2021

CWDSS

(Cauliflower and Weed Detection using Semantic Segmentation)

by

Vishavjeet Singh Kainth ***vishavjeetsk.cs.18@nitj.ac.in***

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to **Professor Anil Kumar** for their able guidance and support in completing my project.

DATE:
01/07/2021

Vishavjeet Singh Kainth
(NITJ-18103096)

Table Of Contents

S.no.	Reports	Page No.
1	Introduction	4-7
2.	Cauliflower Data Set	7-11
3.	Convolutional Neural Network	11-12
4.	Project Outcomes and Results	13-16
5.	Conclusion and References	17-18

CWDSS

Cauliflower and Weed Detection using Semantic Segmentation

Abstract :

Crop Segmentation is an important task in Precision agriculture, where the use of aerial robots with an onboard camera has been contributed to the development of new solution alternatives. We address the problem of cauliflower plant segmentation in top-view RGB images of a crop grown under open-field difficult circumstances of complete lighting conditions and non-ideal crop maintenance practices defined by local farmers. We present a U-net Convolutional Neural Network (CNN) with an encoder-decoder that classify each pixel as cauliflower, weed and background using only raw color images as input. Our approach achieves a mean training accuracy of 96% and validation accuracy of 93% despite the complexity of background.

Keywords:

Convolutional Neural Network, Semantic Segmentation, Aerial Vehicles, Brassica oleracea var. botrytis.

1. Introduction:

Precision Agriculture or Smart Farming aims to increase crop yield, reduce production costs and decrease environmental impact. In this context, an active research area is to identify crops automatically in digital images to classify plants, to monitor its growth or to detect problems of water stress, nutrition or health in cultivated plants. This problem is complicated under open field cultivation due to different factors such as natural lighting, weather and agricultural practices of the farmers. The research carried out so far has been limited only to cases where the open field crops have small plants that are well separated from one another, the color of the soil with respect to the plants is very different and the overlap among leaves of the same plant occurs very rarely. Moreover, the state-of-the-art has focused on annual crops with careful cultivation techniques, without addressing the accurate segmentation case of cauliflower (Brassica oleracea var. botrytis.) perennial plants in an orchard where their particular characteristics cause complex image patterns.

1.1 UNMANNED AERIAL VEHICLES ::

Unmanned Aerial Vehicles (UAVs) (Fig-1) have many characteristics that make them attractive elements for precision agriculture. UAVs can continuously travel large tracts of cultivated land in a short time and they have the capacity to carry light-weight compact sensors to capture information at low altitude. The RGB (Red-Green-Blue) cameras are one of the most used sensors in UAVs since they are relatively cheap, have low energy consumption and are light. It is true that multi-spectral or thermal cameras have been extensively used to vegetation monitoring, however these cameras are more expensive compared to RGB.

Unmanned Aerial Vehicle (UAV)
Datalink
Avionics
Gimball Control
Flight Control Systems
Imaging
Radar

Figure 1: UAV

1.2 CONTRIBUTION OF THIS PROJECT ::

Considering the realistic difficult environment conditions, sometimes it becomes difficult to practice traditional farming procedures and also to monitor that if there is any crop attack or excessive weed growth etc.

We propose a model designed as per U-net convolutional neural network encoder-decoder architecture with 34 layers using ResNet-34, which classifies the image into crop, non-crop regions and weed, non-weed region.

1.3 WHY ARTIFICIAL NEURAL NETWORKS ::

The artificial neural networks are highly robust approximation functions, which when used with convolutional layers have set the state-of-the-art for dealing with different image-related tasks, it is reasonable to expect that they could be used to perform segmentation in such a challenging scenario as ours. Furthermore, an encoder-decoder architecture provides the tools required to map RGB images onto binary images corresponding to segmentation indicator. Thus, it is able to adapt complex functions through U-Nets with ResNet Encoders and cross connections [2].

1.4 U-Net ARCHITECTURE ::

The UNET was developed by Olaf Ronneberger et al. for Bio Medical Image Segmentation. The architecture contains two paths. First path is the contraction path (also called as the encoder) which is used to capture the context in the image. The encoder is just a traditional stack of convolutional and max pooling layers. The second path is the symmetric expanding path (also called as the decoder) which is used to enable precise localization using transposed convolutions. Thus it is an end-to-end fully convolutional network (FCN), i.e. it only contains Convolutional layers and does not contain any Dense layer because of which it can accept image of any size [5].

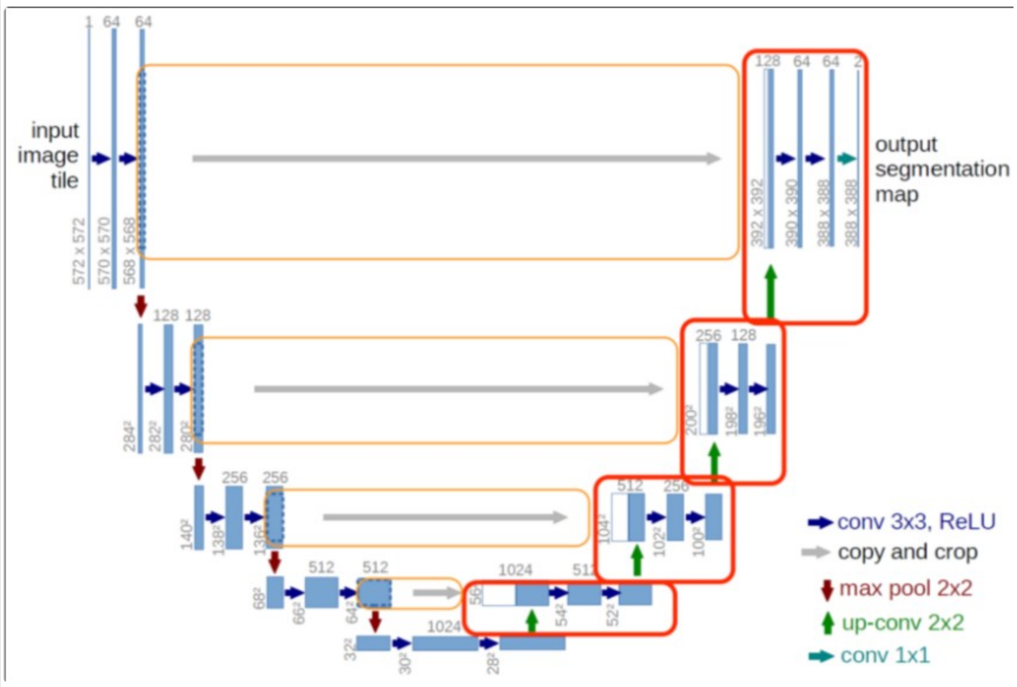


Figure 2: U-Net Architecture

1.5 RESNET (RESIDUAL NEURAL NETWORK) ::

In traditional neural networks, more layers mean a better network but because of the vanishing gradient problem, weights of the first layer won’t be updated correctly through the back-propagation. As the error gradient is back-propagated to earlier layers, repeated multiplication makes the gradient small. Thus, with more layers in the networks, its performance gets saturated and starts decreasing rapidly. Res-Net solves this problem by using the identity matrix. When the back-propagation is done through identity function, the gradient will be multiplied only by 1. This preserves the input and avoids any loss in the information.

ResNet uses a skip connection in which an original input is also added to the output of the convolution block. This helps in solving the problem of vanishing gradient by allowing an alternative path for the gradient to flow through. Also, they use identity function which helps higher layer to perform as good as a lower layer, and not worse.

2. Cauliflower Data-Set:

The raw data is gathered by our instructor’s acquaintance who collected the images by using aerial vehicle ,the RGB camera connected to the vehicle captures up to 16 megapixels images. The data was captured approximately 20 m of altitude above ground level. The plantation so done is in the form of rows of cauliflower as shown in the image [2].

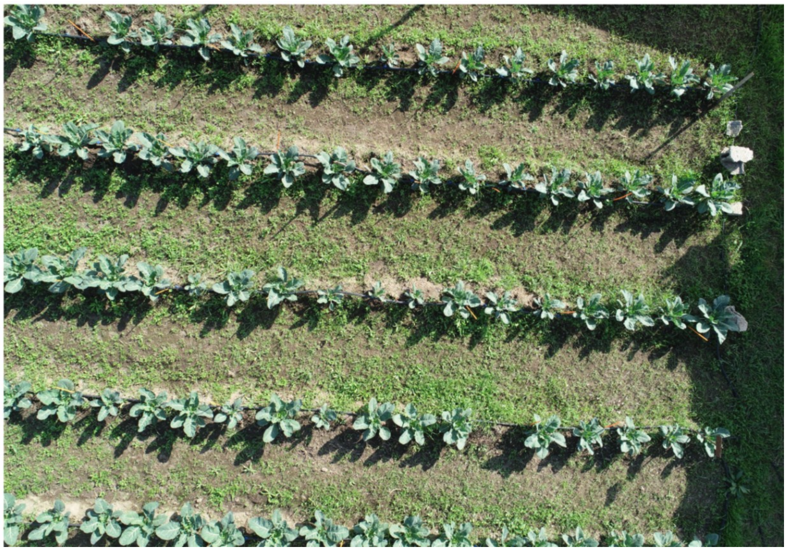


Figure 3: Rows of cauliflower so planted

This data set consist of a total of 30 RGB images. Images are geo-tagged and have a resolution of 5472 x 3648 pixels. Images contain several troubled areas, which we have classified in the following categories:

- Lighting: Plants cast shadows on the ground. Likewise, leaves closest to the ground are covered by shadows originated by the upper leaves. The lower leaves are prone to appear in a colour close to black, while some of the upper leaves of the trees tend to be almost white in colour.

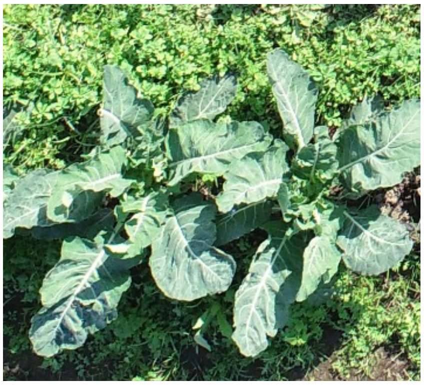


Figure 4: a) Leaves tends to be more whitish in color

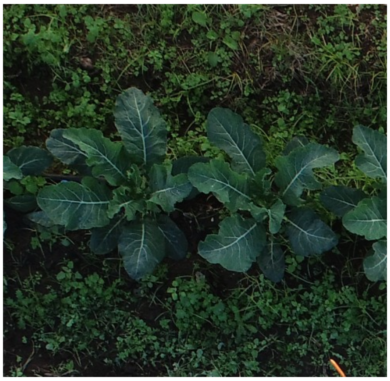


Figure 3: b) Leaves tends to blackish in color while being in shadow.

- Weeds: There is a mixture of broad and narrow leaved weeds on the soil. Also, dry grass is present in different areas of the field.

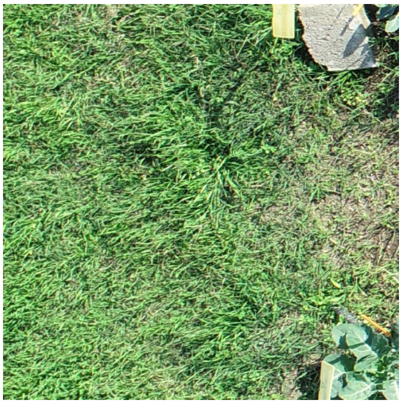


Figure 5: a) Area of the field covered by grass

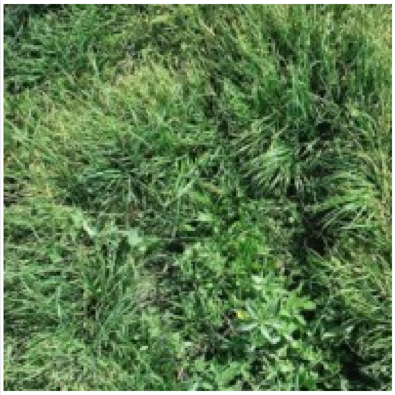


Figure 4: b) Weed

- Camouflaged plants: There are cases where it is difficult to decide whether a pixel belongs to a part of cauliflower plant or not. This situation arises when the cauliflower leaves are on top of a background where green weeds are predominant on soil.

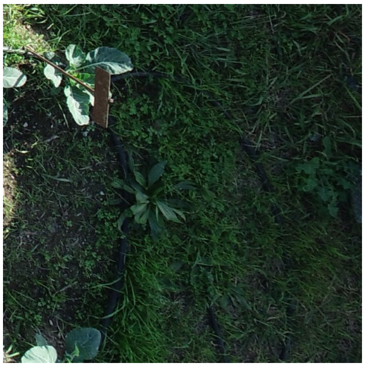


Figure 6: A shaded Camouflaged plant that appears to be cauliflower but actually it is not.

- Residues: Residues include stones, dry branches, objects used by farmers or anything else that is not of interest in crop detection.

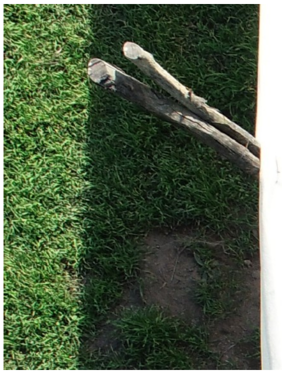


Figure 7: a) Stick used to hold the the white sheet

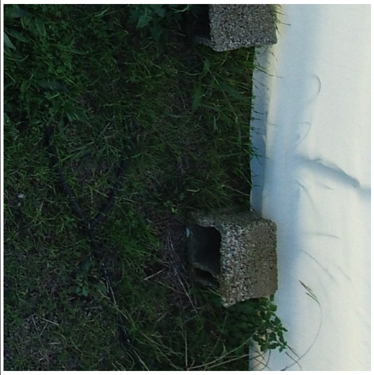


Figure 8: b) A large concrete piece and a white sheet

2.1 IMAGE DIVISION AND RESIZING ::

In our proposed data set, there are only 8 labeled images of 5472 x 3648 pixels. However, a cauliflower leaf in these images is represented, on average, by a region of 100 x 100 pixels. Therefore, they contain a large number of leaves samples subjected to different conditions, with which it is possible to carry out training of our CNN without the need to resort to data augmentation techniques. The image is divided into patches of 912 x 912 pixels along horizontal and vertical axis, generating more input data (Fig-8 and 9). We observe that as the size of the patch increases, the performance improves. Nevertheless, larger patches involve more processing time and the problem of getting a small number of patches per image, that's why we have to further resize the patch size to 128 so that server does not crashes while eorking with the data. Finally, we work with a total of 150 patches with their
respective masks.

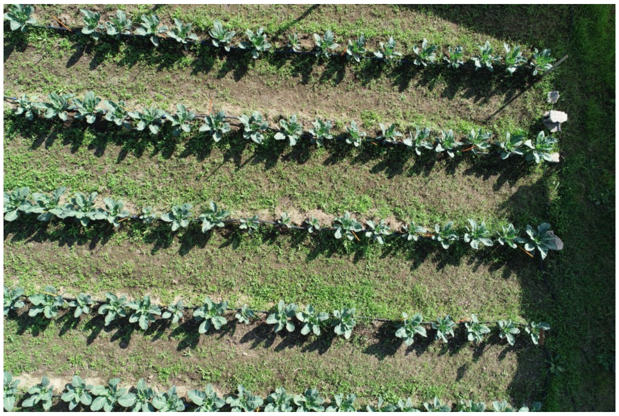


Figure 9: Original Image

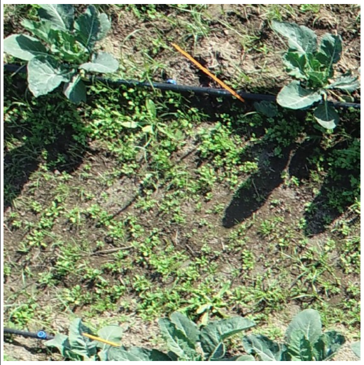


Figure 10: One patch of image by dividing original image into 6x4 folds

2.2 IMAGE ANNOTATION ::

In order to train the model we need corresponding masks for each training image. Thus to achieve this purpose I’ve used Image Annotation tool by 4 Smart Machines [3] (Fig-10). The tool provides output masks in the form of overlays , png, json, etc..

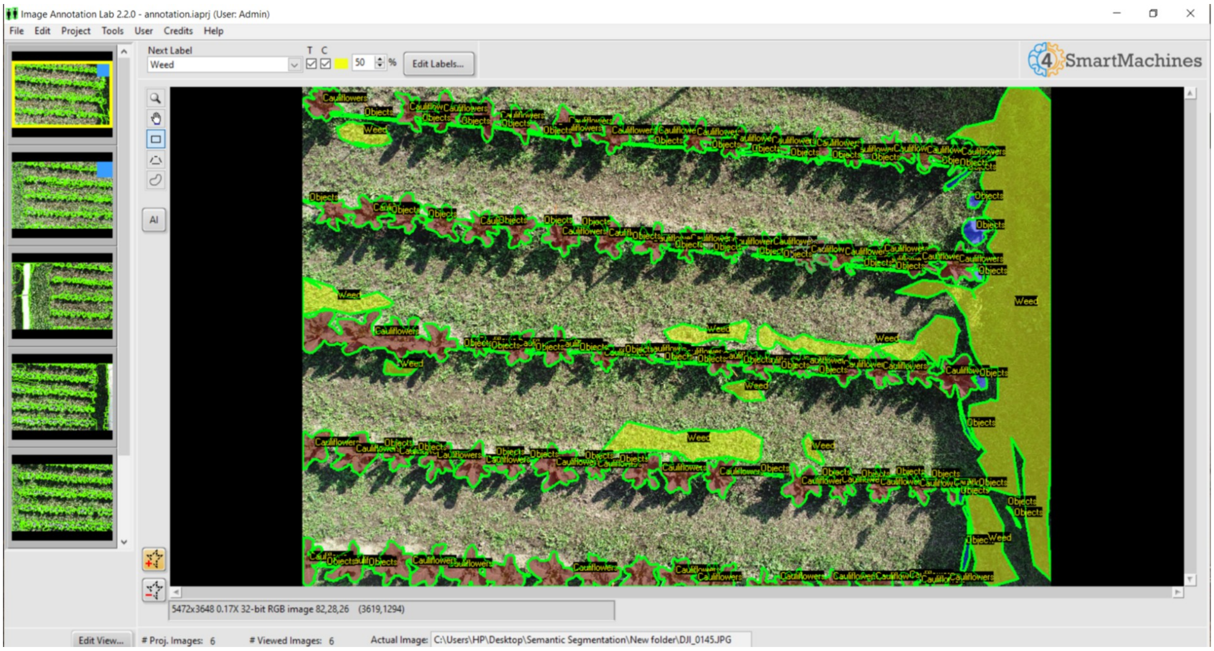


Image Annotation Lab 2.2.0 - annotation.iaprj (User: Admin)
File Edit Project Tools User Credits Help
Next Label
T C
%
Edit Labels...
SmartMachines
AI
Cauliflowers
Objects
Weed
Edit View...

Next Label: Weed ☑ ☑ 50 %

5472x3648 0.17X 32-bit RGB image 82,28,26 (3619,1294)

Proj. Images: 6 # Viewed Images: 6 Actual Image: C:\Users\HP\Desktop\Semantic Segmentation\New folder\DJI_0145.JPG

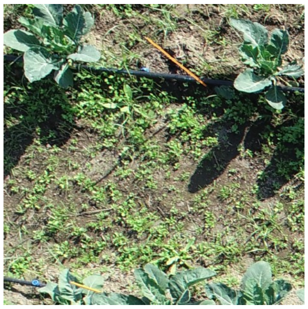
Figure 11: Image Annotation tool used for Labeling the image

3. Convolutional Neural Network :

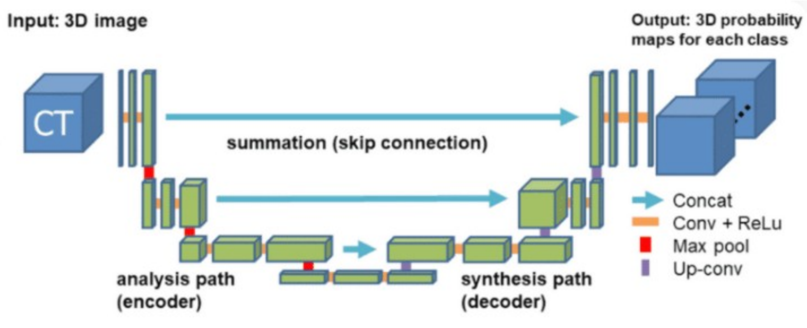
In top-view images of a cauliflower crop most of the leaves are overlapped and present different tonality due to the sunlight and shadows. Also, the leaves can be camouflaged with the weed. Thus, with an approach based on hand-engineered features, the expected result could hardly be obtained. On the contrary, it has been proven that a CNN has the capability to discover effective representations of complex scenes in order to perform good discrimination in different Computer Vision tasks with large image repositories. For these reasons, we decided to explore the CNN models to classify the pixels into crop or non-crop classes and into weed or non-weed classes in order to perform a crop segmentation. In this section, we describe the CNN architecture for the segmentation of cauliflower plants.

3.1 APPROACH::

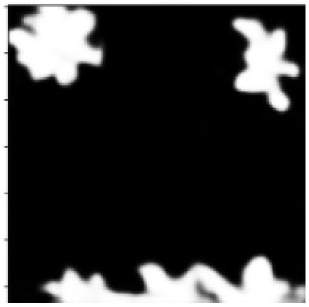
- Our CNN model is inspired by standard U-Net architecture. The architecture framework is used from the semantic segmentation package. Along with this we have used resnet-34 backbone, in order to cope up with vanishing gradient problem.
- U-Net architecture helps to identify the image patch as a whole (classification) and where is the concerned image (localization).
- The encoder phase of U-Net architecture helps for the classification purpose and the decoder phase helps in localization. There are total 34 layers in resnet-34 backbone including the pooling convolutional layers and decoder layers.
- In order to go for perfect loss function, "Binary Cross Entropy Loss"[8] stands a chance apart. 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. In our case 0 stands for the background and 1 stands for the crop (cauliflower) and in case of Weed model 1 stands for weed.
- Thus this model basically provides us the probabilities that how much chances are there that a particular pixel would belong to a particular class.
- To plot the output, I've used cmap as gray, as it is best suited for a binary image.



Input



Network Architecture



Output

For calculating that how much content the image patch contains as of cauliflower or weed, I've also defined a function that basically calculates the proportion of weed and cauliflower in the same patch.

4. Project Outcomes and Results:

To find out the impact of image division on model accuracy, I’ve simply compared the outputs of the test images with their corresponding outputs, and the result is as follows::

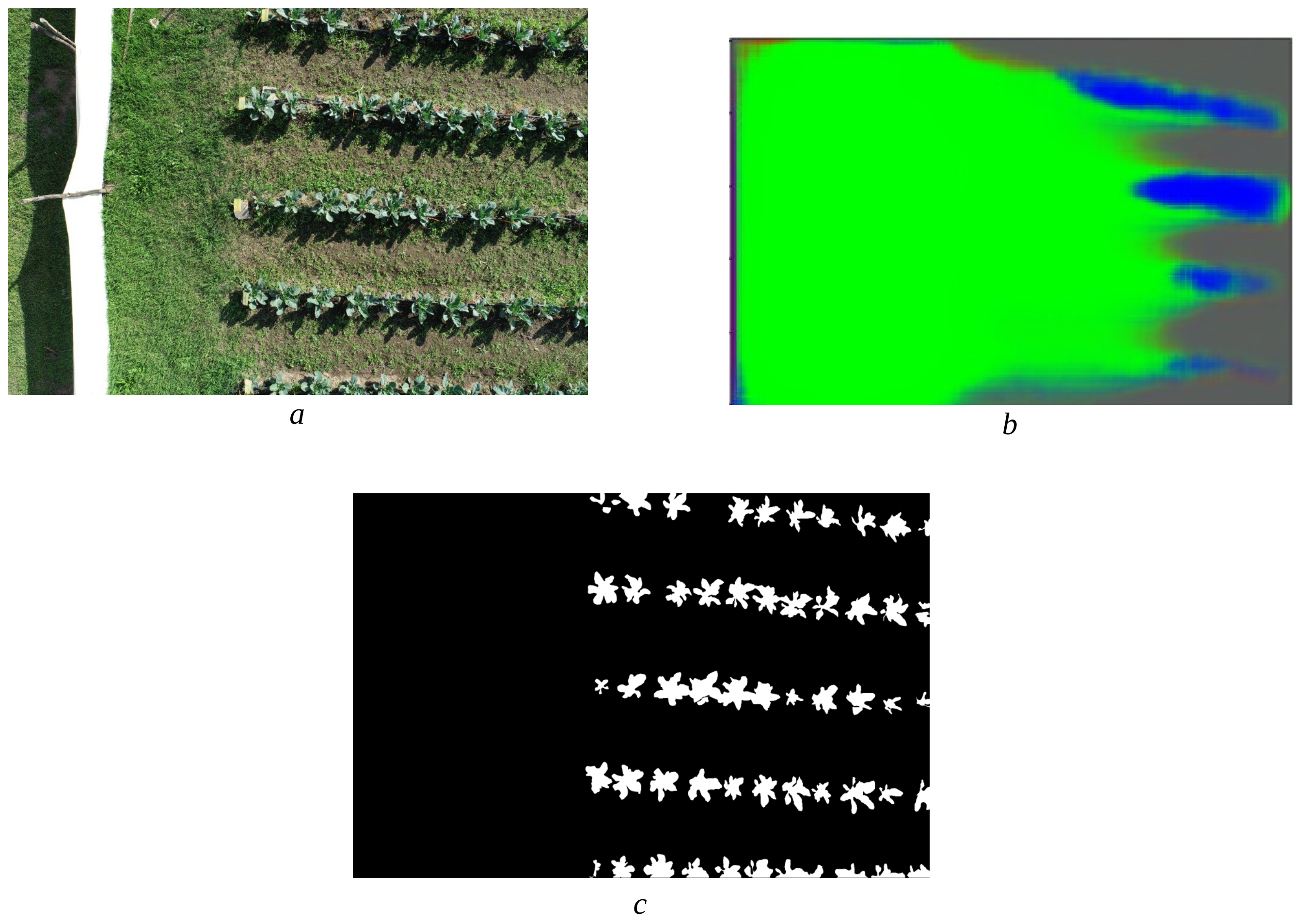
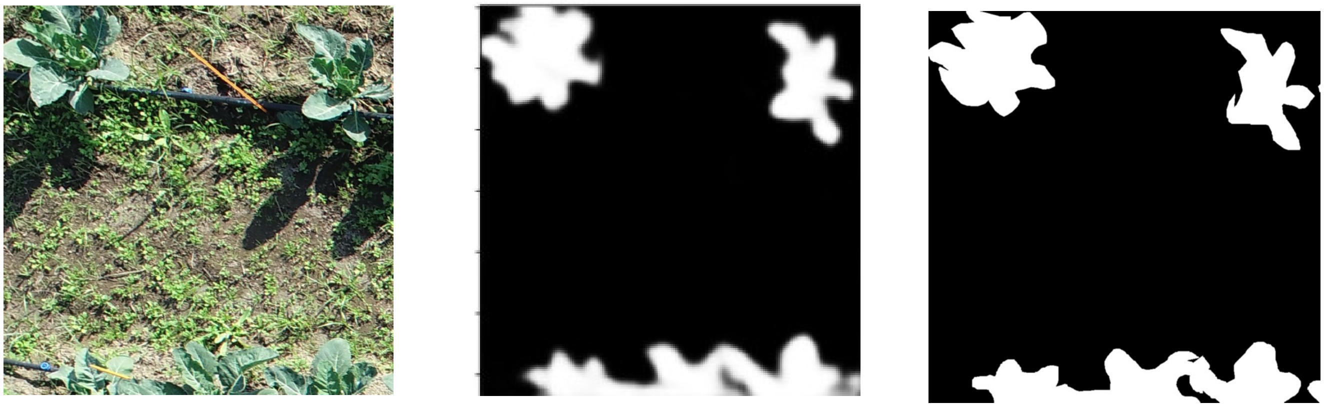


Fig a) The Input image (Without Division) to test the model b) The corresponding output mask c) The ground Truth value of the input image(Expected/Actual mask)

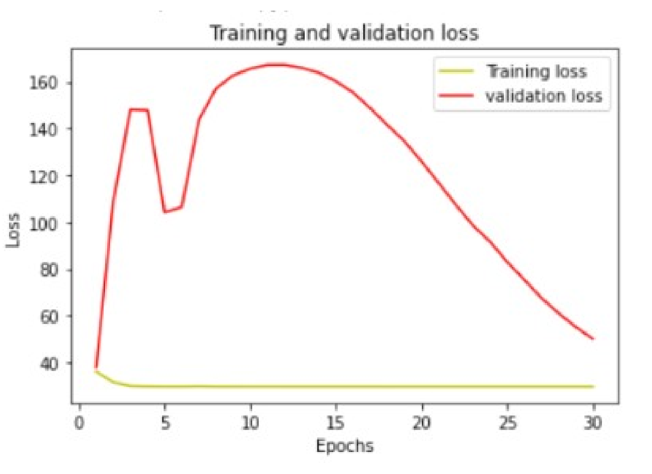


a *b* *c*

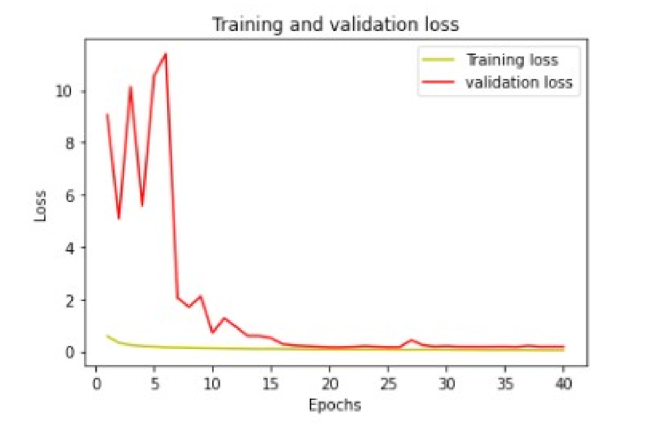
Fig a) The Input image (With Division) to test the model b) The corresponding output mask c) The ground Truth value of the input image(Expected/Actual mask)

Clearly, we can see that the later model outperforms in terms of accuracy and cauliflower detection. This is so because, the model is able to drive more information from the same data as now the size of leaf is comparatively larger in the later case, and thus easy to detect by the model trained on the same sub-images.

Given below are the graphical representations of how our model trains for both above mentioned cases. The graph shows that how validation and training loss decreases through each iteration.



a



b

In above diagram Fig a) corresponds to the model which is trained by using original images and Fig b) corresponds to the model which is trained by using Sub-Images obtained by dividing original images into equal parts.

The performance of weed detection model is shown as below::

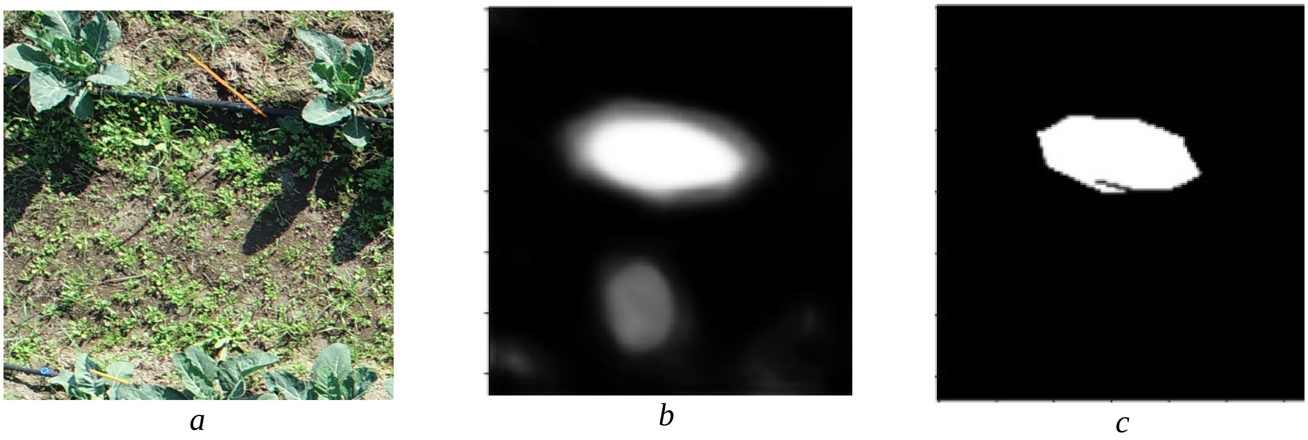


Fig a) The Input image (With Division) to test the model b) The corresponding output mask for weed c) The ground Truth value of the input image(Expected/Actual mask)

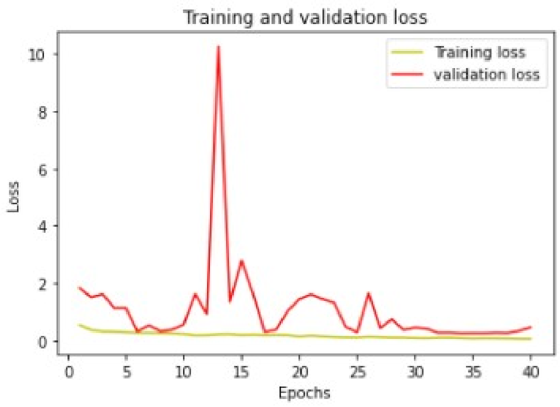


Figure 12: Training and validation loss v/s epochs (No. of iterations)

4.1 ACCURACY ::

The max-accuracy of Cauliflower Detection model and Weed Detection model after 40 iterations is as follows::

Models	Training Accuracy	Validation Accuracy
Cauliflower Detection Model	98.37	94.00
Weed Detection Model	97.97	92.02

We can see that the training accuracy is more than validation accuracy, this means that our model has overfitted the training data. But this overfitting can be controlled by training the same model using more number of instances but less epochs.

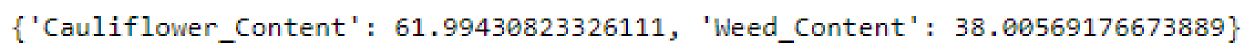
4.2 CONTENT CALCULATION ::

In order to specify relative cauliflower and weed content, I’ve used simple formula as follows::

$$\text{Cauliflower Content} = ((C)/(W+C))*100$$

where C and W stands for the sum of probabilities of each pixel that it belongs to cauliflower and weed respectively.

For the instance image given in figure 10 we have the following output::



```
{'Cauliflower_Content': 61.99430823326111, 'Weed_Content': 38.00569176673889}
```

Figure 13: Cauliflower and Weed content in Sub-Image

5. Conclusion::

We have proposed a cauliflower plant segmentation model based on deep learning and a challenging data set with its ground truth labeled by hand at the pixel level. The data set is of particular interest to smart farming and computer vision researchers. It consists of 30 high resolution aerial images captured by a UAV.
Our approach was based on a U-Net model with an resnnet-34 encoder architecture trained end-to-end. The experimental results showed that our model can be trained in just about 50 min while maintaining its ability to accurately segment the cauliflower plants and weed. The U-Net-based method is adequate to deal with the two-class segmentation problem, even in highly challenging scenarios such as the segmentation of Cauliflower plants and Weed as introduced in this work.

References ::

[1] A 2019 Guide to Semantic Segmentation. Available online: https://heartbeat.fritz.ai/a-2019-guide-to-semantic-segmentation-ca8242f5a7fc (By: *Derrick Mwiti*)

[2] U-Nets with ResNet Encoders and cross connections. Available online: https://towardsdatascience.com/u-nets-with-resnet-encoders-and-cross-connections-d8ba94125a2c (By: *Christopher Thomas BSc Hons. MIAP*)

[3] Image annotation lab by 4Smart Machines. Available online: (https://ial.4smartmachines.com/?gclid=CjwKCAjwz_WGBhA1EiwAUAxIcVZuoiuITKqBgZWzoJoUTq-zMg3xsJ6Vxf2gA1Vb32rFeW2ZIGYJjRoCjrAQAvD_BwE)

[4] My experiment with UNet – building an image segmentation model. Available online(https://analyticsindiamag.com/my-experiment-with-unet-building-an-image-segmentation-model/)

[5] Understanding Semantic Segmentation with UNET. Available online: https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47 (By: *Harshall Lamba*)

[6] Introduction to U-Net and Res-Net for Image Segmentation. Available online: https://aditi-mittal.medium.com/introduction-to-u-net-and-res-net-for-image-segmentation-9afcb432ee2f (By: *Aditi Mittal*)

[7] Segmentation models. Available online: https://segmentation-models.readthedocs.io/en/latest/tutorial.html

[8] Understanding binary cross-entropy / log loss: a visual explanation. Available online: https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a (By: *Daniel Godoy*)

[9] Fig Plant Segmentation from Aerial Images Using a Deep Convolutional Encoder-Decoder Network. Available online: https://www.mdpi.com/2072-4292/11/10/1157 (By: *J. Fuentes-Pacheco 1 , J. Torres-Olivares 2 , E. Roman-Rangel 3 , S. Cervantes 4 , P. Juarez-Lopez 5 , J. Hermosillo-Valadez 6 and J.M. Rendón-Mancha 6.*)