Semantic Weed Mapping Using Multispectral Imaging and Deep Neural Network for Precision Farming

GEXXX Development Engineering Project Report

Gagan Singh 2017eeb1141
Lakshya 2017eeb1149
Sibee Sanchay 2017eeb1170
Vinith Reddy 2017eeb1176
Yajurmani Sharma 2017eeb1180

June 21, 2020

Abstract

In this report we discuss the development of a Deep Neural Network that would facilitate in mapping the locations of weeds(unwanted plants) in an organised agricultural setting. Such mappings are useful in the development of precision farming techniques such as selective herbicide and stomp machines. These developments present an outlook into the future of automated farming.

1 Introduction

We chose [Sa+18b] as our base paper for this report. We intend to present our key observations, insights and learnings from this literature. Weed identification and mapping is one of the most discussed problems in the techno-agricultural sector. This paper will help us understand the verbiage and current state of the art Deep Learning techniques involved in solving such problems.

2 Method

The author of [Sa+18b] used UAV(unmanned aerial vehicles) to map the farm land. The UAVs will have specific spectral cameras attached to them to capture different frequency ranges of radiation reflectance from the ground. The UAVs are also equipped with GPS(global positioning system) and INS(inertial navigation system) to keep track of multiple flights and later help in mapping images from same location.

2.1 Why UAVs?

UAVs industry has been picking up in remote sensing industires because of their agility, because they are more affordable than ever before, have inbuilt tracking systems, can be modularised and equipped with camera(s) of choice.

2.2 The non trivialities of the problem

As the altitude of the UAVs increases, so does the physical distance between the samples on ground, that is ,each pixel in the image will be a sample of points on the ground whose distance will keep on increasing with the altitude. DNNs result in downsampling of the data due to presence of pooling layers. This is an issue as the features that we're interested in are very small in sizze and will be lost because of this downsampling. The multispectral images thus captured will need to be aligned with each other so as to make them an effective multidimensional input for our DNNs. The lateral shift in perspective will have to be accounted for.

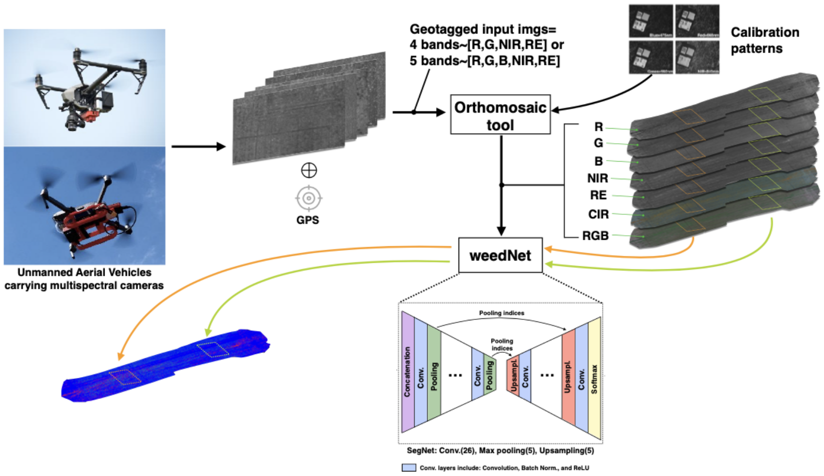


Figure 1: Overall processing pipeline. GPS tagged multispectral images are first collected by multiple UAVs and then passed to an orthomosaic tool with images for radiometric calibration. Multi-channel and aligned orthomosaic images are then tiled into a small portion (480 × 360 pixels, as indicated by the orange and green boxes) for subsequent segmentation with a DNN. This operation is repeated in a sliding window manner until the entire orthomosaic map is covered.[Sa+18b]

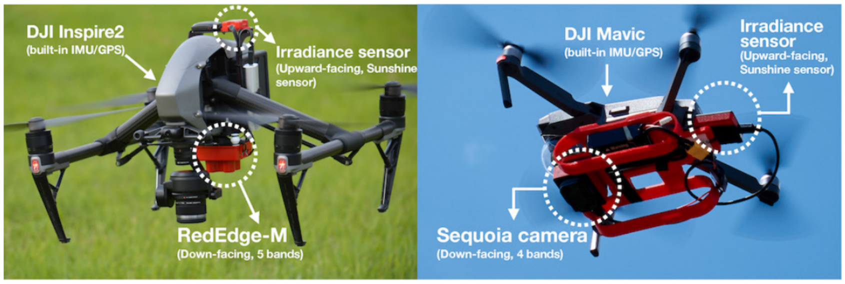


Figure 2: Multispectral cameras and irradiance (Sunshine) sensors' configuration. Both cameras are facing-down with respect to the drone body and irradiance sensors are facing-up.[Sa+18b]

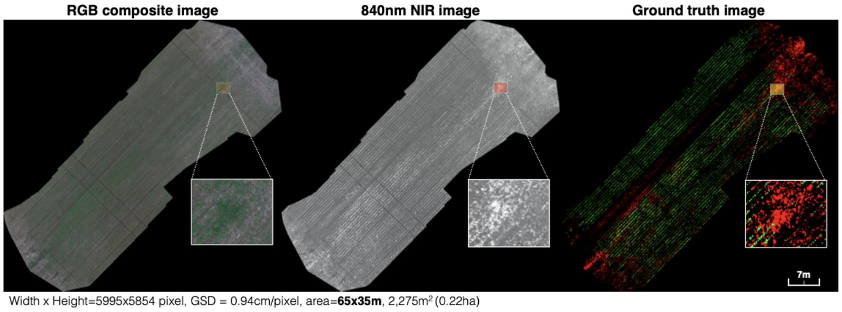

RGB composite image
840nm NIR image
Ground truth image
7m
Width x Height=5995x5854 pixel, GSD = 0.94cm/pixel, area=65x35m, 2,275m² (0.22ha)

Figure 3: An example of the orthomosaic maps used in [Sa+18b]. Left, middle and right are RGB (red, green, and blue). composite, near-infrared (NIR) and manually labeled ground truth (crop = green, weed = red) images with their zoomed-in views. We present these images in order to provide an intuition of the scale of the sugar beet field and quality of data used in this paper.

2.3 How these issues are handled

The author employs a sliding window approach. The final constructed Orthomosaic map is divided into windows whose dimensions in pixels are those of the input of the DNNs.
A point cloud projector will help in conjunction to measure the distance and topography of the region. An algorithm can map the points projected by this projector to achieve this. This data is useful when generating the Orthomosaic maps as they relay the altitude data.

2.4 What are Orthomosaic maps?

Orthomosaic maps are maps generated by stictching together multiple smaller image tiles. The differ from the usual image stitching because unlike that, the content of the orthomosaic maps are supposed to be in a single plane.
The images are pre processed for perspective using keypoint detection and later are stiched together while maintaining the finer details. As mentioned earlier, the pre-processing utilises the point cloud projector topography estimation.

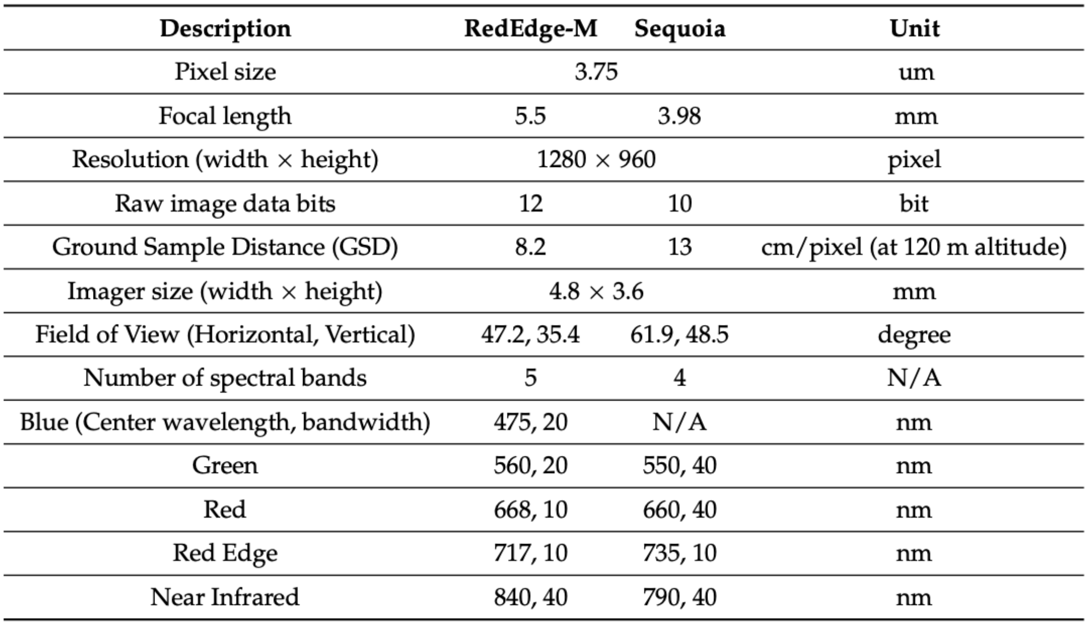
2.5 Advantages of Orthomosaic maps

Orthomosaic maps preserve the etric scale of their content. The precise alignment of multiple spectral samples enable the data to be ready as input to DNNs as multichannel image.

3 Technical Details

3.1 Spectrums of interest

Infrared regions are widely used in visualizing vegetation in remote sensing problems. For our application, the author decided to go with these specific spectrums.



Description	RedEdge-M	Sequoia	Unit
Pixel size	3.75		um
Focal length	5.5	3.98	mm
Resolution (width × height)	1280 × 960		pixel
Raw image data bits	12	10	bit
Ground Sample Distance (GSD)	8.2	13	cm/pixel (at 120 m altitude)
Imager size (width × height)	4.8 × 3.6		mm
Field of View (Horizontal, Vertical)	47.2, 35.4	61.9, 48.5	degree
Number of spectral bands	5	4	N/A
Blue (Center wavelength, bandwidth)	475, 20	N/A	nm
Green	560, 20	550, 40	nm
Red	668, 10	660, 40	nm
Red Edge	717, 10	735, 10	nm
Near Infrared	840, 40	790, 40	nm

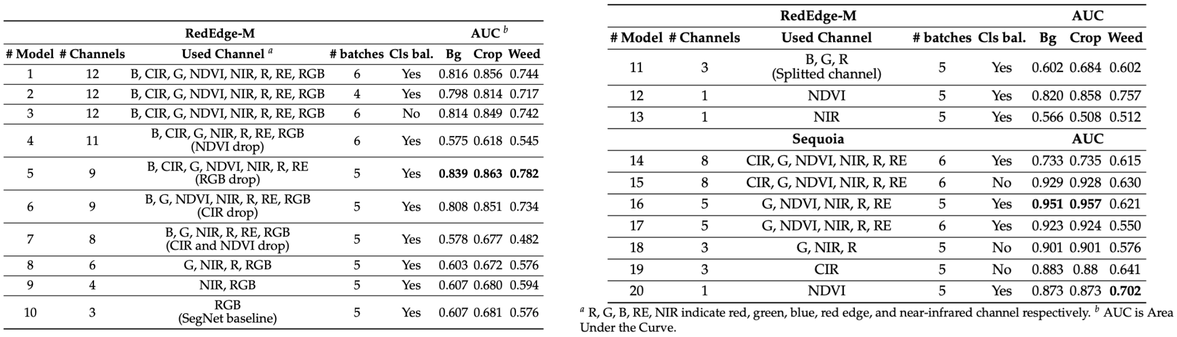
Figure 4: Multispectral camera sensors specifications used in this paper.[Sa+18b]

The channels were combined in these ways:

$$\begin{aligned}\textbf{Individual} &: Red, Green, Blue \\ \textbf{Combined} &: RGB = Red + Green + blue \\ &: ColorinfraredCIR = R + G + NIR \\ \textbf{Normalized Difference Vegitation Index} &: NDVI = \frac{NIR - R}{NIR + R}\end{aligned}$$

Reflectance maps account for pixel value captured and calibrates for camera exposure, gain, black level and vignetting. This process makes the final Orthomosaic map homogenous(i.e. individual tiles cannot be segmented). To obtain highest quality mosaics, we have to take in account the lighting conditions (partial cloud coverage, overcast etc). For this, the UAVs are also equipped with upward facing sunlight sensors that measure sun's orientation and irradience.

4 Experimental Results



		RedEdge-M			AUC [b]		
# Model	# Channels	Used Channel [a]	# batches	Cls bal.	Bg	Crop	Weed
1	12	B, CIR, G, NDVI, NIR, R, RE, RGB	6	Yes	0.816	0.856	0.744
2	12	B, CIR, G, NDVI, NIR, R, RE, RGB	4	Yes	0.798	0.814	0.717
3	12	B, CIR, G, NDVI, NIR, R, RE, RGB	6	No	0.814	0.849	0.742
4	11	B, CIR, G, NIR, R, RE, RGB (NDVI drop)	6	Yes	0.575	0.618	0.545
5	9	B, CIR, G, NDVI, NIR, R, RE (RGB drop)	5	Yes	**0.839**	**0.863**	**0.782**
6	9	B, G, NDVI, NIR, R, RE, RGB (CIR drop)	5	Yes	0.808	0.851	0.734
7	8	B, G, NIR, R, RE, RGB (CIR and NDVI drop)	5	Yes	0.578	0.677	0.482
8	6	G, NIR, R, RGB	5	Yes	0.603	0.672	0.576
9	4	NIR, RGB	5	Yes	0.607	0.680	0.594
10	3	RGB (SegNet baseline)	5	Yes	0.607	0.681	0.576

		RedEdge-M			AUC		
# Model	# Channels	Used Channel	# batches	Cls bal.	Bg	Crop	Weed
11	3	B, G, R (Splitted channel)	5	Yes	0.602	0.684	0.602
12	1	NDVI	5	Yes	0.820	0.858	0.757
13	1	NIR	5	Yes	0.566	0.508	0.512
		Sequoia			AUC		
14	8	CIR, G, NDVI, NIR, R, RE	6	Yes	0.733	0.735	0.615
15	8	CIR, G, NDVI, NIR, R, RE	6	No	0.929	0.928	0.630
16	5	G, NDVI, NIR, R, RE	5	Yes	**0.951**	**0.957**	0.621
17	5	G, NDVI, NIR, R, RE	6	Yes	0.923	0.924	0.550
18	3	G, NIR, R	5	No	0.901	0.901	0.576
19	3	CIR	5	No	0.883	0.88	0.641
20	1	NDVI	5	Yes	0.873	0.873	**0.702**

[a] R, G, B, RE, NIR indicate red, green, blue, red edge, and near-infrared channel respectively. [b] AUC is Area Under the Curve.

Figure 5: Performance evaluation summary for the two cameras with varying input channels.[Sa+18b]

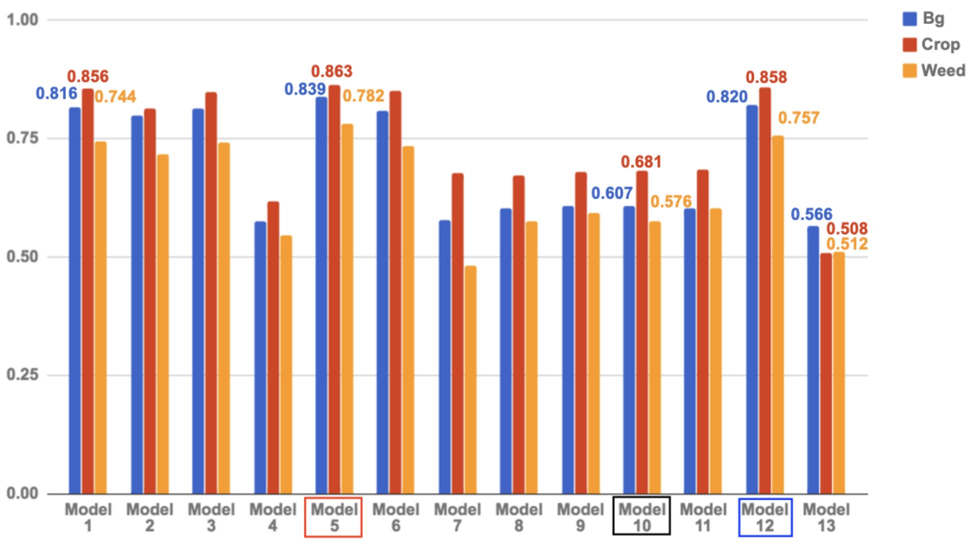


Figure 6: Quantitative evaluation of the segmentation using area under the curve (AUC) of the RedEdge-M dataset. The red box indicates the best model, the black one is our baseline model with only RGB image input, and the blue box is a model with only one NDVI image input.[Sa+18b]

The results record the performance of the classifier with datasets varying in input channels and network hyperparameters

4.1 Experimental Setup

- Multispectral orthomosaic maps are present with their corresponding manually annotated ground truth labels. The classification problem here has three classes background, crop and weed.
- Different datasets are used for training and testing of RedEdge-M (5 channel) and sequoia (4 channel). Datasets can not be combined for training and testing because their multispectral bands are not matched.
- Two-fold data augmentation is done (the input images are horizontally mirrored).

4.2 Performance Evaluation Metric

- The area under the curve (AUC) of a precision-recall curve is used as a metric for performance evaluation where for a given class c.

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \qquad Recall_c = \frac{TP_c}{TP_c + FN_c}$$

- TP_c, TF_c, FP_c, FN_c are true positive, true negative, false positive and false negative classifications for class c.
- The network outputs are the probability of each pixel belonging to each defined class.
- The output is three layered as well due to the three existing classes where each existing class has pixel-wise probability for all pixels.
- These probabilities are converted to binary values based on a threshold. The threshold is incrementally varied from 0 to 1 and precision, recall and AUC is calculated. This helps in getting the optimum threshold.

Note: There are other metrics to calculate dense semantic segmentation, however these metrics either rely on specific thresholds or assign the label with maximum probability among all classes for evaluation. So It can be seen that computing AUC over the probabilistic output gives better classification performance than other metrics.

5 Discussion on Challenges and Limitations

- Obtaining a model which can incorporate different plant maturity stages and several farm fields remains a challenge as visual appearance of plants is constantly changing during their growth.
- There is also difficulty in distinguishing between crops and weed in the early season as they appear quite similar. High quality and high resolution spatiotemporal datasets are required to correct this which is possible for crops but not for weeds as they are more difficult to capture representatively due to the diverse range of species that can be found in a single field.
- Map generation requires a high end GPU to speed up things and it can be further accelerated through hardware or software improvements. The map generated can provide useful information for creating prescription maps for the farm, which can then be transferred to automated machinery. This procedure will minimize the chemical usage and the labor costs (environmental and economical impact) while maintaining the agricultural output of the farm.

6 Future Work

6.1 For dataset acquisition

The author mentions in [Sa+18b] that one of the issues in developing the dataset is that there is high amount of dependence on the age of the crops. 1 month old crops seem much more similar to weeds compared to older crops. We suggest the techniques used in [Sa+18a] to partially overcome this problem. The author of [Sa+18a] suggest that the field be divided into 3 parts and each part be given different doses of herbicide. This enables one part to have no weeds at all, one of them to have moderate amount of weed and the third one to have mostly weed. This method can also help in the generation of ground truth data.

6.2 Incentivising the Network to differentiate crops from weeds.

Multiple face recognition systems use triplet loss to incentivise the network to learn the difference between similiar images (i.e. human faces). It forces the similarity index for images of different objects(people) to be low while at the same time trying to increase it for images of same object. The same can be applied to explicitly make the model differentiate between weeds and crops. furthermore, we can use triplet loss (as used in face recognition) to force the network to learn the difference between crops and weeds.

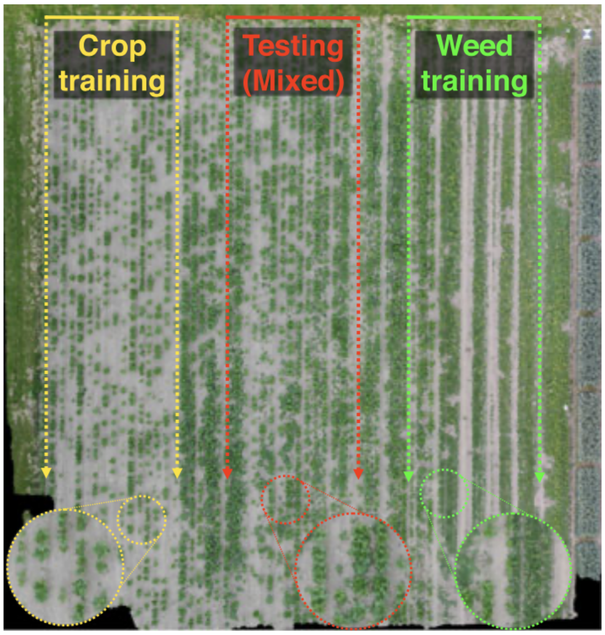


Figure 7: Aerial view of the author's controlled field with varying herbicide levels.The maximum amount of herbicide is applied to the left crop training rows (yellow), and no herbicide is utilized for the right weed training rows (green). The middle shows mixed variants due to medium herbicide usage (red).[Sa+18a]

References

[Sa+18a] I. Sa et al. "weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming". In: *IEEE Robotics and Automation Letters* 3.1 (2018), pp. 588–595.

[Sa+18b] Inkyu Sa et al. *WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming.* 2018. arXiv: 1808.00100 [cs.RO].