DEEP LEARNING FOR REMOTE SENSING IMAGE CLASSIFICATION FOR AGRICULTURE APPLICATIONS

Leila Hashemi-Beni[1*], Asmamaw Gebrehiwot[2]

[1*] Geomatics Program, Department of Built Environment, North Carolina A&T State University - lhashemibeni@ncat.edu
[2] Applied Science and Technology Ph.D. Program, Department of Built Environment, North Carolina A&T State University - aagebrehiwot@aggies.ncat.edu

KEY WORDS: Precision Agriculture, Remote Sensing, U-Net, FCN-8s, Deep Learning, Image Segmentation

ABSTRACT:

This research examines the ability of deep learning methods for remote sensing image classification for agriculture applications. U-net and convolutional neural networks are fine-tuned, utilized and tested for crop/weed classification. The dataset for this study includes 60 top-down images of an organic carrots field, which was collected by an autonomous vehicle and labeled by experts. FCN-8s model achieved 75.1% accuracy on detecting weeds compared to 66.72% of U-net using 60 training images. However, the U-net model performed better on detecting crops which is 60.48% compared to 47.86% of FCN-8s.

1. INTRODUCTION

Weed controlling is one of the main problems that farmers must deal with. Weeds profoundly affect farm productivity by invading crops and smother pastures and significantly decrease the quality of the harvested crops (Milberg et al., 2004). Herbicides are widely used globally to enhance food production; however, it can cause harm to the environment and the ecosystem (Horrigan et al., 2002). In the traditional weeds control approaches, the herbicide is applied over the whole field, even for the area without weeds where no treatment is required. Precision agriculture techniques should regularly monitor crop growth to maximize yield while minimizing the use of resources such as chemicals and reducing the side effects of herbicides on the environment (Duckett et al., 2018). Thus, accurate weeds detection and mapping, and their local treatment are essential steps to improve weed and crop control in modern agriculture.

Recently, unmanned aerial systems (UAVs) have become suitable platforms for acquiring data for crop and weeds monitoring. UAVs can acquire high-resolution imagery at a low cost, and less dependence on weather conditions (Hashemi-Beni et al., 2018, Gebrehiwot et al., 2019, Vinh et al., 2019).

Several methods have been proposed for weed recognition over the last decades. Vegetation index, such as the normalized difference vegetation index (NDVI) is one of the methods to segment weeds in an agricultural field (Dyrmann et al., 2014). The main challenge of this approach is dealing with overlapping plants to separate weeds and crops. Texture-based models have shown great performance in detecting and discriminating plants from images with overlapping leave (Pahikkala et al., 2015). Machine learning-based approaches have gained attention for detecting weeds and crops (Murawwat et al., 2018). Murawwat, et al. (2018) use a support vector machine (SVM) classifier to classify carrot crops and weeds. They used 72 samples for training and eight samples for testing and achieved more than 50% of classification accuracy. The challenge of the traditional ML approach such as SVM or RF classifiers is that feature extraction is not automatic and handcrafted features generation is a time-consuming stage. Some studies prove that unlike conventional machine learning methods, deep learning can efficiently deal with the limitations of handcrafted features for

classifying weed and crops by extracting the features directly from the input data (Lee et al., 2015). Recently, there has been considerable progress in the classification of remote sensing data using deep learning for different applications including agricultural tasks. Some studies used a convolutional neural network (CNN) in agricultural applications such as weed and crop classification (Mortensen et al., 2016, Potena et al., 2016, Di Cicco et al., 2017). Mortensen et al. (2016) used the VGG-16 CNN model to classify weeds using mixed crops of an oil radish plot with barely, weed, stump, grass, and background soil images. Potena et al. (2016) presented a perception system for weed crop classification that uses shallow and deeper CNNs. The shallow CNN is used to detect vegetation, while the deeper one is used to classify weeds and crops. Di et al. (2017) used a SegNet to procedurally generate large synthetic training datasets randomizing the key features of the target environment (i.e., crop and weed species, type of soil, light conditions). Hashemi-Beni, et al (2020) applied the u-net model to detect and discriminate crop and weeds using a small dataset. They applied data augmentation techniques such as random cropping, random rotation, and reflection to improve the classification results.

This research provides a comparison of the U-Net and FCN-8s models for segmenting weeds from Crop/Weed Field Image Dataset (CWFID), which was introduced in Haug et al. (2014).

2. DATASET AND ANNOTATION

The CWFID (Haug et al., 2014) was used to train and test the U-Net classifier in this study. The CWFID dataset comprises field images in top-down view that were collected with an autonomous field robot Bonirob in an organic carrot farm in 2013. The CWFID includes 60 images with a size of 1296 x 966 pixels. All data acquisition was carried out during field tests, and images were acquired while the crop was in the growth stage and one or more true leaves were present. These images were labeled with a vegetation segmentation mask and crop/weed labeled image (In total: 162 and 332 crop and weed plants, respectively). A sample image of this dataset is shown in Figure 1. The CWFID was downloaded from https://github.com/cwfid.
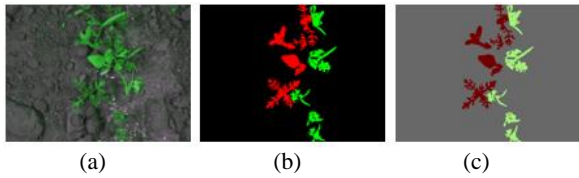
Figure 1. Sample training image. (**a**) training image; (**b**) labeled image, (**c**) the image relabeled using MATLAB tool.

Figure 1 shows one sample example of the training and its corresponding labeled image. The black, green, and red colors represent soil, crops, and weeds, respectively.

3. METHODS AND DATA PROCESSING

3.1 Network Architecture

3.1.1 U-net Architecture: The U-Net is a type of CNN architecture proposed by Ronneberger et al. (2015) for biomedical image segmentation. The U-Net architecture is based on the fully convolutional network, and it was modified to give accurate classification results with fewer training data. As shown in Figure 2, the architecture of U-Net looks like a letter 'U' that justifies its name.
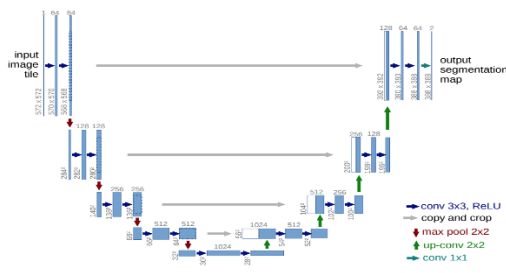

Figure 2. U-Net architecture (Ronneberger et al., 2015)

The U-Net architecture has two structures: a shrinking path (left side) and an expanding path (right side). The shrinking path (also called the encoder) used to extract and capture the context in the image and the expanding path combines the feature maps. Then, more precise predictions of the pixel points on the edge can be obtained. The encoder is just a traditional stack of convolutional and max-pooling layers. The expanding structure (or decoder) is used to enable precise localization using transposed convolutions. Thus, it is an end-to-end fully convolutional network that only contains convolutional layers and does not contain any dense layer because of which it can accept the image of any size.

3.1.2 FCN-8s architecture: The FCN was proposed by Long et al. (2015) to train an end-to-end for semantic segmentation. In this model, VGG16 fully connected based classification layers were replaced by convolutional layers to maintain the 2-D structure of images. VGG-16 is a CNN architecture proposed by Simonyan et al. (2014) to investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. As shown in Figure 3, because of the gradual upsampling of the scoring layer and merging of features from earlier layers, a fine label map is obtained using the FCN-8s. For this study, a fully convolutional neural network (FCN) with a stride of 8 (FCN-8s) model was fine-tuned. The FCN-8s is composed of locally connected layers, such as convolution, pooling, and upsampling, without having any dense layer. This allows reducing the number of parameters and computation time.

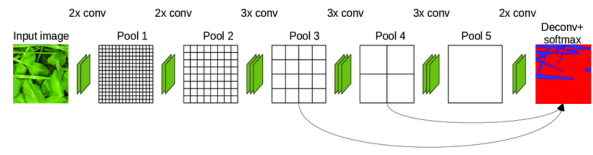Given that all connections are local, FCN-8s can work on any image size.

Figure 3. FCN-8s architecture (Skovsen et al., 2017)

3.2 Training U-Net and FCN-8s

In this study, the U-net and FCN-8s models fine-tuned and trained using Stochastic Gradient Descent (SGD) with a learning rate of 0.001, and a maximum epoch of 16. A 10-fold cross-validation method was used to estimate the ability of the U-net and FCN models on unseen data. For this purpose, we randomly partitioned the training images into 10 equal parts. At each run, the union of 9 parts was put together to form a training set, and the remaining 1-part used as a test set to estimate the classification errors. The above steps are repeated ten times, using a different fold as the testing set each time. Finally, the mean error from all folds was used to estimate the potential of the U-net and FCN-8s models. It took approximately 8 hours to cross-validating U-Net and FCN-8s using a single GPU (NVIDIA Quadro M4000).

4. RESULTS AND DISCUSSION

In this study, a confusion matrix was used to analyze the classification results of the U-net and FCN-8s models. The qualitative classification results of U-Net and FCN-8s are shown in Figure 4, and the detailed results on how the U-Net and FCN-8s models performed for each class (for soil, weeds, and crops classes) are described in Table 1 and Table 2, via the confusion matrices. The U-Net and FCN-8s models achieved an overall accuracy of 75.2% and 72.1%, respectively for separating weeds from crops.
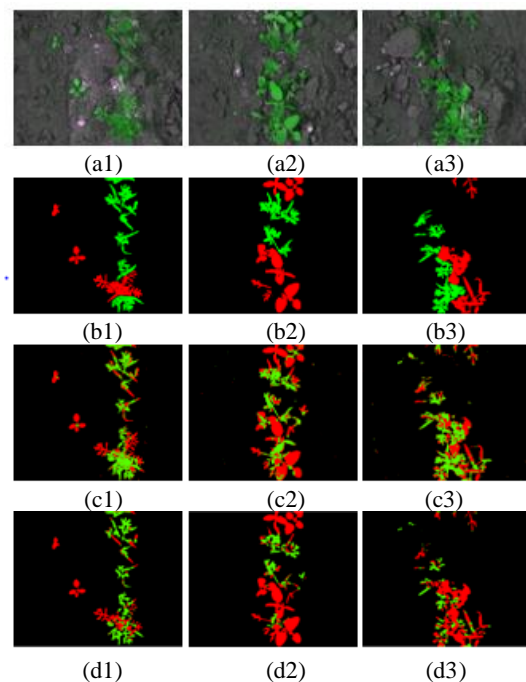

Figure 4. Classification results of U-Net and FCN-8s. (a1-a3) sampe test images; (b1-b3) Ground truth, and (c1-c3) U-net classification results; (d1-d3) FCN-8s results
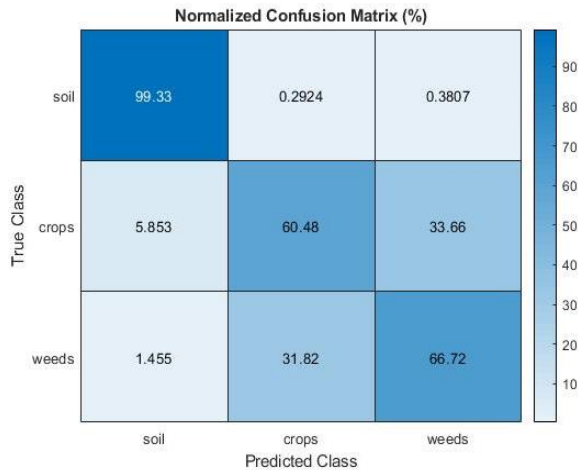
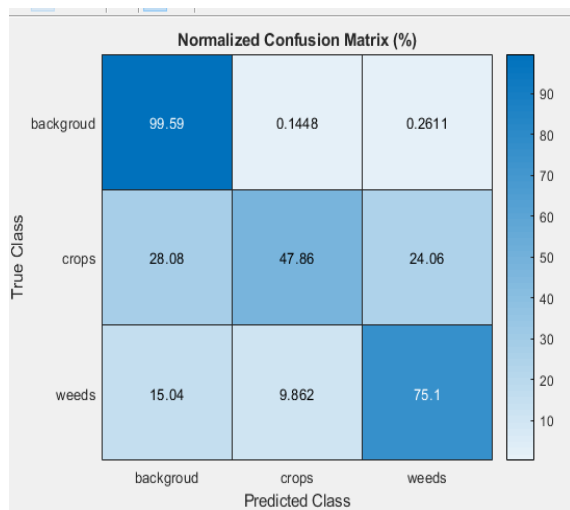Table 1. Confusion matrix of U-Net Classification (unit-percentage)

Table 2. Confusion matrix of FCN-8s Classification(unit-percentage)

Table 1 and Table 2 show the correspondence between the predicted and the validation values. The cells of a confusion matrix represent the percentage of correct and incorrect predictions for all the possible correlations. The cell in the ith row and jth column means the percentage of the ith class samples, which classified into the jth class. The diagonal cell of the matrix contains the number of correctly identified pixels for each class. As the tables demonstrate, the U-Net model achieved the classification accuracy of 99.3% for soil, 60.48% for crops, and 66.72% for weeds, whereas FCN-8s an accuracy of 99.59% for soil, 47.86% for crops, and 75.1% for weeds. The U-net model achieved better overall accuracy (75.2%) compared to FCN-8s (72.1%). However, FCN-8s showed better performance on extracting weeds (75.1%) compared to U-net (66.72%). The results also show that about 33% and 24% of weeds in U-net and FCN-8s, respectively were misclassified as crops. This is because weeds and crops have a similar spectral response, which makes it hard to separate them using U-Net and FCN-8s classifiers solely from optical imagery.

5. CONCLUSION

Mapping the location of the weeds and locally treat those areas are essential to improve weed control capacity. This study compared the performance of two well-known deep learning-based methods to distinguish crops from weeds using high-resolution imagery. The U-Net and FCN-8s models were fine-tuned to classify the CWFID into three classes (soil, crops, and weeds). The fine-tuning and transfer learning technique allowed us to overcome the problem of small dataset size. The U-Net and FCN-8s models achieved an overall accuracy of 75.2% and 72.1%, respectively for classifying the CWFID dataset into three classes using 60 training images. In future research, we will incorporate crop geometry constrains to the model to improve classification accuracy.

REFERENCES

Di Cicco, M., Potena, C., Grisetti, G., Pretto, A., 2017. Automatic model based dataset generation for fast and accurate crop and weeds detection. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 5188-5195.

Duckett, T., Pearson, S., Blackmore, S., Grieve, B., Chen, W.H., Cielniak, G., Cleaversmith, J., Dai, J., Davis, S., Fox, C., From, P., 2018. Agricultural robotics: the future of robotic agriculture. *arXiv preprint arXiv:1806.06762*.

Dyrmann, M., Christiansen, P., 2014. Automated Classification of Seedlings Using Computer Vision: Pattern Recognition of Seedlings Combining Features of Plants and Leaves for Improved Discrimination. Aarhus University.

Gebrehiwot, A., Hashemi-Beni, L., Thompson, G., Kordjamshidi, P., Langan, T. E., 2019. Deep convolutional neural network for flood extent mapping using unmanned aerial vehicle data. *Sensors*, *19*(7), 1486.

Hashemi-Beni, L., Gebrehiwot, A., 2020. Weed and crop discrimination using U-net learning IGRASS.

Hashemi-Beni, L., Jones, J., Thompson, G., Johnson, C., Gebrehiwot, A., 2018. Challenges and opportunities for UAV-based digital elevation model generation for flood-risk management: A case of Princeville, North Carolina. *Sensors*, *18*(11), 3843.

Hashemi-Beni, L, et al. Weed and crop discrimination using U-net learning, IGRASS 2020, In Press.

Haug, S., Ostermann, J., 2014. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In *European Conference on Computer Vision*, 105-116. Springer, Cham.

Horrigan, L., Lawrence, R. S., Walker, P., 2002. How sustainable agriculture can address the environmental and human health harms of industrial agriculture. *Environ Health Perspect, 110*(5), 445-456.

Lee, S. H., Chan, C. S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional neural networks. In *2015 IEEE international conference on image processing (ICIP)*, IEEE, 452-456.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431-3440.

Milberg, P., Hallgren, E., 2004. Yield loss due to weeds in cereals and its large-scale variability in Sweden. *Field Crops Research*, *86*(2-3), 199-209.

Mortensen, A. K., Dyrmann, M., Karstoft, H., Jørgensen, R. N., Gislum, R., 2016, Ju. Semantic segmentation of mixed crops using deep convolutional neural network. In *Proc. of the International Conf. of Agricultural Engineering (CIGR)*.

Murawwat, S., Qureshi, A., Ahmad, S., Shahid, Y., 2018. Weed Detection Using SVMs. *Engineering, Technology & Applied Science Research*, *8*(1), 2412-2416.

Pahikkala, T., Kari, K., Mattila, H., Lepistö, A., Teuhola, J., Nevalainen, O. S., Tyystjärvi, E., 2015. Classification of plant species from images of overlapping leaves. *Computers and Electronics in Agriculture*, *118*, 186-192.

Potena, C., Nardi, D., Pretto, A., 2016. Fast and accurate crop and weed identification with summarized train sets for precision agriculture. In *International Conference on Intelligent Autonomous Systems*, 105-121. Springer, Cham.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*, 9351, 234-241.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Skovsen, S., Dyrmann, M., Mortensen, A. K., Steen, K. A., Green, O., Eriksen, J., Karstoft, H., 2017. Estimation of the botanical composition of clover-grass leys from RGB images using data simulation and fully convolutional neural networks. *Sensors*, *17*(12), 2930.

Vinh, K., Gebreyohannes, S., Karimoddini, A., 2019. An Area-Decomposition Based Approach for Cooperative Tasking and Coordination of UAVs in a Search and Coverage Mission. In *2019 IEEE Aerospace Conference*, IEEE, 1-8.