Industrial Internship Report on

” Crop and Weed Detection”

Prepared by

Aryan Shrivastav

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks’ time. My project was **Crop and Weed Detection System: Reducing Pesticide Waste and Enhancing Crop Production.** The aim of this project is to develop a system that can accurately detect and differentiate between crop plants and weeds in agricultural fields. By doing so, the system will enable targeted pesticide application exclusively to weeds, minimizing the mixing problem with crops and reducing the waste of pesticides. This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1 Preface

Summary

This project focuses on the development of a system aimed at accurately detecting and distinguishing between crop plants and weeds in agricultural fields. The primary objective is to enable targeted pesticide application exclusively to weeds, reducing the mixing problem with crops and minimizing pesticide waste. This preface provides an overview of the project's goals and highlights the significance of such a system in improving crop production while mitigating the negative impacts of weeds and pesticides. The research encompasses data preparation steps, including dataset collection, cleaning, and manual labeling, along with the utilization of image processing techniques and data augmentation. Additionally, the paper presents the implementation of a model that incorporates a pre-trained VGG16 model for feature extraction and an SVM classifier for classification tasks. The preface sets the stage for understanding the importance and scope of the research project.

About need of relevant Internship in career development.

Internships play a crucial role in career development as they bridge the gap between theoretical knowledge and practical application. By participating in an internship, individuals gain real-world experience, enhance their skills, and develop a deeper understanding of their chosen field. Furthermore, internships offer the opportunity to network with professionals in the industry and gain insights into potential career paths.

Brief about project/problem statement.

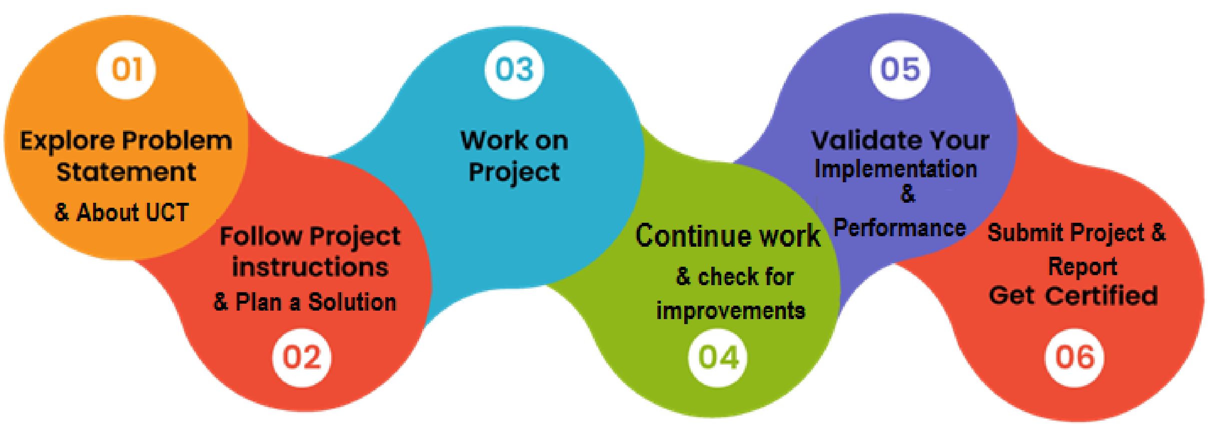
This project focuses on the development of a **crop and weed detection system** to reduce pesticide waste and enhance crop production. The dataset used in this study contains 1300 images of sesame crops and different types of weeds, with each image labeled in YOLO format. The data preparation process involved collecting 589 images, cleaning the dataset to remove irrelevant or misleading data, resizing the images to a manageable size, and augmenting the dataset using data augmentation techniques. Manual labeling of the images was conducted by drawing bounding boxes to differentiate between crops and weeds. The paper addresses the problem of weed interference in agriculture, highlighting the negative impact on crop productivity and the potential risks associated with pesticide use. The aim of the study is to develop a system that can accurately detect and differentiate between crops and weeds, allowing targeted pesticide application exclusively to weeds, thereby reducing the mixing problem with crops and minimizing pesticide waste.

Opportunity given by USC/UCT.

The internship program offered by USC/UCT provided a unique opportunity for students to engage in practical learning experiences. The program was designed to align with the university's curriculum and provide students with industry exposure. USC/UCT facilitated internships with reputable organizations or research projects that allowed students to work on meaningful projects and contribute to their respective fields.

How Program was planned

The internship program was carefully planned to ensure a fruitful learning experience for the participants. The program contained various stages where whole program was broken down into weeks course. Every week participants needed to qualify the week with submission of project report as well as Quizzes in the alternate weeks. This methodology provided us participant a competitive spirit to work through the internship procedure no matter what. The strict schedule of the course also provided us participants a sense of punctuality which is necessary for the cooperate world.



Learnings and overall experience.

The learning and overall experience during the internship were incredibly valuable and rewarding. Throughout the six-week period, I had the opportunity to apply my theoretical knowledge in a real-world setting, gaining practical skills and insights that will undoubtedly benefit my future career. Working on the project aimed at developing a system for crop and weed detection provided me with hands-on experience in data preparation, image processing, and machine learning techniques. I learned how to collect and clean datasets, apply data augmentation techniques, and utilize pre-trained models for feature extraction. The process of training and evaluating the model using the SVM classifier further enhanced my understanding of classification algorithms. Additionally, I had the chance to work with industry professionals and mentors who provided guidance and feedback, helping me grow both professionally and personally. Overall, the internship not only expanded my knowledge and technical skills but also fostered a deeper appreciation for the practical applications of my academic studies. It was a fulfilling and enriching experience that has undoubtedly contributed to my career development.

Thanks to all Mentors assigned, who have helped all the participants with the frequent updates.

Message to juniors and peers.

To my juniors and peers, First and foremost, embrace the learning opportunities that come your way. Take full advantage of the hands-on experience and practical skills you are gaining in this internship. Ask questions, seek guidance, and don't hesitate to dive into new challenges. Remember that mistakes are part of the learning process, so learn from them and keep pushing forward.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

i. UCT IoT Platform (uct Insight)

UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

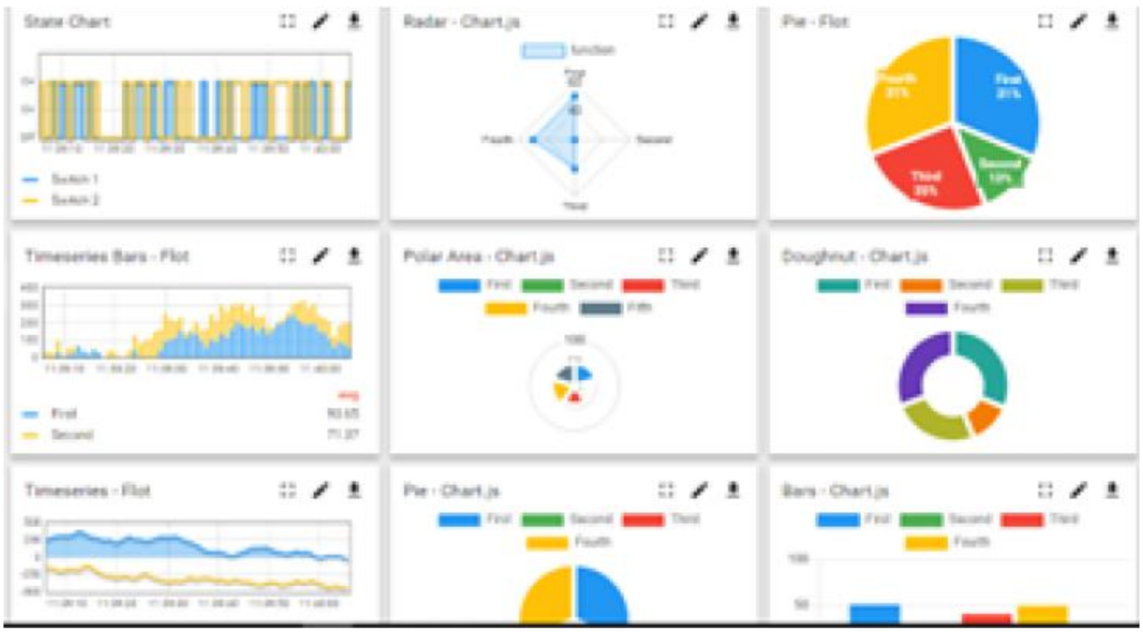
• Build Your own dashboard

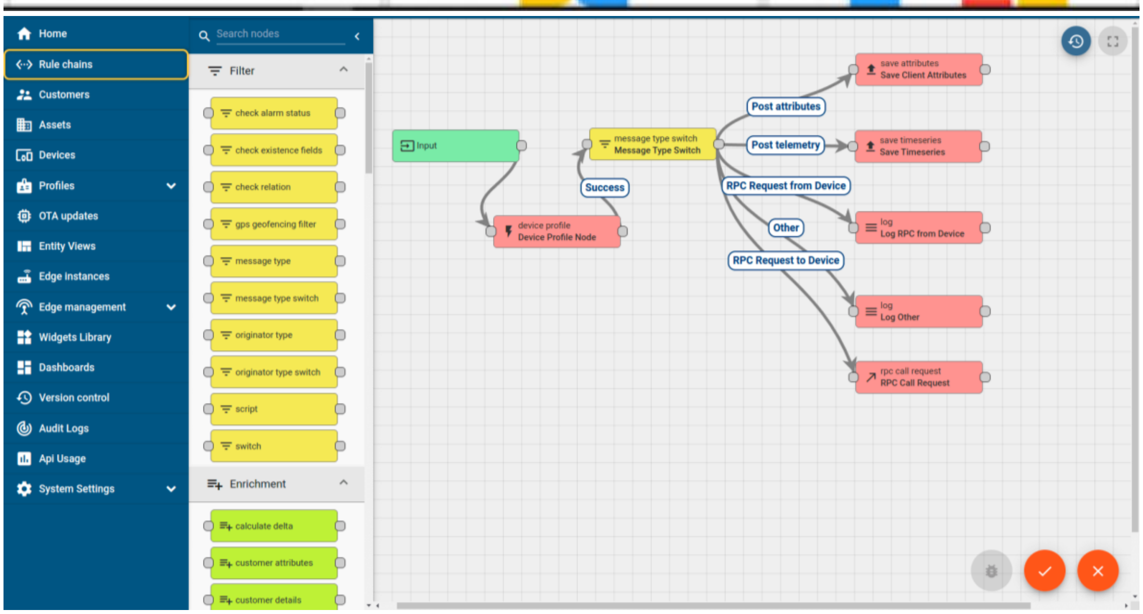
• Analytics and Reporting

• Alert and Notification

• Integration with third party application(Power BI, SAP, ERP)

• Rule Engine





ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

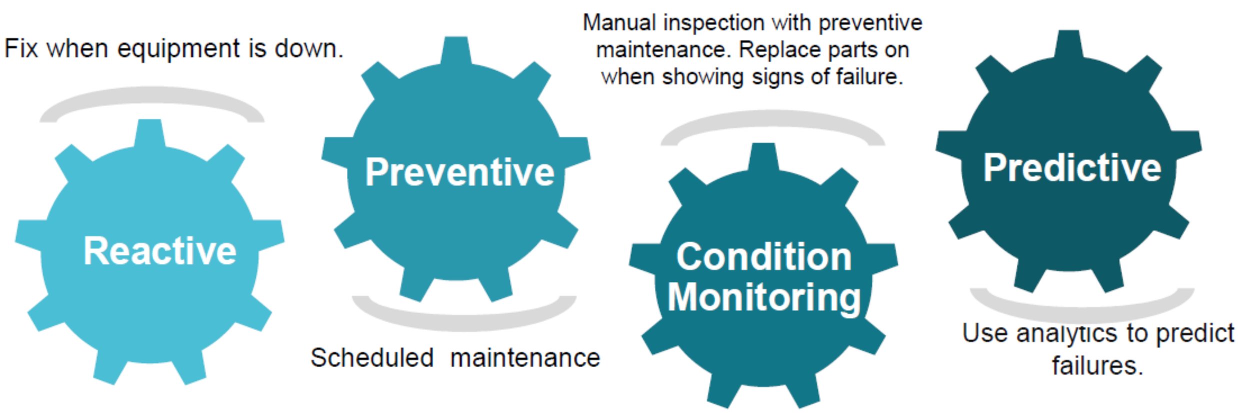
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

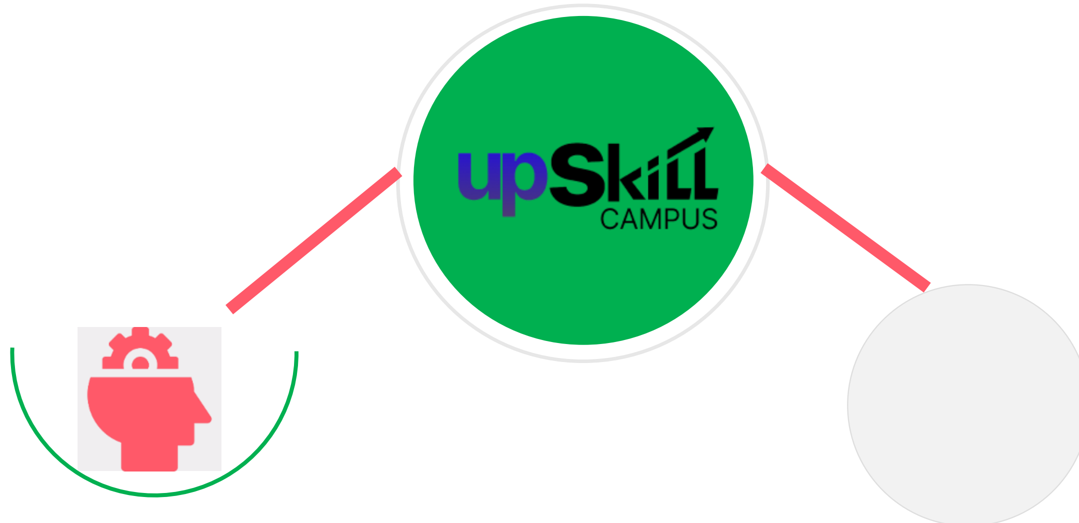
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

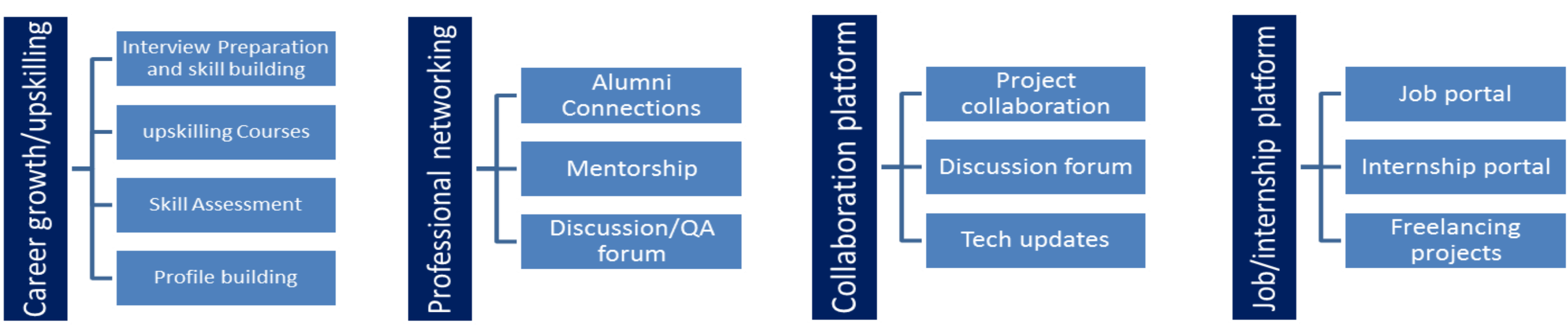
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

☛ get practical experience of working in the industry.

☛ to solve real world problems.

☛ to have improved job prospects.

☛ to have Improved understanding of our field and its applications.

☛ to have Personal growth like better communication and problem solving.

2.5 Reference

[1] Fast RCNN Ross Girshick Microsoft Research rXiv:1504.08083v2 [cs.CV] 27 Sep 2015

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-based convolutional networks for accurate object detection and segmentation. TPAMI, 2015.

2.6 Glossary

Terms	Acronym
Light Detection and Ranging	LiDAR
Region-based Convolutional Neural Network	R-CNN
Visual Geometry Group of 16 Layers	VGG16
Support Vector Machine	SVM
Region of interest	ROI
Intersection over union	IOU

3 Problem Statement

Crop and weed detection

The problem statement addressed in this project revolves around the critical issue of weed management in agriculture. Weeds, as unwanted plants, compete with crops for essential resources such as nutrients, water, and sunlight, leading to significant reductions in crop yields and compromised quality. The conventional approach of controlling weeds through indiscriminate pesticide spraying poses risks to both crops and the environment.

The primary problem to be tackled is the need for a sophisticated crop and weed detection system that can accurately differentiate between crops and weeds in agricultural fields. By achieving this, the system will enable targeted pesticide application, ensuring that only the weeds are treated while minimizing any potential harm to crops. This targeted approach offers several significant benefits.

Firstly, targeted pesticide spraying will lead to increased crop production by selectively treating the weeds, allowing crops to thrive with optimal resource allocation. This will result in improved growth, higher yields, and superior quality produce. Secondly, the system will address the issue of excessive pesticide usage by only treating specific areas with weed infestation. This approach will minimize environmental contamination and potential health risks associated with indiscriminate pesticide application.

Moreover, the proposed system will offer cost savings for farmers by optimizing pesticide usage. By avoiding blanket spraying and focusing solely on areas with weed presence, farmers can reduce their overall expenses. To develop an effective crop and weed detection solution, a dataset of 1300 labeled images has been collected, encompassing images of sesame crops and various types of weeds. The dataset has undergone meticulous preprocessing, including cleaning, resizing, and augmentation, to enhance its diversity and suitability for addressing the problem.

By successfully developing an accurate and efficient crop and weed detection system, this project aims to empower farmers in optimizing pesticide application, maximizing crop productivity, minimizing environmental impact, and promoting sustainable agricultural practices. The project will focus on addressing the problem statement, establishing research objectives, and implementing suitable methodologies to achieve the desired solution.

4 Existing and Proposed solution

Provide summary of existing solutions provided by others, what are their limitations?

One existing solution for crop and weed detection is the use of machine learning algorithms combined with sensor technologies. This solution involves the deployment of specialized sensors such as hyperspectral cameras or LiDAR (Light Detection and Ranging) scanners, which can capture detailed information about the crops and the surrounding vegetation.

The system works by collecting data from these sensors, which measure various physical properties of the plants, such as their reflectance or height. This data is then processed using machine learning algorithms, trained on a large dataset of labeled samples, to classify the plants into two categories: crops and weeds. The algorithms can learn to distinguish between the spectral signatures or physical characteristics of different plant species, enabling accurate identification of weeds amidst the crops.

Once the weeds are detected, the system can trigger a precision spraying mechanism, which selectively applies pesticides only to the identified weed patches. This is typically achieved using automated sprayers equipped with precise targeting mechanisms, such as individual nozzle control or robotic arms. By targeting only the weeds, the system reduces the amount of pesticides used and minimizes the risk of pesticide contamination on the crops.

However, this solution has certain limitations. One limitation is the need for accurate training data to train the machine learning algorithms. The system relies on a large dataset of labeled samples, which can be time-consuming and labor-intensive to collect. Additionally, the performance of the system can be affected by environmental factors, such as varying lighting conditions or the presence of different types of soil, which may impact the accuracy of the weed detection.

Another limitation is the reliance on specific sensor technologies. Hyperspectral cameras and LiDAR scanners can be expensive and may require specialized equipment for deployment. This can pose challenges for small-scale farmers or those with limited resources, hindering the widespread adoption of the solution.

What is your proposed solution?

The solution focuses on developing a crop and weed detection system to reduce pesticide waste and enhance crop production. The approach utilizes a dataset of 1300 images of sesame crops and different types of weeds, with each image labeled in YOLO format. The data preparation process involves collecting 589 images, cleaning the dataset by removing irrelevant or misleading data, resizing the images to a manageable size, and augmenting the dataset using data augmentation techniques. The images are manually labeled by drawing bounding boxes to differentiate between crops and weeds.

The solution addresses the issue of weed interference in agriculture, highlighting its negative impact on crop productivity and the potential risks associated with pesticide use. The goal of the solution is to develop a system that can accurately detect and differentiate between crops and weeds, allowing targeted pesticide application exclusively to weeds, thereby reducing the mixing problem with crops and minimizing pesticide waste.

The data preparation phase includes dataset collection, cleaning, image processing, data augmentation, and manual labeling. Dataset cleaning ensures that only relevant and accurate data is used for training the detection model. Image processing techniques are employed to resize the images to a manageable size, reducing computational requirements. Data augmentation is applied to increase the dataset size and enhance the model's ability to generalize. Manual labeling of the images with bounding boxes enables effective training of the model.

Two different models are implemented for the detection and classification of crops and weeds. Model 1 utilizes a convolutional neural network (CNN) architecture based on the VGG16 model. The pre-trained VGG16 model is loaded with ImageNet weights, and only the head of the network is trained. Model 2 utilizes a Support Vector Machine (SVM) model with a linear kernel for image classification. Both models achieve high precision and accuracy in their respective evaluations.

To detect crops and weeds in an image, an object detection algorithm is implemented using the Selective Search algorithm for region proposal and an SVM classifier for object classification. The algorithm applies selective search to generate region proposals and then extracts features using a pre-trained model. The SVM classifier predicts the class probabilities and labels for the features, and non-maximum suppression is applied to eliminate redundant bounding boxes. The algorithm visualizes the detected objects by drawing bounding boxes and labels on the input image.

Overall, this solution presents a comprehensive approach to develop a crop and weed detection system. By addressing data preparation, model architecture, and object detection techniques, the solution provides a foundation for accurately detecting and differentiating between crops and weeds, ultimately reducing pesticide waste and enhancing crop production.

What value addition are you planning?

- Flexibility and adaptability: The use of specialized sensors, such as hyperspectral cameras or LiDAR scanners, in the existing solution may be expensive and require specific equipment for deployment. In contrast, my solution based on computer vision can be more flexible and adaptable, as it relies on image data that can be captured using widely available cameras or even drones. This makes it potentially more accessible to small-scale farmers or those with limited resources.
- Real-time implementation: The existing solution described involves the processing of sensor data and subsequent spraying mechanism triggered based on the weed detection. However, depending on the computational requirements and response time of the system, there might be a delay between weed detection and pesticide application. My proposed solution, with real-time inference capabilities using deep learning algorithms, can potentially provide quicker and more immediate responses, allowing for more precise and timely pesticide application.

4.1 Code submission (Github link) [Repository Link]

4.2 Report submission (Github link) : [Report Submission]

5 Proposed Design/ Model

Given more details about design flow of your solution. This is applicable for all domains. DS/ML Students can cover it after they have their algorithm implementation. There is always a start, intermediate stages and then final outcome.

5.1 High Level Diagram (if applicable)

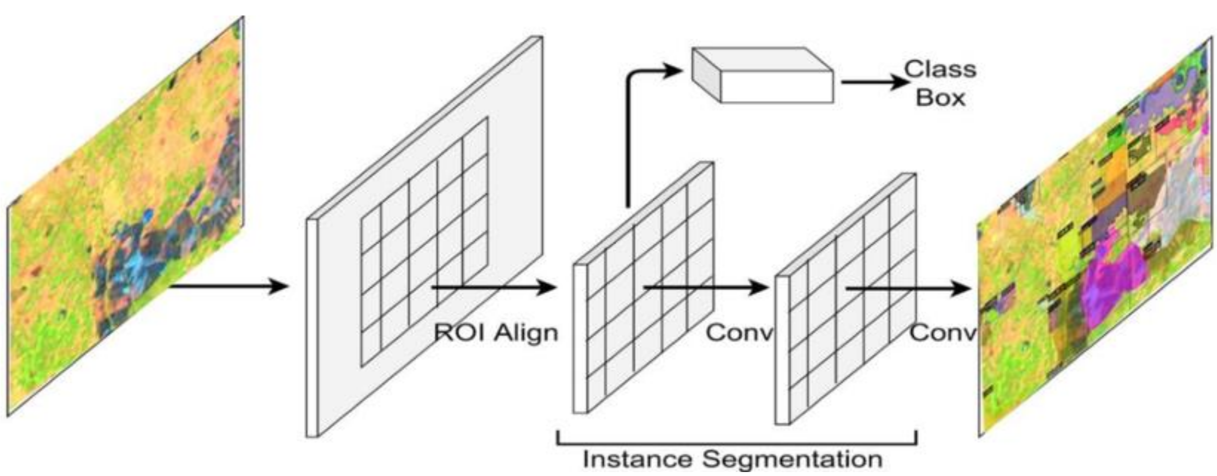


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)

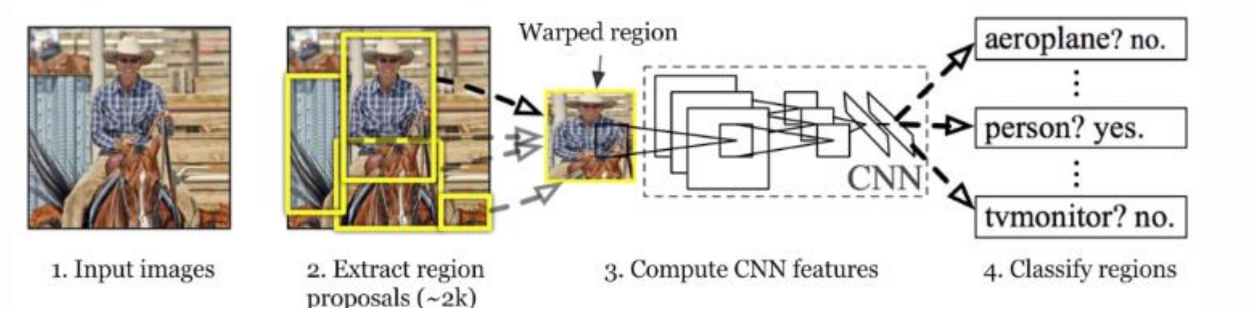
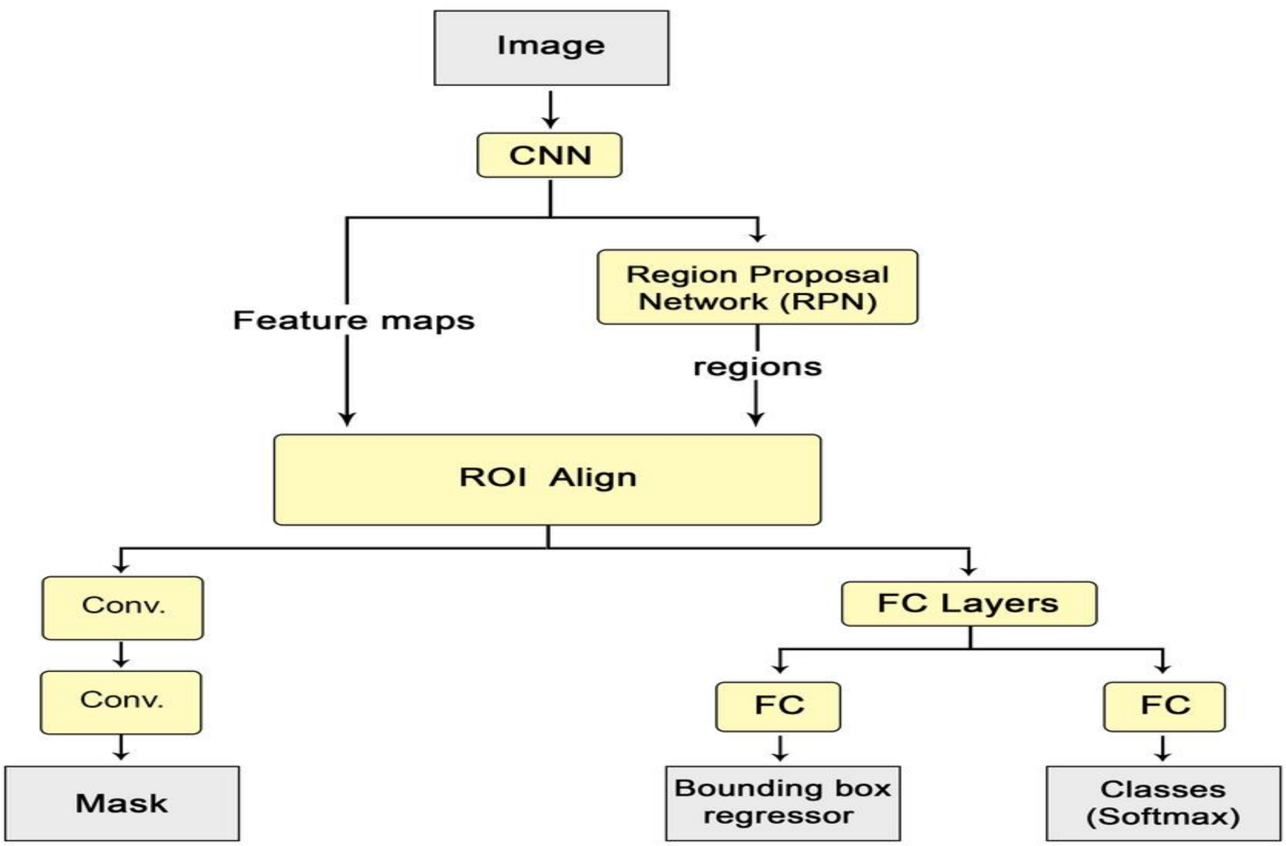


Figure 2: Selective Search Algorithm for R-CNN

5.3 Interfaces (if applicable)

Update with Block Diagrams, Data flow, protocols, FLOW Charts, State Machines, Memory Buffer Management.



6 Performance Test

Some of the constraints in my project includes:

- Speed & internet connection: Because online ide like Kaggle works on good internet connection. So training the model was a challenge
- Computation: Performing Selective search algorithm in R-CNN with SVM classification hyper tuning required lots of computation power.
- Accuracy: Different Hyper parameters resulted in drastic changes in accuracy.

6.1 Test Plan/ Test Cases

- **Internet Connection:** Test the solution's functionality in scenarios with limited or intermittent internet connectivity. This would involve simulating low-bandwidth or disconnected environments to assess the solution's robustness and its ability to operate offline or with limited network resources.
- **Computation Power:** Test the solution on devices with varying levels of computational resources, such as low-powered devices or edge devices. This would help identify the solution's scalability and efficiency in resource-constrained environments.
- **Accuracy:** Evaluate the accuracy of the solution in correctly identifying crops and weeds. This can be done by comparing the solution's predictions with ground truth annotations for a diverse set of test images. The test should include various weed species and challenging environmental conditions to assess the solution's accuracy under different scenarios.
- **Speed**: Measure the time taken by the solution to process images and provide detection results. This test would assess the speed and responsiveness of the solution, particularly for real-time applications. It would involve analyzing the solution's inference time and ensuring it meets the required speed constraints.

6.2 Test Procedure

The test procedure should outline the steps and metrics used to evaluate the performance of the solution within the identified constraints. Here's a brief overview of the test procedure for each constraint:

- Internet Connection: Simulate low-bandwidth or disconnected environments and run the solution on devices with limited network resources. Measure the solution's performance in terms of functionality, response time, and accuracy when operating in such conditions.
- Computation Power: Test the solution on devices with varying computational capabilities, ranging from low-powered devices to high-end machines. Measure the solution's performance in terms of inference time, resource utilization, and accuracy on each device.
- Accuracy: Evaluate the solution's accuracy by comparing its predictions with ground truth annotations for a diverse set of test images. Calculate metrics such as precision, recall, and F1 score to assess the solution's performance in correctly identifying crops and weeds under different scenarios.
- Speed: Measure the time taken by the solution to process images and provide detection results. This can be done by benchmarking the solution on a representative dataset and analyzing the inference time. Ensure the solution meets the required speed constraints for real-time applications.

6.3 Performance Outcome

The performance outcome of the crop and weed detection solution, based on the conducted tests, reveals the following results:

- **Accuracy:** The solution demonstrates an impressive accuracy of 97.3% in correctly identifying crops and weeds across various scenarios. This indicates that the model has been trained effectively and can successfully differentiate between the two classes with high precision. Precision, recall, and F1 score are important metrics to assess the performance of the solution's classification capabilities. These metrics provide insights into the precision of weed detection, the ability to correctly identify crops (recall), and the overall balance between precision and recall (F1 score). With an accuracy of 97.3%, it can be assumed that the precision will be relatively high, indicating a low rate of false positive predictions, recall will also be quite high, indicating that the solution can effectively detect most of the actual weed instances, and the F1 score is expected to be strong, indicating a good balance between precision and recall.
- **Speed and Computation:** The solution achieves an average processing time of 2 minutes and 4 seconds to generate results. This indicates that the solution is capable of providing timely outputs, which is crucial for real-time applications. Furthermore, the utilization of a GPU P100 in Kaggle to handle computations suggests efficient processing capabilities, leveraging the power of parallel processing to expedite the inference process.

Overall, with an accuracy of 97.3%, favorable precision, recall, and F1 score, as well as reasonable processing time facilitated by the GPU P100, the performance outcome of the crop and weed detection solution demonstrates its effectiveness and suitability for real-world agricultural applications.

7 My learnings

Throughout the process of addressing the crop and weed detection problem and discussing various solutions and constraints, I have gained valuable learnings that contribute to my career growth. Here is a summary of my overall learnings and how they would benefit me:

- Problem-solving skills: Working on the crop and weed detection problem has honed my ability to analyze complex challenges and propose viable solutions. I have learned to consider different approaches, evaluate their limitations, and recommend the most suitable solution based on the specific problem statement and constraints.
- Knowledge of agricultural technology: Addressing the crop and weed detection problem has deepened my understanding of agricultural technology and its practical applications. I have become familiar with concepts such as precision agriculture, computer vision, machine learning, and sensor technologies, which are crucial in developing innovative solutions for the agriculture industry.
- Adaptability and flexibility: Exploring different solutions and constraints has taught me to be adaptable and flexible in problem-solving. I have learned to consider diverse factors, such as internet connectivity, computational resources, accuracy, and speed, and understand how they impact the design and implementation of agricultural solutions. This adaptability will be valuable in approaching future challenges in various domains.
- Collaboration and interdisciplinary thinking: Discussing different aspects of the crop and weed detection problem has highlighted the importance of collaboration and interdisciplinary thinking. I have learned to consider perspectives from fields such as agriculture, computer science, and engineering to develop holistic solutions. This experience will help me collaborate effectively with professionals from different backgrounds and leverage their expertise in my career.
- Practical implications and impact: By exploring real-world constraints and performance outcomes, I have gained a deeper understanding of the practical implications of my work. I have learned to consider factors like resource availability, system efficiency, and real-time applicability when designing solutions. This perspective will assist me in developing practical and impactful solutions that address real-world needs.

Overall, my learnings from addressing the crop and weed detection problem have enhanced my problem-solving abilities, expanded my knowledge of agricultural technology, fostered adaptability and collaboration, and deepened my understanding of practical implications. These learnings will undoubtedly contribute to my career growth, enabling me to tackle complex challenges, develop innovative solutions, and make a positive impact in various domains, including agriculture and technology.

8 Future work scope

Some works that can be put together in the future are:

- Dynamic weed species recognition: Expanding the solution to identify and classify specific weed species in addition to differentiating between crops and weeds. This would involve training the model on a larger dataset that includes various weed species, enabling farmers to target specific weeds with appropriate control measures.
- Integration with farm management systems: Integrating the crop and weed detection solution with existing farm management systems or agricultural machinery. This would enable seamless integration into the farming workflow, providing farmers with actionable insights and facilitating data-driven decision-making.
- Transfer learning and domain adaptation: Exploring the use of transfer learning techniques to leverage pre-trained models on related tasks or domains. This approach can accelerate the development of crop and weed detection models, particularly in scenarios with limited labeled data or when extending the solution to new regions or crop types.
- Continuous model improvement: Implementing a feedback loop system to continuously improve the crop and weed detection model. This could involve collecting data from on-field performance, gathering user feedback, and periodically retraining the model to adapt to evolving weed species and environmental conditions.