Grenze International Journal of Engineering and Technology, January Issue

Comparative Analysis of U-Net and U-Net++ for Semantic Segmentation of Crop and Weed in Resource Constrained Environments

Sharmila Meinam[1], Kishorjit Nongmeikapam[2] and N. Basanta Singh[3]
[1,3]Department of Electronics and Communication Engineering, Manipur Institute of Technology, Manipur University, Manipur, India
Email: sharmila_meinam@yahoo.com, basanta_n@rediffmail.com
[2]Department of Computer Science and Engineering, Indian Institute of Information Technology Manipur, India
Email: kishorjit@iiitmanipur.ac.in

***Abstract*— In precision agriculture, effective weed management requires precise semantic segmentation of crops and weeds. Optimization of agricultural yield largely depends on the accurate discrimination between crops and weeds. This study presents a comparative analysis of effect of architectural complexity on the performance of deep learning models, specifically U-Net and U-Net++ under limited data and hardware constraints. The publicly available CWFID dataset has been used to train the U-Net and U-Net++ architectures. The dataset undergoes standard pre-processing steps followed by data augmentation to expand the limited dataset which eventually improves the model generalization. Model evaluation has been performed by using different metrics such as pixel accuracy, F1-score, recall, precision and mIoU. Experimental results demonstrate that in comparison to the U-Net++ model which achieved an overall pixel accuracy of 98.16 %, the U-Net architecture obtained an overall pixel accuracy of 98.33 %. Also U-Net outperformed U-Net++ in terms of F1-score, mIoU and recall. This study high-lights the potential of CNN-based segmentation techniques to support sustainable weed management and con-tribute to the advancement of precision agriculture under conditions of constraint computation and limited data availability.**

***Index Terms*— semantic segmentation, augmentation, CNN, deep learning, precision agriculture.**

I. Introduction

The integration of deep learning into precision agriculture has resulted in a significant advancement in the automation of a number of tasks such as monitoring crop health, detection of plant diseases and weed management [1, 2]. In precision farming, weed management is considered as a critical aspect. If left uncontrolled, weeds can drastically reduce productivity of crops and quality. Crop–weed segmentation is one of the most impactful applications in precision agriculture. Effective crop–weed segmentation facilitates site-specific herbicide application (precision spraying). This minimizes the excessive use of herbicide thus minimizing the use of chemicals and its impact in the environment [3]. This ultimately will result in an overall increase in the agricultural productivity. Accurate crop-weed segmentation enables a model to distinguish between crop and weed portions at the pixel level thus forming the backbone of intelligent farming systems.

Grenze ID: 01.GIJET.12.1.201_1

Conventional methods of image processing and classical machine learning approaches have shown limited robust-ness in real-world agricultural environments. Variations in illumination, occlusions caused by overlapping plants, irregular patterns in weed growth, complex soil backgrounds and crop-weed similarities often degrade the perfor-mance of these conventional methods. Such methods are not robust to the diverse field conditions due to their sensi-tivity to lighting changes, image noise, and morphological variation in weeds which ultimately limits their generaliza-bility and accuracy [4].

Over the years, deep learning based techniques, mainly convolutional neural networks (CNNs) have come up as an important tool for segmentation tasks. Deep learning architectures automatically learn rich features from training data, making them more resilient to field variability. Among CNN-based models, the U-Net architecture and its ver-sions have been widely adopted for semantic segmentation in agriculture. Researchers have proposed enhanced net-work architectures based on U-Net's encoder–decoder design for further improvement of segmentation perfor-mance. One such advanced model is U-Net++ [5], an extension of U-Net that introduces nested and dense skip pathways to better fuse multi-scale features across encoder and decoder layers. This architectural refinement in U-Net++ is intended to bridge the semantic gap between low-level and high-level feature maps, leading to more precise segmentation outputs. Many earlier studies have reported that U-Net++ can outperform U-Net in tasks such as fine-grained weed detection and early-stage plant segmentation. However, as we demonstrate in this study, U-Net outper-forms U-Net++ on the CWFID dataset under data-limited and CPU-only training conditions. This highlights the im-portance of task-specific evaluation, especially when deploying deep models in constrained agricultural settings such as limited dataset size, training conditions etc. CWFID dataset offers high-resolution field images with pixel-level annotations of crop vs. weed [6].

In this article, the related work has been presented in section II. Section III discusses the methodology. The experi-mental set-up has been presented in section IV. Next, section V discusses and analyses the experimental results. Lastly, section VI presents the conclusion and the future work.

II. Related Work

Till date, mainly two learning-based methods have been adopted to discriminate between weeds and crops: classical machine learning approaches [7, 8] and deep learning techniques. Crops and weeds being visually very similar, it poses a challenge for classical machine learning methods in extracting and selecting discriminative features. Deep learning techniques, specifically CNNs have gained wide popularity in crop–weed segmentation tasks. Deep learning architectures automatically learn complex and hierarchical features from raw image data. In comparison to classical methods, they offer superior accuracy and robustness which makes them more suitable for real-world agricultural applications. Gao, J. et al. [9] designed a novel methodology for classification of weed and maize. They extracted a total of 185 spectral features. PCA was applied to reduce the redundancy of the features extracted. A Random Forest classifier was then trained on three different feature combinations, with hyperparameter tuning to select the optimal model. Using 30 important spectral features, the optimized Random Forest classifier could achieve mean correct classification rates of 1.0 for Zea mays, a value of 0.752 for Cirsium arvense, 0.691 for Rumex, and 0.789 for Con-volvulus arvensis. Ma, X. et al. [10] proposed a segmentation approach based on the SegNet architecture [11], a fully convolutional network with a symmetric encoder-decoder structure. The methodology was designed to classify rice seedlings, weeds, and background in RGB images captured from paddy fields. The dataset was manually annotated into three classes and class imbalance was solved by a class weighting strategy. The proposed SegNet model could effectively handle small and irregularly shaped seedlings of rice and weeds. The model obtained a classification accuracy of 92.7%, compared to 89.5% for FCN and 70.8% for U-Net. Bosilj, P. et al. [12] explored the potential of transfer learning among deep learning classifiers for different types of crop. Such an approach can reduce both re-training time and manual annotation efforts when adapting to a new type of crop. The segmentation framework thus proposed, referred to as SegNet-Basic, is a simpler version of the SegNet architecture. Evaluation of classification performance across three distinct crop-type datasets containing various weeds has been conducted in this study. Model accuracy is compared using fully annotated pixel-level data versus partially labelled data. Through their study, the authors claim that transfer learning across crop types is feasible and could achieve up to 80% reduction in training time. Gallo, I. et al. [13] introduced a new dataset called Chicory Plant (CP) which contains over 3,000 UAV-captured RGB images and 12,113 annotations for weed detection, specifically Mercurialis annua. The object detection algorithm, YOLOv7 was used for evaluation on both the CP dataset and a public dataset, Lincoln beet. The study reported that YOLOv7 outperformed other YOLO versions. YOLOv7 could achieve strong results on the CP dataset as well as improved detection performance on the LB dataset. The study illustrates that YOLOv7 is effec-tive for weed detection and emphasizes the need for large scale, annotated weed datasets for real-time applications. A multi-task CNN segmentation model, MTS-CNN has been proposed in [14] for detection of crops and weeds. Their experiments used three types of datasets: rice seedling and weed dataset, the BoniRob dataset and CWFID dataset. Their results demonstrated that their approach could achieve higher accuracy compared with the advanced methods.

III. Methodology

A. Dataset and Pre-processing

Crop–Weed Field Image Dataset (CWFID) employed in our study comprises high-resolution RGB images of agricul-tural fields captured under natural lighting conditions. Pairing of each RGB image with a corresponding annotation mask is done that provides pixel-level class labels for three semantic categories: crop, weed, and soil. All the RGB images and their associated annotation masks in the dataset were resized to 256×256 pixels so as to ensure compati-bility with deep learning models and optimize computational efficiency. This uniform resizing to 256×256 pixels standardizes the input dimensions across the dataset and facilitates batch processing during training. Additionally, normalization of the pixel intensity values of the RGB images were performed in the range [0, 1] to improve conver-gence during optimization. Also, a verification step has been performed to ensure correct alignment between each image and its mask. The proposed methodology is presented in Fig. 1.

B. Mask Conversion And Label Encoding

The original annotation masks in the CWFID dataset are colored RGB images, where each class is visually represent-ed by a distinct color code with soil being black, crop being green and weed being red. These color-coded masks were systematically converted to single-channel integer-encoded label masks to make it compatible with multi-class semantic segmentation models. In the single-channel integer-encoded label masks, each pixel is assigned a distinct class index: 0 = soil, 1 = crop and 2 = weed. The resulting integer label masks are compatible with loss functions such as Categorical cross-entropy loss or Dice loss used in multi-class segmentation tasks.

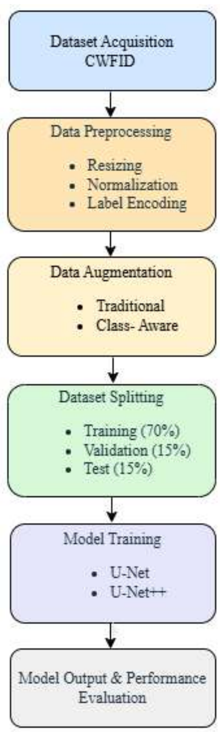

Fig. 1. Proposed methodology framework

C. Data Augmentation

To expand the CWFID dataset, a comprehensive two-stage augmentation strategy was employed. This strategy will enhance the generalization capability of the segmentation models during training.

Stage 1: Traditional Augmentation Techniques

Conventional image transformation techniques were applied uniformly to both the RGB images and their corre-sponding label masks using deterministic random seeds for spatial consistency. These augmentations include: hori-zontal and vertical flips, random rotations (±15–30 degrees), brightness and contrast adjustments, hue and saturation shifts and Gaussian blur. These augmentations enhanced the model's robustness to variations in natural environment commonly observed in field imagery.

Stage 2: Class-Aware Augmentation Techniques

Apart from the conventional image transformations, class-aware augmentations were incorporated to specifically target the semantic regions of interest: soil, crop and weed in the images. Augmentations included in this category are: crop brightness jitter, soil contrast modulation and localized occlusion patches. The purpose of introducing the class-aware techniques is to ensure that the model learned meaningful context-specific features instead of over-fitting on fixed color or texture cues. Thus original dataset comprising of 60 image-masks pairs were expanded to 780 image-masks pairs through the combined application of both augmentation stages.

D. Dataset Splitting

The augmented 780 image–mask pairs were divided into three subsets: training set, validation set and testing dataset. The training set consisted of 546 images (70%), validation set consisted of 117 images (15%) and test set consisted of 117 images (15%). During training, back-propagation algorithm was employed for optimizing the model parame-ters by utilizing the training dataset. The test set was finally employed for final performance evaluation of both the segmentation models under unseen conditions. To ensure reproducibility of experimental results all images and masks pairs in the data-set were pre-shuffled using a fixed random seed before splitting. The same dataset splits were consistently utilized for both U-Net and U-Net++ models to enable a fair comparative analysis.

E. Model Architectures

The evaluation and performance comparison of two CNN architectures: U-Net and U-Net++ designed for semantic segmentation tasks has been presented in this study. Both these models are based on an encoder–decoder frame-work. These architectures are widely adopted by researchers for pixel-wise classification problems, particularly in biomedical and agricultural imaging domains.

U-Net Architecture

The U-Net [15] is a Convolutional Neural Network architecture consisting of a symmetric encoder–decoder structure with skip connections that link each encoder block to its corresponding decoder block. The encoder progressively reduces the spatial dimensions by successive convolution and down-sampling operations while capturing high-level contextual features. The skip connections directly transfer high-resolution, fine-grained spatial information from the encoder block to the decoder. This facilitates in reconstruction of the detailed segmentation masks. It would be diffi-cult to precisely recover the spatial information that was lost during down-sampling without these skip connections. The widespread recognition of U-Net architecture is due to its effectiveness in tasks involving limited training data owing to its efficient use of spatial context and hierarchical features [16]. Fig. 2 depicts the U-Net architecture.

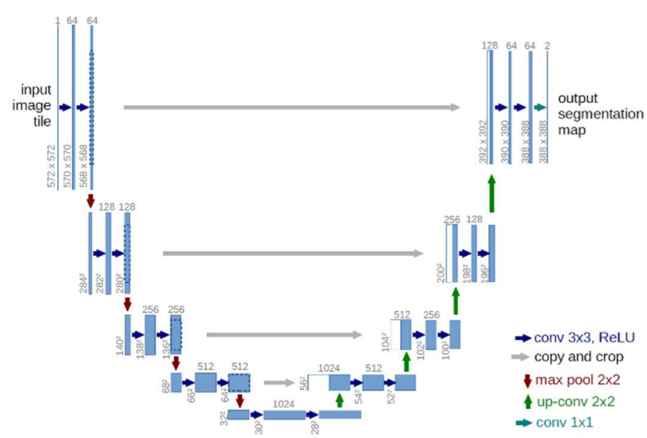

Fig. 2. U-Net architecture [15].

U-Net++ Architecture

As depicted in Fig. 3, the U-Net++ (also called Nested U-Net) [5] is an improved version of the original U-Net archi-tecture for semantic segmentation tasks. Instead of a single skip connection per level in U-Net model, the U-Net++ architecture consists a series of nested and dense convolutional blocks along the skip pathways. These convolution-al blocks reduce the semantic gap between the low-level features of the encoder and the high-level abstractions of the decoder improving feature fusion. The dense connections within these nested blocks also facilitates better flow of gradients during training process which eventually results in efficient learning. As a result, U-Net++ is particularly effective in complex segmentation scenarios, such as medical imaging [17, 18] or agricultural tasks [19], where pre-cise boundary delineation and context-aware classification are critical.

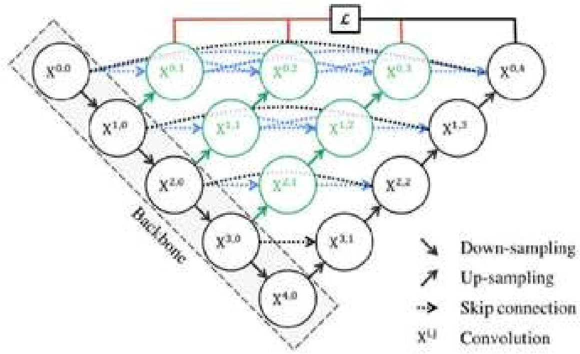

Fig. 3. U-Net++ architecture [5].

Backbone Initialization

Both the U-Net and U-Net++ models were implemented by using ResNet-34 as their encoder backbone [20]. This approach is to take the advantages of transfer learning to ensure faster model convergence. These encoders were initialized with weights already trained on the ImageNet dataset, which is a standard practice in deep learning to leverage generalized feature representations learned from large-scale datasets [21]. This approach enables the models to start from a rich set of generic features instead of learning from scratch, thus accelerating convergence and improving performance [22, 23]. Such an approach is beneficial when training on relatively small datasets like CWFID where over-fitting is a concern and high generalization is critical [24]. The segmentation head for both the models has been configured to output three semantic classes - soil, crop and weed. The Softmax activation function was applied during prediction to generate per-pixel probability maps for multi-class segmentation.

F. Loss Function

Class imbalance [25] often occurs in semantic segmentation using agricultural images whereby certain classes such as soil tend to dominate the pixel distribution compared to the less represented classes like weed. To effectively address such a challenge and improve the performance of pixel-wise classification, a hybrid loss function has been adopted [26]. The total loss during training, denoted as L_{total} is equal to the sum of the dice loss and the cross-entropy loss and is defined by (1). Dice coefficient loss is calculated by (2).

$$L_{total} = L_{Dice} + \lambda\, L_{CrossEntropy}\,. \tag{1}$$

where L_{Dice} = Dice co-efficient loss

$L_{CrossEntropy}$ = Cross-entropy loss

λ = a weight factor balancing the contributions of the two losses.

$$L_{Dice} = 1 - \frac{1}{C}\sum_{c=1}^{C} \frac{2\sum_{n=1}^{N} y_{n,c}\, p_{n,c}}{\sum_{n=1}^{N} y_{n,c} + \sum_{n=1}^{N} p_{n,c}}\,. \tag{2}$$

The cross-entropy loss awards penalty for incorrect pixel level class predictions while encouraging higher probabilities for the correct class. The cross-entropy loss for multi-class problems is defined by (3) as

$$L_{CrossEntropy} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} w_c\, y_{n,c}\, log\left(p_{n,c}\right)\,. \tag{3}$$

where w_c = weight for class c.

In-order to mitigate the effects of class imbalance, class weights were integrated into the cross-entropy loss function. These weights give higher significance to the minority classes such as weed and crop. This strategy helps reduce the model's tendency to make predictions in favour of the majority class such as soil in our study thereby enhancing its ability to accurately detect less represented regions such as weed and crop in the agricultural field.

IV. Experimental Set-Up

All the experiments were performed using a uniform training set-up. This ensures a fair comparison in evaluating the performance of both the models on the CWFID dataset. Python programming language was used to develop and train all the models with PyTorch serving as the deep learning framework along with the segmentation_models_pytorch (SMP) library. Both the models were trained using a consistent set of hyperparameters to ensure stable and effective learning. The hyperparameters given in Table I are used to train U-Net and U-Net++ semantic segmentation models. Throughout the whole training process, a fixed learning rate was used without scheduling since the empirical results demonstrated stable convergence. The training of the models were performed on a CPU-based system equipped with Intel-based Processor and 16GB RAM. Even though there was no GPU support, the use of a small input resolution (256×256) and a compact model backbone (ResNet-34) has resulted in feasible training durations though longer training times were required.

TABLE I. HYPERPARAMETERS USED TO TRAIN U-NET AND U-NET++ ARCHITECTURE.

Encoder Architecture	ResNet-34
Optimizer	Adam optimizer
Loss function	Combined dice loss + weighted cross entropy loss
Initial Learning Rate	1×10^{-3}
Batch size	8
Number of epochs	200
Number of classes	3(soil, crop, weed)
Activation function	Softmax (applied during inference)

V. Results and Discussion

A. Quantitative Metrics

Post training, the performance evaluation of the segmentation models were conducted using standard metrics: accu-racy, recall, F1-score, precision and mean Intersection-over-Union (mIoU). Equation (4) – (9) presents the corre-sponding equations for the evaluation metrics. Here, TP is the True Positive, TN is the True Negative, FP is False Positive and FN is the False Negative for a particular class. Table II summarizes the evaluation results for both the U-Net and U-Net++ architectures.

$$\text{Accuracy} = \frac{TP+}{TP+TN+FP+FN}\ . \tag{4}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{6}$$

$$\text{F1} - \text{Score} = \frac{2\text{ x Precision x Recall}}{\text{Precision}+\text{Recall}} \tag{7}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{8}$$

$$mIoU = \frac{1}{C}\sum IoU \tag{9}$$

where C is the number of classes and C = 3: (soil, crop, weed) in our study.
As observed from Table II, both the models demonstrated strong segmentation performances across all evaluation metrics. U-Net achieved a slightly higher overall pixel accuracy of 98.33% compared to U-Net++ with a pixel accu-racy value of 98.16% and recall of 92.26% compared to a value of 90.72% for U-Net++, which indicates its ability to accurately identify a high proportion of crop and weed pixels. The value of

precision for U-Net is 90.43% while for U-Net++ the corresponding value is 90.77 % indicating that U-Net++ has better control over false positives. In terms of F1-score, U-Net had a slightly higher F1-score value of 91.30% as compared to U-Net++ with a value of 90.73%. Similarly, the mIoU score for U-Net, 84.54% was better as compared to U-Net++ with a value of 83.63% indicating improved average overlap between predicted and ground truth segments across the three classes: soil, crop and weed.

As demonstrated by these results, U-Net architecture slightly outperformed U-Net++ in this specific application and dataset. Several factors such as relatively small dataset size, less diversity in training samples and CPU based training environment might have been responsible for this outcome. The U-Net architecture with its simpler structure may have generalized better under the above mentioned constraints thereby avoiding overfitting and converging more efficiently. The findings of our study demonstrates that deeper or more complex architectures may not always yield enhanced performance particularly under constraint conditions or specific dataset. This emphasizes the importance of context-specific model selection and the need for careful benchmarking before adopting more complex architec-tures.

B. Visual Results

A visual inspection of the predicted segmentation masks was conducted for selected test samples. Here, each image has been presented as a triplet: original RGB input image, ground truth annotation and the model's predicted mask. Fig. 4 presents a qualitative comparison between the ground truth annotations and the segmentation outputs for U-Net and U-Net++ for three representative test images. Column (a) represents the RGB input images, column (b) rep-resents the ground truth annotations, the predicted masks from U-Net and U-Net++ models are presented in column (c) and (d) respectively. The semantic classes: soil, crop and weed are visualized in blue, green and red respectively. From these test images we can draw the inference that U-Net demonstrates superior performance in preservation of object boundaries, better continuity in crop and weed regions and fewer misclassifications, particularly in thin and detailed structured areas and the occluded regions. These visual results align with our quantitative findings, where U-Net achieved higher overall accuracy, mIoU, F1-score, and recall compared to U-Net++ on the test set.

TABLE II. EVALUATION RESULTS FOR U-NET AND U-NET++ ARCHITECTURES.

Model	Accuracy (%)	F1-Score (%)	Recall (%)	Precision (%)	mIoU (%)
U-Net	98.33	91.30	92.26	90.43	84.54
U-Net++	98.16	90.73	90.72	90.77	83.63

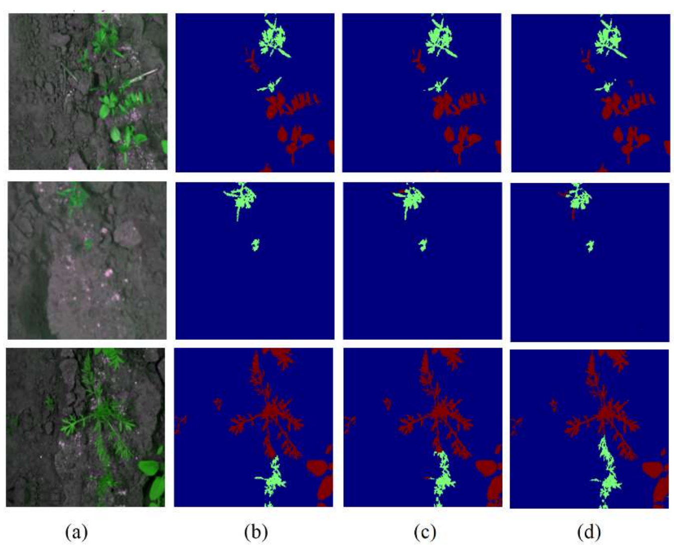

Fig. 4. Segmentation results of the U-Net and U-Net++ models.

C. Training Curves

The assessment of the training dynamics of both the models were done by plotting epoch-wise trends of training and validation accuracy, loss, and mIoU across 200 epochs. These plots provide valuable information about the conver-gence behaviour of the models, generalization capability and possible tendencies of overfitting. Fig. 5

and Fig. 6 shows the plots of training and validation accuracy, loss and mIoU over 200 epochs for both the U-Net and U-Net++ models respectively. As observed from these curves, both models converges within about 100 epochs. However, training of the models was continued till 200 epochs for further fine-tuning and marginal improvements. The corre-sponding curves for U-Net exhibit more stability and were balanced across all metrics of the training and validation phases. U-Net++ exhibited greater fluctuations and significant signs of overfitting as shown by the larger gaps be-tween training and validation metrics and the rising validation loss.

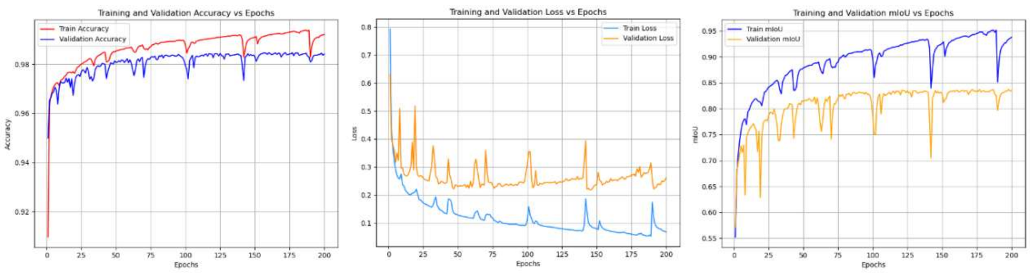

Fig. 5. U-Net model training and validation curves

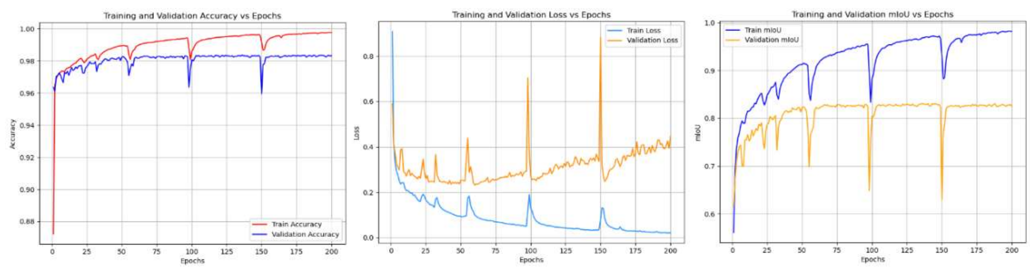

Fig. 6. U-Net++ model training and validation curves.

D. Confusion Matrix

Fig. 7 presents the normalized confusion matrices for U-Net and U-Net++ models on the test set which highlights their per-class classification performance. Both models could identify soil pixels with high accuracy, with U-Net++ achieving 99.00% and U-Net 98.94%. However, U-Net is superior in discriminating between crop and weed classes. The misclassification of crop as soil (11.80%) and weed as soil (9.24%) is higher in U-Net++ as compared to U-Net which reduces these errors to 7.61% and 6.04% respectively. Additionally, the recall values for both crop (84.87%) and weed (92.96%) is slightly better for U-Net in comparison to U-Net++ whose corresponding values are 83.31% and 89.86% respectively. From these results we can conclude that although the generalization is better for U-Net++ for the dominant background class (soil), the overall segmentation accuracy in vegetation regions is better for U-Net. The results obtained supports our findings that U-Net outperforms U-Net++ in distinguishing crop and weed regions in particular which is crucial for precise weed control in precision agriculture.

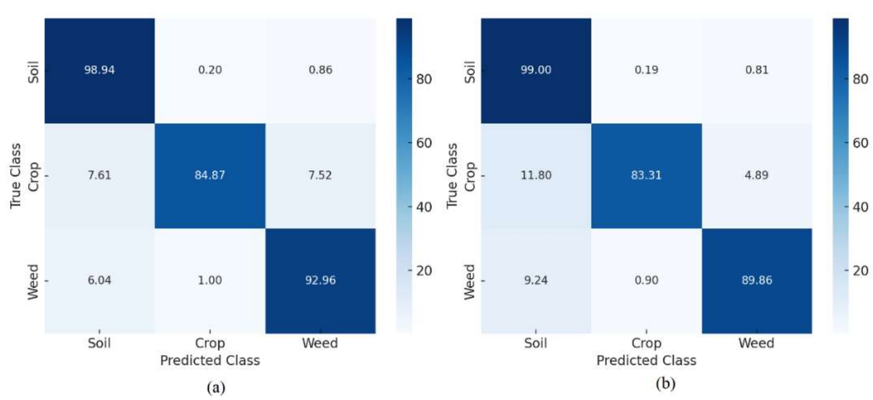

Fig. 7. Normalized Confusion matrix for (a) U-Net and (b) U-Net++ models.

VI. CONCLUSION AND FUTURE WORK

A comparative analysis of U-Net and U-Net++ architectures for semantic segmentation of crop–weed imagery using the CWFID dataset has been presented in this study. This study compares how the complexity of U-Net and U-Net++ architectures affects segmentation accuracy, boundary precision, and generalization under constrained data and hardware conditions. Though U-Net++ is an advanced variant of U-Net with nested skip connections, its added com-plexity did not yield significant performance improvements in our experiments. The model rather exhibited a higher tendency to overfit. The results demonstrated that U-Net consistently outperformed U-Net++ in terms of overall ac-curacy, F1-score, mIoU and recall. These findings suggest that simpler architectures like U-Net may offer a more effective balance of performance and efficiency when training data is limited and computational resources are con-strained. Our future experiments will involve testing the trained models on agricultural datasets such as Sugar Beets 2016 and DeepWeeds which exhibits greater variation in crop type, weed species etc. To make it applicable in smart farming systems, our future work will also focus on deploying the trained models to resource-constrained edge de-vices.

REFERENCES

[1] Murad, N. Y., Mahmood, T., Forkan, A. R. M., Morshed, A., Jayaraman, P. P., & Siddiqui, M. S., "Weed De-tection using Deep Learning: A Systematic Literature Review," Sensors, vol. 23, no.7, p. 3670, 2023.

[2] Shin, J., Mahmud, M. S., Rehman, T. U., Ravichandran, P., Heung, B., & Chang, Y. K., "Trends and Prospect of Machine Vision Technology for Stresses and Diseases Detection in Precision Agricul-ture," AgriEngineering, vol. 5, no.1, pp. 20-39, 2022.

[3] Fathipoor, H., Shah-Hosseini, R., & Arefi, H., "Crop and Weed Segmentation on Ground-Based Images using Deep Convolutional Neural Network," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial In-formation Sciences, vol. 10, pp.195-200, 2023.

[4] Khan, S. D., Basalamah, S., & Lbath, A., "Weed–Crop Segmentation in Drone Images with a Novel Encoder–Decoder Framework Enhanced via Attention Modules," Remote Sensing, vol. 15, no. 23, p. 5615, 2023.

[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A Nested U-net Architecture for Medical Image Segmentation," in Deep learning in medical image analysis and multimodal learning for clinical deci-sion support: 4th international workshop, pp. 3–11, 2018.

[6] Ahmad, S., Chen, Z., Ikram, S., & Ikram, A., "AI-Enabled Vision Transformer for Automated Weed Detection: Advancing Innovation in Agriculture," International Journal of Advanced Computer Science & Applications, vol.15, no.12, 2024.

[7] Chen, Y., Wu, Z., Zhao, B., Fan, C., & Shi, S., "Weed and Corn Seedling Detection in Field Based on Multi Feature Fusion and Support Vector Machine," Sensors, vol. 21, no.1, p. 212, 2020.

[8] Tufail, M., Iqbal, J., Tiwana, M. I., Alam, M. S., Khan, Z. A., & Khan, M. T., " Identification of Tobacco Crop Based on Machine Learning for a Precision Agricultural Sprayer," IEEE access, vol. 9, pp. 23814-23825, 2021.

[9] Gao, J., Nuyttens, D., Lootens, P., He, Y., & Pieters, J. G., " Recognising Weeds in a Maize Crop using a Ran-dom Forest Machine-Learning Algorithm and Near-Infrared Snapshot Mosaic Hyperspectral Imagery," Biosys-tems engineering, vol. 170, pp. 39-50, 2018.

[10] Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y., & Xing, X., "Fully Convolutional Network for Rice Seed-ling and Weed Image Segmentation at the Seedling Stage in Paddy Fields," PloS one, vol. 14, no. 4, p. e0215676, 2019.

[11] Badrinarayanan, V., Kendall, A., & Cipolla, R., "Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481-2495, 2017.

[12] Bosilj, P., Aptoula, E., Duckett, T., & Cielniak, G., "Transfer Learning Between Crop Types for Semantic Seg-mentation of Crops versus Weeds in Precision Agriculture," Journal of Field Robotics, vol. 37, no. 1, pp. 7-19, 2020.

[13] Gallo, I., Rehman, A. U., Dehkordi, R. H., Landro, N., La Grassa, R., & Boschetti, M., "Deep Object Detection of Crop Weeds: Performance of YOLOv7 on a Real Case Dataset from UAV Images," Remote Sensing, vol. 15, no. 2, p. 539, 2023.

[14] Kim, Y. H., & Park, K. R., "MTS-CNN: Multi-task Semantic Segmentation-Convolutional Neural Network for Detecting Crops and Weeds," Computers and Electronics in Agriculture, vol. 199, p. 107146, 2022.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmenta-tion," in Proceedings of MICCAI, vol. 9351, pp. 234–241, 2015.

[16] Dimitrovski, I., Spasev, V., Loshkovska, S., & Kitanovski, I., "U-net Ensemble for Enhanced Semantic Seg-mentation in Remote Sensing Imagery," Remote Sensing, vol. 16, no. 12, p. 2077, 2024.

[17] Zhao, C., Shuai, R., Ma, L., Liu, W., & Wu, M., "Segmentation of Skin Lesions Image Based on U-Net++," Multimedia Tools and Applications, vol. 81, no. 6, pp. 8691-8717, 2022.

[18] Micallef, N., Seychell, D., & Bajada, C. J., "Exploring the U-net++ Model for Automatic Brain Tumor Segmen-tation," IEEE Access, vol. 9, pp. 125523-125539, 2021.

[19] Sahoo, P., Paikaray, P., Sahu, T. C., Patnaik, R. R. K., & Barik, R. C., "Cascading Computer Vision and Resnet Families with U-Net++ for Plant Leaf Disease Detection and Classification: A Novel Deep Learning Model," In International Conference on Computational Intelligence in Pattern Recognition, pp. 435-451, 2024.
[20] Saviolo, A., Bonotto, M., Evangelista, D., Imperoli, M., Lazzaro, J., Menegatti, E., & Pretto, A., "Learning to Segment Human Body Parts with Synthetically Trained Deep Convolutional Networks," In International Con-ference on Intelligent Autonomous Systems, pp. 696-712, 2021.
[21] Girshick, R., "Fast R-cnn.," In Proceedings of the IEEE international conference on computer vision pp. 1440-1448, 2015.
[22] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J., "Convolu-tional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," IEEE transactions on med-ical imaging, vol. 35, no. 5, pp. 1299-1312, 2016.
[23] He, K., Girshick, R., & Dollár, P., "Rethinking Imagenet Pre-training," In Proceedings of the IEEE/CVF inter-national conference on computer vision, pp. 4918-4927, 2019.
[24] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson, "How Transferable are Features in Deep Neural Networks?," Advances in Neural Information Processing Systems, vol. 27, pp. 3320–3328, 2014.
[25] Das, M., & Bais, A., "DeepVeg: Deep Learning Model for Segmentation of Weed, Canola, and Canola Flea Beetle Damage," IEEE access, vol. 9, pp. 119367-119380, 2021.
[26] Yeung, M., Sala, E., Schönlieb, C. B., & Rundo, L., "Unified Focal Loss: Generalising Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation," Computerized Medical Imaging and Graphics, vol. 95, p. 102026, 2022.