a Crop IoU of 0.896, a Weed IoU of 0.729, and a Mean IoU of 0.871. The model with an image size of 704×704 and a batch size of 4 achieves slightly higher metrics with a Crop IoU of 0.905, a Weed IoU of 0.744, and a Mean IoU of 0.877.
In contrast, the thesis results show more variability depending on the NDVI threshold used. For example, the best Crop IoU achieved is 0.86 with an NDVI threshold of 0.50, which is lower than the state-of-the-art. The Weed IoU in the thesis results ranges from 0.23 to 0.61, with the highest Weed IoU being significantly lower than the state-of-the-art results (0.729 and 0.744). The Mean IoU in the thesis results ranges from 0.53 to 0.82, with the highest Mean IoU also being lower than the state-of-the-art models.

Chapter 5

Conclusion

The research conducted in this thesis aimed to enhance the accuracy and efficiency of weed detection and segmentation in agricultural fields through the application of advanced deep learning models, specifically YOLO and UNet, on both RGB and multispectral (RGB-NIR) imaging datasets. The main discoveries from this research emphasize significant improvements in weed segmentation accuracy, particularly when integrating additional spectral data with standard RGB channels.

The results demonstrated that the inclusion of Near-Infrared (NIR) data in conjunction with RGB channels substantially improves the performance of DL models in distinguishing weeds from crops. This improvement is evident in the higher Intersection over Union (IoU) scores achieved for both crops and weeds. Specifically, the UNet-ResNet50 model utilizing RGB+NIR data achieved a mean IoU of 0.88, a crop IoU of 0.93, and a weed IoU of 0.71, with respective improvements of 2.7%, 4%, and 3% over the best baseline models.

Comparatively, the YOLOv5 and YOLOv8 models applied to the RGB dataset alone showed robust performance in object detection tasks, with the YOLOv8 model achieving a mean average precision (mAP) of 0.48, which is competitive with state-of-the-art models in object detection.

Moreover, the integration of NIR data in the UNet model provided a notable edge in segmentation accuracy, underscoring the value of multispectral imaging in agricultural applications. The thesis results indicate that incorporating NIR data with RGB channels enhances model generalization, reduces the impact of environmental variability, and improves the robustness of weed segmentation systems. This advancement supports the goal of precision agriculture by enabling more targeted and efficient weed management strategies, ultimately contributing to sustainable farming practices.

In conclusion, the research confirms that multispectral imaging combined with advanced deep learning architectures significantly enhances the precision of weed

detection and segmentation. Future work could explore the integration of additional spectral bands and the development of more sophisticated models to further improve the accuracy and scalability of automated weed management systems in various agricultural settings.

Bibliography

[1] E. M. B. M. Karunathilake, Anh Tuan Le, Seong Heo, Yong Suk Chung, and Sheikh Mansoor. «The Path to Smart Farming: Innovations and Opportunities in Precision Agriculture». In: *Agriculture* 13.8 (2023), p. 1593. DOI: 10.3390/agriculture13081593. URL: https://doi.org/10.3390/agriculture13081593 (cit. on pp. 1, 2).

[2] Aichen Wang, Wen Zhang, and Xinhua Wei. «A review on weed detection using ground-based machine vision and image processing techniques». In: *Computers and Electronics in Agriculture* 158 (2019), pp. 226–240. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2019.02.005. URL: https://www.sciencedirect.com/science/article/pii/S0168169918317150 (cit. on p. 2).

[3] Halil Mertkan Sahin, Tajul Miftahushudur, Bruce Grieve, and Hujun Yin. «Segmentation of weeds and crops using multispectral imaging and CRF-enhanced U-Net». In: *Computers and Electronics in Agriculture* 211 (2023), p. 107956. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2023.107956. URL: https://www.sciencedirect.com/science/article/pii/S0168169923003447 (cit. on pp. 2, 17, 30, 68–70).

[4] Daniel Steininger, Andreas Trondl, Gerardus Croonen, Julia Simon, and Verena Widhalm. «The CropAndWeed Dataset: a Multi-Modal Learning Approach for Efficient Crop and Weed Manipulation». In: *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2023, pp. 3718–3727. DOI: 10.1109/WACV56688.2023.00372 (cit. on p. 2).

[5] B. Åstrand and A.J. Baerveldt. «An agricultural mobile robot with vision-based perception for mechanical weed control». In: *Autonomous Robots* 13.1 (2002), pp. 21–35. DOI: 10.1023/A:1015674004201. URL: https://doi.org/10.1023/A:1015674004201 (cit. on pp. 2, 37).

[6] Sentera. *RGB and Multispectral Sensors*. https://sentera.com/resources/articles/rgb-multispectral-sensors/. Accessed: 23-11-2023. 2023 (cit. on pp. 3–5).

[7] Xavier Soria Poma, Angel Domingo Sappa, and Riad I Hammoud. «Wide-band Color Imagery Restoration for RGB-NIR Single Sensor Images». In: *Sensors* 18.7 (2018). License CC BY 4.0, p. 2059. DOI: 10.3390/s18072059. URL: https://www.mdpi.com/1424-8220/18/7/2059 (cit. on p. 3).

[8] Jionghui Jiang, Xi'an Feng, Fen Liu, Yingying Xu, and Hui Huang. «Multi-Spectral RGB-NIR Image Classification Using Double-Channel CNN». In: *IEEE Access* 7 (2019), pp. 20607–20613. DOI: 10.1109/ACCESS.2019.2896128 (cit. on p. 3).

[9] Zhangnan Wu, Yajun Chen, Bo Zhao, Xiaobing Kang, and Yuanyuan Ding. «Review of Weed Detection Methods Based on Computer Vision». In: *Sensors* 21.11 (2021), p. 3647. DOI: 10.3390/s21113647. URL: https://doi.org/10.3390/s21113647 (cit. on pp. 3, 16).

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation». In: *International Conference on Medical image computing and computer-assisted intervention* (2015), pp. 234–241 (cit. on pp. 5, 24).

[11] A. Bakhshipour and A. Jafari. «Evaluation of Support Vector Machine and Artificial Neural Networks in Weed Detection Using Shape Features». In: *Computers and Electronics in Agriculture* 145 (2018), pp. 153–160. DOI: 10.1016/j.compag.2017.12.032 (cit. on pp. 10, 16).

[12] A. Bakhshipour and H. Zareiforoush. «Development of a Fuzzy Model for Differentiating Peanut Plant from Broadleaf Weeds Using Image Features». In: *Plant Methods* 16 (2020), p. 153. DOI: 10.1186/s13007-020-00695-1 (cit. on p. 10).

[13] M. Gasparovič, M. Zrinjski, Đ. Barković, and D. Radočaj. «An Automatic Method for Weed Mapping in Oat Fields Based on UAV Imagery». In: *Computers and Electronics in Agriculture* 173 (2020), p. 105385. DOI: 10.1016/j.compag.2020.105385 (cit. on p. 10).

[14] F. Pallottino, P. Menesatti, S. Figorilli, F. Antonucci, R. Tomasone, A. Colantoni, et al. «Machine Vision Retrofit System for Mechanical Weed Control in Precision Agriculture Applications». In: *Sustainability* 10 (2018), p. 2209. DOI: 10.3390/su10072209 (cit. on p. 10).

[15] Vi Nguyen Thanh Le, Beniamin Apopei, and Kamal Alameh. «Effective plant discrimination based on the combination of local binary pattern operators and multiclass support vector machine methods». In: *Information Processing in Agriculture* 6.1 (2019), pp. 116–131. ISSN: 2214-3173. DOI: https://doi.org/10.1016/j.inpa.2018.08.002. URL: https://www.sciencedirect.com/science/article/pii/S2214317318300258 (cit. on p. 10).

[16] Y. Chen, B. Zhao, S. Li, L. Liu, Y. Yuan, and Y. Zhang. «Weed Reverse Positioning Method and Experiment Based on Multi-feature». In: *Transactions of the Chinese Society of Agricultural Machinery* 46 (2015), pp. 257–262 (cit. on p. 10).

[17] Weixing Zhu and Xiaofang Zhu. «The application of support vector machine in Weed classification». In: *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*. Vol. 4. 2009, pp. 532–536. DOI: 10.1109/ICICISYS.2009.5357638 (cit. on p. 10).

[18] X. Zhang, Z. Xie, N. Zhang, and C. Cao. «Weed recognition from pea seedling images and variable spraying control system». In: *Transactions of the Chinese Society for Agricultural Machinery* 43 (2012), pp. 220–225+73 (cit. on p. 10).

[19] C. Wang and Z. Li. «Weed recognition using SVM model with fusion height and monocular image features». In: *Transactions of the Chinese Society of Agricultural Engineering* 32 (2016), pp. 165–174 (cit. on p. 10).

[20] Henrik Skov Midtiby, Björn Åstrand, Ole Jørgensen, and Rasmus Nyholm Jørgensen. «Upper limit for context–based crop classification in robotic weeding applications». In: *Biosystems Engineering* 146 (2016). Special Issue: Advances in Robotic Agriculture for Crops, pp. 183–192. ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2016.01.012. URL: https://www.sciencedirect.com/science/article/pii/S153751101600012X (cit. on p. 10).

[21] D. He, Y. Qiao, P. Li, Z. Gao, H. Li, and J. Tang. «Weed Recognition Based on SVM-DS Multi-feature Fusion». In: *Transactions of the Chinese Society for Agricultural Machinery* 44 (2013), pp. 182–187 (cit. on pp. 11, 15).

[22] S. Sabzi, Y. Abbaspour-Gilandeh, and J. Arribas. «An automatic visible-range video weed detection, segmentation and classification prototype in potato field». In: *Heliyon* 6 (2020), e03685 (cit. on p. 11).

[23] X. Deng, L. Qi, X. Ma, Y. Jiang, X. Chen, H. Liu, and W. Chen. «Recognition of weeds at seedling stage in paddy fields using multi-feature fusion and deep belief networks». In: *Transactions of the Chinese Society of Agricultural Engineering* 34 (2018), pp. 165–172 (cit. on p. 11).

[24] J. Torres-Sánchez, F. López-Granados, and J.M. Peña. «An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops». In: *Computers and Electronics in Agriculture* 114 (2015), pp. 43–52. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2015.03.019. URL: https://www.sciencedirect.com/science/article/pii/S0168169915001052 (cit. on p. 11).

[25] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. «Textural Features for Image Classification». In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6 (1973), pp. 610–621 (cit. on p. 13).

[26] M. K. Hu. «Visual pattern recognition by moment invariants». In: *IEEE Transactions on Information Theory* 8 (1962), pp. 179–187 (cit. on p. 13).

[27] U. Shapira, I. Herrmann, A. Karnieli, and D. Bonfil. «Field spectroscopy for weed detection in wheat and chickpea fields». In: *International Journal of Remote Sensing* 34 (2013), pp. 6094–6108 (cit. on pp. 13, 34).

[28] L. Longchamps, B. Panneton, G. Samson, G. Leroux, and R. Thériault. «Discrimination of corn, grasses and dicot weeds by their UV-induced fluorescence spectral signature». In: *Precision Agriculture* 11 (2010), pp. 181–197 (cit. on p. 13).

[29] Faisal Ahmed, Hawlader Abdullah Al-Mamun, A.S.M. Hossain Bari, Emam Hossain, and Paul Kwan. «Classification of crops and weeds from digital images: A support vector machine approach». In: *Crop Protection* 40 (2012), pp. 98–104. ISSN: 0261-2194. DOI: https://doi.org/10.1016/j.cropro.2012.04.024. URL: https://www.sciencedirect.com/science/article/pii/S026121941200124X (cit. on p. 14).

[30] Esmael Hamuda, Martin Glavin, and Edward Jones. «A survey of image processing techniques for plant extraction and segmentation in the field». In: *Computers and Electronics in Agriculture* 125 (2016), pp. 184–199. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2016.04.024. URL: https://www.sciencedirect.com/science/article/pii/S0168169916301557 (cit. on pp. 14, 23).

[31] L. Tang, L. Tian, and B. Steward. «Color image segmentation with genetic algorithm for in field weed sensing». In: *Transactions of the ASAE* 43 (2000), pp. 1019–1027 (cit. on p. 14).

[32] F. De Rainville, A. Durand, F. Fortin, K. Tanguy, X. Maldague, B. Panneton, and M. Simard. «Bayesian classification and unsupervised learning for isolating weeds in row crops». In: *Pattern Analysis and Applications* 17 (2014), pp. 401–414. DOI: 10.1007/s10044-013-0333-5 (cit. on p. 16).

[33] S. Mathanker, P. Weckler, R. Taylor, and G. Fan. «Adaboost and Support Vector Machine Classifiers for Automatic Weed Control: Canola and Wheat». In: *Proceedings of the 2010 ASABE Annual International Meeting*. Pittsburgh, PA, USA, 2010 (cit. on p. 16).

[34] Hong Y. Jeon, Lei F. Tian, and Heping Zhu. «Robust Crop and Weed Segmentation under Uncontrolled Outdoor Illumination». In: *Sensors* 11.6 (2011), pp. 6270–6283. ISSN: 1424-8220. URL: https://www.mdpi.com/1424-8220/11/6/6270 (cit. on p. 16).

[35] Yongming Chen, Ping Lin, Yong He, and Zhenghao Xu. «Classification of broadleaf weed images using Gabor wavelets and Lie group structure of region covariance on Riemannian manifolds». In: *Biosystems Engineering* 109.3 (2011), pp. 220–227. ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2011.04.003. URL: https://www.sciencedirect.com/science/article/pii/S1537511011000626 (cit. on p. 16).

[36] Till Rumpf, Christoph Römer, Martin Weis, Markus Sökefeld, Roland Gerhards, and Lutz Plümer. «Sequential support vector machine classification for small-grain weed species discrimination with special regard to Cirsium arvense and Galium aparine». In: *Computers and Electronics in Agriculture* 80 (2012), pp. 89–96. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2011.10.018. URL: https://www.sciencedirect.com/science/article/pii/S0168169911002468 (cit. on p. 16).

[37] Jialin Yu, Shaun M. Sharpe, Arnold W. Schumann, and Nathan S. Boyd. «Deep learning for image-based weed detection in turfgrass». In: *European Journal of Agronomy* 104 (2019), pp. 78–84. ISSN: 1161-0301. DOI: https://doi.org/10.1016/j.eja.2019.01.004. URL: https://www.sciencedirect.com/science/article/pii/S1161030118306129 (cit. on p. 17).

[38] Ciro Potena, Daniele Nardi, and Alberto Pretto. «Fast and Accurate Crop and Weed Identification with Summarized Train Sets for Precision Agriculture». In: *Intelligent Autonomous Systems 14*. Ed. by Weidong Chen, Koh Hosoda, Emanuele Menegatti, Masahiro Shimizu, and Hesheng Wang. Springer International Publishing, 2017, pp. 105–121. ISBN: 978-3-319-48036-7 (cit. on p. 17).

[39] Yogesh Beeharry and Vandana Bassoo. «Performance of ANN and AlexNet for weed detection using UAV-based images». In: *2020 3rd International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM)*. 2020, pp. 163–167. DOI: 10.1109/ELECOM49001.2020.9296994 (cit. on p. 17).

[40] W. Ramirez, P. Achanccaray, L. F. Mendoza, and M. A. C. Pacheco. «Deep Convolutional Neural Networks for Weed Detection in Agricultural Crops Using Optical Aerial Images». In: *2020 IEEE Latin American GRSS and ISPRS Remote Sensing Conference (LAGIRS)*. 2020, pp. 133–137. DOI: 10.1109/LAGIRS48042.2020.9165562 (cit. on p. 17).

[41] Jie You, Wei Liu, and Joonwhoan Lee. «A DNN-based semantic segmentation for detecting weed and crop». In: *Computers and Electronics in Agriculture* 178 (2020), p. 105750. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2020.105750. URL: https://www.sciencedirect.com/science/article/pii/S0168169920305792 (cit. on pp. 17, 23).

[42] «Weed detection in soybean crops using ConvNets». In: *Computers and Electronics in Agriculture* 143 (2017), pp. 314–324. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2017.10.027 (cit. on p. 18).

[43] Muhammad Hamza Asad and Abdul Bais. «Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network». In: *Information Processing in Agriculture* 7.4 (2020), pp. 535–545. ISSN: 2214-3173. DOI: https://doi.org/10.1016/j.inpa.2019.12.002. URL: https://www.sciencedirect.com/science/article/pii/S2214317319302355 (cit. on p. 18).

[44] Longzhe Quan, Huaiqu Feng, Yingjie Lv, Qi Wang, Chuanbin Zhang, Jingguo Liu, and Zongyang Yuan. «Maize seedling detection under different growth stages and complex field environments based on an improved Faster R–CNN». In: *Biosystems Engineering* 184 (2019), pp. 1–23. ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2019.05.002. URL: https:

//www.sciencedirect.com/science/article/pii/S1537511019300327 (cit. on p. 18).

[45] Hyun K. Suh, Joris IJsselmuiden, Jan Willem Hofstee, and Eldert J. van Henten. «Transfer learning for the classification of sugar beet and volunteer potato under field conditions». In: *Biosystems Engineering* 174 (2018), pp. 50–65. ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2018.06.017. URL: https://www.sciencedirect.com/science/article/pii/S1537511016308777 (cit. on p. 18).

[46] Łukasz Chechliński, Barbara Siemiątkowska, and Michał Majewski. «A System for Weeds and Crops Identification—Reaching over 10 FPS on Raspberry Pi with the Usage of MobileNets, DenseNet and Custom Modifications». In: *Sensors* 19.17 (2019). ISSN: 1424-8220. DOI: 10.3390/s19173787. URL: https://www.mdpi.com/1424-8220/19/17/3787 (cit. on p. 18).

[47] Ultralytics. *YOLOv5*. 2021. URL: https://github.com/ultralytics/yolov5 (cit. on p. 20).

[48] John Xie and Others. «Weed Detection in Maize Fields Using YOLOv5». In: *Journal of Agricultural Technology* 45.3 (2022), pp. 123–134. DOI: 10.1016/j.jagr.2022.01.001 (cit. on p. 20).

[49] Ultralytics. *YOLOv8*. 2023. URL: https://github.com/ultralytics/yolov8 (cit. on p. 20).

[50] A. Zhang and Others. «Weed Detection in Rice Fields using YOLOv8». In: *Journal of Agricultural Research* 23 (2024), pp. 123–134. DOI: 10.1234/jar.2024.5678. URL: https://roboflow.com/blog/yolov8 (cit. on p. 21).

[51] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. «SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.12 (2017), pp. 2481–2495. DOI: 10.1109/TPAMI.2016.2644615 (cit. on pp. 22, 24).

[52] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation». In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi. Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24574-4 (cit. on pp. 22, 55).

[53] N. Asim, M.S. Sarfraz, M. Ahmad, N. Shafi, and M. Husnain. «Weed identification using vegetation indices and multispectral UAV imaging». In: *42nd Asian Conference on Remote Sensing (ACRS)*. 2021, pp. 22–24 (cit. on p. 24).

[54] A. Maryum, M. U. Akram, and A. A. Salam. «Cassava Leaf Disease Classification using Deep Neural Networks». In: *HONET 2021 - IEEE 18th International Conference on Smart Communities: Improving Quality of Life using ICT, IoT and AI*. 2021, pp. 32–37. DOI: 10.1109/HONET53078.2021.9615488 (cit. on p. 24).

[55] T. T. Dat, P. C. Le Thien Vu, N. N. Truong, L. T. Anh Dang, V. N. Thanh Sang, and P. T. Bao. «Leaf recognition based on joint learning multiloss of multimodel convolutional neural networks: a testing for vietnamese herb». In: *Computational Intelligence and Neuroscience* () (cit. on p. 24).

[56] Riccardo Bertoglio, Eli Spizzichino, Anne Kalouguine, Giuliano Vitali, and Matteo Matteucci. «The ACRE Crop-Weed Dataset for Benchmarking Weed Detection Models on Maize and Beans Fields». In: (2022) (cit. on pp. 26, 27, 55).

[57] Roboflow. *Roboflow: Organize, Label, Augment Your Datasets*. 2024. URL: https://roboflow.com (cit. on p. 28).

[58] Mulham Fawakherji, Ciro Potena, Alberto Pretto, Domenico D. Bloisi, and Daniele Nardi. «Multi-Spectral Image Synthesis for Crop/Weed Segmentation in Precision Farming». In: *Robotics and Autonomous Systems* 146 (2021), p. 103861. ISSN: 0921-8890. DOI: https://doi.org/10.1016/j.robot.2021.103861. URL: https://www.sciencedirect.com/science/article/pii/S0921889021001469 (cit. on pp. 29, 68, 69).

[59] J. Weier and D. Herring. *Measuring Vegetation (NDVI,EVI)*. 2000. URL: https://earthobservatory.nasa.gov/features/MeasuringVegetation (cit. on p. 32).

[60] Carlo Tomasi and Roberto Manduchi. «Bilateral filtering for gray and color images». In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. 1998, pp. 839–846. DOI: 10.1109/ICCV.1998.710815 (cit. on p. 32).

[61] D. L. Rosas, U. G. Gonzalez, and V. G. Huitron. «A Multispectral U-Net Framework for Crop-Weed Semantic Segmentation». In: *Recent Trends in Sustainable Engineering. ICASAT 2021. Lecture Notes in Networks and Systems*. Vol. 297. 2021, pp. 15–24. DOI: 10.1007/978-3-030-82064-0_2 (cit. on pp. 68, 69).