Good morning. My name is Pham Gia Phuc.maham Welcome and thank you for coming to my bachelor internship defense on the topic of Crop/Weed Analysis on UAV Images.

(This internship study was conducted at USTH ICTLab under the supervision of Dr. Tran Giang Son. Before starting, I would like to express my sincere thanks to Dr. Tran Giang Son for his support and help during the implementation of this study.)

The presentation contains 5 parts: Introduction, Background, Materials and Methods, Results and discussion and Conclusion.

So, to begin with, crop and weed are the terms which related to the agriculture. Agriculture is the practice of cultivating plants and livestock. It is the key development in the rise of the human civilization, which created food surpluses that enabled people to live in cities and ensure the quality of human life. On the other hand, after years of rapid development, technology, especially information technology, nowadays, plays a really important role in all aspects of life, including agriculture. And so, the term Precision Agriculture was born. Precision agriculture is an approach to farm management, which uses information technology to ensure the crop and soil receive exactly what they need for optimum health and productivity, thereby maximizing the profitability, sustainability and protection of the environment. But in order to be effective, precision agriculture needs a steady amount of data. And this is where UAVs come into play.

Unmanned Aerial Vehicle or UAV, also known as drone, is an aircraft without any human pilot, crew or passengers on board. UAVs also are a component of an Unmanned Aircraft System, which includes adding a ground-base controller and a system of communications with. In precision agriculture, the UAV provides more up-to-date information than traditional methods, and allowing even greater precision. Compares to other tools like satellite or unmanned ground vehicles, UAV usually has high spatial resolution, be able to collect data all time whenever it is needed, not sensitive to clouds with accurate position and be able to access to the difficult areas.

However, only raw data from UAV is not enough effective but needs to be processed. And the best way for this problem is using deep learning model, which the potential approach is semantic segmentation.

To handle the problem, in this internship, I have studied about precision agriculture, research in semantic segmentation, study about developing a deep learning model for weed mapping that utilizes multispectral images and implemented a model which might be able to assist scientists and farmers in deciding which algorithms to employ to accurately analyzing different crop/weed locations.

Well, about the background, the first thing to know is Semantic Segmentation. It is a computer vision task which specific regions of an image will be labelled. It is aimed to assign a class to each pixel in an image, that corresponds to the concept being represented. This process is known as dense prediction since every pixel in the image is forecasted. One crucial point to keep in mind is that it only considers the category of each pixel and not instances of the same class. In other words, the segmentation map does not automatically differentiate two objects of the same category in the input image as independent ones. It is useful for a variety of tasks, including Precision Agriculture.

Besides, the term Convolution Operation, which is the most important topic in the field of image processing, is the process of transforming an image, which applying a kernel over each pixel and its local neighbors across the entire image. In this operation, two arrays can be merged by multiplying them, which could be of different sizes, but the dimensions should be the same.

So close, by choosing the maximum pixel value from each 2x2 block of the input feature map, the term Max Pooling operation makes the feature map smaller so that the network has fewer parameters. This aims to eliminate unnecessary information and save only the most crucial details from each zone.

Together with convolution, this kind of operation is called down-sampling.

On the other hand, Semantic Segmentation produces more than just a class label or a set of bounding box parameters. In actuality, the output is a comprehensive, high-resolution image with pixel classification. So, in order to recover the information, it is necessary to up sample the image, or change a low-resolution to high-resolution image. This is where Transposed Convolution enters the picture.

Transposed Convolution is a technique for performing up-sampling of an image with learnable parameters. In essence, it is the exact reverse of a conventional convolution to reproduce high-resolution images. This operation is called up-sampling.

It is now ok to turn to chapter three, materials and methods.

In this internship, I have researched about 2018 Weed Map dataset. This dataset is public in ASL datasets, which being gathered from sugar beet field in Switzerland and Germany by two quadcopter UAV. The original dataset is divided into 129 directories with 18746 png pictures taken in various spectral.

Aside from, this study model is processed with the ICTLab infrastructure with some important information shown in the table on slide.

However, due to the limitation of this internship, I have filtered some data from original 2018 Weed Map for only RGB images and images from mask and groundtruth directories, which some examples are shown on the slide. The new dataset is stored into 3 directories and 1669 files as shown on the slide.

To implement the model, I have researched UNet. Unet is first introduced for biological image segmentation. With a tweaked and extended architecture that allows it to operate with less training photos and provide more accurate segmentation, it is based on fully convolutional neural networks. By downsampling, copy and crop, max pooling and upsampling, the goal of segmentation can be achieved.

In the training process, to evaluate the effectiveness of the model, I have used Intersection over Union metrics or IoU, which also known as Jaccard index. This metric is used to measure the accuracy of an object detector on a particular dataset. In this metric, the predicted and groundtruth mask overlap acts as both the nominator and the denominator in the equation. By dividing them, the IoU is determined.

IoU ranges from 0 to 1, the higher value, the more precise the model is.

The model is then trained with 100 epochs with 10 steps per each epoch, the batch size is 32, early stopping after 10 continuous epochs if validation loss does not improve, reduce the learning rate when IoU has stopped improving after 5 epochs. Adam is used to optimize the model.

After training, I have got the result of 76.79% for IoU and 74.95% for dice coefficient score. As shown in the figures, the scores are grown up gradually from around 0.5%. In the opposite site, the losses decreased from more than 90% to lower than 30%. This means that the model is not over or underfitting.

To ensure the performance of the model, I have tested it with only RGB image then compare it to the ground truth image. The results are varied from 62% to 86% due to the difference of details in the images. However, average result is still more than 70%.

I have also compared the proposed model with some similar article as shown on the slide, and I believe that the proposed model is quite good.

In conclusion, in this internship, I have completed the objective, studied about precision agriculture and semantic segmentation, research and implement the work of weed segmentation in UAV images, and created a model which achieve about 76% accuracy.

In the future…

That ‘s the end of my defense presentation. Thanks.