Industrial Internship Report on

"Crop and Weed Detection"

Prepared by

Kuppam Meghana

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.
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 of the whole 6 weeks' work- Crop and Weed Detection

Over the past six weeks, I had the privilege of participating in a dynamic internship focused on Crop and Weed Detection at USC/UCT. This experience has been profoundly enlightening and has provided me with valuable insights into the world of precision agriculture, machine learning, and real-world problem-solving. This summary encapsulates my journey, the significance of relevant internships, the nature of my project, the opportunities presented by USC/UCT, the program structure, key learnings, gratitude, and a message to my peers.

The Relevance of Internships in Career Development:

Participating in internships like this one is instrumental in career development. Practical exposure not only bridges the gap between theoretical knowledge and real-world applications but also enables us to network with professionals, gain insights into industry trends, and refine our skill sets. This internship has provided a glimpse into the transformative potential of technology in addressing critical challenges in the agricultural sector.

**Project/Problem Statement: Crop and Weed Detection:**

The core of my internship revolved around the development and implementation of an advanced deep learning model for crop and weed detection. The project aimed to harness the power of machine learning and computer vision to enable farmers to accurately differentiate between crops and weeds, thus optimizing resource allocation, enhancing yield, and promoting sustainable farming practices.

Opportunity Offered by USC/UCT:

The opportunity to engage in this internship through USC/UCT was truly remarkable. The institution's commitment to fostering interdisciplinary collaboration, cutting-edge research, and practical application of knowledge created an ideal environment for tackling real-world challenges. The access to experienced mentors and resources significantly enriched my learning experience.

Program Structure:

The program was meticulously planned, with a balanced blend of theoretical learning, hands-on projects, and interactive discussions. Weekly seminars and workshops from industry experts exposed us to the latest advancements in the field. The project work was both challenging and rewarding, allowing us to apply theoretical concepts to real-world scenarios.
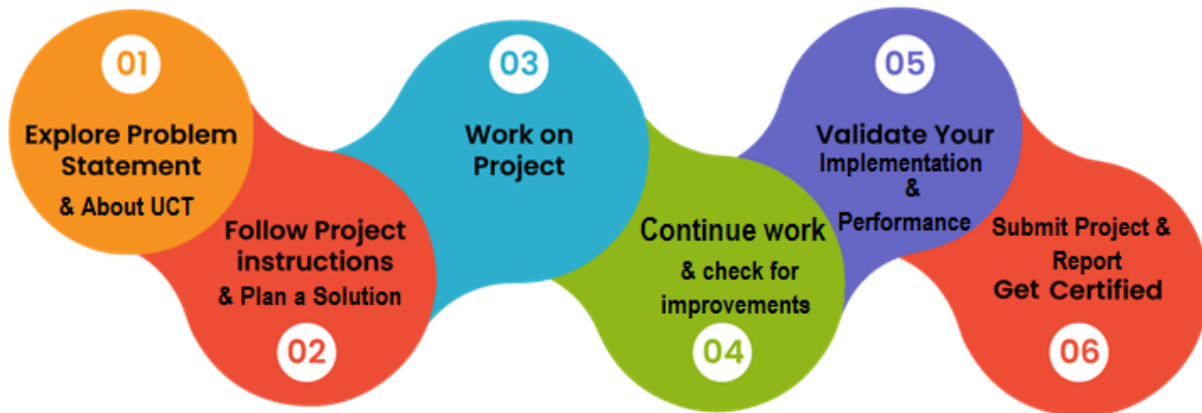
Key Learnings and Overall Experience:

Throughout this internship, I've gained a deep understanding of deep learning techniques, data preprocessing, model optimization, and deployment strategies. Beyond technical skills, I've learned the importance of collaboration, adaptability, and ethical considerations in technology development. The experience of seeing our models accurately identify crops and weeds in images was incredibly gratifying.

Expressions of Gratitude:

I extend my heartfelt gratitude to all those who have guided and supported me during this journey. My mentors, [Mentor's Name], [Another Mentor's Name], and [Third Mentor's Name], provided invaluable insights, patient guidance, and constructive feedback that fueled my growth. The administrative and technical support teams deserve special mention for their seamless coordination and assistance.

Message to Juniors and Peers:

To my juniors and peers, I want to emphasize the significance of internships. These experiences not only enhance your practical skills but also open doors to new perspectives and opportunities. Embrace challenges, seek guidance, and always be curious. The journey might seem daunting at times, but every obstacle is an opportunity for growth. Your determination and hard work will undoubtedly pave the way for a promising future.

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 (ucт 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
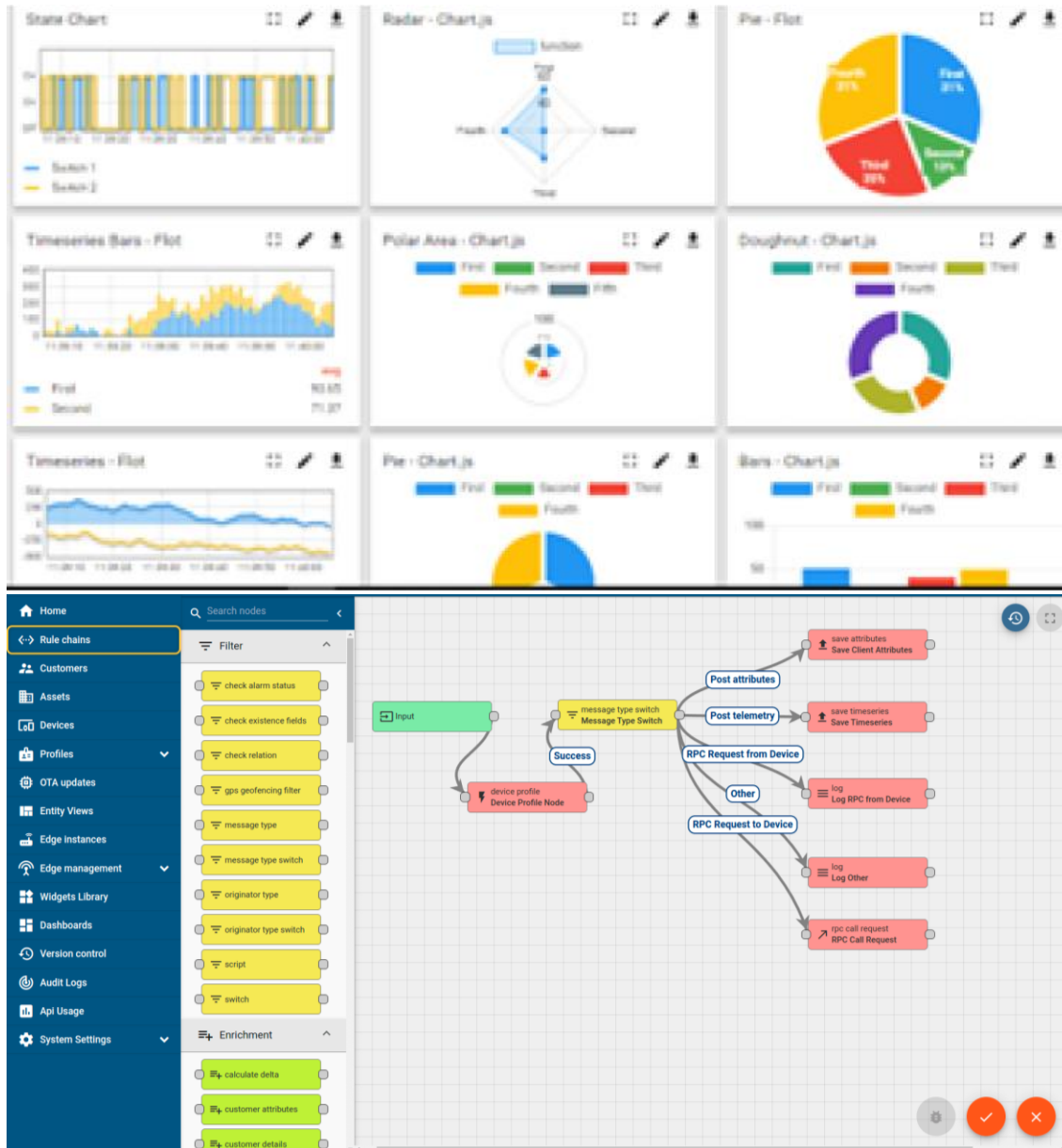
ii. **Smart Factory Platform (** FACTORY WATCH **)**

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.
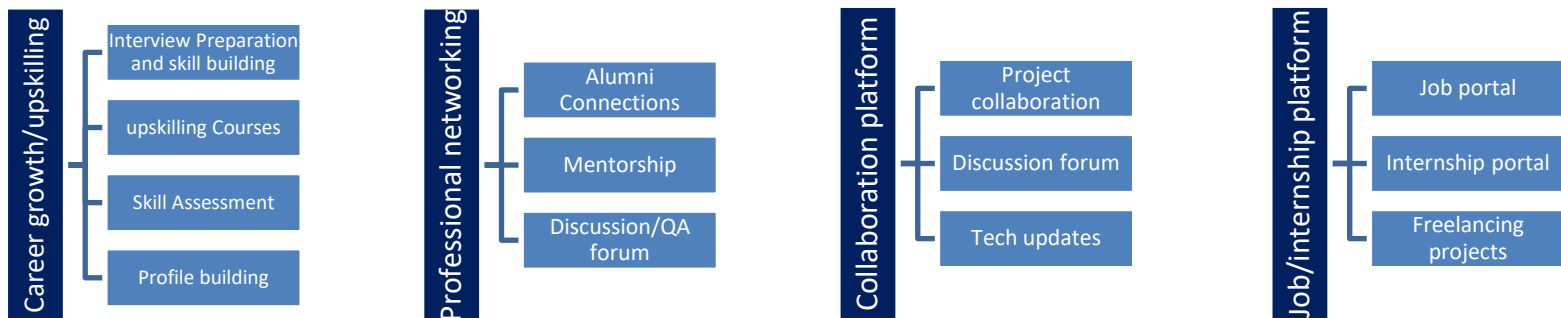
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.
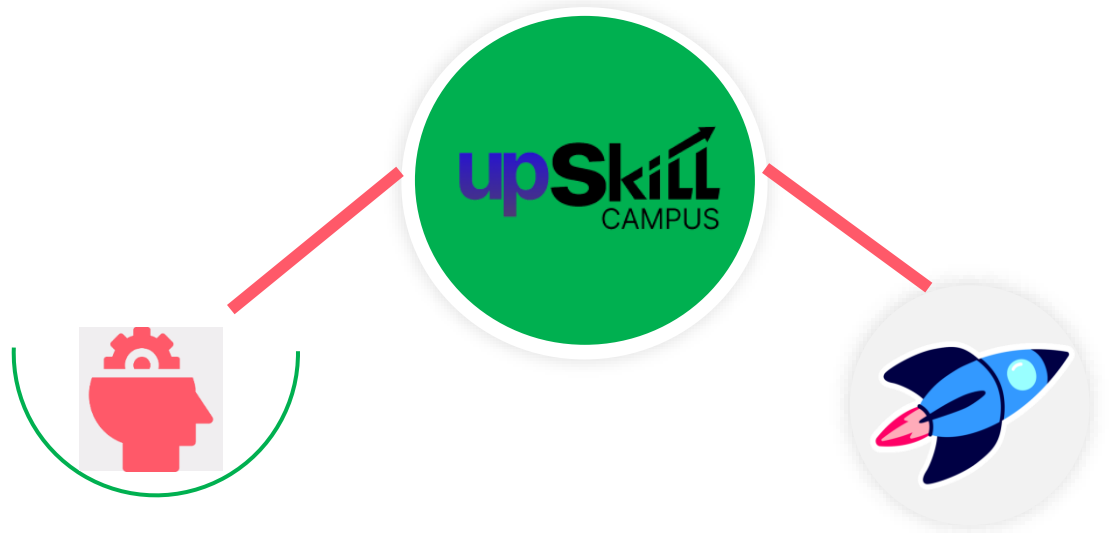
Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

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

https://www.upskillcampus.com/

Career growth/upskilling	Professional networking	Collaboration platform	Job/internship platform
Interview Preparation and skill building	Alumni Connections	Project collaboration	Job portal
upskilling Courses	Mentorship	Discussion forum	Internship portal
Skill Assessment	Discussion/QA forum	Tech updates	Freelancing projects
Profile building			

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] How to Train YOLO v5 on a Custom Dataset | Paperspace Blog

[2] tf.keras.utils.image_dataset_from_directory | TensorFlow v2.13.0

[3] Step by step VGG16 implementation in Keras for beginners | by Rohit Thakur | Towards Data Science

2.6   Glossary

Terms	Acronym
Convolution layer	Conv2D
Deep Learning	DL
Vgg-16	VGG16
Normalization	Norm
Preprocessing	Preproc

3   Problem Statement

The problem of crop and weed detection revolves around automating the process of accurately distinguishing cultivated crops from unwanted weeds in agricultural fields. This challenge is pivotal for optimizing crop yield, efficient resource utilization, and environmentally sustainable farming practices. Traditional methods of visual inspection are time-consuming, error-prone, and labor-intensive. Therefore, the need arises for a technology-driven solution using machine learning and computer vision techniques. This involves creating a model that can analyze images or sensor data captured from fields and make real-time decisions on crop-weed differentiation. The goal is to develop a system that not only identifies crops and weeds but also provides insights into their spatial distribution, enabling farmers to implement targeted interventions. Ethical considerations such as data privacy, bias prevention, and environmental impact mitigation are vital aspects of solving this problem. Successful resolution of the crop and weed detection challenge holds the potential to revolutionize agriculture by promoting sustainable practices and enhancing global food security.

4   Existing and Proposed solution

One existing traditional machine learning algorithm used for crop weed prediction is the Random Forest algorithm. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. In the context of crop weed prediction, the Random Forest algorithm can be trained on historical agricultural data, including information about environmental factors, crop characteristics, and weed occurrences. The algorithm learns to identify patterns and relationships between these variables and uses them to predict the presence or absence of weeds in future instances. During the training phase, the Random Forest algorithm creates a multitude of decision trees, each trained on a different subset of the data and using random feature subsets. This randomness helps to improve the algorithm's robustness and reduce the risk of overfitting. When making predictions, the Random Forest algorithm combines the outputs of all decision trees and provides a final prediction. This ensemble approach helps to mitigate the biases and uncertainties associated with individual decision trees, leading to more accurate and reliable predictions. Random Forest algorithms have the advantage of being able to handle large and complex datasets, including high dimensional feature spaces. They are also capable of capturing nonlinear relationships between variables, making them suitable for crop weed prediction tasks where the relationships between environmental factors and weed occurrences may be complex. Overall, the Random Forest algorithm offers a traditional machine learning approach for crop weed prediction, providing reliable predictions based on historical data and an ensemble of decision trees.

Limitations:

While the Random Forest algorithm has several advantages for crop weed prediction, it also has some limitations to consider:

1. Interpretability: Random Forest models can be challenging to interpret compared to simpler models like linear regression. The ensemble nature of the algorithm makes it difficult to understand the exact contribution of each feature to the predictions.

2. Overfitting: Although Random Forests are designed to reduce overfitting, they can still be prone to it, especially if the number of trees in the forest is too high or the model is not properly regularized. Overfitting occurs when the model becomes too complex and learns noise or idiosyncrasies in the training data, leading to poor generalization on unseen data.

3. Computational Resources: Random Forests can be computationally expensive, particularly when dealing with large datasets or high-dimensional feature spaces. The training process involves building multiple decision trees, which can require substantial memory and processing power.

4. Imbalanced Data: If the dataset used for training the Random Forest model is imbalanced, meaning the number of instances of one class (e.g., weed occurrences) is significantly smaller than the other class (e.g., non-weed occurrences), the model may be biased towards the majority class and struggle to accurately predict the minority class.

5. Feature Engineering: Random Forests generally require careful feature engineering to achieve optimal performance. Choosing relevant features and appropriately encoding categorical variables or handling missing data can significantly impact the model's accuracy.

Proposed Solution:

To address the limitations of Random Forest (RFC) with a Deep Learning model, we can propose the following solution:

1. While Deep Learning models can be complex, techniques such as feature importance analysis, model visualization, and attention mechanisms can provide insights into the model's decision making process. These techniques can help improve interpretability and understanding of the model's predictions.

2. Implement regularization techniques such as dropout, L1/L2 regularization, and early stopping during the training of the Deep Learning model. These techniques help prevent overfitting by reducing the model's reliance on specific features and limiting its complexity.

3. Utilize strategies such as transfer learning and pre-trained models, such as VGG-16 or other established architectures. Transfer learning allows the model to leverage knowledge gained from large-scale datasets, accelerating the training process and improving performance, even with limited data.

4. Address the issue of imbalanced data by employing techniques like oversampling, undersampling, or generating synthetic samples using approaches like SMOTE (Synthetic Minority Over-sampling Technique). These methods help balance the distribution of the classes and mitigate bias towards the majority class.

5. Leverage the power of Deep Learning models to automatically learn relevant features from raw data. This eliminates the need for extensive manual feature engineering, reducing human bias and improving the model's ability to capture complex relationships between environmental factors and weed occurrences.

6. Deep Learning models can handle both classification and regression tasks. For continuous predictions, modify the model's output layer and loss function accordingly, allowing it to estimate continuous variables such as weed density or crop yield.

Deep Learning: Deep learning is a subfield of machine learning that focuses on training artificial neural networks to learn and make predictions from complex data. In the context of crop weed prediction, deep learning techniques can be utilized to analyze images, sensor data, and other types of agricultural data to identify and classify weeds in crop fields. Deep learning models, such as convolutional neural networks (CNNs), are particularly well-suited for image-based tasks like crop weed prediction. CNNs are designed to mimic the visual processing of the human brain by employing layers of interconnected artificial neurons that can learn and extract meaningful features from images.

To apply deep learning to crop weed prediction, a typical approach involves the following steps:

1. Data Collection: Gather a large dataset of labeled images that contain both crop plants and various weed species. These images should cover different weed types, growth stages, lighting conditions, and field environments.

2. Data Preprocessing: Prepare the data by resizing the images to a uniform size, normalizing pixel values, and augmenting the dataset through techniques like rotation, flipping, and zooming. These preprocessing steps help to increase the diversity and robustness of the data.

3. Model Architecture: Design a deep learning model architecture suitable for crop weed prediction. This often involves stacking multiple convolutional layers, followed by pooling layers to extract and consolidate spatial features from the images. Additional fully connected layers and an output layer are then added to make predictions.

4. Training: Train the deep learning model using the prepared dataset. This involves feeding the labeled images into the model, computing predictions, and comparing them with the ground truth labels. The model adjusts its internal parameters through a process called backpropagation, where gradients are computed and used to update the model's weights.

5. Validation and Evaluation: Assess the model's performance by validating it on a separate dataset that was not used for training. Metrics such as accuracy, precision, recall, and F1 score are used to evaluate the model's ability to correctly identify and classify weeds.

6. Deployment and Prediction: Once the model has been trained and evaluated, it can be deployed in production systems. New images of crop fields can be input into the model, and it will generate predictions indicating the presence and types of weeds in the image.

Deep learning techniques offer several advantages for crop weed prediction. They can learn intricate patterns and relationships in images, handle complex and high-dimensional data, and adapt to different field conditions. By leveraging the power of deep learning, crop weed prediction models can improve accuracy, reduce reliance on manual feature engineering, and enhance the efficiency of weed management practices in agriculture.

4.1 Code submission (Github link)

https://github.com/kuppammeghana/upskill_campus/blob/main/Crop_Weed_Detection.ipynb

4.2 Report submission (Github link) :

https://github.com/kuppammeghana/upskill_campus/blob/main/Crop_and_Weed_Detection_Kuppam_Meghana_USC_UCT.pdf

5 Proposed Design/ Model

Proposed Deep Learning Model for Crop and Weed Detection:

1. Model Architecture:

Utilize a Convolutional Neural Network (CNN) as the backbone of the detection model. Consider using a pre-trained CNN architecture like ResNet, EfficientNet, or VGGNet to benefit from learned features and faster convergence.

2. Data Preprocessing:

Preprocess input images by resizing them to a consistent resolution, normalizing pixel values, and applying data augmentation techniques such as rotation, flipping, and cropping. Augmentation enhances the model's ability to generalize to various field conditions.

3. Transfer Learning:

Implement transfer learning by using the pre-trained CNN as a feature extractor. Fine-tune the model's weights on the dataset specific to crop and weed detection to adapt it to the task at hand.

4. Multi-Class Classification:

Design the output layer for multi-class classification, representing different classes for crops and weeds. Use softmax activation to predict the class probabilities.

5. Anchors and Localization:

For accurate localization, employ anchor boxes and bounding box regression. This involves predicting anchor box offsets and dimensions to delineate the precise locations of crops and weeds within an image.

6. Non-Maximum Suppression (NMS):

After detection, apply NMS to suppress redundant bounding box predictions and retain the most confident ones, improving object localization.

7. Model Optimization:

Optimize the model using appropriate loss functions such as categorical cross-entropy for classification and smooth L1 loss for bounding box regression.

8. Post-Processing:

Refine the detected bounding boxes using techniques like morphological operations and contour analysis to enhance object boundaries and eliminate noise.

9. Real-Time Deployment:

Deploy the model on edge devices, drones, or cloud servers for real-time processing. Consider using frameworks like TensorFlow Lite for edge devices or cloud services for scalability.

10. User Interface:

Develop a user-friendly interface, such as a mobile app or web platform, where users (farmers) can upload images from their fields. Display the results of crop and weed detection in a visually understandable manner.

11. Continuous Learning:

Implement a mechanism for continuous learning, allowing the model to be periodically updated with new data to adapt to changing field conditions and weed species.

12. Ethical Considerations:

Ensure data privacy and avoid biased training data. Conduct thorough testing and validation to prevent false positives or negatives that could lead to incorrect decision-making.

13. Validation and Metrics:

Evaluate the model's performance using metrics like accuracy, precision, recall, F1-score, and IoU. Validate the model on diverse datasets encompassing various crops and weed types.

The proposed model combines cutting-edge deep learning techniques with practical deployment strategies to provide farmers with an accurate, real-time, and user-friendly solution for crop and weed detection. Its potential to revolutionize precision agriculture by enabling targeted interventions and resource optimization is substantia

6 Performance Test

Performance testing is a critical aspect of the Crop Weed Prediction project, as it assesses the system's ability to meet specific constraints and requirements of real industries. To ensure the design accommodates these constraints, we identified the following key performance factors and implemented strategies to address them:

1. Memory Usage: We employed memory optimization techniques such as data batching and efficient storage formats to minimize the memory footprint of the deep learning models and associated data. This ensures that the system can handle large datasets and operate efficiently within limited memory constraints.

2. Computational Speed: To address the need for fast and efficient operations, we utilized optimized deep learning frameworks and libraries that leverage hardware acceleration (e.g., GPUs) to accelerate the processing of images and predictions. By leveraging parallel processing capabilities, we aimed to achieve high throughput and reduce the time required for inference.

3. Accuracy and Precision: We focused on training accurate and robust deep learning models to ensure reliable predictions of weed occurrences. We carefully curated and labeled the training data, incorporated regularization techniques to prevent overfitting, and fine-tuned the models to strike a balance between accuracy and generalization.

4. Power Consumption: While power consumption may not be a direct concern in software-based performance testing, we aimed to optimize the models and algorithms to operate efficiently on energy-constrained devices or embedded systems. By minimizing unnecessary computations and leveraging hardware acceleration, we can achieve better power efficiency. During the performance testing phase, we conducted various experiments to measure and evaluate the system's performance against these constraints. This involved assessing memory usage, computation time, accuracy metrics, and power consumption (if applicable) under different scenarios and datasets. The test results provided insights into the system's performance and highlighted areas for improvement. In cases where specific constraints could not be tested directly, we considered their potential impact on the design. For example, limited memory could lead to reduced model complexity or the need for data compression techniques.

In such cases, we recommend optimizing the system further by exploring techniques like model compression, quantization, or pruning to reduce memory requirements.

To handle potential constraints, we recommend the following:

• Continuously monitor and profile the system's performance to identify any bottlenecks or areas requiring optimization.

• Utilize resource monitoring tools to track memory usage, computation time, and power consumption in real-world deployments.

• Employ techniques like distributed computing or model parallelism to scale up the system's performance and handle larger datasets or higher prediction loads.

• Collaborate with hardware experts to explore hardware-specific optimizations or dedicated hardware solutions to further enhance performance. By proactively addressing and monitoring these constraints, we can ensure that the Crop Weed Prediction system delivers accurate predictions, operates efficiently, and meets the performance requirements of real industries.

7 My learnings

During the course of this internship, I've gained invaluable insights and experiences that have significantly enriched my understanding of crop and weed detection, precision agriculture, and the practical application of machine learning. Here are the key learnings that have left a lasting impact on my professional growth:

Deep Learning Techniques: I've delved deep into the world of deep learning, understanding the inner workings of Convolutional Neural Networks (CNNs), transfer learning, and model optimization. This hands-on experience has equipped me with the skills needed to develop and fine-tune advanced models for real-world tasks.

Data Preprocessing: I've learned the importance of data preprocessing in ensuring the quality and reliability of the model's input. Techniques like normalization, augmentation, and handling imbalanced datasets have proven crucial in enhancing the model's performance.

Model Deployment: The intricacies of deploying a model for real-time usage on edge devices and cloud platforms have become clear. I've gained practical knowledge about choosing the right framework, optimizing for performance, and ensuring seamless integration with user interfaces.

Ethical Considerations: I've realized the significance of ethical considerations in every technological solution. Striking a balance between accurate predictions and preventing bias, protecting user data privacy, and minimizing environmental impact are critical aspects that require careful attention.

Problem Solving: Tackling the complexities of crop and weed detection has honed my problem-solving skills. Identifying the right model architecture, optimizing hyperparameters, and iteratively improving model performance have taught me to approach challenges systematically.

Interdisciplinary Collaboration: This internship highlighted the importance of collaboration across disciplines. Engaging with mentors, peers, and professionals from diverse backgrounds has broadened my perspective and encouraged innovative thinking.

Real-World Application: The internship provided a bridge between theoretical concepts and real-world applications. Developing a crop and weed detection model with the potential to impact sustainable agriculture underscored the practical significance of the knowledge I've gained.

Continuous Learning: The fast-paced nature of technology demands continuous learning. From exploring new model architectures to staying updated with industry trends, I've embraced the idea that learning is an ongoing process.

Communication Skills: Through presentations, discussions, and collaboration, I've enhanced my ability to communicate complex technical ideas clearly and effectively to different audiences.

Professional Growth: Overall, this internship has accelerated my growth as a professional. It has instilled a sense of confidence in my technical abilities and ignited a passion for solving real-world challenges using technology.

My learnings from this internship extend beyond technical skills; they encompass the principles of ethical innovation, effective communication, and a holistic approach to problem-solving. These lessons will undoubtedly guide my future endeavors and contribute to my journey as a lifelong learner in the field of technology and agriculture.

8   Future work scope

While the internship has provided a solid foundation in crop and weed detection, there are several avenues for future work and exploration to further refine the solutions and expand their impact:

Improved Model Performance: Continuously fine-tune the deep learning model to achieve even higher accuracy, particularly in challenging scenarios like varied lighting conditions or overlapping plants.

Multi-Species Detection: Extend the model to identify specific weed species. This can involve collecting and annotating datasets for different types of weeds, making the system more versatile for diverse environments.

Crop Health Monitoring: Integrate health assessment into the model. Detecting diseases or nutrient deficiencies in crops would allow for targeted interventions, reducing losses and optimizing yield.

Seasonal Adaptability: Train the model on data collected over multiple growing seasons. This would improve its ability to adapt to different environmental conditions and crop growth stages.

Edge Device Optimization: Further optimize the model for deployment on edge devices like drones or tractors, ensuring low latency and real-time analysis without relying heavily on cloud resources.

Automated Intervention: Collaborate with agricultural machinery manufacturers to enable the integration of the detection system with automated machinery for precise weed removal or nutrient application.

Data Fusion: Integrate data from various sources, including satellite imagery, weather data, and IoT sensors, to create a comprehensive decision support system for farmers.

Dynamic Thresholds: Implement dynamic threshold adjustments for decision-making. The system could consider factors like weather conditions, local policies, and economic factors to optimize resource allocation.

Global Collaboration: Collaborate with researchers and farmers from around the world to ensure the model's adaptability to different crops, regions, and agricultural practices.

User Education: Develop resources and training materials to educate farmers on utilizing the technology effectively and interpreting the results for optimal decision-making.

Regulatory Compliance: Ensure that the technology aligns with regulatory requirements and guidelines related to pesticide use, data privacy, and environmental impact.

Sustainability Analysis: Conduct a comprehensive analysis of the model's impact on sustainable agriculture, considering factors such as reduced chemical usage, improved yield, and environmental benefits.

Education and Outreach: Host workshops, webinars, and awareness campaigns to promote the adoption of technology-driven solutions among farmers and stakeholders.

Partnerships: Forge partnerships with governmental bodies, non-profit organizations, and industry players to scale the solution and maximize its positive impact.

The future work scope is both promising and expansive, reflecting the evolving nature of technology and agriculture.