CROP AND WEED SEGMENTATION ON GROUND-BASED IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORK

H. Fathipoor [1] *, R. Shah-Hosseini [1], H. Arefi [2]

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Iran - (hafez.fathipoor, rshahosseini)@ut.ac.ir
[2] School of Technology, Department of Geoinformatics and Surveying, Mainz University of Applied Sciences, Germany - hossein.arefi@hs-mainz.de

Commission IV, WG IV/3

KEY WORDS: Precision Agriculture, Deep Learning, Weed Management, Semantic Segmentation, CNN, U-Net, U-Net++

ABSTRACT:

Weed management is of crucial importance in precision agriculture to improve productivity and reduce herbicide pollution. In this regard, showing promising results, deep learning algorithms have increasingly gained attention for crop and weed segmentation in agricultural fields. In this paper, the U-Net++ network, a state-of-the-art convolutional neural network (CNN) algorithm, which has rarely been used in precision agriculture, was implemented for the semantic segmentation of weed images. Then, we compared the model performance to that of the U-Net algorithm based on various criteria.
The results show that the U-Net++ outperforms traditional U-Net in terms of overall accuracy, intersection over union (IoU), recall, and F1-Score metrics. Furthermore, the U-Net++ model provided weed IoU of 65%, whereas the U-Net gave weed IoU of 56%. In addition, the results indicate that the U-Net++ is quite capable of detecting small weeds, suggesting that this architecture is more desirable for identifying weeds in the early growing season.

1. INTRODUCTION

A monitoring system in precision agriculture is of fundamental significance to increase crop productivity (Fathipoor et al., 2019), and weed management is one of the critical elements of this system. Competing with plants for water and nutrients, weed adversely affects crop yield quality and takes a deleterious toll on crop production (Wang et al., 2019). Therefore, identifying and eliminating weeds is an essential step in precision agriculture. In this regard, many efforts have been made by farmers to counter the threat posed by weeds. However, conventional agriculture practices are laborious and inefficient, spraying herbicides uniformly to the whole field (Wang et al., 2019). Moreover, the remaining herbicides result in environmental pollution, which pose a serious threat to human health (Khan et al., 2020).

To deal with this problem, site-specific weed management (SSWM) was proposed, which involves spraying the correct dose of herbicide (depending on the density of the weed patches or the species composition) in the right locations (Jensen et al., 2012). But, SSWM requires advanced autonomous weed detection systems. Accordingly, automated weed segmentation, a labor-saving process, is crucial to reducing the detrimental effects of herbicides or pesticides through localizing weeds precisely in agricultural fields (Pretto et al., 2021). In this context, several weed detection methods based on traditional image processing, such as decision trees (Deng et al., 2014), support vector machine (SVM) (Ishak et al., 2008), and random forest (Fletcher et al., 2016) are introduced to differentiate weeds from crops. In these techniques, pixels are classified into crop and weed classes based on extracted features, such as color and texture (Rico-Fernández et al., 2019). However, feature extraction in these methods

greatly depends on numerous parameters, such as weed density and lighting conditions, which hinders the performance of these algorithms in complex situations (Abdalla et al., 2019). Hence, there is a need to build efficient and robust modules to identify and recognize weeds.

In recnet years, thanks to the advancement in computing power coupled with a rise in the amount of data, deep learning based methods such as convolutional neural networks (CNNs) offer a promising step toward managing weeds and pests more efficiently (Wu et al., 2021). Having been widely used recently, semantic segmentation algorithms based on CNN, such as fully convolutional network (FCN), SegNet, U-Net, and DeepLabV3, make it possible to segment weeds from crops with high accuracy. These algorithms generally are a fully convolutional network that often involves an encoder-decoder scheme, extracting features of input images and then up-sampling to the size of the original image. For instance, a study on crop/weed segmentation used an encoder-decoder deep learning architecture, which utilized different vegetation indices as inputs to improve performance, and the best mean segmentation accuracy of 96.12% was obtained (Wang et al., 2020). In another similar study using an RGB image dataset of carrot-weed, SegNet architecture was employed for semantic segmentation of images (Lameski et al., 2017). In another study of weed segmentation, the performance of SegNet was compared with that of U-Net based on a dataset of canola fields, and the authors showed SegNet had higher accuracy in their dataset (Asad et al., 2020).

DeeplabV3 is another complex and powerful network with satisfactory performance in semantic segmentation studies. In a

---

* Corresponding author

study of weed segmentation using aerial images, DeepLabv3 outperforms SegNet and U-Net with the highest accuracy of 0.89 and 0.81 in terms of area under the curve (AUC) and F1-score, respectively (Ramirez et al., 2020). (Khan et al., 2020) proposed a cascaded encoder-decoder network to segment precisely crop and weed, with fewer training parameters. Based on their architecture, four small networks were used to predict crop and weed in two stages independently and performed more accurately than U-Net, FCN-8s, SegNet, and DeepLabv3. However, SegNet and Deeplab V3 require more training data in comparison with U-Net (Zou et al., 2021). With the limited accessibility to large datasets for training models, the U-Net architecture that can be trained on small datasets is highly advantageous (Bousias Alexakis et al., 2020). For instance, in (Hashemi-Beni et al., 2020), the authors used 60 images of a carrot field and reached an accuracy of 60.48% with the U-Net network for weed semantic segmentation. Lately, researchers have tried to enhance the performance of the U-Net model by modifying this architecture. In an effort to discriminate weeds from other classes including the soil and crops, a modified VGG-UNet was implemented, which gave a desirable accuracy of intersection over union (IoU) of 92.91% (Zou et al., 2021).

Although being the state-of-the-art models for image segmentation, modified U-Nets have two main limitations. Firstly, it is hard to reach the best accuracy achievable with the model because of its uncertainty and variability of optimal depth. Secondly, the skip connection scheme is inefficient due to the gap between pathways of corresponding convolutional encoder-decoder blocks (Bousias Alexakis et al., 2020; Zhou et al., 2019).

The U-Net++ is a new architecture designed to overcome previous drawbacks and has a more robust performance in semantic segmentation (Zhou et al., 2019). For example, in a study of change detection in an urban environment, the performance of the U-Net++ and the U-Net architectures were evaluated by different loss functions and metrics (Bousias Alexakis et al., 2020). The authors showed that U-Net++ architecture with BCE-Dice Loss function provides better results than the U-Net. The U-Net++ network is based on nested and dense skip connections, which has rarely been used in agricultural tasks. Hence, this paper mainly aims to address the weed management task by automatic crop/weed segmentation from high-resolution images using the U-Net++ model. The methodology is described in section 2, followed by the results presented in section 3, and finally, section 4 includes the conclusion.

2. METHODOLOGY

2.1 Dataset

In this study, a public carrot-weed dataset was utilized in order to train and test our models (Lameski et al., 2017). The dataset contains 39 RGB images with a dimension of 3264×2448 pixels, acquired by a 10 MegaPixel phone camera. This dataset is complex data in which weeds are highly overlapped with plants, making segmentation a challenging task. In addition, it is an imbalanced dataset, meaning that the number of weed pixels is much less than those of other classes, which hinders the performance of classification algorithms. Pixel-level annotations with three classes: soil, carrots, and unspecified weeds are also provided for this dataset. Some images of this dataset are shown in Figure 7 (a).

2.2 Network Architecture

2.2.1 U-Net Architecture: U-Net is a modified fully convolutional network with a 'U' shape architecture in which the output image has the same size as the input image (Figure 1). The main difference between traditional fully convolutional networks (FCNs) and U-Net is that U-Net refills lost information in edges and localizes features more accurately by constantly extracting and combining the high-resolution features of the down-sampling parts to the corresponding up-sampling block (Bousias Alexakis et al., 2020; Hashemi-Beni et al., 2020).

2.2.2 U-Net++ architecture: U-Net++, also called Nested U-Net, is based on the U-Net network introduced to enhance the performance of U-Net. The motivation behind introducing U-Net++ is to make the optimization problem of the model easier and achieve more accurate results by densifying the connectivity and aggregating various depths of U-Nets (Figure 2). To bridge the gap between encoder and decoder sub-networks, skip pathways have been re-designed. Besides, deep supervision has been added, making the model more flexible by making a balance between performance and speed (Zhou et al., 2019).
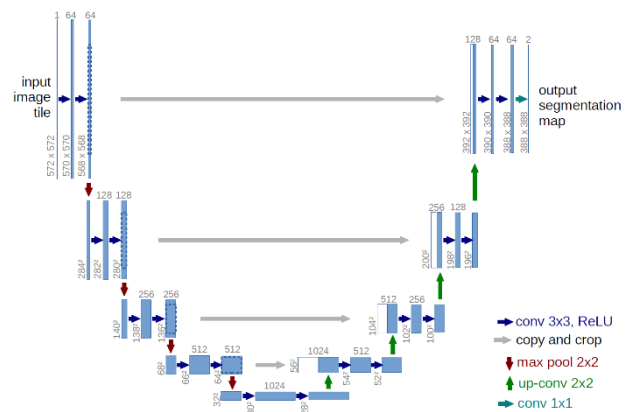

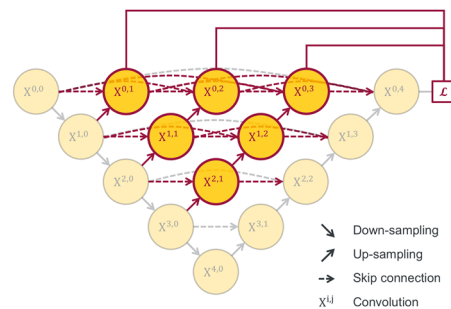**Figure 1**. U-Net architecture (Ronneberger et al., 2015).

Figure 2. U-Net++ architecture (Zhou et al., 2019).

2.3 Training Networks

The Google Colab framework was employed to implement the models in this research. For the sake of computational efficiency, all images were resized into 128×128 pixels. In addition, among 39 images, 27, 3, and 8 images were utilized as training, validation, and testing, respectively. The networks were tuned using the "Adam" optimizer with a learning rate of 1×10^{-4}.

Furthermore, the loss function for the U-Net++ model was a weighted combination of categorical cross-entropy and dice coefficient loss. The reason behind using the hybrid loss function

is to fully exploit what both functions provide; on the one hand, cross-entropy has smoother gradients; on the other, dice coefficient handles properly imbalanced dataset (Zhou et al., 2019). The hybrid loss function can be calculated with the following equations:

$$\mathcal{L} = \mathcal{L}_{cce} + \lambda\,\mathcal{L}_{dice}, \qquad (1)$$

where $\quad \mathcal{L}_{cce}$ = categorical cross-entropy loss
$\mathcal{L}_{dice}$ = dice coefficient loss
λ = weight that balances the two losses

Categorical cross-entropy loss can be computed based on equation 2:

$$\mathcal{L}_{cce} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} y_{n,c}\,\log\hat{y}_{n,c}, \qquad (2)$$

where $\quad$ C = number of classes
N = number of samples within one batch
$y_{n,c}$ = class label (if label is C, $y_{n,c}$ equals to 1; otherwise is 0) and $n \in [1,\dots,N]$
$\hat{y}_{n,c}$ = probability of sample n being correctly classified as c

Moreover, the formula for calculating dice coefficient loss is given in the following:

$$\mathcal{L}_{dice} = 1 - \sum_{c=1}^{C}\frac{2\sum_{n=1}^{N} y_{n,c}\times\hat{y}_{n,c}}{\sum_{n=1}^{N} y_{n,c}^2 + \sum_{n=1}^{N}\hat{y}_{n,c}^2} \qquad (3)$$

2.4 Quantitative Assessment

To evaluate the segmentation results and compare the performance of models, five popular criteria, including IoU, accuracy (Acc), precision (Pre), recall (Re), and F1-Score, were calculated for each class and then averaged. These metrics were computed based on four variables, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which were derived from confusion matrixes between the predictions and the ground truth maps via the following equations. To describe these variables, take weed class for instance. TP represents the number of pixels correctly classified as weed, TN denotes the number of pixels correctly classified as non-weed, FP stands for the number of pixels classified as weed that were not actually weed, and FN represents the number of pixels incorrectly classified as non-weed classes.

$$\text{IoU} = \frac{\sum\text{TP}}{\sum\text{TP}+\sum\text{FN}+\sum\text{FP}} \times 100\% \qquad (4)$$

$$\text{Acc} = \frac{\sum\text{TP}+\sum\text{TN}}{\sum\text{TN}+\sum\text{FP}+\sum\text{FN}} \times 100\% \qquad (5)$$

$$\text{Pr} = \frac{\sum\text{TP}}{\sum\text{TP}+\sum\text{FP}} \times 100\% \qquad (6)$$

$$\text{Re} = \frac{\sum\text{TP}}{\sum\text{TP}+\sum\text{FN}} \times 100\% \qquad (7)$$

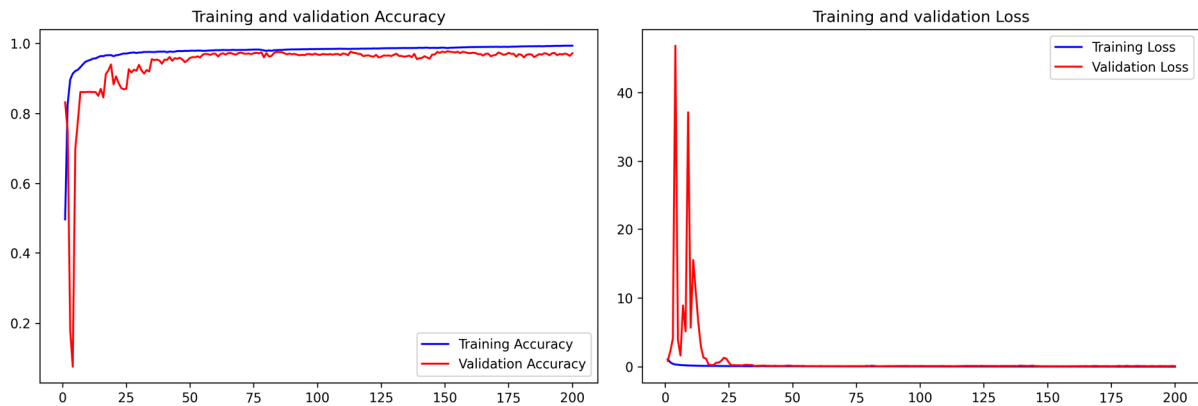$$\text{F1-Score} = 2 \times \frac{\text{Pr}\times\text{Re}}{\text{Pr}+\text{Re}} \times 100\% \qquad (8)$$

Figure 3. The accuracy and loss during training U-Net model.
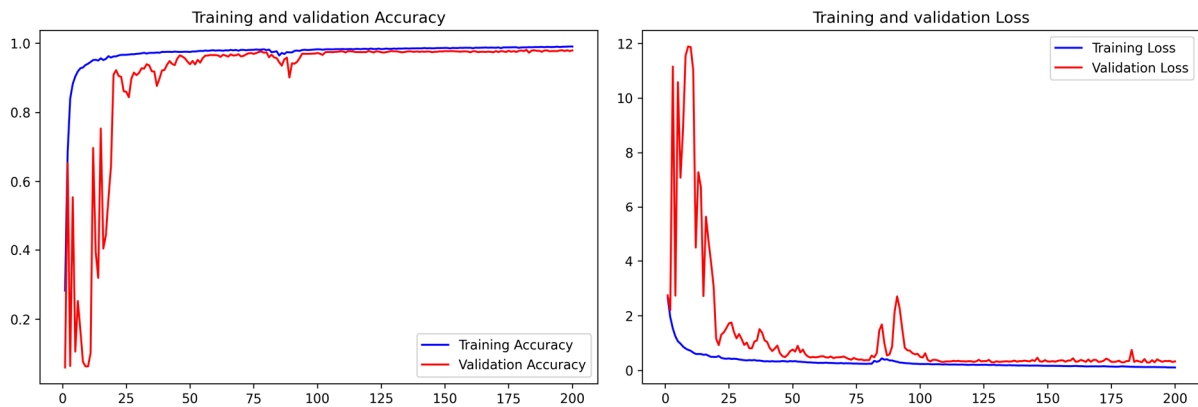

Figure 4. The accuracy and loss during training U-Net++ model.

3. RESULTS AND DISCUSSION

This section provides results of the semantic segmentation of the U-Net and the U-Net++ networks. Figure 3 and Figure 4 show the loss and average class accuracy for 200 iterations during the training process of the U-Net and the U-Net++ networks. According to these figures, there is a very neglectable performance improvement beyond 100 iterations, suggesting that the networks were trained enough for this segmentation task. Also, it should be noted that in the U-Net++ model the convergence process took much longer than in the U-Net network because of the more complex loss function used in the U-Net++ model.

Figure 5 and Figure 6 provide the information of the confusion matrix between segmentation results and ground truth for both models. Accordingly, in the U-Net++, the TP values, the proportion of pixels correctly predicted, for three classes of weed, plant, and soil are 83.45%, 86.28%, and 98.96%, respectively. By comparing these values with the corresponding values achieved by the U-Net, we can infer that the U-Net++ had much better performance in classifying weed, even though it had quite similar performance as the U-Net in classifying plants and soil. In fact, in the U-Net model, there were a lot of weeds wrongly predicted as plants (15%), but the U-Net++ network overcame this problem to a great extent. On the other hand, the percentage of plants that were mistakenly classified as weeds in the U-Net network was almost twice in the U-Net++ model. Furthermore, in the U-Net model, more plant and weed pixels were wrongly classified as soil class.
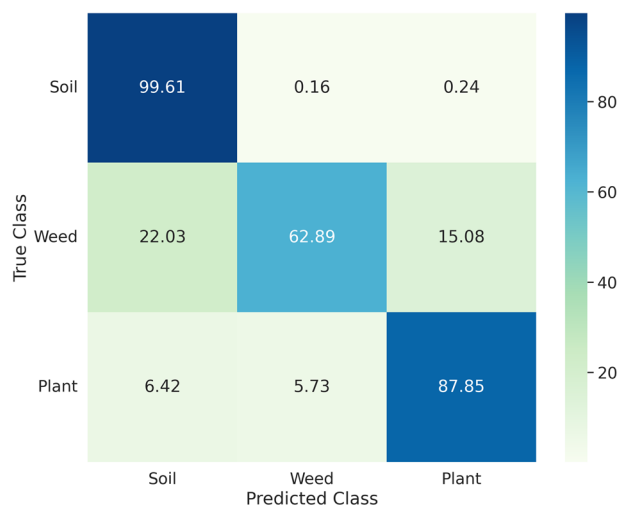
Figure 5. Normalized confusion matrix of the U-Net model.

In Figure 7 parts (c) and (d), some parts of the qualitative results of both models are shown. As just mentioned, it is evident in this figure that some weeds were recognized wrongly as plants by the U-Net, while they were identified correctly by the U-Net++. This poor performance of the U-Net happened more particularly in complex parts of the image where weeds and plants are mixed (e.g., the first and third images in Figure 7). In addition, unlike the U-Net network, the U-Net++ represented the high ability to identify tiny weeds. This is because of the aggregating multi-depth structure of the U-Net++ that makes it more powerful to

segment weeds of various sizes. This advantage becomes greatly important, especially at the beginning of the growing season when young plants and weeds start to germinate. Therefore, detecting and removing weeds at this stage will be highly beneficial for young plants to flourish.
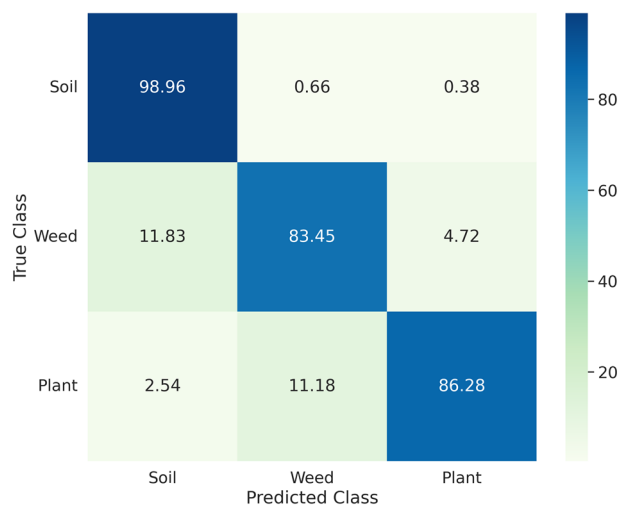
Figure 6. Normalized confusion matrix of the U-Net++ model.

The detailed results of the evaluation of both network architectures based on five well-known metrics are given in Table 1. Accordingly, the U-Net++ had higher mean IoU, Acc, Re, and F1-Score metric values, outperforming traditional U-Net in our semantic segmentation task. Although Pre decreased somewhat compared with that of the U-Net, the Re was better in the U-Net++, that is, the U-Net++ model segmented weed more aggressively and correctly classified more weeds at the expense of misclassification of some plants as weeds.

| Segmentation method	IoU (%)	Acc (%)	Pre (%)	Re (%)	F$_1$ Score (%)
U-Net++	81.31	97.02	88.84	89.56	89.20
U-Net	77.51	96.59	90.19	83.45	86.69

Table 1. Performance evaluation of segmentation models.

Since segmenting weed is more important than other classes, this study intends to assess the accuracy of identifying per class individually. For this purpose, IoU values for three classes of weed, carrot, and soil are given in Table 2. U-Net++ model provided an IoU of 97.97% for soil, 80.80% for crops, and 65.13% for weeds, which are higher than the accuracy obtained from U-Net. In general, IoU value for each class represents the ability of models to correctly classify the corresponding class; the higher value for a given class, the better performance the model has for separating that class. Given the considerable improvement in weed IoU, U-Net++ proved much better in weed segmentation than U-Net.

Architecture	Weed IoU	Plant IoU	Soil IoU
U-Net++	65.14	80.81	97.98
U-Net	56.38	78.53	97.62

Table 2. Comparison of IoU for per class in models.
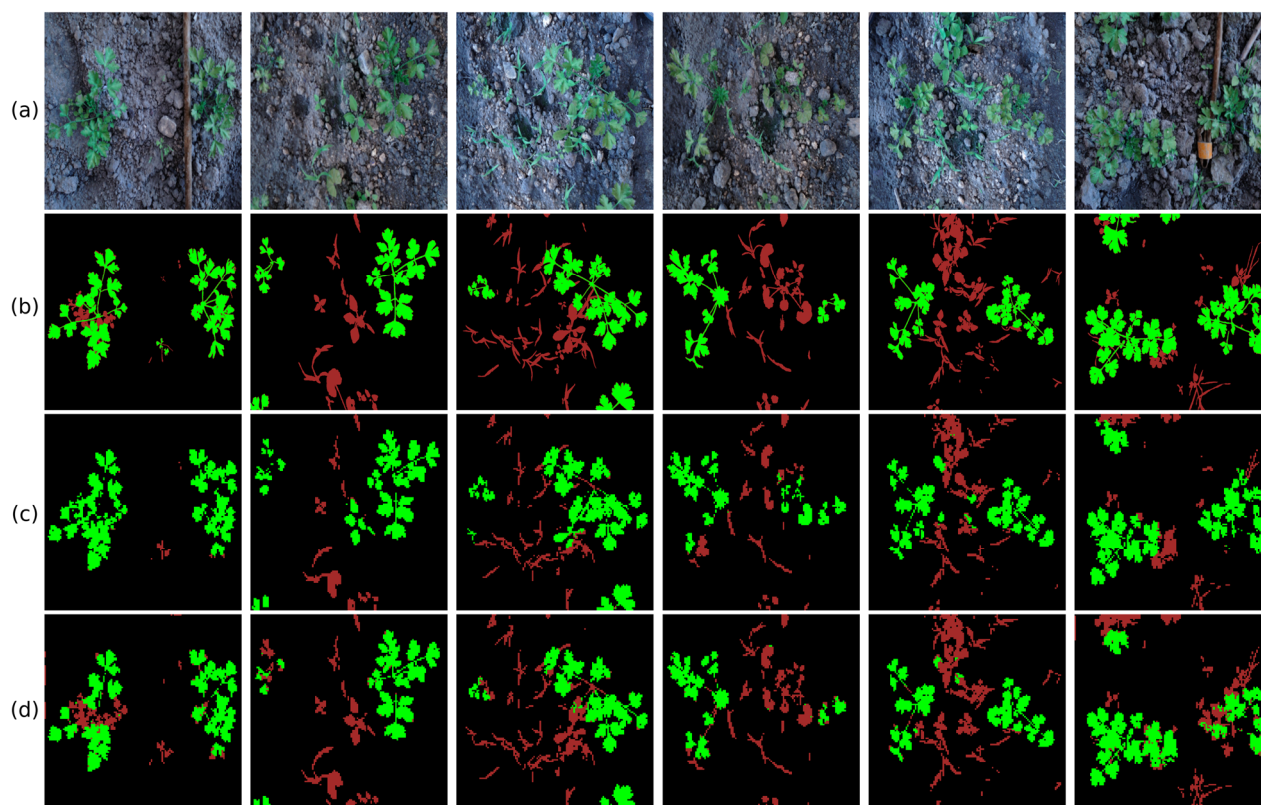
Figure 7. Segmentation results of models. (a) Samples of original images. (b) Ground truth. (c) Segmentation results of U-Net. (d) Segmentation results of U-Net++.

4. CONCLUSION

The advent of deep learning algorithms has provided an unprecedented opportunity to pinpoint and thus eliminate weeds more efficiently. With the aim of pixel-wise semantic segmentation of a weed-carrot dataset, including 39 images, this study employed U-Net and U-Net++ as two kinds of advanced deep convolutional networks. The results show that the U-Net++ provides better performance than the U-Net in terms of overall accuracy, mean IoU, recall, and F1-Score metrics. Most importantly, the U-Net++ model performed better in complex parts of images where weeds were mixed with plants. In addition, the U-Net++ was notably more effective than the U-Net in weed segmentation based on weed IoU. Overall, this paper demonstrated that the U-Net++ network architecture has a high potential for crop/weed segmentation, especially at the beginning of the growing season, leading to high profitability and cost reduction in agricultural management. In future research, we aim to focus on enhancing the proposed algorithm through data augmentation using generative adversarial networks (GANs).

REFERENCES

Abdalla, A., Cen, H., Wan, L., Rashid, R., Weng, H., Zhou, W., He, Y., 2019: Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure. *Computers and Electronics in Agriculture*, 167, 105091.

Asad, M. H., Bais, A., 2020: Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Information Processing in Agriculture*, 7(4), 535–545.

Bousias Alexakis, E., Armenakis, C., 2020: Evaluation of unet and unet++ architectures in high resolution image change detection applications. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, 1507-1514. doi: 10.5194/ISPRS-ARCHIVES-XLIII-B3-2020-1507-2020.

Deng, W., Huang, Y., Zhao, C., Wang, X., 2014: Discrimination of Crop and Weeds on Visible and Visible/Near-Infrared Spectrums Using Support Vector Machine, Artificial Neural Network and Decision Tree. *Sensors & Transducers*, 26, 26–34.

Fathipoor, H., Arefi, H., Shah-Hosseini, R., Moghadam, H., 2019: Corn forage yield prediction using unmanned aerial vehicle images at mid-season growth stage. *Journal of Applied Remote Sensing,* 13(3), 034503.

Fletcher, R. S., Fletcher, R. S., 2016: Using Vegetation Indices as Input into Random Forest for Soybean and Weed Classification. *American Journal of Plant Sciences*, 7(15), 2186–2198.

Hashemi-Beni, L., Gebrehiwot, A., 2020: Deep learning for remote sensing image classification for agriculture applications. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIV-M-2–2020*, 51–54. doi: 10.5194/ISPRS-ARCHIVES-XLIV-M-2-2020-51-2020.

Ishak, A. J., Mustafa, M. M., Tahir, N. M., Hussain, A., 2008: Weed detection system using support vector machine. *2008 International Symposium on Information Theory and Its Applications,* 1–4. doi: 10.1109/ISITA.2008.4895454.

Jensen, H. G., Jacobsen, L. B., Pedersen, S. M., & Tavella, E., 2012: Socioeconomic impact of widespread adoption of precision farming and controlled traffic systems in Denmark. *Precision Agriculture*, 13(6), 661–677.

Khan, A., Ilyas, T., Umraiz, M., Mannan, Z. I., Kim, H., 2020: CED-Net: Crops and Weeds Segmentation for Smart Farming Using a Small Cascaded Encoder-Decoder Architecture. *Electronics,* 9(10), 1602.

Lameski, P., Zdravevski, E., Trajkovik, V., Kulakov, A., 2017: Weed Detection Dataset with RGB Images Taken Under Variable Light Conditions. *Communications in Computer and Information Science*, 778, 112–119.

Pretto, A., Aravecchia, S., Burgard, W., Chebrolu, N., Dornhege, C., Falck, T., Fleckenstein, F. V., Fontenla, A., Imperoli, M., Khanna, R., Liebisch, F., Lottes, P., Milioto, A., Nardi, D., Nardi, S., Pfeifer, J., Popovic, M., Potena, C., Pradalier, C., … Nieto, J., 2021: Building an Aerial-Ground Robotics System for Precision Farming: An Adaptable Solution. *IEEE Robotics and Automation Magazine*, 28(3), 29–49. doi: 10.1109/MRA.2020.3012492.

Ramirez, W., Achanccaray, P., Mendoza, L. F., Pacheco, M. A. C., 2020: Deep Convolutional Neural Networks for Weed Detection in Agricultural Crops Using Optical Aerial Images. *2020 IEEE Latin American GRSS and ISPRS Remote Sensing Conference(LAGIRS)*, 133–137. doi: 10.1109/LAGIRS48042.2020.9165562.

Rico-Fernández, M. P., Rios-Cabrera, R., Castelán, M., Guerrero-Reyes, H. I., Juarez-Maldonado, A., 2019: A contextualized approach for segmentation of foliage in different crop species. *Computers and Electronics in Agriculture*, 156, 378–386.

Ronneberger, O., Fischer, P., Brox, T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical image computing and computer-assisted intervention.* 9351, 234–241. doi: 10.1007/978-3-319-24574-4_28.

Wang, A., Xu, Y., Wei, X., Cui, B., 2020: Semantic Segmentation of Crop and Weed using an Encoder-Decoder Network and Image Enhancement Method under Uncontrolled Outdoor Illumination. *IEEE Access*, 8, 81724–81734.

Wang, A., Zhang, W., Wei, X., 2019: A review on weed detection using ground-based machine vision and image processing techniques. *Computers and Electronics in Agriculture*, 158, 226–240.

Wu, Z., Chen, Y., Zhao, B., Kang, X., Ding, Y., 2021: Review of Weed Detection Methods Based on Computer Vision. *Sensors,* 21(11), 3647.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J., 2019: UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, 39(6), 1856–1867. doi: 10.48550/arxiv.1912.05074

Zou, K., Chen, X., Wang, Y., Zhang, C., Zhang, F., 2021: A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. *Computers and Electronics in Agriculture*, 187, 106242.