SEMANTIC SEGMENTATION OF AERIAL FARMLAND IMAGES USING DEEP LEARNING ARCHITECTURE

By: Hisham Parol

URN: 6614207

A dissertation submitted in partial fulfilment of the requirements for the award of

Master of Science in Data Science

Department of Computer Science

August 2021

Supervised by: Dr Alireza Tamaddoni Nezhad

DECLARATION

I declare that this dissertation is my own work and that the work of others is acknowledged and indicated by explicit references.

Hisham Parol.
August 2021.

ABSTRACT

Agriculture is very crucial for the UK economy. The application of technology in farming has expanded dramatically over the past few years. Innovations in agriculture are defining the direction of the future of the agri-tech sector. This drastic improvement in research in the area of agriculture provokes the objectives and motivation to implement an effective deep learning-based semantic segmentation model for the detection of weed clusters. Here aerial images are captured by the agricultural drones and these pictures are fed to advanced convolutional neural networks to analyse the agricultural patterns. Unlike the common segmentation tasks, agricultural patterns are very complex as well as specific to crop types. Thus little development has been made to merge computer vision and agriculture due to the lack of suitable agricultural image datasets and their complex nature. Furthermore, semantic segmentation of aerial farmland images requires judgment over large images with advanced annotation sparsity. These are not present in the common segmentation tasks. That said, the Agriculture vision challenge competition presents a large-scale aerial image dataset for the semantic segmentation task. This project is primarily focused on weed clusters. As a pilot study of aerial agricultural pattern classification, various deep learning-based segmentation models are studied and implemented. This project selected three popular models to fit agricultural data and results were evaluated. Techniques such as image augmentation using both geometric and colour transformations, transfer learning and hyper-parameter tuning are executed to analyse the improvements in the model results. The final evaluation metric used here is the mean intersection over union and the results are compared with an advanced model developed by the top three participants in the agricultural vision challenge 2020. After comparing the results, we have made a comprehensive analysis to explain why the model's performance is high or low for this particular dataset. This experiment highlights the importance of applications of image segmentation technology and how they make it more useful for agriculture in daily life. It will enable us to take full advantage of this technology in monitoring the real-time farmlands and identifying patterns to quickly respond and take action. Doing this can reduce the usage of herbicides in farmlands thus reducing labour costs and environmental impacts.

Keywords: Deep learning, Semantic Segmentation, Data Augmentation, Transfer Learning, Convolutional Neural Networks, TensorFlow, Keras, DeepLab, U-Net, FCN.

ACKNOWLEDGMENT

I would like to thank my Supervisor, Dr Alireza Tamaddoni Nezhad for the detailed guide during the development of this piece of work. I am grateful to all of those with whom I have had pleasure to work during this project- to my family, friends and colleagues, I thank you so much.

Hisham Parol

1 TABLE OF CONTENTS

1. LIST OF FIGURES

2. ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
IoT	Internet of Things
CNN	Convolutional Neural Networks
DCNN	Deep Convolutional Neural Network
FCN	Fully Convolutional Network
mIoU	Mean intersection over union
ASPP	Atrous Spatial Pyramid Pooling
GPU	Graphics processing unit
VM	Virtual Machine
API	Application programming interface
SVM	Support Vector Machine
ANN	Artificial Neural Network
RGB	Red Green Blue
HSV	Hue Saturation Value
RDSE	Residual Dense with Squeeze and Excitement
GNN	Graph Neural Network
SCG	Self constructing graph

2 CHAPTER ONE- INTRODUCTION

Today, the world is addressing the challenges of population, climate change and food shortage. It is estimated that the world population will rise to 8.5 billion by 2030. With the rapid growth of population, food consumption worldwide also increases. A rapid escalation in food production is necessary to keep up with the world's food demand. Agricultural industry has evolved so far and with the arrival of digital technology and Artificial Intelligence we can bring an unprecedented level of control and better decision making that will generate disruptive improvements in agricultural practises.

The digital revolution is bringing new technology and data-driven solutions to the field of agricultural farming. Managing farms using technologies such as drones, IoT and AI has significantly increased. Semantic segmentation is one of the challenging tasks in computer vision. Much of the success in this field is related to the modern development of cutting edge technology in Deep Learning and the introduction of ImageNet - a large dataset used to identify patterns and classify images. Recent works in image segmentation of medical images and self-driving cars have shown outstanding results. While computer vision technologies have drawn significant attention in smart-farming, the progress in this field is relatively slower due to a lack of appropriate datasets to fuel the research. For example, unlike other datasets, identification of weeds from aerial farm images require high-level expertise in annotating the labels and identifying the weed patterns based on geographical and climatic conditions. These challenges are not present in common datasets.

It is important to point out that this project is inspired by the Agriculture Vision Prize Challenge 2020 - which aims to encourage research in developing novel and effective algorithms for agricultural pattern recognition from aerial images. While the original challenge competition contains 6 types of annotations, this project is focused on comprehensive analysis and comparison of deep learning methods for weed segmentation from large aerial farmland images. Additionally, various pre-processing methods (Data Augmentation), CNN networks and hyper parameter tuning will be investigated and applied in order to evaluate the performance and improve the results.

To perform the implementation of deep neural networks, the dataset was requested from the Vision for Agriculture and they provided a Challenge dataset - a subset of the Agricultural-Vision Dataset

containing Aerial Farmland images. Moreover, various Deep Neural architectures such as U-Net, DeepLab and FCN used for semantic segmentation were reviewed, executed and compared.

Unlike the common segmentation tasks involving identifying the objects with definite shapes and features, this project aims to use a challenging dataset, data augmentation techniques and neural network models to identify irregular patterns and unusual shapes in the aerial farmland images. The objective of this project is to explore various methods and compare the results with top performers in the First Agricultural Vision Challenge. The evaluation metrics used to compare the performance in this project are Mean Intersection over Union (mIoU) used in the original challenge competition.

2.1 AIMS AND MOTIVATION

As mentioned in the above section, this project aims to explore a complex dataset and state of art deep neural networks to improve the performance and get a result comparable with the Top performers.

The motivation to undertake this project comes from my interest in computer vision applications, to do research on innovative smart farming projects and develop something that can help farmers to cultivate crops sustainably as well as reduce greenhouse emissions. Initially, this idea was mentioned in my Statement of Purpose (SOP) to do a Master's in Data Science. And I continued to implement this project with the knowledge gained during my Master's Degree.

The first aim of this project is to examine the recent works in the field of smart farming, especially weed detection and discuss the challenges in the existing technology. This may help us to understand the broader perception of technologies used, the accuracy level of current systems and the difficulties in the implementation part.

The second aim is to suggest various deep neural networks used for segmentation tasks that have shown excellent results in other datasets, tuning those models with our datasets and evaluating their performance. The main goal is to identify weeds from crops, that is, classifying each pixel in the image to either weed or background. For this purpose, various encoder-decoder based architectures are used and explore the advantages and disadvantages of each one for this specific task.

Finally, this project aims to understand the importance of pre-processing technologies and hyper parameters and discuss how each one of them affects the final evaluation results. This aim explores the various data-augmentation methods to enhance the object from the background, the hyper-parameter tuning to reduce the losses and improve the results, which is comparable with top performers in the agricultural vision challenge.

2.2 PROJECT OBJECTIVES

1. Discuss and identify the traditional farming methods used to identify weeds and their challenges.
2. Literature review on related works done in this area.
3. Explore architectures used for semantic segmentation.
4. Design, build and implement deep learning models.
5. Comparison of model performance.
6. Discuss the future scope and improvements.

2.3 PROJECT SCOPE AND CONTEXT

In this section, we will explain the overall scope and context of this project and how this project is related to improving real-world challenges and problems. Traditional farming as a convention has been in the past and will persist in upcoming times to be a manual, physical and labour-intensive industry.

Recent technologies and growing abundance of data can bring efficiency to work management, effective handling with adequate attention to productivity and quality, and resource-efficient approach. This is an era where technology is employed even in the farming/ agriculture sector. “Smart Farming” is an emerging evolving concept that focuses on managing farms with the help of technologies like Artificial Intelligence (AI), Machine Learning, Big Data, Internet of Things (IoT), Drones, robotics, etc. to improve productivity, reduce waste, enhance quality and quantity of products, while efficiently managing the cost. Smart Farming refers to the exertion of modern technologies and information into agriculture and is also referred to as the 4.0 Green Revolution. Smart Farming systems help to realize the determined efficiencies in quantity and quality while minimizing cost.

According to a survey conducted, around 80% of farmers in the USA and possibly up to 24% in Europe use Smart Farming Tool (SFT). These percentages are the established proven fact that Smart Farming Tools are contributing to the enhancement and development of the farming sector. As mentioned, AI technologies assist in enhancing the quality of crops. AI systems are being used in precision farming for identifying diseases in plants, weeds and pests. Farmers are also taking to the sky to monitor and observe the farm. From farm drones, AI-enabled cameras can capture images of the entire farm and observe real-time to detect the problem areas and formulate potential improvements. With the help of images gathered from cameras mounted on drones or satellite images, they use computers to analyse those images to identify which crops are surrounded by weeds. This is a time-saving technique, where farmers need to concentrate on only those plants with weeds, spray pesticides on these specific crops and thus reduce the usage of pesticides. With the availability of aerial farm images, notable results and methods were implemented in agricultural vision competitions.

The main focus of this project is implementing AI solutions and comparing the results of various approaches. Inspirations for semantic segmentation can be drawn from methods aimed to develop object detection. Recent works in semantic segmentation have shown excellent results. For example, the U-Net model based on encoder-decoder architecture is widely used in medical imaging technology, identification of cancer tumours etc. The DeepLab series uses Atrous Spatial Pyramid Pooling (ASPP) to encode multi-scale contextual information. Whereas the MobileNet architecture uses depth-wise separable convolutions to build lightweight models. Pre-Trained models have been used to quickly train complex models and extract general features without training the models from the beginning. These technologies can potentially be transferred to identify weeds in agricultural farms to incorporate smart farming to improve yield and reduce environmental impacts.

This project uses the vision dataset to train a model using state-of-art deep learning networks. The model then predicts each pixel of test data into either crop or weeds. Post-processing of predicted images draws a bounding line across the weed clusters, thus making it easier to analyse the images. This project also discusses the future improvements and scope.

2.4 RESOURCES

In order to complete this project, the following resources were used:

Integrated Development Environment (IDE): Google Colab Pro - A specialized version of the Jupyter Notebook provided by Google. This runs on the Cloud Platform and provides priority access to faster GPUs such as T4 and P100 GPUs. Also, Colab Pro provides access to high memory VMs. This helped to train the neural network without exhausting the resources.

Data Storage: Google Drive - Initially used Amazon S3, However, due to cost constraints, I moved to Google Drive. Drive provides 30 GB of free memory without a subscription. Also, Drive can be easily mounted to Colab making it easier to access data quickly.

Framework: This project is built using the TensorFlow framework - an open-source library developed by Google specifically used for training and inference of deep learning networks.

1. **Keras**: Keras is used as an interface for TensorFlow. This API developed by Google is used for implementing neural networks
2. **TensorBoard**: This is used for visualization and tracking the evaluation metrics such as loss functions and train and validation datasets.

Data: The data needed for this project was provided by Agricultural-Vision- an independent research board for the research and development of computer vision technology for agriculture. The dataset is

large scale and high quality images of aerial farmlands for the advanced study of agricultural patterns. The dataset was precisely annotated by professional agronomists with a strict quality assurance process. The proposed agricultural vision contains 94,986 samples of images and the subset of this dataset containing 12,901 samples was used for challenge competition. Same was used for this project as well. The dataset contains nine types of annotations - double plant, dry down, endow, nutrient deficiency, planter skip, storm damage, water, waterway and weed cluster. In this project only weed clusters are used for the segmentation task. These images were captured by special mounted camera on Drones flown over various corn and soybean fields around Illinois and Iowa (USA).

Data Privacy: According to the terms and conditions for the usage of data by IntelinAir, the User shall not use this dataset for any purpose other than internal and non-commercial academic purposes. Sell, license, transfer or redistribute the Dataset to any third party (except to research employees) and use the Dataset to develop any commercial product or service is restricted. (Screenshot attached in Appendix)

2.5 REPORT STRUCTURE

In order to make it more comfortable and simpler for the user to read through, the report is organized and structured. The diagram below illustrates the structured approach of this project.

To meet the project objectives, the report contains a total of five Chapters. Each Chapter is outlined below with a brief description:

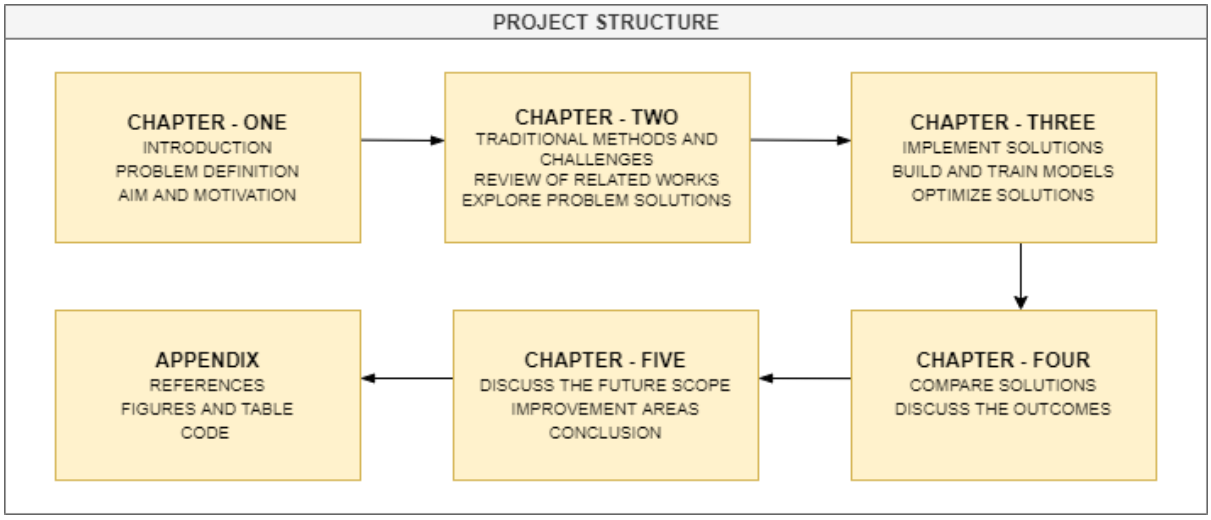


FIGURE 1. *Project Structure*

Chapter One- Introduction - A brief description about problem definition, resources and report structure.

Chapter Two - Analysis of traditional weed control methods, their drawbacks and literature review on common segmentation models and related works in agricultural vision challenge competition.

Chapter Three - Design, build and implement models. Further performance improvement techniques.

Chapter Four - Comparison of model performance, results and discussion.

Chapter Five - Conclusion and Future scope of this project.

3 CHAPTER TWO- LITERATURE REVIEW

3.1 INTRODUCTION

In the United Kingdom, the focus on farming and food security has become an intense topic post BREXIT. The issue of climate change and its environmental impacts has been a major subject nowadays. Agriculture is one of the major sources of greenhouse gas and contributes to the Greenhouse Effect and Climate Change. According to the International Panel of Climate Change (IPCC, 2013) agriculture accounts for almost 25% of Greenhouse Gas (GHG).With the UK government's commitment to reduce the greenhouse gas emissions to net-zero by 2050, British agriculture has a role in adapting to sustainable farming and tackling climate change. One of the fundamental areas of activity in sustainable farming is the use of technology and data. Now the agricultural industries are transforming to implement artificial intelligence technologies, use of data to control weeds, detect diseases, monitor real-time activities on farmland and so on.

Farmers are using AI technologies for precision farming where robots are targeted to reduce the number of pesticides used in farms by precisely identifying the position of weeds and distinguishing crops from the weeds. The precision agriculture systems will mitigate leaching problems as well as the emission of greenhouse gases.

In this chapter, some background insight will be provided on Conventional and Machine Learning weed identification methods and their challenges, Evolution of Deep Learning methods in Agricultural sectors. Related semantic segmentation works, models and their performance. Finally, a literature review on the top two models implemented for the agricultural vision challenge competition 2020 were discussed. The purpose of this chapter is to establish familiarity with the current research and

implementations in this field and technology used in this project in a much more detailed way in order to understand the scope of this project.

3.2 SMART FARMING

To understand the application of this project in the real-world, it is necessary to become familiar with smart farming, its key objectives and progress. Smart Farming has the potential to distribute more valuable and long-lasting, viable agriculture production by employing emerging and updated technologies. The objective of Smart farming is to improve economic rewards and reduce environmental impact. From the farmer's viewpoint, Smart Farming can contribute and aid the farmers with an added advantage in productive and efficient utilization of resources, betterment of decision making, and enhanced management with the minimal capitalization of costs. Smart Farming also encourages additional benefits in the form of environmental concerns such as ensuring sufficient utilization of pesticides and herbicides by accurately classifying the infected crops. Preferably, instead of walking in the field all the time, with the assistance of smart farming tools, the farmers can investigate the field activity from any smart tablet or mobile phone. In the cyber-physical system, cloud platforms are used by farmers to get notified of the latest information and also help to continuously learn the neural network models. The most recent generation of hyperspectral images supplied by Sentinel satellites, GPS technology, and local UAV drones can be used to fetch data and monitor real-time farm activities with precise locations.

One of the most attractive elements of Smart Farming Tools (SFT) is that it can be befitting irrespective of the size and scale of farming - can be pertinent in large scale as well as small scale farming. It can benefit every common and conventional farming like family farming, organic farming, etc., can be modified and customised according to the necessities, demand, transparency required by the customers and market in large. This review aims to get insight into advanced deep learning applications related to smart farming especially focused on automatic weed detection.

Figure 2 summarizes the concept of smart farming for weed detection in the context of a cyber-physical system. The smart devices connected to the internet provide a dashboard with well-informed data about farm land activity from drones and sensors. The robot can play an important role in controlling weeds and spraying herbicides on farms. Analysis and decision making is increasingly assisted by machines, thus making the whole system almost autonomous.

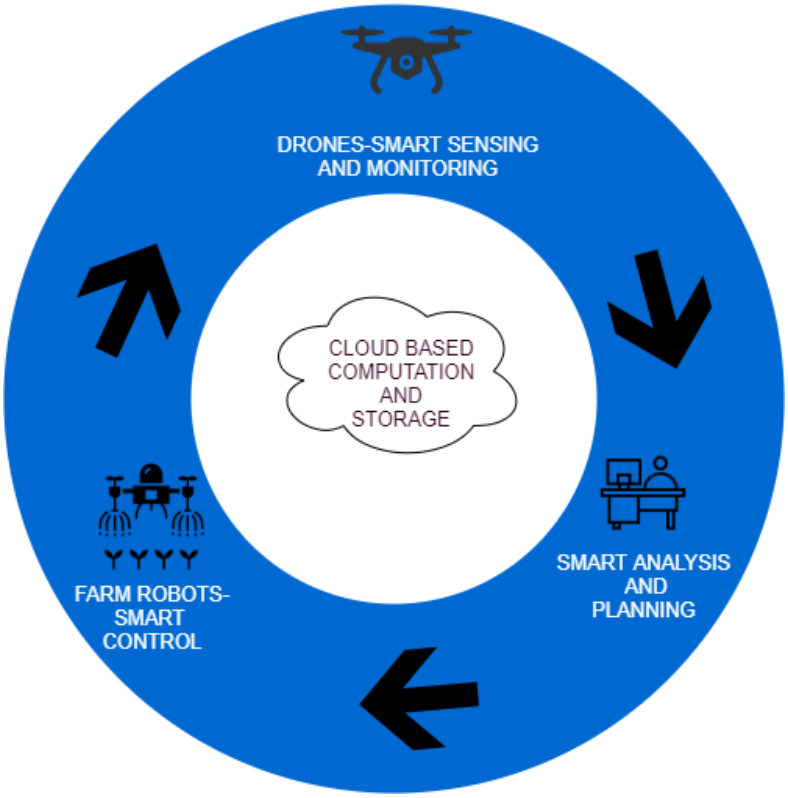


Figure 2: *Smart Farming (Cyber-Physical system)*

3.3 TRADITIONAL AGRICULTURAL WEED CONTROL PRACTICES

Traditional manual control of weeds in large farmlands is practically time-consuming as well as expensive labour tasks. According to the study of Gianessi and Reigner (2007) [6], the usage of herbicides in the state of Mississippi had a $10M saving as compared to manual control of weeds. The herbicides were successful in controlling weeds. However, they also have harmful impacts on the environment and health risks, for plants and animals. To tackle this problem, a number of researchers are nowadays following several methods where the main goal is to minimize the harmful effect of herbicide application. One solution is site-specific spraying, where the weed clusters are detected, herbicide spray is minimized and only applied where necessary. The methods to identify these sites are a major area of research in agriculture and are investigated by different researchers. In order to make this method operational, machine vision was often a choice before the introduction of Deep Learning methods.

3.3.1 CONVENTIONAL CONTROL METHODS

Weeds are highly harmful to the growth and production of crops. Conventional weed control system sprays the entire field with a uniform level of herbicides. This has a severe consequence on agriculture as well as adding cost. Large quantities of herbicides not only affect the growth of crops but also have an impact on the environment.

The traditional way of distinguishing weeds is with the help of an expert field guide, manuals, and taxonomic keys to the agricultural weeds in that particular area. Experts collect specimens from the field and examine them closely including stem, flower, roots etc.

Some of the characteristics that help identify some weeds include:

- The appearance of spines, prickles, thorns or stinging hairs
- Milky fluid or sap when stem or leaves are sliced
- Presence of a leaf cover enclosing the stem at each node
- Stems square in cross-section

Hand rouging or pulling is a widely used technique for directly pulling the weeds and removing. Tools such as docks, hoe, thistles and ragwort are used to manually remove the weeds after identification. In this methodology, the speed of harvesting is reduced and the cost of elimination of weeds is more expensive in farmland, larger numbers of labourers are used to remove weeds, which degrades the production flow.

3.3.2 MACHINE LEARNING APPROACH

In the early stage, Weed recognition tasks were implemented using Image Processing Technology and Machine Learning, achieving the purpose of weed detection. Traditional Machine Learning systems need a small sample size and short training time, thus have a low demand for GPU. To deliver precise spraying, a key concern that should be answered is how to realize real-time precise detection and identification of crops and weeds. This intelligent technology mainly consists of two parts:

1. Image Processing stage - A series of Image Processing techniques to extract the shallow features of weeds and crops.
2. Classifier - A Machine Learning based Classifier to train and classify images (eg: SVM, Random Forest algorithms).

Most of the Machine Learning Vision based technology used for weed detection utilizes the feature differences between plant leaves and weeds to distinguish them. The crops and weeds are classified by analysing the texture, shape, colour or spectral features of the image. Since these methods have low computation costs, they can be incorporated in agricultural machinery at a low cost, providing an efficient method. SVMs and Artificial Neural Networks (ANNs) have been widely used classifiers in crop and weed classification. The SVM can convert the data to the high dimension space through nonlinear transformation and is able to handle high-dimensional data. SVM is perfect for working with

data having more features. It has good performance in dealing with small-sample data and nonlocal minimum problems. ANNs can classify untrained data and have a powerful learning ability. Other algorithms such as K-nearest neighbour (KNN), random forest, naive Bayesian algorithm, Bayesian classifier, and AdaBoost are also used in traditional methods.

Several works were submitted by researchers in Machine Vision technology for weed classification for a specific type of crop or region. One outstanding example is the research done by Le et al., which determined the difference between corn and single species of weeds on the basis of Local Binary Pattern (LBP) texture features and Support Vector Machine (SVM). Another study by Chen et al. and the team proposed a multi-feature weed reverse location method. This was used in the soybean field and the features were extracted on the basis of shape and colour. By using the shape and texture features, Zhu et al. proposed a classification method to identify five kinds of weeds in farmland. Some scholars proposed a technique for weed extraction based on R-B colour difference features. Using comparative analysis of the grey distribution of each component in the colour space of RGB, HSV and HIS. Zhang et al. proposed a weed classifier in a field at the pea seedling stage. Additionally, plant height and location information was used to improve identification accuracy.

However, traditional ML techniques are simple to understand and many advancements have been made, most of them are verified in low-density images. Changing the lighting conditions, clustering and occlusion in a natural environment impact the accuracy of weed detection and localization. Thus they need to design features manually for specific situations and have a high dependency on image acquisition, pre-processing methods, and the quality of feature extraction.

Year	**Data**	**Algorithm**	**Accuracy**	**References**
2016	Grape Leaves	Combining HOG features with SVM	83.50%	[9]
2018	Sugar Beets and weeds	Using three shape features with SVM and ANN	93.33%	[10]
2017	Different plant leaves	Shape and texture features extracted	92.51%	[11]

TABLE 1: *Traditional Machine Learning methods and their accuracy*

3.3.3 CHALLENGES IN TRADITIONAL METHODS

As mentioned in the previous section, following the traditional way of uniformly spraying herbicides in farmland can result in health and environmental risks and disadvantages that may affect the quality of crops as well.

Here are a few key insights of various challenges and risks of traditional methods and disadvantages of machine vision technology for large farmlands.

3.3.3.1 Risks of Traditional methods:

- Repeated use of specific herbicides may develop resistance to the chemicals and the weeds will no longer respond to the properties of chemical herbicides.
- Overuse of herbicides persists in the soil for a long time, which causes permanent impairment on future vegetation growth.
- Most of these herbicides pose significant health risks to both humans and animals when they come in contact with the skin, inhaled or ingested.
- Prolonged use of herbicides and fertilizers can cause the water bodies near farmlands to become saturated with dissolved nutrients such as nitrates and phosphates.
- Dissolved nutrients in water bodies affect aquatic life in rivers and lakes, where they are overstimulated.

Having said that, controlled use of weed on farmlands is preferred. If not managed properly, this can result in increasing the crop production cost and also destroys the natural environment. Thus identification of the exact location of weeds using Digital technology is necessary to solve the risks of traditional methods.

3.3.3.2 Disadvantages of Machine Vision Technology.

The Implementation of weed detection technology using machine vision has achieved its purpose in specific crops and weeds. They are low cost and easy to implement and have shown notable accuracy in crop and weed classification tasks. However, they are inadequate for large-scale active detection and classification of weeds in a natural environment

- Machine learning methods depend on a large amount of texture information in crops and weeds. Thus, making it less reliable in complex natural scenarios such as obscured weeds and crops, high weed density and overlapping.
- Shape Feature: This mainly includes shape parameters, region-based descriptors, and contour-based descriptors. But the shapes can be distorted by many factors such as disease, human damage,

insects etc. In lab testing, most of the research is conducted in ideal scenarios and these special features are not considered. Thus making it less accurate in the field environment.

- Spectral Features: This learns the colour features of leaves. When the spectral reflectance of weeds and crops are remarkably distinguished, they can be classified. However in the real environment, the reflectivity of plants changes due to the amount of light they absorb. Although research has achieved encouraging results in cases of distinctive spectral bands, the accuracy is low in the situations where spectral difference is not obvious.

In summary, scholars have focused on reducing the usage of herbicides using digital technology and improving classifiers based on machine vision. The features of crops and weeds are of great significance to improve accuracy. These techniques can be used in small-scale fields and do not require high hardware. However, in large farmlands, we need to look for deep learning algorithms that can extract multiscale and multidimensional spatial semantic feature information of weeds through Convolutional Neural Networks (CNNs).

3.4 DEEP LEARNING APPROACH

At present, many smart farming tools such as plant disease detection, species identification, weed detection and crop yield prediction use Deep Learning technology. With the advancements in cloud computing, the computation cost has reduced significantly, and the computing power of GPU has remarkably improved. This has helped to achieve good results in weed detection and classification methods based on deep CNNs.

3.4.1 FROM IMAGE CLASSIFICATION TO SEGMENTATION

Deep learning has been very prosperous when working with image data to solve computer vision problems such as Image Classification, Object Detection and Semantic Segmentation. Semantic segmentation in the Agricultural industry is inspired by methods aimed at detecting common object segmentation. In the image classification problem, we are interested in identifying the class labels of all the objects present in that image. The outcome is classified labels. Then we moved a step further to know the position of those objects in that image with a bounding box. This is called Object Detection. And finally comes the Image Segmentation where we accurately find the exact boundary of objects in an image. Image segmentation is a process of classifying each pixel in an image to a certain class. Recently, two fine-grained new segmentation models - instance segmentation and panoptic segmentation, have appeared as the new research objectives. Instance segmentation aims to detect and delineate each distinct object of interest appearing in an image. For example, as shown in Fig. 3, each plant is labelled with different colours, and each label denotes an instance of plant species. Taking the

task a step further, panoptic segmentation has the highest goal. It combines two distinct concepts used to segment images namely, - semantic segmentation and instance segmentation**.** It assigns a semantic label and an instance label to each pixel. Compared with traditional requirements of Image Classification, these two tasks are more challenging, as they focus not only on pixel-level classification but also on separating pixels and assigning labels to them.

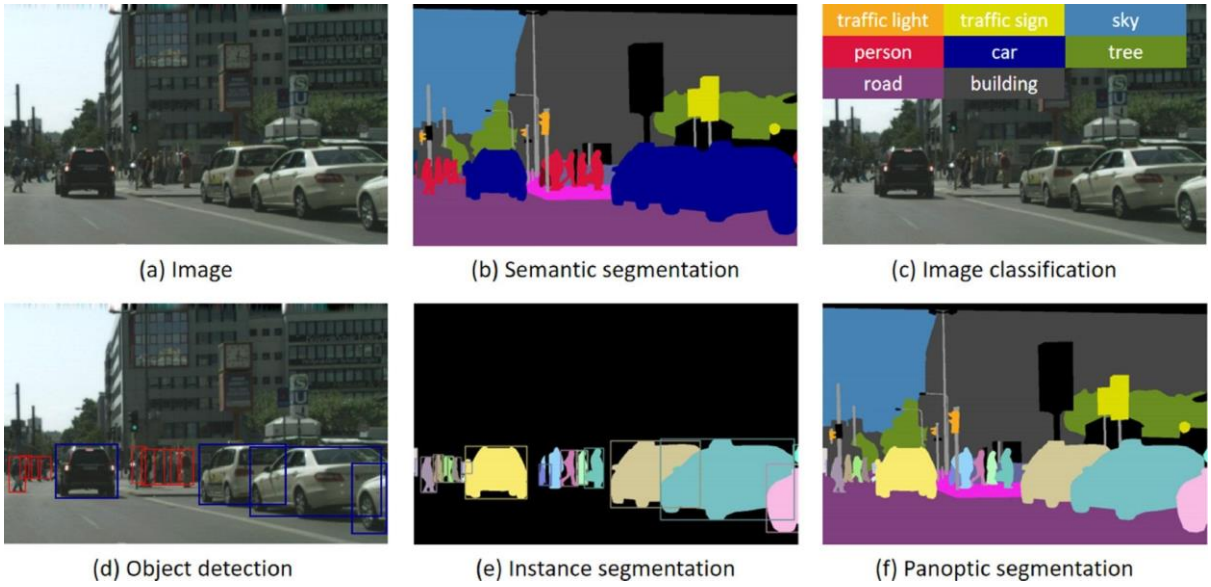


Figure 3: *Examples of Computer Vision tasks*

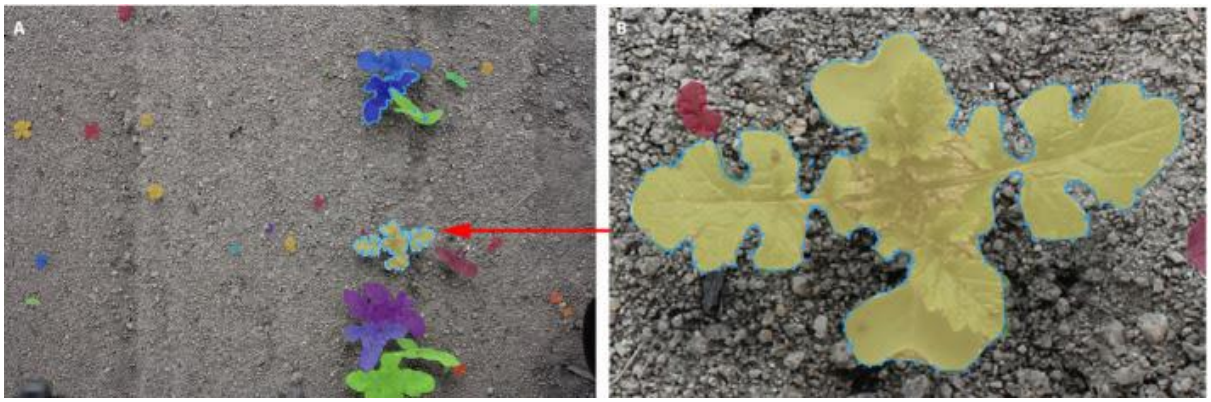


Figure 4: *Output from the instance segmentation convolutional neural network.*

3.4.2 CONCEPT OF ENCODER-DECODER NETWORK

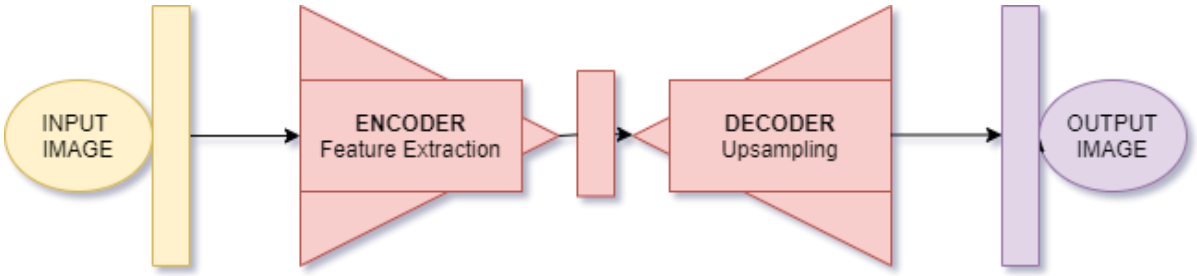


Figure 5: *Basic blocks of encoder-decoder architecture*

In this section, the most important concept used in semantic segmentation is explained in detail. In recent years, the convolutional neural network (CNN) has made remarkable achievements in semantic segmentation. Nowadays most semantic segmentation networks are based on the concept of encoder-decoder architecture, where it has an encoder side to extract the feature vectors and a decoder side for recovering feature map resolution.

In regular Image Classification Deep Convolutional Neural Network (DCNN) models, it takes images as input and outputs a single value representing the label of that image. It has four main operations, namely - Convolutions, activation function, pooling, and fully connected layers. When we pass the image through these four layers it presents a feature vector with probabilities of each class. In this network, we assign a single label to an entire image. In classification problems, we don't care much about spatial location. Only the presence of a class label is determined. But in segmentation, it is very important to preserve spatial information. Here we want to categorize each pixel in that image. Understanding images at the pixel level is important here. Hence regular DCNN Models are not suitable. These models reduce spatial characteristics which are critical in semantic segmentation. Thus instead of having pooling and fully connected layers, we can set up a convolution layer having a stride of 1 and the same padding. This preserves the input dimension and spatial information. However, this approach adds another disadvantage to the performance and cost - High memory and computation requirements.

To ease that problem, an encoder-Decoder architecture is introduced for semantic segmentation tasks. This network usually has 3 main components - Convolutions, down sampling and up sampling. On the encoder side, the network performs down sampling to perform deeper convolutions without requiring more memory. This part looks like a regular DCNN without fully connected layers. We can also use

pre-trained models to extract features on the encoder side. Down sampling in neural networks can be done by using convolutional striding or pooling. The output of the first stage is a compressed feature vector with smaller spatial dimensions. Then we feed this compressed feature vector to the up sampling stage to reconstruct our original size. The goal is to increase the spatial dimensions so that the output is the same size as the original image. Here we use transpose convolutions to convert deep and narrow vectors to wider and shallow ones. Some of the popular networks implemented based on encoder-decoder architectures are FCNs, U-Net, SegNet etc. Experiments prove that the encoder-decoder architecture has achieved a good performance in many segmentation datasets.

3.4.3 CHALLENGES IN DEEP LEARNING SEGMENTATION TASKS

Today, the performance of semantic segmentation has been greatly improved by using deep learning techniques. Availability of high computing power and storage memory has helped to direct research in image processing using deep neural networks. Having said that, Semantic segmentation is a challenging task in computer vision. Compared to traditional Machine Learning, the Deep-learning based methods have shown extraordinary improvement in outcomes. A large number of novel methods have been proposed.

In the deep layers, features are more semantic aware. On the other hand, in shallow layers, the extracted features are more aware of details such as strong edges. The appropriate coordination of these two features can significantly boost the semantic segmentation performance.

Recently, the emergence of deep learning technology has greatly promoted semantic segmentation research. For example, Long et al. introduced the pioneering Fully Convolutional Network (FCN), which dramatically improved the segmentation accuracy. With the help of Transfer Learning, models trained on large image datasets can be reused and act as a backbone for new implementations. The availability of annotated PASCAL VOC and Cityscapes Datasets has been widely used as a benchmark for segmentation tasks. Using the pre-trained models, features extracted from huge image datasets can be used again, encouraging more research in this field.

3.4.3.1 The key challenges of semantic segmentation are:

- The huge computation costs hinder the applications in some real-time situations.
- The ways to enhance the model efficiency while keeping the segmentation accuracy level.
- The balance between Accuracy and Efficiency: The balance between accuracy and efficiency are both important from the evaluation point of view. The gain is still a contradictory subject in many semantic segmentation tasks. For example, Models such as PSPNet, DeepLab etc. have high accuracy but tend to have low efficiency. On the other hand, SegNet and U-Net have low accuracy but high efficiency.

- Lack of High-Quality Data: For training, we need high-quality data to get an accurate result. However, for tasks such as agricultural images, segmentation and annotating labels require quality experts and have to do pixel-level annotation. This is inevitably a laborious and time-consuming task.
- Weeds specific to region and crops: As mentioned in section 4.3.1, traditional weed detection requires a manual for weeds specific to that region and crops. Weeds have different characteristic properties for different crops and regions. Thus, the model should be trained for various types of weed and crop combinations. So we can only build an application-specific model and cannot build a general model for weed detection.

3.5 RELATED WORKS IN SEMANTIC SEGMENTATION METHODS

3.5.1 U-NET ARCHITECTURE

One of the famous Encoder-Decoder based neural network architectures for semantic segmentation called U-Net was published in 2015 Medical Image Computing and Computer-Assisted Intervention MICCAI by Ronneberger et al for Biomedical Image segmentation [18]. U-Net has won the Grand Challenge for computer-automated Detection of Caries in Bitewing Radiography at ISBI 2015, and the Cell Tracking Challenge at ISBI 2015 on the two most challenging transmitted light microscopy categories (Phase contrast and DIC microscopy) by an extensive margin.

Usually, the convolutional neural networks were used for classification tasks; there was a need for segmentation tasks. But there was a limitation due to the size of the available training sets and the size of the network for segmentation. Here the output to an image is an image itself with each pixel classified into different classes. In many biomedical tasks, the desired output should include localization. U-Net is a modified version of the Convolutional Neural Network aimed to work with very few training images and more precise segmentation. The main idea is to construct a contracting network by successive layers, and the pooling layers are replaced by up sampling operations. Thus, these layers increase the resolution of the output. In order to maintain localization, high-resolution features are combined from the respective contracting paths. The architecture of the U-Net model is explained below

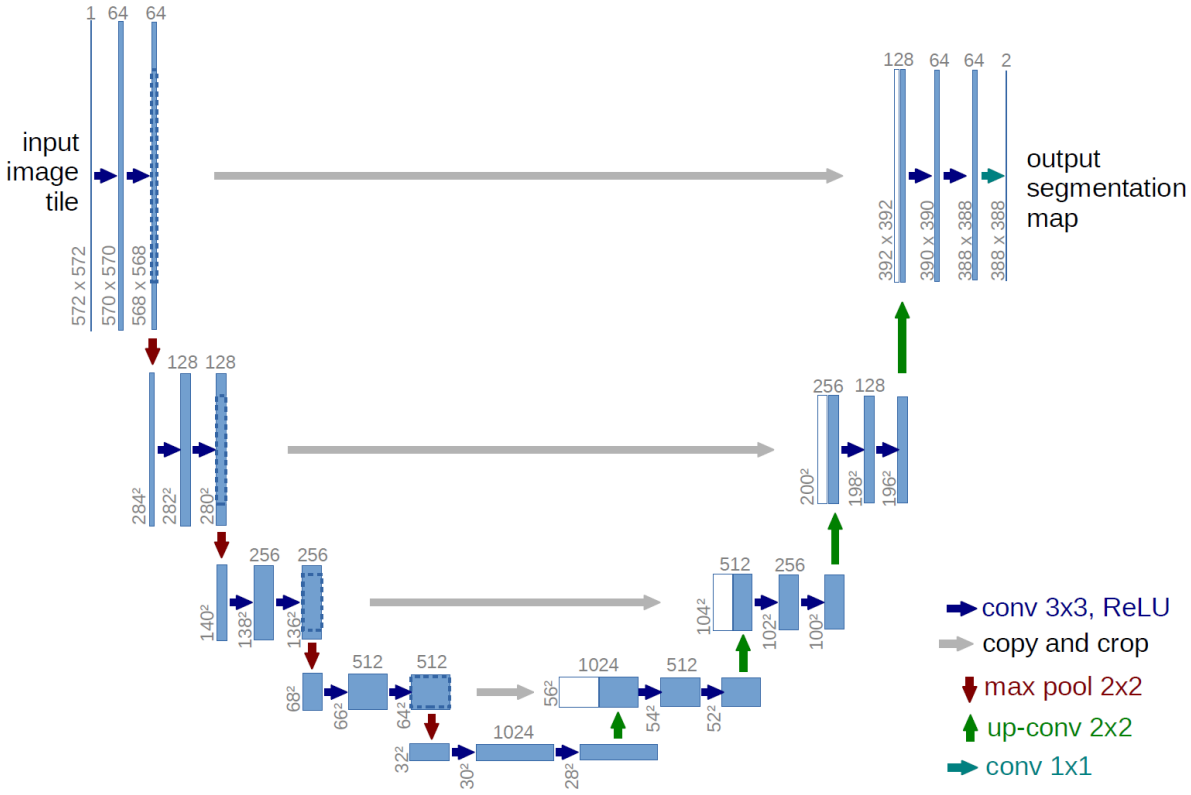


Figure 6: *U-Net architecture*

3.5.1.1 MODEL ARCHITECTURE

As the name suggests. U-Net has a U -Shaped architecture, where the first part is a contraction path and is followed by an expansion path. U-Net utilises dense skip connections to fully leverage the features from each layer. The contraction path is also called the encoder side and the expansion path is called the decoder. The characteristics from each layer in the contraction path are connected to the symmetrical layers in the expansion path. The network architecture is shown in the figure below.

This network does not have any fully connected layers, instead it only uses the valid part of each convolution.

Contraction path

The encoder is a stack of convolutional and max-pooling layers.

- There are two inputs to the convolutional layers - A 3D volume image of size height*width*channels and a set of k filters having size f*f*channels, where f is typically 3 or 5.
- The output image of the convolutional layer is also a 3D volume image of size $n_{OUT} * n_{OUT} *$channel.

Where, $n_{OUT} = \left[\frac{n+2p-k}{s}\right] + 1$

n = Number of input features

n_{OUT} = Number of output features

k = Convolutional kernel size

p =Convolutional padding size

s = Convolutional stride size

- The second layer is the max-pooling layer. This is used to reduce the size of the feature map so that we have fewer parameters.
- The max-pooling selects the maximum pixel value from the region.
- The size of the filter and strides are two essential hyper-parameters in the max-pooling operation.
- Consecutive two times of 3×3 Conv and 2×2 max pooling is made in each contraction stage. This will not only reduce the size of feature maps but also extract more advanced features.

Expansion Path

- One important distinguishing feature of U-Net from normal CNNs is the presence of up sampling or expansion path.
- This stage allows the network to propagate context information to high-resolution layers.
- Transposed convolution or deconvolution is a technique to perform up sampling
- The input of this block is a low-resolution image and outputs a high-resolution image.
- Consecutive 2×2 Up-Conv and two times of 3×3 Conv are done to recover the size of the segmentation map.
- However, in this process, we can get advanced features, but we also lose the localization information.
- Thus to avoid this, after each up-convolution block, we have a concatenation of feature maps that are at the identical level. This provides the localization information from the contraction path to the expansion path.
- Finally, we get an image having the same size as the input image with each pixel being classified based on its class properties.

3.5.1.2 REAL WORLD APPLICATIONS OF THE U-NET MODEL

U-net was developed originally for medical image analysis that can accurately segment images using a limited amount of training data. It is concluded that the U-Net model is certainly a ground-breaking and important deep learning method. These models have been extensively used in medical image analysis and have shown excellent results. There are many variants of the U-Net model as well. This section familiarises with the successful real-world application of u-net models.

- The novel coronavirus (COVID-19) pandemic has created a tremendous global medical crisis. To combat the increasing cases of infection and deaths, the medical imaging community has initiated research in various deep learning methods, including U-Net for the diagnosis of COVID-19. The images of chest CT scans were used for detecting coronavirus.
- Endoscopy is an invasive imaging procedure where the imaging equipment is inserted into an organ or cavity to take photos. The U-net model has been applied to endoscopy images for segmentation of polyps in the gastrointestinal tract, colon objects, and detection of laryngeal leucoplakia.
- U-net has been used on OCT for segmentation of retinal layers, blood vessels, fluid regions, and drusen.
- The U-net model has been used for many medical image analyses such as the segmentation of blood vessels in digital subtraction angiography (DSA), white matter tract segmentation in diffusion tensor imaging, iris segmentation in iris imaging, and tumour detection in mammograms.
- Microsoft partnered with Land O'Lakes SUSTAIN, which collaborates with farmers to help them develop sustainability results using the most advanced deep learning technologies. U-net was used for map labelling tasks such as labelling waterways, terraces, water and sediment control basins and field borders.

3.5.2 DEEPLAB

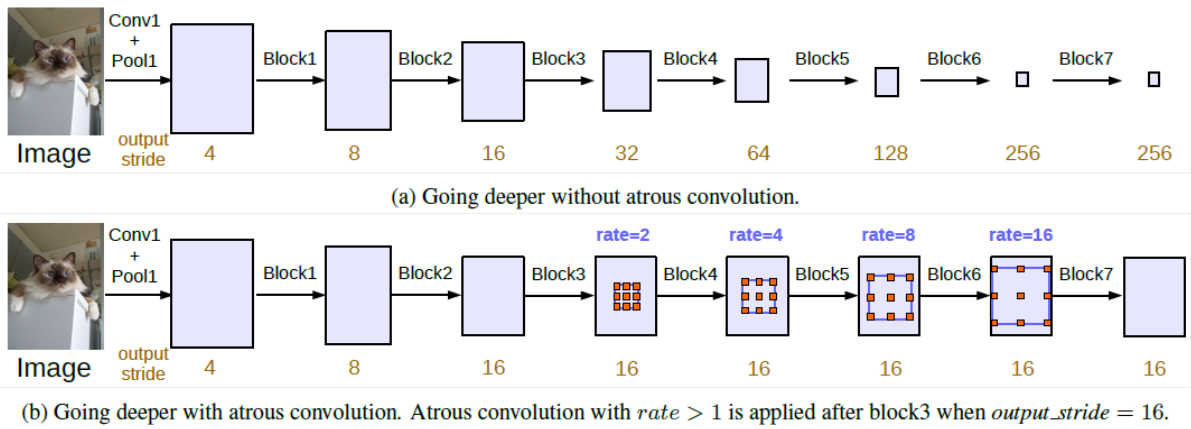
Invented by Google and the most advanced model of the DeepLab series, DeepLabv3+ extends DeepLabv3 by combining a simple yet efficient decoder module to improve the segmentation results, especially along object boundaries. The previous version - DeepLabv3 employs several parallel atrous convolutions with different rates to capture the contextual information at multiple scales. The encoder-decoder architecture used in U-net has been proved to be useful in recovering spatial information. Both of these concepts are combined In DeepLabv3+ making it a superior model in the DeepLab series. The DeepLab v3+ offers an architecture for controlling signal decimation and learning multi-scale contextual features.

The main three components of DeeplabV3+ are Atrous convolution and Atrous Spatial Pyramid Pooling (ASPP), Encoder and Decoder

3.5.2.1 MODEL ARCHITECTURE

Atrous Convolution:

In normal Deep Convolutional Neural Networks (DCNNs), the input feature map becomes smaller from traversing through the network, and the specific information is decimated, making segmentation challenging. To solve this problem, the dilation rate is introduced. For example, the figure 7 illustrates the DCNNs with r = 1 (normal) and r > 1 network.



(a) Going deeper without atrous convolution.

(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Figure 7: *(a) Network without dilation rate (b) Network with dilation rate*

When the dilation rate is equal to 1, it behaves like a regular convolution. But, if we set the dilation factor >1, it has the effect of enlarging the convolution kernel.

Mathematically, atrous convolution is defined by:

$$y[i] = \sum_k x[i + r.k]w[k]$$

Where *i* is the location on the output *y* and filter *w*. *r* is the atrous rate or dilation rate.

By changing *r*, we can adaptively alter the filter's field-of-view. Thus, this method proposes an efficient mechanism to control the field of view and decides the most suitable trade-off between accurate localization and context assimilation.

Atrous Spatial Pyramid Pooling (ASPP)

ASPP is introduced from deeplabv2. This block adds a series of atrous convolutions with different dilution rates, due to the presence of objects at multiple scales. There are several methods proposed to

handle this problem such as Image Pyramid, Encoder-Decoder, Spatial Pyramid Pooling etc. This model uses Atrous Spatial Pyramid Pooling where parallel atrous convolution layers with different rates capture multi-scale information.

As the sampling rate becomes larger, that is when the dilution rate value becomes close to the feature map size, then the filter instead of capturing the whole image context, only degenerates a simple filter having the centre filter weight, For example, applying 3*3* filter to a 65*65 feature map with different dilution rates, at some point the 3*3 filter will degenerate to a simple 1*1 filter. To surmount this problem, a Global Average Pooling (GAP) is incorporated into the last feature map of the model and fed the resulting image-level feature to a 1*1 convolution with 256 filters with batch normalization.

ASPP model contains four parallel operations - one *1x1* convolution and three *3x3* convolutions with dilation rates = (6, 12, 18)

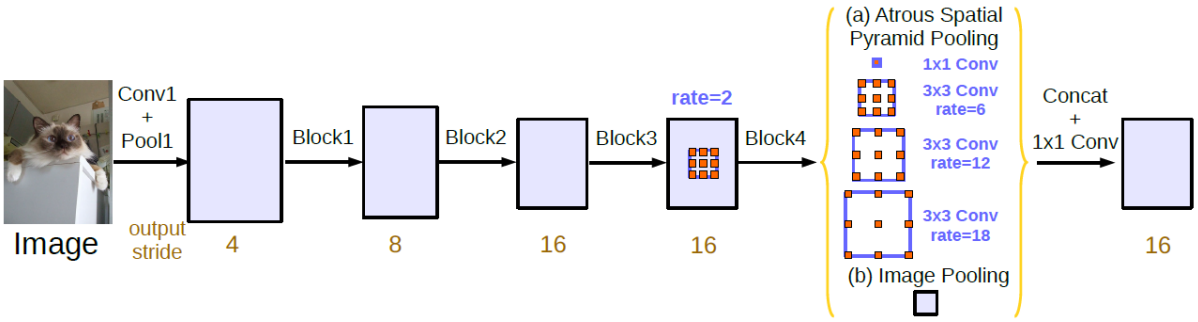


Figure 8: *DeepLab model with Atrous Spatial Pyramid Pooling*

Encoder:

Encoder extracts the feature information of an image. Several models such as DeepLab, Xception and VGGNet can be used as an encoder to extract the features. DeepLab augments the Atrous Spatial Pyramid Pooling module, which examines convolutional features at various scales by applying atrous convolution with different rates, with the image-level features.

Decoder:

The encoder features are first bilinearly up sampled by a factor of 4 and then concatenated with the corresponding low-level features having the same spatial dimensions. To reduce the number of channels, a 1*1 convolution on the low-level features is applied before the concatenation. After the concatenation, A 3 × 3 convolution is applied to refine the features. Then another simple bilinear up sampling is applied by a factor of 4.

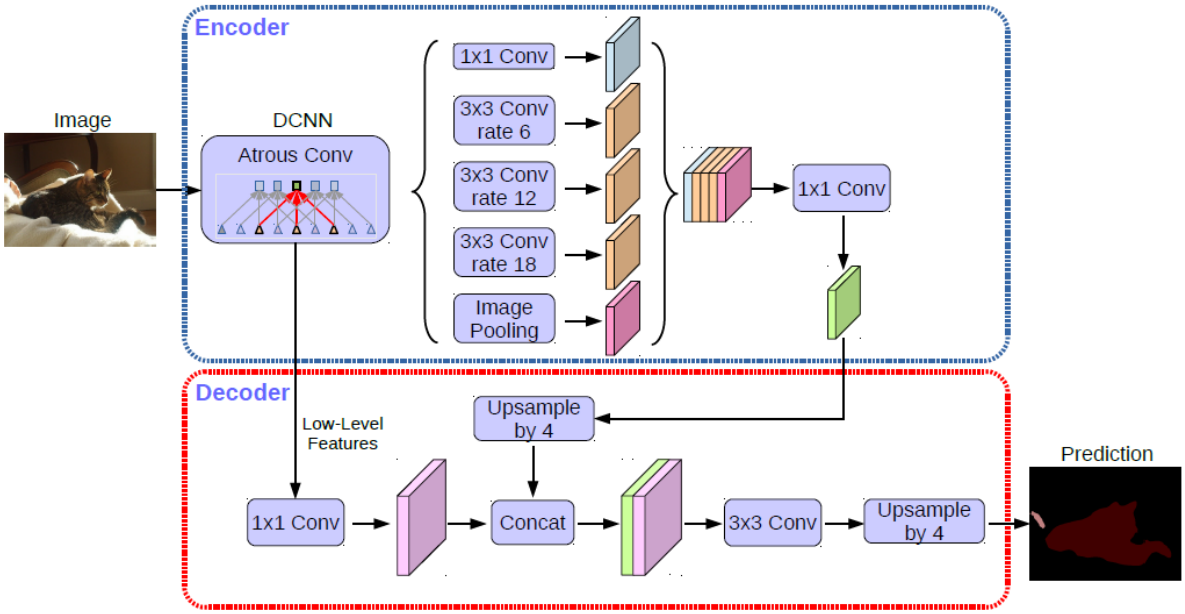


Figure 9: *DeepLabV3+ architecture*

3.5.3 RELATED WORKS FROM AGRICULTURAL VISION CHALLENGE COMPETITION

In this section, the methods and implementation of the first agricultural vision challenge 2020 is explained in detail. The goal of the agricultural vision challenge is to promote and encourage research in the development of aerial agricultural pattern recognition tasks with a challenge dataset. Around 57 teams from various countries compete to implement a deep learning method to identify agricultural patterns. This section presents a review of notable submissions by top two teams from the leader board of the Agricultural vision challenge and discusses their motivation and methodologies[1].

3.5.3.1 MODEL ARCHITECTURE

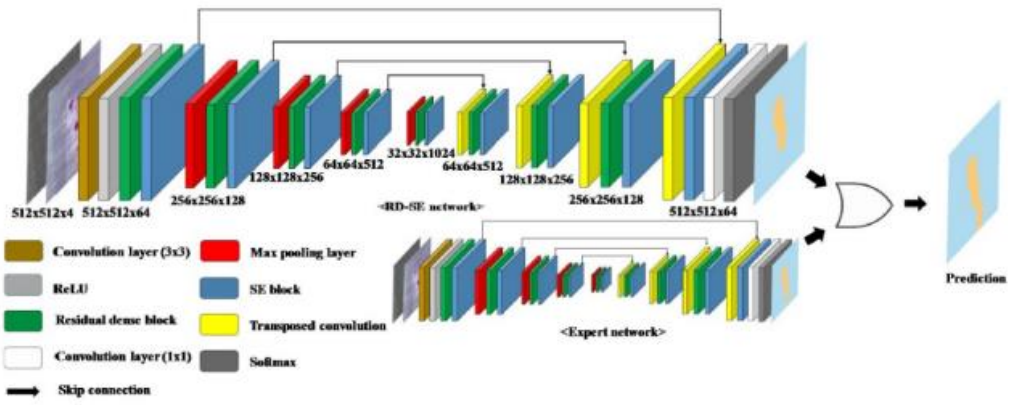


Figure 10: *Residual DenseNet with Expert Network Architecture.*

This architecture was implemented by Team DSSC - Hyunseong Park, Junhee Kim, Sungho Kim. They have implemented a U-Net based architecture called Residual Dense with a Squeeze and Excitation (S&E) Block (RDSE). The whole network is termed as RDSE with the Expert network. Along with the base RDSE model, expert networks were also used to identify less frequent agricultural patterns. RDSE addresses the problem of low resolution in original images. This network explores the hierarchical features from all the convolutional layers to counterbalance the spatial losses due to low resolution. Residual Dense Blocks (RDB) and skip connections allow direct connections from the state of preceding blocks to all the layers of current RDB, leading to a contiguous memory (CM) mechanism.

RDN mainly consists of four parts: Shallow feature Extraction Net, Residual Dense Blocks (RDB), Dense Feature Fusion (DFF) and up-sampling net (Decoder):

The Shallow Feature Extraction (SFE) net consists of one convolutional layer and a ReLu layer to extract shallow features. The output is then fed as input to the residual dense block.

The first convolutional layer extract features F_{-1} from input denoted by:

$$f_{-1} = h_{sfe1}(i_{lr})$$

Where h_{sfe1} denotes a convolutional operation and i_{lr} is input layer and if we have D residual blocks, the output of the last residual block

$$F_d = h_{RDB} * d(f_{d-1})$$

RDB consists of five convolutional layers with kernel size 3*3 and a batch normalization layer.

Dense Feature Fusion includes Global Feature Fusion (GFF) and Global Residual Learning (GRL). DFF takes all the features from the previous layers and is represented by F_{df}.

Where $F_{df} = H_{DF}(F_{-1}, F_0, F_1, \ldots, F_D)$,

H_{DF} is a composite function

Local Residual Learning (LRL): In each RDB Blocks there are 5 convolutional layers. LRL was introduced in RDB to improve the performance and information flow.

Up- Sampling Net: Transposed Convolution is used in the decoder side to get back the original image size.

Apart from Residual Dense Net, Squeeze and Excitation blocks are also introduced to understand the channel interdependencies. The outputs from the residual layers are fed into the SE Layer. First, the feature maps of each channel are squeezed to a single numeric value. Afterwards, it is fed through a two-layer neural network, giving it a linear scalar of how relevant each one is. In SE Block, each channel is squeezed to a single numeric value using Average Pooling. A fully connected layer followed by the ReLu function adds necessary non-linear complexity. Then, a second Convolutional layer followed by sigmoid activation gives each channel a smooth gating function. Adding SE Blocks can improve the model performance while reducing the computation cost. To address the less frequent classes in the Agricultural-Vision Dataset (i.e., Planter skip and Standing Water) an additional Expert Network is implemented.

3.5.3.2 RESULTS

The challenge competition used mean Intersection-over-Union (mIoU) as the main quantitative evaluation metrics. This is popularly used for semantic segmentation tasks. Thus the performance is measured using the mIoU metric.

The Team DSSC secured first place with mIoU value of 56.9% for segmenting weed clusters and an average mIoU of 63.9% (The average mIoU is computed by (true_positive) / (prediction + target - true_positive), averaged across all classes.) Team SCG_Vision also achieved an impressive result in the competition. The mIoU measured for weed identification was 53.8% and an average mIoU of 60.8%. Overall among the 57 participating teams around the globe, 7 leading teams were selected for publishing papers.

3.5.4 SELF CONSTRUCTING GRAPH METHOD

Team SCG Vision represented by Qinghui Liu, Michael C. Kampffmeyer, Robert Jenssen, Arnt B. Salberg from Norwegian Computing Center, UiT The Arctic University of Norway [1] proposed Self Constructing Graph (SCG) module combined with Graph Convolutional Network (GCN) for the segmentation of agricultural dataset. The team received the fourth position in the Agricultural Vision challenge competition in 2020. Unlike Graph Neural Network (GNN) that rely on manually built prior knowledge paths, the Self Constructing Graph(SCG) uses the learnable latent variables to produce embedding and to self-construct the underlying graphs straight from the input features. SCG can automatically obtain the optimized non-local context graphs from complex-shaped objects in aerial imagery. This is necessary for our agricultural dataset.

In the proposed SCG Network, SCG is followed by Graph Convolutional Networks to update the node features along the edges of the graph. SCG model architecture contains a CNN-based feature extractor,

for example - customized ResNet50 output 1024-channel, three SCG-GCN modules to extract features at multiple views and a fused output block, where the element-wise sum is projected back to 2D maps for final prediction.

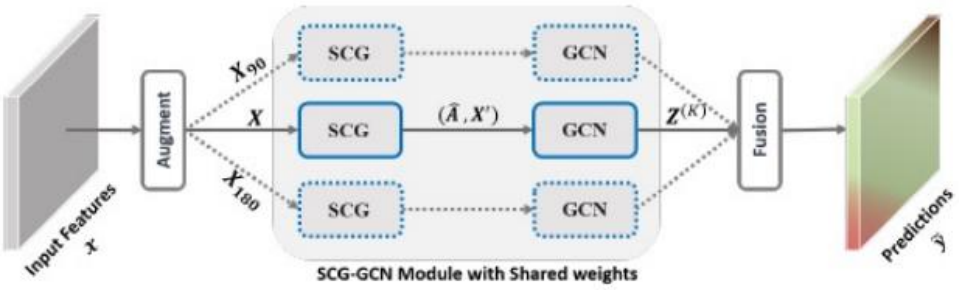


Figure 11: *Self constructing graph with GCN Module*

An adaptive class reweighting loss is designed to overcome the class imbalance problem in this dataset. Further, a positive-negative class balanced function is adopted to accommodate for negative samples. The SCG Team achieved a performance of 53.8% mIoU in the challenge competition.

4 CHAPTER THREE – EXPERIMENTATION

4.1 INTRODUCTION

In Chapter 2, I have discussed various segmentation models and related works in the segmentation tasks and in the context of agricultural vision datasets. In this Chapter, I will show a much more defined way of implementing a custom deep learning model, selecting hyperparameters for optimized results improving accuracy, and finally comparing various deep networks that can efficiently segment weed clusters from aerial farm images. The key background and requirements for implementing the segmentation models are gathered in the previous chapters. To recap, in the first chapter, it discusses the growing demand of digital technology in the agricultural sector and deep learning based precision farming to control the weeds. Chapter one also describes the resources and data used for implementing this project. From the literature review in chapter two, it has seen the various works in the segmentation tasks and methodologies proposed by researchers in this particular dataset with their results. The custom models based on U-Net, pre-trained transfer learning on DeepLabV3+ architecture and FCN models are implemented in this chapter. The results of these models are evaluated, compared under different hyper-parameter conditions to achieve an optimal performance for this particular task. This chapter also

examines the hyper-parameters and its impact on model performance. The results and evaluation metrics such as mean Intersection over Union (mIoU) are compared and discussed in the next chapter.

To summarize, this chapter addresses the following objectives:

- Data Pre-processing techniques.
- Building Semantic segmentation models.
- Hyper-Parameter Tuning.

4.2 PROJECT BREAKDOWN STRUCTURE

The figure below shows the proposed data pipeline used for implementation of this project:

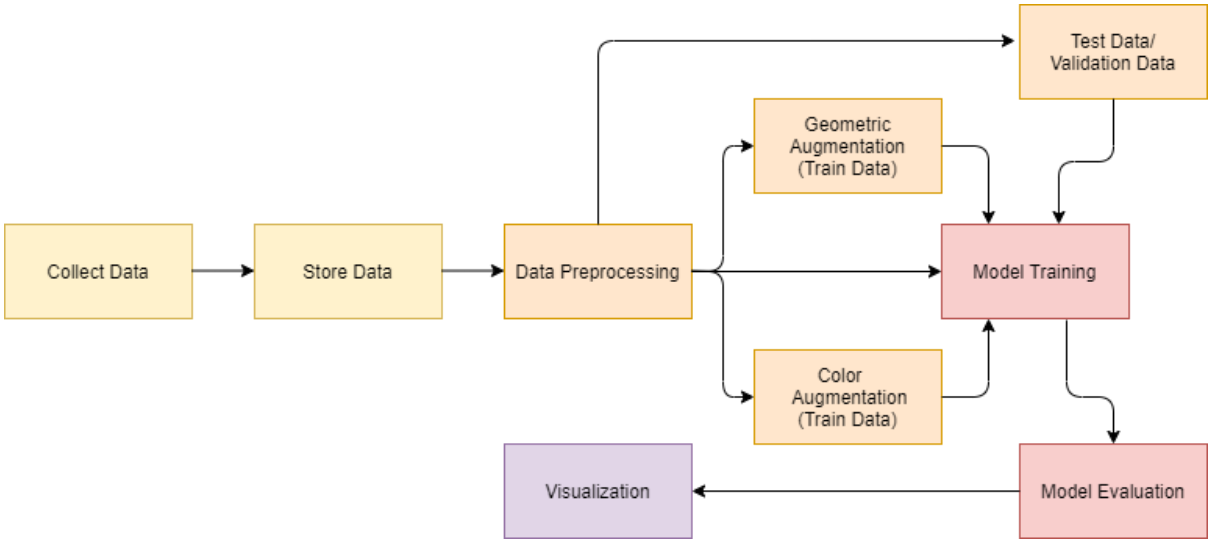


Figure 12: *Proposed data pipeline*

4.3 DATA ACQUISITION

Aerial farmland images were captured by cameras mounted on drones and flown over fields in the US during the growing seasons between 2017 and 2019. Each field image contains four channels - Near Infrared (NIR), Red, Green and Blue. The images were captured by Nikon D850, Canon SLR, Nikon D800E and WAMS cameras taken at a resolution of 10cm/pixel to 20cm/pixel. The acquired images were extremely large. For instance, the average size of 10875 × 3303 pixels and the largest field image collected is 33571×24351 pixels in size. This could be a challenging task for deep learning implementation. For this reason, the images were subsampled by cropping annotations with a window size of 512 × 512 pixels.

4.4 DATA COLLECTION AND STORAGE

As mentioned in Chapter One, the Dataset was collected from the Agricultural Vision and the team sent the data as a zip file. The Image dataset contains 21,061 aerial farmland images with a size of around 3.5 GB and a dimension of 512 x 512. The folder contains RGB Images, Grey scale Images, Masks and Boundaries for different classes.

The data was initially stored in the local system and the model was trained using the local Jupyter IDE. Due to computation and memory limitations, the resources were exhausted while training the model. Secondly, the Amazon S3 bucket was used to store data and the model was trained on Google Colab. However, considering the cost factor, I looked forward to a different approach. Finally, the data was uploaded in Google Drive and directly mounted on Google Colab Pro, which provides high speed RAM and processing power required for training the model.

Figure 13: *Input Image*

Figure 14*: Ground truth*

4.5 DATA STATISTICS

4.5.1 ANNOTATION AREA

Agricultural patterns have irregular shapes and sizes. For example, the weed clusters may appear in either small patches or enormous areas on the land. Thus the concentration of weed clusters is different for different samples. To examine the class imbalance, here the pixels belonging to the weed clusters are analysed. The graph below represents the total number of pixels for weed clusters in the dataset. We can see that the majority of images have only a small portion (around 50k pixels) covered by weeds out

of approximately 700k pixels. (512*512*3=786,432 pixels). While only a small portion of images have large portions covered by weeds. This indicates extreme class imbalance.

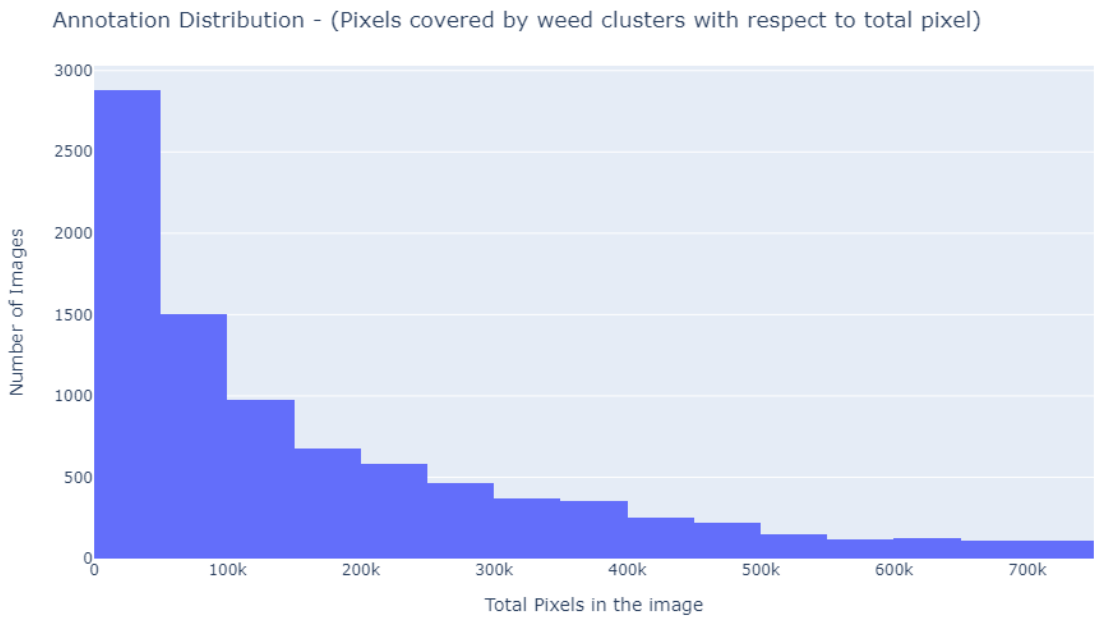


Figure 15: *Annotation distribution*

4.5.2 PROPORTION OF ZERO CLASS IMAGES

Analysing training data, we found out that not all the images had weed clusters on them. Some of the images fall into different annotation categories which are not examined in this project. The graph below represents the number of images having weed clusters and the ones without. Thus for training purposes, we found 8633 images having weed clusters. Other images were ignored and not used for training the model.

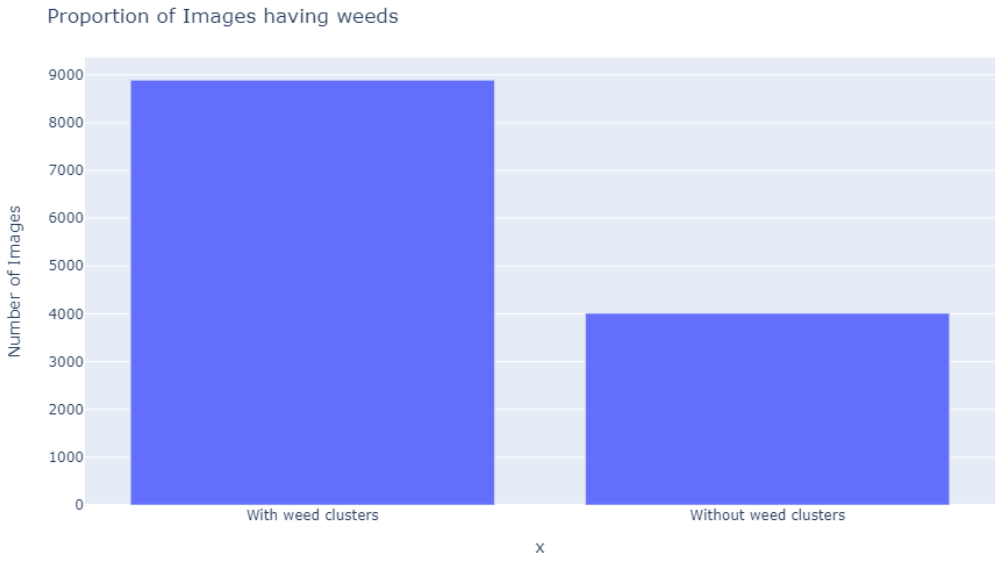


Figure 16: *Proportion of images with and without weed clusters*

4.6 DATA PRE-PROCESSING

Because the original dataset contains different classes such as Cloud shadow, double plant, Planter skip, Standing Water, Waterway and Weed cluster, this project is concentrating on weed clusters. Pre-processing requires defining the correct path for train, validation and test images as well as their mask images. The boundary images are ignored because of time limits. In this experiment, the validation and train images containing only weed clusters are selected for training.

The data file structure is shown below:

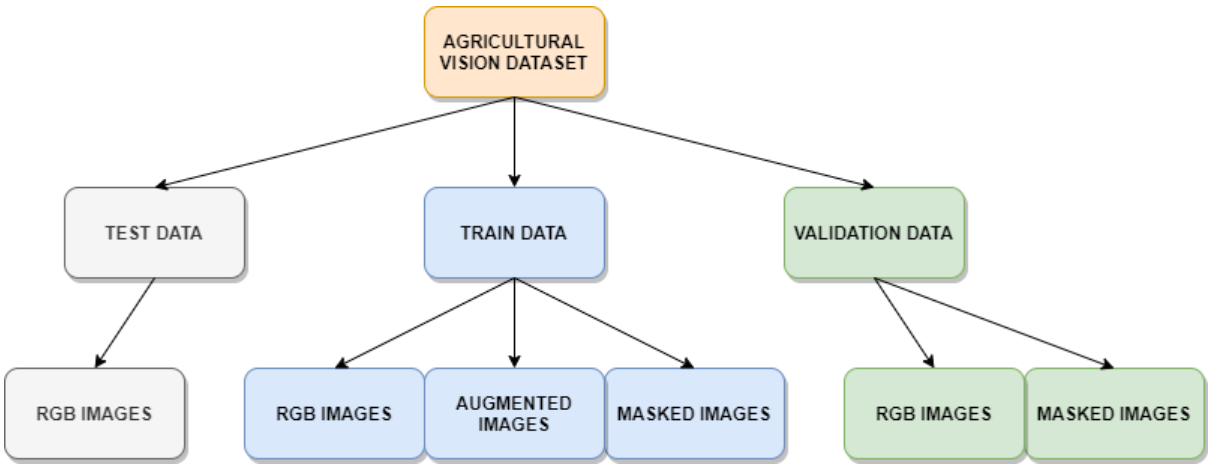


Figure 17: *File structure of data*

The following methods were performed in the pre-processing stage:

4.6.1 CHOOSING IMAGES WITH WEED CLUSTERS

In this process, Masked Images in the 'weed_cluster' folder were analysed to study the presence of weeds. After careful analysis, it is found that the images that don't have weed clusters are fully background images. The figure below shows the masks, one with weeds and the other without weeds. The black area represents the background (crops) and white represents the annotated weed clusters. As seen below, the Images that do not have weeds are completely black. Converting images to NumPy arrays, and analysing the pixel information, all the pixel values in the black images are of value 255. Thus a python code is implemented to select only images having different pixel values (black and white pixels) in their corresponding NumPy arrays. The path to these selected images is stored in a text file (.txt) for further analysis and modelling.

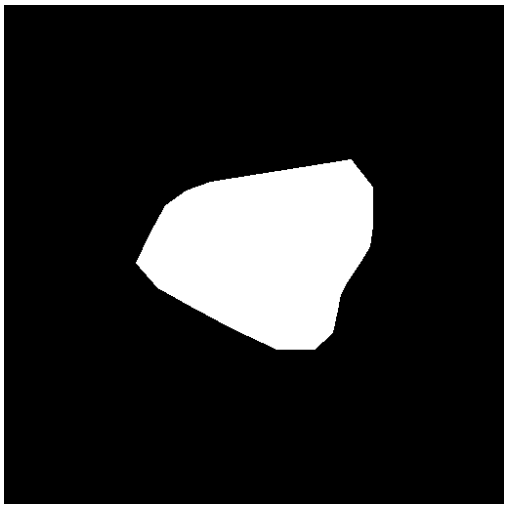


Figure 18: *With weed*

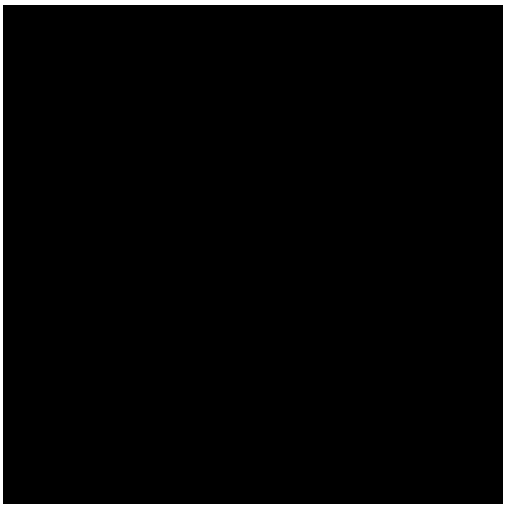


Figure 19: *Without weed*

4.6.2 PREPARE PATHS OF INPUT IMAGES AND TARGET MASKS

From step 5.6.1, Target images were chosen and their paths were stored in the Target text file. Target paths in this text file were mapped to corresponding Input Images stored in the 'rgb' folder and the paths to those were stored in a different text file named Input text file. These two files were used as path definitions of input and target images used for training the models.

Input text file= List of input image paths.

Target text file = List of Target image paths.

4.6.3 PREPARE A SEQUENCE CLASS TO LOAD AND VECTORIZE BATCHES OF DATA

Batch size is an important hyperparameters in deep learning to estimate the error gradient. The number of examples used from the training dataset defines the batch size. The sequence class is used to generate train and validation datasets based on batch size. The train and validation data are categorized into different batches and fed into the model.

4.7 DATA AUGMENTATION

This technique is used to increase the diversity of training sets by applying random transformations. In this Dataset, Augmentation plays an important role to avoid overfitting. Data Augmentation incorporates a suite of systems that enhance the size and quality of training datasets such that high-grade Deep Learning models can be developed using them. It is reasonable to produce a variety of images with different locations, orientations, scale and brightness. That is able to accurately recreate the conditions faced by the farmers in acquiring aerial images of farms. The image augmentation algorithms proposed in this project include geometric transformations and colour space augmentations.

4.7.1 GEOMETRIC TRANSFORMATION:

Choice of geometric transformation is primarily based on understanding the augmentation in the context of the safety of the application. Safety refers to the likelihood of preserving the label post transform. For example, cropping, rotation, flipping are safe on agricultural field datasets. These augmentations address the broad scope of challenges in the cultivation of corn and soybeans. This data augmentation technique is used for improving the quality of the dataset with real-world weeds captured by drones. Key assumptions include capturing aerial images at various angles, depths and directions.

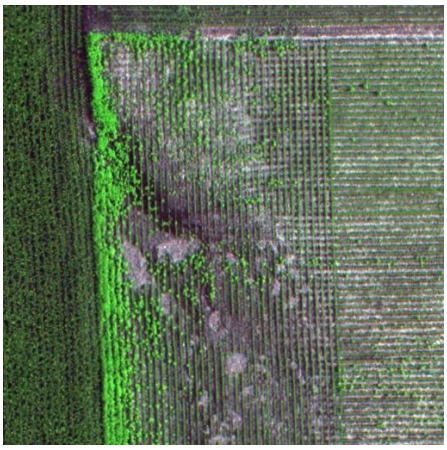


Figure 20: *Input image*

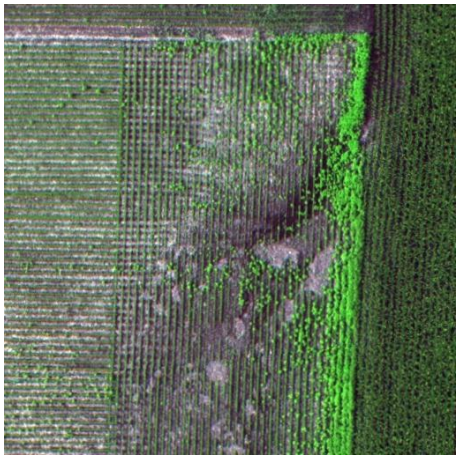


Figure 21: *Horizontal flip*

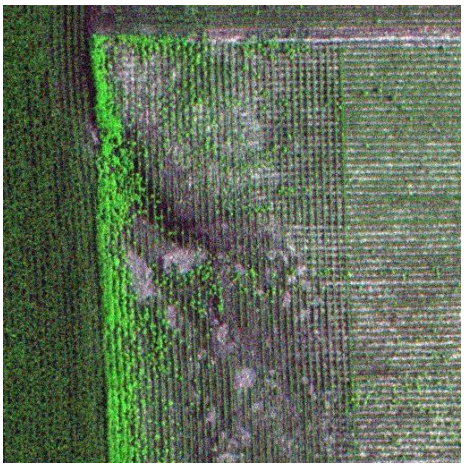


Figure 22: *Vertical flip*

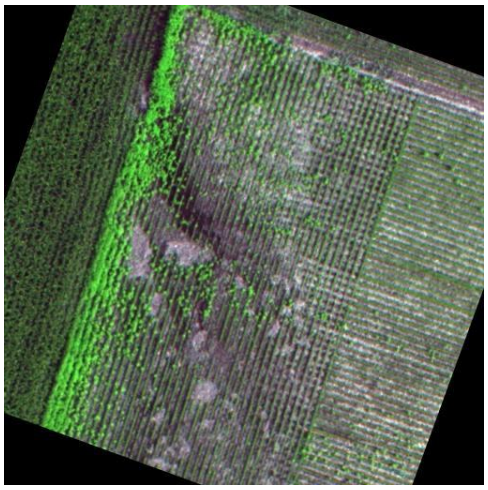


Figure 23*: Rotated image*

4.7.2 COLOUR AUGMENTATION:

Performing transformations in the colour channels space is another approach that is very reasonable to achieve. Digitally, Image data is encoded as a tensor of the dimension (height × width × colour channels). Colour augmentation includes isolating a single colour channel such as R, G, B. Histogram equalization for adjusting the contrast, Controlling the Brightness, Saturation and Hue. These techniques address various environmental conditions in the real world such as sunny weather (High brightness), cloud shades, nature of soils etc. Colour augmentation is used to stimulate a brighter or darker environment by decreasing or increasing the pixel values respectively.

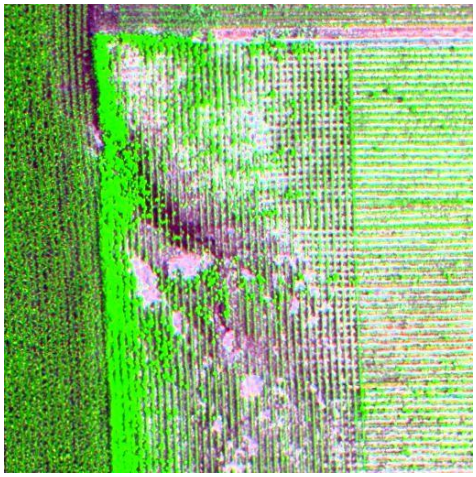


Figure 24: *High Brightness*

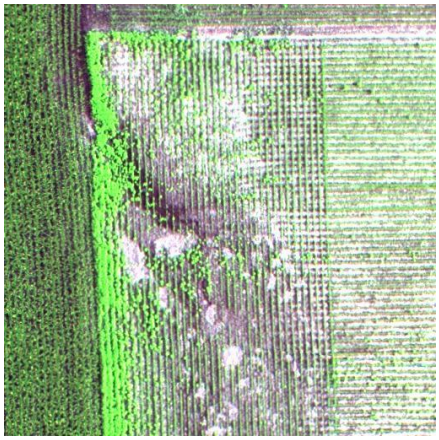


Figure 25*: High contrast*

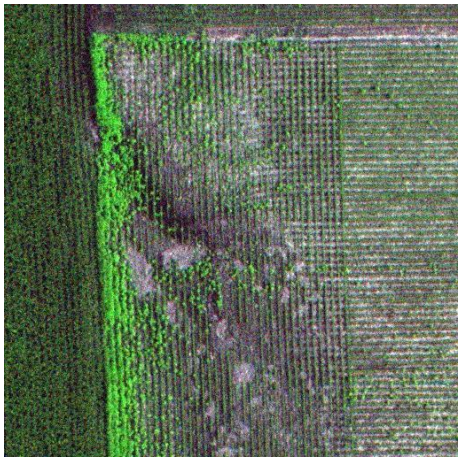


Figure 26: *Noise*

In this experiment, eight modes of image transformation are used. The diagram below illustrates the pathway of each mode in detail. Initially, the images are geometrically transformed, and then colour augmentation techniques are performed.

4.8 DATA SPLITTING

After augmentation, images are fed to the models for training. A model is stated to be a good classifier if it can certainly segment any unseen data. However, that is not always the case in real scenarios. The model may fail to predict the new data. This is called overfitting. This is a common problem in deep learning. To address this, Dataset is split into separate training, validation and test data. Validation data is used to validate the model performance on unknown data and test its generalization property. A concept called cross-validation. The obtained dataset from agricultural vision had already split data into train, test and validation. The generated dataset thus contains farmland images with a 6/2/2 train/Val/test ratio.

4.9 MODEL CHOICE ASSUMPTIONS

The choice of model for training agricultural dataset is based on the following assumptions.

- The dataset used for training the model is huge, however adding the augmentation makes it a large-scale dataset. Thus a lightweight model such as U-Net is proposed, given the time and resource constraints.
- The dataset is highly unbalanced. In some images, the weed clusters are sparse. Thus proper optimization techniques and class weights are assigned to tackle this problem.

- To handle the problem of segmenting objects at multiple scales, the Atrous convolution method is employed. This convolution is used in the DeepLab series in cascades or in parallel to capture multi-scale contexts by adopting multiple atrous rates. Thus the latest version of DeepLab is investigated in this project.
- Used a pre-trained model to transfer learned features from already available huge datasets. Implemented Transfer Learning in FCN and DeepLabV3+.
- It was assumed the aerial farmland images had the characteristics of high spatial resolution and low spectral resolution.

4.10 BUILD AND TRAIN MODELS

In this project, the dataset is trained on prominent and widely used deep learning-based semantic segmentation models such as FCN, U-Net and DeepLabv3+. Pre-trained models such as VGG16 and Xception are used in the encoder side of FCN and DeepLabv3+. These models have achieved excellent results on some public competitions like the PASCAL VOC-2012 semantic segmentation task and thus performing these models for our agricultural datasets to recognize patterns in farmlands.

The first model employed here is Fully Convolutional Network (FCN). They employ locally connected layers such as convolution, pooling and up sampling. In a regular CNN network, if the fully connected layers at the end are replaced by a convolutional layer. We get coarse spatial features as output instead of vectors. This can be further up sampled to get the segmented image.

One of the biggest challenges in using pre-trained models that are trained on huge image datasets such as ImageNet, COCO etc. is that these are usually trained on fixed input image sizes typically ranging from 224*224*3 to 512*512*3. To select pre-trained models trained on Images having different inputs, the FCN model has been chosen. FCN is one model that doesn't have a dense layer instead it has a 1*1 convolution layer that performs the task of dense layers in CNN. This helped FCN to take away restrictions on variable input and use different pre-trained models.

The architecture is built using Keras - A high-level API for Google's TensorFlow library. Keras is a popular choice for deep learning as it is highly integrated into TensorFlow. TensorFlow also offers various tools for visualization, debugging, running models on browsers etc. Building a model in Keras is straightforward and can be implemented using fewer lines of code unlike other libraries such as PyTorch. Keras is ideal for building complex and advanced models using its functional API and layers. The Keras architecture of the proposed three models is illustrated in the diagram below.

4.11 TRANSFER LEARNING

Transfer learning is a method, where a model used for a particular task is reused to train a second model. Given the computation and time resources of deep learning models, Transfer learning is used as a starting point to allow rapid progress and improve performance.

In this methodology, a base network is trained on a base dataset and target. The model tends to learn the general features such as edge detection, shape, size etc. and then repurpose the learned features and transfer them to the second network to be trained. This model will work as a generic network of the real world. Assuming that the general features of the base model may align with the features of weed such as shape, size and colour, pre-trained models trained on PASCAL VOC 2011 and Cityscaper Image datasets were chosen. Feature extraction methodology was used to customize transfer learning. A new classifier trained from scratch was used on top of the pre-trained model. The base network contains general features used for classifying images while the final classification part of the pre-trained model is specific to the features of the agricultural farmland images.

The U-Net architecture does not use any pre-trained models. In FCN architecture, an encoder extracts features that are used by a decoder to reproduce the segmented images. The pre-trained VGG16 model is used here for feature extraction on the encoder side. While on DeepLabV3+ architecture, pre-trained Xception and MobileNetV2 models were used as a base feature extractor.

In this work, the following backbone models were used as the base model:

VGG16: Used in FCN Network 'Very Deep Convolutional Networks for Large-Scale Image Recognition" Pretrained on PASCAL VOC 2011 dataset

Xception: Used in DeepLabV3+ pre-trained on PASCAL VOC 2011 and City Scrapes

MobileNetV2: Used in DeepLabV3+ pre-trained on PASCAL VOC 2011 and City Scrapes.

4.12 HYPER-PARAMETER TUNING

When training a deep learning model, the main goal is to get the best performance on validation datasets. This represents how well the model generalizes. When dealing with this experiment, evaluating the right hyperparameters to get a model that is both precise and generalized is a challenging task. Evaluating the objective function can be expensive, as training 12000 images of size 512*512*3 takes a long time and trying out different hyperparameters may take several days.

To optimize the hyperparameters, a hyperparameter search space is defined. Examples of hyperparameters include batch size, data augmentations to be used, dropout rate, learning rate, epochs, convolutional layer filter size, pooling type etc. In the hyperparameter search space, a config dictionary

is introduced that contains the hyperparameters used for training. Rather than going through an exhaustive search space, lists of the most common values for hyperparameters are used in the config dictionary. The choice function selects a random variable from the given most common values and models are trained and compared on different sets of hyperparameter combinations.

4.12.1 OPTIMIZERS

Optimizers are methods used to alter the weights, learning rates and other attributes of a neural network. The goal of the optimizer is to find a local minimum point to reduce the loss.

Optimizers are algorithms used to change the attributes such as weights and the learning rate of neural networks to reduce the losses. A comprehensive comparative analysis of various optimizers specifically used in the context of semantic segmentation is performed. Performance was analysed on three different state-of-art optimizers such as Adaptive Momentum (Adam), Root Mean Square Propagation (RMS Prop) and Stochastic Gradient Descent (SGD).

4.12.2 LEARNING RATE:

Learning rate is another important hyperparameter to determine the performance of neural nets. A higher learning rate causes undesirable divergent behaviour in the loss function. While a smaller learning rate takes the training a long time. To find the optimal learning rate, a simple experiment is performed, where the learning rate is gradually decreased by a factor in each epoch. The loss function is analysed for each epoch, and the optimal learning rate zone is where a quick drop in loss function is observed. The best learning rate is associated with the steepest drop in loss

4.12.3EVALUATION METRICS:

In this project, mean Intersection-over-Union (mIoU) is used as a main quantitative evaluation metric, which is one of the most commonly used measures in semantic segmentation datasets. The mIoU is computed as:

$$\text{mIoU} = \frac{1}{C}\sum \frac{Area(P(c) \cap T(c))}{Area(P(c) \cup T(c))}$$

Where c is the number of annotation types (c = 2 in this dataset, with 1 pattern + background), P(c) and T(c) are the predicted mask and ground truth mask of class c respectively.

For pixels in each predicted image, a prediction of a masked area label will be counted as a correct pixel classification for that label, and a prediction that does not contain any ground truth labels will be counted as an incorrect classification for all ground truth labels.

5 CHAPTER FOUR - EVALUATION, DISCUSSIONS AND RESULTS ANALYSIS

5.1 INTRODUCTION

In the previous chapter, I have trained the model with a proposed architecture. Model is trained on training data and validated by validation data. But how well can this model achieve quality results? In other words, evaluation of the model is essential to estimate the performance and how strongly this model predicts on generalized data.

In this section, I present the evaluation methods, results and discuss the training hyperparameters. In this context of agriculture, correctly identifying the patterns is crucial, as it informs the farmer with potential weed clusters on their farmlands. The goal of this model is to classify every pixel in the image to either weed or background. However, segmentation may work well for a particular pattern but may miss or misclassify different examples. Hence, a well-defined metric will help to rank the model. The accuracy - proportion of correct classification among all the classification- is not the main focus of this project. As seen from the 'Data Statistics' section in Chapter 3, the proportion of weed and background is highly unbalanced. That said, appropriate evaluation metrics for semantic segmentation are considered to effectively classify the pixel.

Furthermore, this chapter highlights the results of three models implemented in Chapter three and value the quality under different circumstances.

5.2 EVALUATION APPROACH

The following outlines the evaluation methodology used in this experiment.

- Choice of Evaluation Metrics
- Benchmarking the Models.
- Models Parameters
- Performance of Models
- Hyper-Parameter tuning
- Results evaluation and discussions

5.3 CHOICE OF EVALUATION METRICS

5.3.1 PIXEL ACCURACY

Pixel accuracy is the simplest and most commonly used evaluation metric for semantic segmentation. Pixel accuracy reports the proportion of pixels in an image that were correctly classified. For example, suppose we have an image, where the weed clusters occupy a small portion. The majority of pixels belong to the background or crop. Pixel accuracy may give great accuracy even in the case that weeds were not identified. As seen from the 'Data Statistics' section in Chapter 3, the proportion of weed and background is highly unbalanced. Thus it is important to note that, majority of weed clusters make up only a small portion of the image. This makes pixel accuracy a completely misleading metric for this application.

5.3.2 MEAN INTERSECTION OVER UNION (mIoU)

As we mentioned in the above section, PA is not a good metric for measuring performance. One of the most successful ways to measure the quality of segmentation problems is by using the concept of sets.

The mean intersection over union is defined as the size of the intersection divided by the size of the union. In other words, it is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. The range of this metric is from 0-100%, where zero denotes no overlapping. It measures the similarities and differences of sample sets. This metric works well in our application. For example, the main focus of this metric is to identify the overlapping of images. Even in the case of an unbalanced dataset, if the predicted value doesn't overlap the ground truth, the mIoU value would be zero, thus making it a suitable measurement for calculating the performance.

5.4 BENCHMARKING THE MODELS

As mentioned earlier, this is an academic model project of the Agriculture-Vision Challenge Competition in conjunction with IEEE/CVF-CVPR Conference 2020 to encourage research and work towards a vision in agriculture. Currently, the results of the First Agriculture-Vision were published. The best example for judging the comparative performance of the recommended model would ideally be based on the top three performing submissions in the challenge leader board. This has provided an approach for independent validation and performance results of the experiment conducted. This is also an ideal benchmark for the stated architecture, as this project is in compliance with the terms and methods mentioned in the competition. Figure 27 shows the results of the first Agriculture-Vision challenge

Challenge leaderboard

Team	mIoU (%)	Background	Cloud shadow	Double plant	Planter skip	Standing water	Waterway	Weed cluster
Hyunseong	63.9	80.6	56.0	57.9	57.5	75.0	63.7	56.9
seungjae	62.2	79.3	44.4	60.4	65.9	76.9	55.4	53.2
yjl912.2	61.5	80.1	53.7	46.1	48.6	76.8	71.5	53.6
ddcm	60.8	80.5	51.0	58.6	49.8	72.0	59.8	53.8
RodrigoTrevisan	60.5	80.2	43.8	57.5	51.6	75.3	66.2	49.2
SYDu	59.5	81.3	41.6	50.3	43.4	73.2	71.7	55.2
agri	59.2	78.2	55.8	42.9	42.0	77.5	64.7	53.2
Tennant	57.4	79.9	36.6	54.8	41.4	69.8	66.9	52.0
celery03.0	55.4	79.1	38.9	43.3	41.2	73.0	61.5	50.5
stevenwudi	55.0	77.4	42.0	54.4	20.1	69.5	67.7	53.8
PAII	55.0	79.9	38.6	47.6	26.2	74.6	62.1	55.7
agrichallenge1.2	54.6	80.9	50.9	39.3	29.2	73.4	57.8	50.5
hui	54.0	80.2	41.6	46.4	20.8	72.8	64.8	51.4

Figure 27: *Results of the first Agriculture-Vision challenge*

The models selected for comparison are Hyunseong (Residual DenseNet with Expert Network for Semantic Segmentation – 56.9%), Team SYDu (55.2%) and TeamPAII (55.7%)

5.5 MODEL PARAMETERS

Model parameters are divided into fixed and variable parameters. Some of the parameters are fixed within their respective models and no methods were implemented to change those parameters. These are required by the models to make predictions and define the skill of the model. The table below shows the fixed parameters with their respective models

Model	Total params	Trainable params	Non-trainable params
FCN	134,460,074	134,460,074	0
U-NET	31,055,427	31,043,651	11,776
DeepLabV3+	2,141,762	2,108,674	33,088

Table 2: *Illustrates the trainable parameters of each model*

For optimizing the results, some of the parameters are being changed. These are external to the model and are specific to the given task. Estimating the best combination of hyperparameters will help to optimize the performance. Hyperparameter tuning is analogous to the settings of the algorithm that can be tailored to enhance the results.

The table below shows the default hyperparameters used for tuning in this experiment

Hyper-parameters	Value
Number of epochs	200
Batch size	8
Optimizer	SGE
Learning rate	0.01

Table 3: *Default hyperparameter used in this experiment*

5.6 PERFORMANCE OF MODELS

Evaluation of the trained model was performed by visual examination of performance metrics and how they changed by altering the hyperparameters and model architecture. As mentioned in Chapter 2, different models have been put forward to perform the task of segmenting weeds. Along with that various hyperparameters were investigated. The main difficulty in using Convolutional Neural Networks to solve this particular task is that weed experts identify the weeds by interpreting their species root, shape and size while deep learning methods rely on the available dataset and annotated images to map input and output. These images have complex weed structures, thus making the model segment weeds by the least performance margin. After running models with different combinations of hyperparameters, the model mIoU is reaching around 40%-45%.

Figure 28*: Snapshot of training FCN Model*

Figure 29: *Snapshot of training U-Net Model*

5.7 RESULTS OF HYPER-PARAMETER TUNING

It is important to say that at this level little is known about the exact impact of hyper-parameters on the performance of models. In the previous section, we have trained models using the default hyperparameter setting and used an epoch of 200. The strategy of this stage is based on a trial and error approach, where we customize hyperparameters, expecting to see improvement in results. Following parameters have been selected for this process - number of epochs, batch size, selection of optimizers, loss functions, learning rates and evaluation metrics. Why and what hyperparameters were used for this project is mentioned in Chapter 3. Config dictionary is used as a hyperparameter search space where we define hyperparameters and possible valid configurations. However, in the later stage, parameters were defined manually and tested to see the performance of each setting.

Hyper-parameters	Valid Configuration
Number of epochs	100,200
Batch size	8,16,32,64
Optimizers	SGE, Adam RMSProp
Learning rate	0.01, 0.001

Table 4: *Table shows the hyperparameters used and their valid configurations*

5.7.1 NUMBER OF EPOCHS

Epoch defines the number of times the learning algorithm will work through the entire training dataset. Finding the right epoch number is an experimental task. This could be done by analysing the validation loss after each epoch. As the model starts training, both validation and training loss decreases. Validation loss saturates at a point and then starts to increase. The epoch where the validation loss saturates is the optimal epoch count for this particular dataset.

5.7.2 BATCH SIZE

Number of samples processed before the model gets updated is known as batch size. On the one extreme, using a batch size equal to the entire training dataset allows convergence to the global minima of the loss function. However, a large batch size requires more computation power and training becomes slower. On the other hand, using a small batch size cannot guarantee global minima but converges to a good solution. The experimental results using two different batch sizes are shown in the table below.

Model	Batch size	mIoU
FCN	8	0.4268
FCN	16	0.4311
U-NET	8	0.4278
U-NET	16	0.4137
DeepLabv3+	8	0.4314
DeepLabv3+	16	0.4465

Table 5: *Performance of models vs batch size*

5.7.3 SELECTION OF OPTIMIZERS

Optimization strategies are effective for diminishing losses and providing the best concrete outcomes possible. The default values of optimizer parameters are:

SGE	Learning rate = 0.01, decay rate=5**(-4), momentum=0.9, nesterov=True
Adam	Learning rate=0.01, beta_1=0.9,beta_2=0.999,epsilon=1e-07,amsgrad=False
RMSProp	Learning rate = 0.01, rho = 0.9, momentum=0.0,epsilon=1e-07

Table 6: *Default values of optimizers*

The performance of three different optimizers are shown below:

SGE	0.4222
Adam	0.4321
RMSProp	0.4079

Table 7: *Performance of optimizers for U-Net model*

5.7.4 LOSS FUNCTIONS

The choice of loss/objective function is extremely important while designing complex image segmentation based deep learning architectures as they instigate the learning process of algorithms. In this dataset, we use weeds as the point of interest and the rest is classified as background. Known as a binary classification problem. Binary cross-entropy is a good loss function for binary segmentation.

Binary Cross-entropy is defined as a measure of the difference between two probability distributions for a given random variable or set of events.

Binary Cross-Entropy is defined as:

$$L_{BCE}(y, \hat{y}) = -(y\log(\hat{y}) + (1 - y)\log(1 - \hat{y}))$$

Here, $\hat{y}$ is the predicted value by the prediction model.

Where y is the label (1 for weeds and 0 for background) and p(y) is the predicted probability of the point being weed for all N points.

Binary cross entropy compares each of the predicted probabilities to actual class output which can be either 0 or 1. It then calculates the score that penalizes the probabilities based on the distance from the expected value. That means how close or far from the actual value.

5.7.5 LEARNING RATE (OPTIMIZER)

By default, the learning rate of all optimizers are set to 0.01. However, as mentioned in chapter 3, to find an optimum learning rate, the learning rate is gradually decreased by a factor in each epoch.

5.8 RESULTS, EVALUATION AND DISCUSSION

The results suggest that the DeepLabv3+ model with Xception backbone and pascal_voc weights outperformed the other two models to some extent. Mean intersection over union (mIoU) of the pre-trained models in the study was found to be 0.4672 with a batch size of 16 and Adam optimizer. The batch size is set to 16 due to resource constraints, as higher batch size leads to computation expense.

In addition, it is assumed that the deficiency of complex farmland images in the data set in which pre-trained models were used may have caused a contrary impression on evaluation metrics. In using the data augmentation techniques, it is observed that there is an improvement in the generalization, which is potentially the reason the overall performance of the best model was updated.

This study is to perform the semantic segmentation model and compare the proposed models with the top performers in the agricultural vision challenge competition. The table presents the results of the proposed model, FCN, U-Net, DeepLabv3+ and the custom models from the challenge competition. Moreover, the performance of the proposed model did not exceed the level of top performers - . However, it has achieved comparable results. Pre-trained models give better results than training networks from scratch. Optimizing the parameters and hyperparameter tuning has improved the results slightly.

Models	mIoU
U-Net	0.428
FCN	0.422
DeepLabv3+ (MobileNetv2 and PASCAL)	0.441
DeepLabv3+ (MobileNetv2 and Cityscapes)	0.434
DeepLabv3+ (Xception and PASCAL)	0.467
DeepLabv3+ (Xception and Cityscapes)	0.452
Hyunseong	0.569
Team SYDu	0.552
Team PAII	0.557

Table 8: *Performance of various models*.

Results of the experiment show that the performance of the model varied slightly. Even though the model architectures and their concepts were different, there is no significant gap in the results of the three proposed models. That said, there should be methods to efficiently segment the weeds. While the proposed models have outstanding results in various other tasks such as medical imaging segmentation, city scrapers and common object segmentation, it is assumed that the complex features of agricultural patterns might need advanced high-performance models. Unlike other segmentation, the agricultural patterns for different crops may change significantly. Thus, adding a new challenge in finding the custom models for particular weeds.

Overall, the original motivation describes the examination of various segmentation models, comparing the results and improving the models by tuning hyperparameters and augmenting data. The initial focus was on data preparation and building the models. Secondly on the evaluation and performance improvement tasks. However, the focus shifted a bit in the data augmentation part. The proposed augmentation includes eight different modes. But, not all the images were augmented. This is due to the resource constraint as augmenting images took more memory space and exhausted the resources. That said, the main areas of research including a literature review on related models, building a deep learning-based neural network, and understanding the hyperparameters were addressed. During the implementation, lots of effort went into writing the code in python and using TensorFlow and Keras. One of the motivations was to learn about the state-of-art convolutional neural networks using the TensorFlow framework. With great community support, working in TensorFlow and Keras was relatively smooth and straightforward.

It was found that further studies are required to explore this situation in more detail. Future work will also direct the improvement of the model with complex weed images.

6 CHAPTER FIVE - CONCLUSIONS AND FUTURE SCOPE

In this chapter, the main points of this project are summed up and the future scope and areas of improvements are discussed.

6.1 CONCLUSION

In this work, we have compared various semantic segmentation techniques and their implementations. The motivation behind this research was to investigate various deep learning-based models used in segmentation tasks. To begin with, the presence of weeds in farmland raises the cost of agriculture and also carries away the vital nutrients needed for the crops. This is one of the major challenges encountered in farming nowadays. However, the traditional method used to tackle the weeds uses high levels of herbicides and fertilizers which has an adverse influence not only on the quality of production but also on the environment. In most cases, traditional agricultural methods are the cause of major carbon emissions. That results in climate change. Using Artificial Intelligence technology in agriculture can be seen as a developing area of research. That said, this project aims to build a deep learning-based solution for farmers to precisely identify the weed location by using semantic segmentation technology. This could help future farming to control the use of herbicides. The solution can be found by using an AI model, and I aim to identify suitable models here. The results were compared and improved using various hyperparameter tuning techniques.

Correspondingly, the answers to this hypothesis are spread across the report chapters starting from the introduction section in chapter one where the following are needed to be understood:

The project begins with the importance of smart farming and the vital requirement of weed controls to cope with changing climatic patterns. This section also explained the project aims, motivation and objectives and the resources used in this experiment.

Secondly, the literature review explains the traditional methods used to classify specific crops, their scope and challenges. With the implementation of machine learning, a new approach had revolutionized agriculture that used the image processing methods to extract the features and fed them into the machine learning-based classifiers. Although the results of this model had outstanding values, they are susceptible to specific environmental conditions. With the advancement of cloud computing, low-cost

GPUs and the availability of farm images, technological implementations in agriculture have seen further improvement in recent days. The deep learning-based approach plays an important role in extracting features that are crucial to identify the shapes, colour and size of a variety of weeds. In that case, various semantic segmentation models have evolved from common object detection tasks. The introduction of ImageNet data sets made notable improvements in the research domain of segmentation

models. Models such as U-Net have significant results in medical imaging. Having said that, this project aims to review common segmentation models, their architectural designs, implementations and challenges.

Thirdly, this project proposed three models -U-Net, FCN and DeepLabV3+ to test the performance of agricultural data. For this purpose, we developed a deep learning-based pipeline to build a classification model and evaluate the results. Most of the work in this section is done using python language, Keras API and TensorFlow framework. At that point, the investigation is required to look into the image data provided by agricultural vision, progress with the pre-processing and data augmentation tasks. Detailed analysis of CNN models and Keras implementation is done here.

Fourthly, the models were run on the TensorFlow framework and TensorBoard is used to visualize the progress of models. Performance is evaluated using mIoU metrics, as these metrics measure the intersection areas instead of accuracy, which apparently gives us wrong information. In view of improving the results of the proposed model, eight modes of data augmentation methods were implemented. However, not all the datasets were augmented due to memory constraints. Finally, various hyperparameters were adjusted to get refined outcomes from the proposed model.

To sum up, as depicted in the results, the best model, which is DeepLabv3+ has shown a result of 0.4672 - an equivalent of detecting 46.72% of areas covered by weed clusters. Compared with the result of the challenge competition, Team Hyunseong where they used Residual DenseNet with Expert Network and has achieved a result of 56.9%.

6.2 FUTURE SCOPE

Computer vision applications in agricultural farmlands are growing on a positive path and will be developed intelligently. Throughout the analysis, it is found that the existing technology and challenges explore the future opportunities to form references for researchers. The deficiency of aerial image data and the complexity of patterns are the major challenges faced in the research area. However, with the availability of advanced image capturing and drone technology more images can be captured to build a database addressing agricultural patterns under different environmental conditions and crop types. Advancement in convolutional neural networks could bring a state of the art model for precisely classifying the patterns. This task is labelled as a dynamic problem and more data is required to include various areas of challenges in segmentation. Therefore, it should be someone's research job to timely update the latest data and work on the implementation of an application useful for farmers. In the future, farmers can use cloud-based mobile applications to analyse their farm's real-time information and make decisions based on the finding of artificial intelligence technology.

7 REFERENCES

[1] Chiu, M., Xu, X., Wang, K., Hobbs, J., Hovakimyan, N., Huang, T., Shi, H., Wei, Y., Huang, Z., Schwing, A., Brunner, R., Dozier, I., Dozier, W., Ghandilyan, K., Wilson, D., Park, H., Kim, J., Kim, S., Liu, Q., Kampffmeyer, M., Jenssen, R., Salberg, A., Barbosa, A., Trevisan, R., Zhao, B., Yu, S., Yang, S., Wang, Y., Sheng, H., Chen, X., Su, J., Rajagopal, R., Ng, A., Huynh, V., Kim, S., Na, I., Baid, U., Innani, S., Dutande, P., Baheti, B., Talbar, S. and Tang, J., 2021. *The 1st Agriculture-Vision Challenge: Methods and Results*. [online] arXiv.org. Available at: <https://arxiv.org/abs/2004.09754> [Accessed 18 August 2021].

[2] Lmb.informatik.uni-freiburg.de. 2021. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. [online] Available at: <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/> [Accessed 13 August 2021].

[3] Medium. 2021. *UNet Line by Line Explanation*. [online] Available at: <https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5> [Accessed 31 August 2021].

[4] Ronneberger, O., Fischer, P. and Brox, T., 2021. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1505.04597> [Accessed 31 August 2021].

[5] Sabzi, S., Abbaspour-Gilandeh, Y. and Arribas, J., 2021. *An automatic visible-range video weed detection, segmentation and classification prototype in potato field*.

[6] Gianessi, L. and Reigner, N., 2021. *The Value of Herbicides in U.S. Crop Production*.

[7] College of Agricultural Sciences. 2021. *Conventional Weed Control*. [online] Available at: <https://agsci.oregonstate.edu/mes/sustainable/onion/best-management-practices-weed-control/conventional-weed-control> [Accessed 31 August 2021].

[8] Wu, Z.; Chen, Y.; Zhao, B.; Kang, X.; Ding, Y. Review of Weed Detection Methods Based on Computer Vision. Sensors 2021, 21, 3647. https://doi.org/10.3390/ s2111364

[9] Ma, Y.; Feng, Q.; Yang, M.; Li, M. Wine grape leaf detection based on HOG. Comput. Eng. Appl. 2016, 52, 158–161.

[10] Bakhshipour, A.; Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. Comput. Electron. Agric. 2018, 145, 153–160.

[11] Zheng, Y.; Zhong, G.; Wang, Q.; Zhao, Y.; Zhao, Y. Method of Leaf Identification Based on Multi-feature Dimension Reduction. Trans. Chin. Soc. Agric. Mach. 2017, 48, 30–37.

[12] Image Segmentation – Towards Data Science. 2021. *Image Segmentation – Towards Data Science*. [online] Available at: <https://towardsdatascience.com/tagged/image-segmentation> [Accessed 31 August 2021].

[13] Hao, S., Zhou, Y. and Guo, Y., 2021. *A Brief Survey on Semantic Segmentation with Deep Learning*.

[14] Champ, J., Mora-Fallas, A., Goëau, H., Mata-Montero, E., Bonnet, P. and Joly, A., 2021. *Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots*.

[15] Long, J., Shelhamer, E. and Darrell, T., 2021. *Fully Convolutional Networks for Semantic Segmentation*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1411.4038> [Accessed 31 August 2021].

[16] Medium. 2021. *Understanding Semantic Segmentation with UNET*. [online] Available at: <https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47> [Accessed 31 August 2021].

[17] Link.springer.com. 2021. [online] Available at: <https://link.springer.com/content/pdf/10.1007%2F978-3-319-24574-4_28.pdf> [Accessed 31 August 2021].

[18] 2019, M., M. Peters, T., H. Staib, L., M. Peters, T., H. Staib, L., Hill, U., Georgia, U., University, W., University, Y., Strasbourg, U., Intelligence, U., Hill, U., University, W., AG, S. and Springer, C., 2021. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019 | SpringerLink*. [online] Link.springer.com. Available at: <https://link.springer.com/book/10.1007/978-3-030-32239-7> [Accessed 31 August 2021].

[19] GitHub. 2021. *models/research/deeplab at master · tensorflow/models*. [online] Available at: <https://github.com/tensorflow/models/tree/master/research/deeplab> [Accessed 31 August 2021].

[20] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A., 2021. *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1606.00915> [Accessed 31 August 2021].

[21] Medium. 2021. *Review: DeepLabv3+—Atrous Separable Convolution (Semantic Segmentation)*. [online] Available at: <https://sh-tsang.medium.com/review-deeplabv3-atrous-separable-convolution-semantic-segmentation-a625f6e83b90> [Accessed 31 August 2021].

[22] Liu, Q., Kampffmeyer, M., Jenssen, R. and Salberg, A., 2021. *Self-Constructing Graph Convolutional Networks for Semantic Labeling*. [online] arXiv.org. Available at: <https://arxiv.org/abs/2003.06932> [Accessed 31 August 2021].

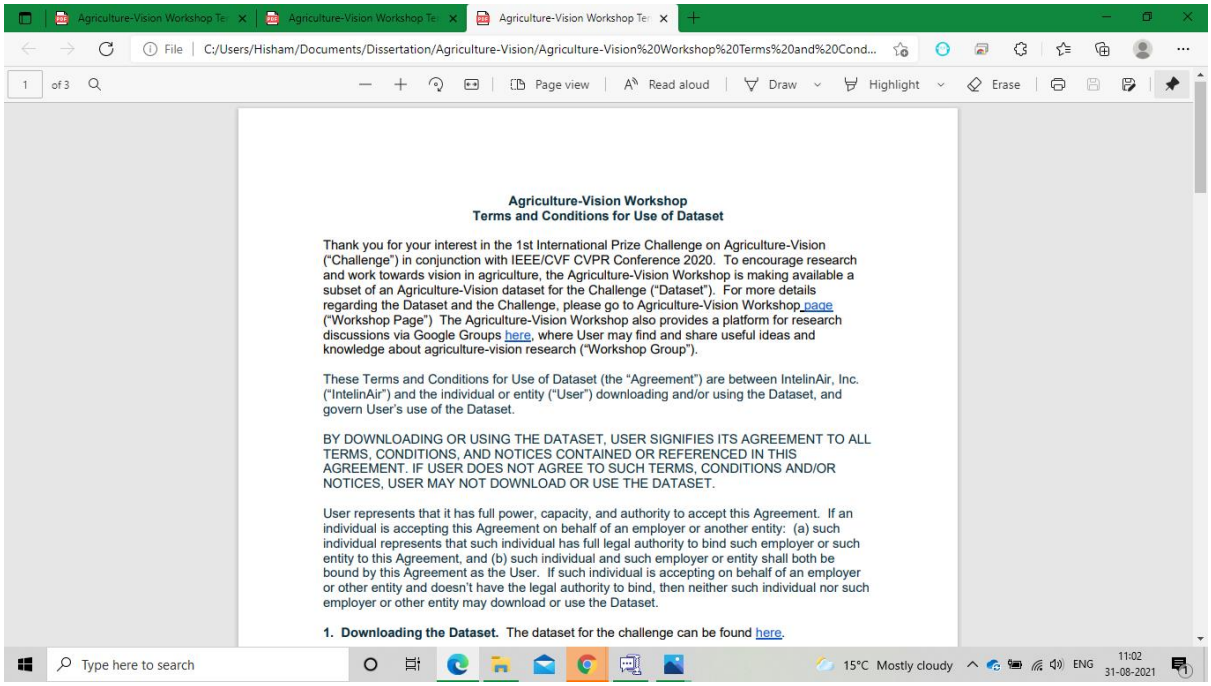
[23] Shorten, C. and Khoshgoftaar, T., 2021. *A survey on Image Data Augmentation for Deep Learning*.

8 APPENDIX

The link to the Google Colab code:

https://colab.research.google.com/drive/1CIycg9ds5EF_8khNdB1e2ZROLpYitEhq#scrollTo=K3qc8b0XH7Fr

Screenshot of Terms and Conditions for use of dataset



Agriculture-Vision Workshop
Terms and Conditions for Use of Dataset

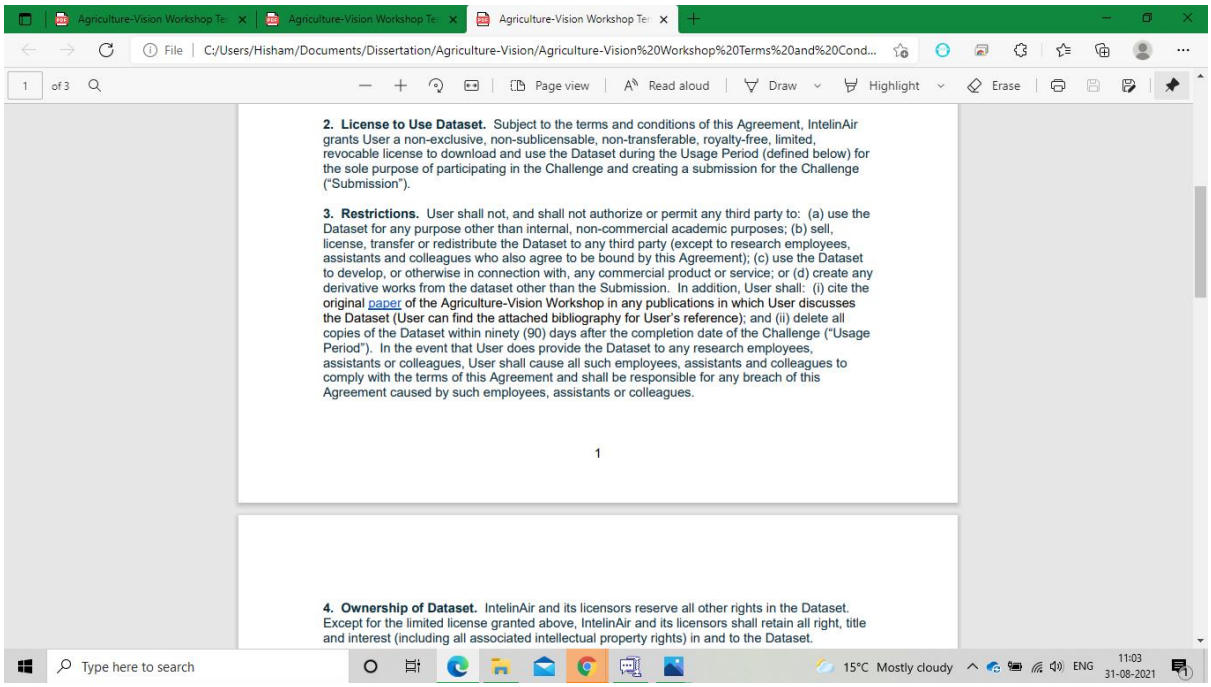
Thank you for your interest in the 1st International Prize Challenge on Agriculture-Vision ("Challenge") in conjunction with IEEE/CVF CVPR Conference 2020. To encourage research and work towards vision in agriculture, the Agriculture-Vision Workshop is making available a subset of an Agriculture-Vision dataset for the Challenge ("Dataset"). For more details regarding the Dataset and the Challenge, please go to Agriculture-Vision Workshop page ("Workshop Page") The Agriculture-Vision Workshop also provides a platform for research discussions via Google Groups here, where User may find and share useful ideas and knowledge about agriculture-vision research ("Workshop Group").

These Terms and Conditions for Use of Dataset (the "Agreement") are between IntelinAir, Inc. ("IntelinAir") and the individual or entity ("User") downloading and/or using the Dataset, and govern User's use of the Dataset.

BY DOWNLOADING OR USING THE DATASET, USER SIGNIFIES ITS AGREEMENT TO ALL TERMS, CONDITIONS, AND NOTICES CONTAINED OR REFERENCED IN THIS AGREEMENT. IF USER DOES NOT AGREE TO SUCH TERMS, CONDITIONS AND/OR NOTICES, USER MAY NOT DOWNLOAD OR USE THE DATASET.

User represents that it has full power, capacity, and authority to accept this Agreement. If an individual is accepting this Agreement on behalf of an employer or another entity: (a) such individual represents that such individual has full legal authority to bind such employer or such entity to this Agreement, and (b) such individual and such employer or entity shall both be bound by this Agreement as the User. If such individual is accepting on behalf of an employer or other entity and doesn't have the legal authority to bind, then neither such individual nor such employer or other entity may download or use the Dataset.

1. Downloading the Dataset. The dataset for the challenge can be found here.

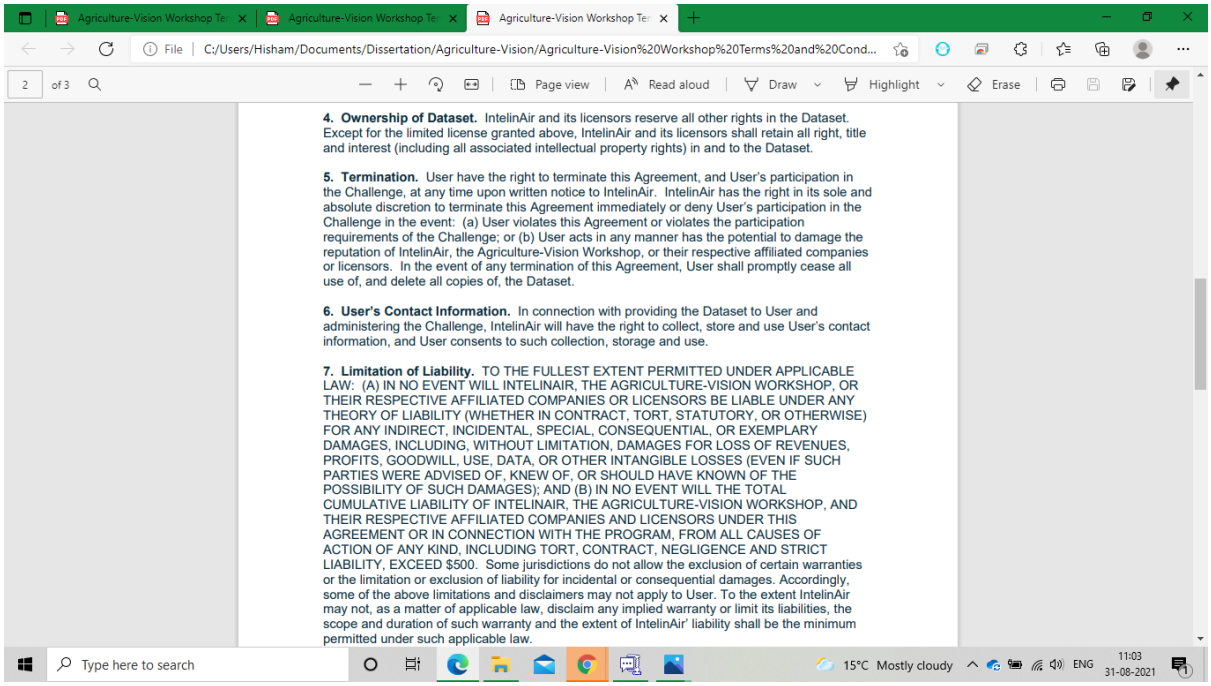


2. License to Use Dataset. Subject to the terms and conditions of this Agreement, IntelinAir grants User a non-exclusive, non-sublicensable, non-transferable, royalty-free, limited, revocable license to download and use the Dataset during the Usage Period (defined below) for the sole purpose of participating in the Challenge and creating a submission for the Challenge ("Submission").

3. Restrictions. User shall not, and shall not authorize or permit any third party to: (a) use the Dataset for any purpose other than internal, non-commercial academic purposes; (b) sell, license, transfer or redistribute the Dataset to any third party (except to research employees, assistants and colleagues who also agree to be bound by this Agreement); (c) use the Dataset to develop, or otherwise in connection with, any commercial product or service; or (d) create any derivative works from the dataset other than the Submission. In addition, User shall: (i) cite the original paper of the Agriculture-Vision Workshop in any publications in which User discusses the Dataset (User can find the attached bibliography for User's reference); and (ii) delete all copies of the Dataset within ninety (90) days after the completion date of the Challenge ("Usage Period"). In the event that User does provide the Dataset to any research employees, assistants or colleagues, User shall cause all such employees, assistants and colleagues to comply with the terms of this Agreement and shall be responsible for any breach of this Agreement caused by such employees, assistants or colleagues.

1

4. Ownership of Dataset. IntelinAir and its licensors reserve all other rights in the Dataset. Except for the limited license granted above, IntelinAir and its licensors shall retain all right, title and interest (including all associated intellectual property rights) in and to the Dataset.

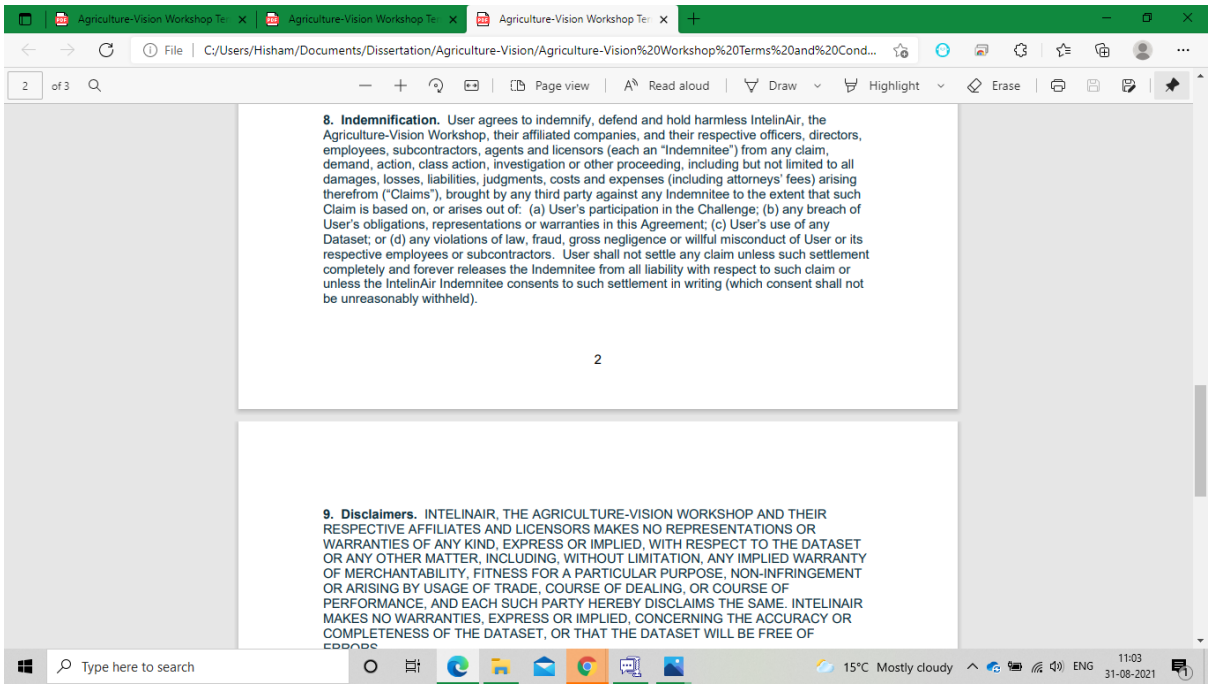


4. Ownership of Dataset. IntelinAir and its licensors reserve all other rights in the Dataset. Except for the limited license granted above, IntelinAir and its licensors shall retain all right, title and interest (including all associated intellectual property rights) in and to the Dataset.

5. Termination. User have the right to terminate this Agreement, and User's participation in the Challenge, at any time upon written notice to IntelinAir. IntelinAir has the right in its sole and absolute discretion to terminate this Agreement immediately or deny User's participation in the Challenge in the event: (a) User violates this Agreement or violates the participation requirements of the Challenge; or (b) User acts in any manner has the potential to damage the reputation of IntelinAir, the Agriculture-Vision Workshop, or their respective affiliated companies or licensors. In the event of any termination of this Agreement, User shall promptly cease all use of, and delete all copies of, the Dataset.

6. User's Contact Information. In connection with providing the Dataset to User and administering the Challenge, IntelinAir will have the right to collect, store and use User's contact information, and User consents to such collection, storage and use.

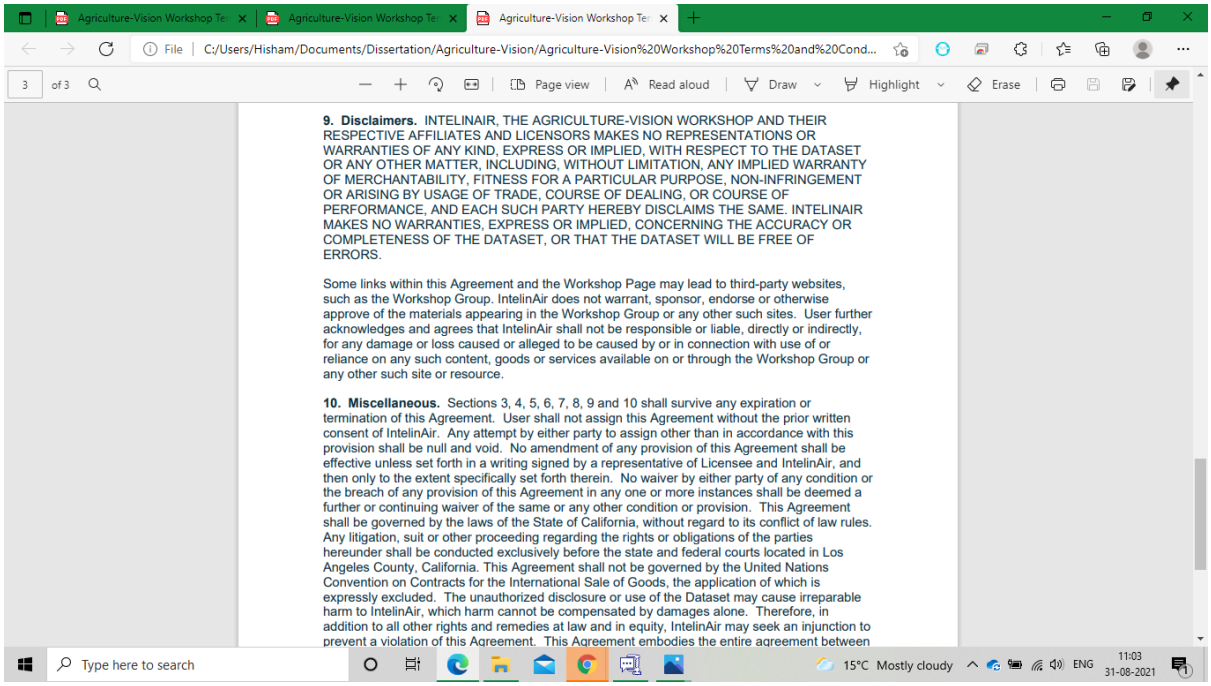
7. Limitation of Liability. TO THE FULLEST EXTENT PERMITTED UNDER APPLICABLE LAW: (A) IN NO EVENT WILL INTELINAIR, THE AGRICULTURE-VISION WORKSHOP, OR THEIR RESPECTIVE AFFILIATED COMPANIES OR LICENSORS BE LIABLE UNDER ANY THEORY OF LIABILITY (WHETHER IN CONTRACT, TORT, STATUTORY, OR OTHERWISE) FOR ANY INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL, OR EXEMPLARY DAMAGES, INCLUDING, WITHOUT LIMITATION, DAMAGES FOR LOSS OF REVENUES, PROFITS, GOODWILL, USE, DATA, OR OTHER INTANGIBLE LOSSES (EVEN IF SUCH PARTIES WERE ADVISED OF, KNEW OF, OR SHOULD HAVE KNOWN OF THE POSSIBILITY OF SUCH DAMAGES); AND (B) IN NO EVENT WILL THE TOTAL CUMULATIVE LIABILITY OF INTELINAIR, THE AGRICULTURE-VISION WORKSHOP, AND THEIR RESPECTIVE AFFILIATED COMPANIES AND LICENSORS UNDER THIS AGREEMENT OR IN CONNECTION WITH THE PROGRAM, FROM ALL CAUSES OF ACTION OF ANY KIND, INCLUDING TORT, CONTRACT, NEGLIGENCE AND STRICT LIABILITY, EXCEED $500. Some jurisdictions do not allow the exclusion of certain warranties or the limitation or exclusion of liability for incidental or consequential damages. Accordingly, some of the above limitations and disclaimers may not apply to User. To the extent IntelinAir may not, as a matter of applicable law, disclaim any implied warranty or limit its liabilities, the scope and duration of such warranty and the extent of IntelinAir' liability shall be the minimum permitted under such applicable law.



8. Indemnification. User agrees to indemnify, defend and hold harmless IntelinAir, the Agriculture-Vision Workshop, their affiliated companies, and their respective officers, directors, employees, subcontractors, agents and licensors (each an "Indemnitee") from any claim, demand, action, class action, investigation or other proceeding, including but not limited to all damages, losses, liabilities, judgments, costs and expenses (including attorneys' fees) arising therefrom ("Claims"), brought by any third party against any Indemnitee to the extent that such Claim is based on, or arises out of: (a) User's participation in the Challenge; (b) any breach of User's obligations, representations or warranties in this Agreement; (c) User's use of any Dataset; or (d) any violations of law, fraud, gross negligence or willful misconduct of User or its respective employees or subcontractors. User shall not settle any claim unless such settlement completely and forever releases the Indemnitee from all liability with respect to such claim or unless the IntelinAir Indemnitee consents to such settlement in writing (which consent shall not be unreasonably withheld).

2

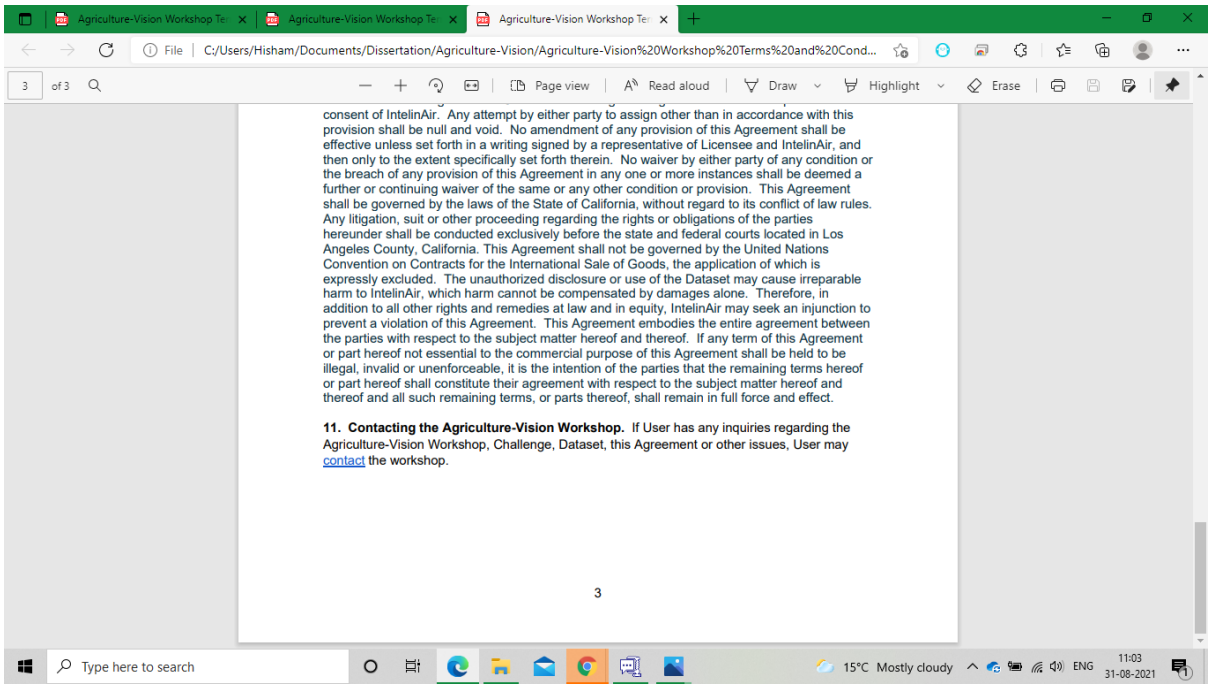
9. Disclaimers. INTELINAIR, THE AGRICULTURE-VISION WORKSHOP AND THEIR RESPECTIVE AFFILIATES AND LICENSORS MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, WITH RESPECT TO THE DATASET OR ANY OTHER MATTER, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT OR ARISING BY USAGE OF TRADE, COURSE OF DEALING, OR COURSE OF PERFORMANCE, AND EACH SUCH PARTY HEREBY DISCLAIMS THE SAME. INTELINAIR MAKES NO WARRANTIES, EXPRESS OR IMPLIED, CONCERNING THE ACCURACY OR COMPLETENESS OF THE DATASET, OR THAT THE DATASET WILL BE FREE OF



9. Disclaimers. INTELINAIR, THE AGRICULTURE-VISION WORKSHOP AND THEIR RESPECTIVE AFFILIATES AND LICENSORS MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, WITH RESPECT TO THE DATASET OR ANY OTHER MATTER, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT OR ARISING BY USAGE OF TRADE, COURSE OF DEALING, OR COURSE OF PERFORMANCE, AND EACH SUCH PARTY HEREBY DISCLAIMS THE SAME. INTELINAIR MAKES NO WARRANTIES, EXPRESS OR IMPLIED, CONCERNING THE ACCURACY OR COMPLETENESS OF THE DATASET, OR THAT THE DATASET WILL BE FREE OF ERRORS.

Some links within this Agreement and the Workshop Page may lead to third-party websites, such as the Workshop Group. IntelinAir does not warrant, sponsor, endorse or otherwise approve of the materials appearing in the Workshop Group or any other such sites. User further acknowledges and agrees that IntelinAir shall not be responsible or liable, directly or indirectly, for any damage or loss caused or alleged to be caused by or in connection with use of or reliance on any such content, goods or services available on or through the Workshop Group or any other such site or resource.

10. Miscellaneous. Sections 3, 4, 5, 6, 7, 8, 9 and 10 shall survive any expiration or termination of this Agreement. User shall not assign this Agreement without the prior written consent of IntelinAir. Any attempt by either party to assign other than in accordance with this provision shall be null and void. No amendment of any provision of this Agreement shall be effective unless set forth in a writing signed by a representative of Licensee and IntelinAir, and then only to the extent specifically set forth therein. No waiver by either party of any condition or the breach of any provision of this Agreement in any one or more instances shall be deemed a further or continuing waiver of the same or any other condition or provision. This Agreement shall be governed by the laws of the State of California, without regard to its conflict of law rules. Any litigation, suit or other proceeding regarding the rights or obligations of the parties hereunder shall be conducted exclusively before the state and federal courts located in Los Angeles County, California. This Agreement shall not be governed by the United Nations Convention on Contracts for the International Sale of Goods, the application of which is expressly excluded. The unauthorized disclosure or use of the Dataset may cause irreparable harm to IntelinAir, which harm cannot be compensated by damages alone. Therefore, in addition to all other rights and remedies at law and in equity, IntelinAir may seek an injunction to prevent a violation of this Agreement. This Agreement embodies the entire agreement between



consent of IntelinAir. Any attempt by either party to assign other than in accordance with this provision shall be null and void. No amendment of any provision of this Agreement shall be effective unless set forth in a writing signed by a representative of Licensee and IntelinAir, and then only to the extent specifically set forth therein. No waiver by either party of any condition or the breach of any provision of this Agreement in any one or more instances shall be deemed a further or continuing waiver of the same or any other condition or provision. This Agreement shall be governed by the laws of the State of California, without regard to its conflict of law rules. Any litigation, suit or other proceeding regarding the rights or obligations of the parties hereunder shall be conducted exclusively before the state and federal courts located in Los Angeles County, California. This Agreement shall not be governed by the United Nations Convention on Contracts for the International Sale of Goods, the application of which is expressly excluded. The unauthorized disclosure or use of the Dataset may cause irreparable harm to IntelinAir, which harm cannot be compensated by damages alone. Therefore, in addition to all other rights and remedies at law and in equity, IntelinAir may seek an injunction to prevent a violation of this Agreement. This Agreement embodies the entire agreement between the parties with respect to the subject matter hereof and thereof. If any term of this Agreement or part hereof not essential to the commercial purpose of this Agreement shall be held to be illegal, invalid or unenforceable, it is the intention of the parties that the remaining terms hereof or part hereof shall constitute their agreement with respect to the subject matter hereof and thereof and all such remaining terms, or parts thereof, shall remain in full force and effect.

11. Contacting the Agriculture-Vision Workshop. If User has any inquiries regarding the Agriculture-Vision Workshop, Challenge, Dataset, this Agreement or other issues, User may contact the workshop.

3