Deep Learning-Based Semantic Segmentation of Paddy Field Weeds: A Comparative Study of Multi-Model Performance and Perspectives for Smart Agriculture

Cheng-Yu Wen[1*], Bing-Wei Lin[2]

[1,2]Dept. of Computer Science and Information Engineering, National Taitung University, Taitung, Taiwan

*Corresponding Author: littlewen6@gmail.com

DOI: https://doi.org/10.30211/JIC.202503.009

Submitted: Apr. 19, 2025 Accepted: Jun. 10, 2025

ABSTRACT

This paper presents a comparative analysis of six semantic segmentation models for rice seedling and weed identification: SegNet, EfficientNet-SegNet, MobileNet-SegNet, DeepLabV3-ResNet50, DeepLabV3-ResNet101, and SegFormer. Motivated by the need for automated solutions in precision agriculture to reduce labor, improve crop management, and reduce herbicide usage, this study addresses the challenges of identifying morphologically similar and spatially mixed weeds and rice seedlings. The aim was to evaluate the segmentation performance and deployment feasibility under class imbalance and complex background conditions typical of paddy field imagery. The metrics include pixel accuracy, mean Intersection over Union (mIoU), per-class IoU, model complexity, and inference efficiency. SegNet achieved the highest mIoU (0.7204) and outperformed CED-Net (0.7105). MobileNet-SegNet balances the accuracy and speed, whereas SegFormer delivers competitive accuracy with the lowest parameter count and FLOPs. Paired t-tests ($p < 0.0001$) confirmed the statistical significance of the performance differences, offering practical insights for selecting models in resource-constrained agricultural settings.

Keywords: Semantic segmentation, Weed detection, Precision agriculture, Model efficiency, Deep learning

1.Introduction

Weed proliferation in paddy fields remains a critical obstacle to agricultural productivity, particularly in Taiwan, where rice is the primary food source. The presence of weeds not only compromises crop yield, but also deteriorates rice quality. Conventional control strategies such as the use of chemical herbicides may pose serious risks to ecological systems and human health. Consequently, the integration of precision agriculture techniques, especially deep-learning-based semantic segmentation, has emerged as a promising and environmentally sustainable approach for accurate weed detection and management in rice cultivation [1].

Precision agriculture is a technology- and information-based farm management system aimed at accurately monitoring field climate and soil conditions and assessing the effectiveness of cultivation practices and their impact on crop growth. Its goals include reducing resource waste, stabilizing production quality, generating economic profit, and protecting the ecological environment, thereby promoting sustainable development of agriculture [2].

Weed control plays a vital, yet challenging, role in paddy field management. Weeds compete with rice for sunlight, water, and nutrients, leading to reduced crop yield. Common highly competitive weeds include *Monochoria vaginalis*, *Echinochloa crus-galli*, and *Sagittaria trifolia* [3]. Traditional weed control methods primarily rely on the manual removal and application of chemical herbicides. However, these methods face challenges, such as labor intensity, environmental pollution, and negative impacts on non-target organisms.

With the advancement of artificial intelligence (AI) technologies, deep learning has shown increasing value in agricultural image analysis. Deep learning techniques can process and analyze various agricultural data sources, including remote sensing imagery, meteorological data, soil information, and crop growth conditions. These capabilities contribute to optimizing agricultural production, enabling precise resource management, and improving crop yields. Semantic segmentation, a subfield of deep learning, can classify every pixel in an image, allowing for the precise identification of crops and weeds. This provides a novel technical approach for weed management in paddy fields [4].

Ma et al. (2019) proposed a SegNet-based model and provided a high-quality dataset for agricultural image processing [5]. Khan et al. (2020) introduced a cascaded encoder-decoder network, CED-Net, which was further improved by existing methods and significantly enhanced the mean Intersection over Union (mIoU) [6]. The motivation and objective of this study was to investigate the performance of both mainstream and state-of-the-art segmentation models in distinguishing rice seedlings from weeds. By comparing the pixel accuracy and mean IoU across different models, this study aims to provide practical insights for future applications in agricultural automation. Ultimately, this will contribute to the advancement of smart agriculture, enhance production efficiency, and support sustainable agricultural development.

2. Dataset Description

2.1 Data Source

This study utilized the publicly available semantic segmentation dataset of rice seedlings and weeds provided by Ma et al. (2019) [7]. The dataset consisted of RGB images captured in paddy fields located in Jiangmen, Guangdong Province, China. Each image was manually annotated into three categories: background, rice seedling, and weed.

Class 1 (rice seedlings): Rice sprouts grown in paddy fields.

Class 2 (Background): This includes water surfaces, soil, and miscellaneous objects.

Class 3 (weeds): Non-rice plants within the paddy field.

Class 4 (Other): Rare or ambiguous regions that do not clearly belong to these three categories.

In this study, only the first three classes—rice seedlings, background, and weeds— were considered for analysis because they are directly relevant to the target task of rice weed discrimination. Pixels labeled as Class 4 were excluded during model training and evaluation to ensure that the focus remained on agriculturally significant categories.

2.2 Data Processing

The original image resolution was 3648 × 2048 pixels, which is relatively large and may lead to memory exhaustion when fed directly into neural networks, potentially hindering the training process. To address this issue, each image was cropped into eight smaller patches with a resolution of 912 × 1024 pixels. This cropping strategy yielded 224 valid images that were used for both training and testing. Each cropped image is paired with its corresponding ground-truth annotation map, ensuring accurate pixel-level labeling for the segmentation task.

To facilitate the generalization ability of the model, the dataset was randomly split into a training set and testing set at an 80:20 ratio. Specifically, 180 images were designated for training, whereas the remaining 44 images were used for testing. Given that background pixels represent over 80% of the total pixels in the images, there exists a significant class imbalance, which could negatively affect the model's performance, especially in correctly identifying smaller, less frequent classes, such as rice seedlings and weeds. To mitigate this, a weighted cross-entropy loss function was employed, which assigned higher weights to underrepresented classes. This adjustment ensured that the model was more sensitive to these minority classes during training. Furthermore, to optimize the convergence speed and model stability, mean-standard deviation normalization was applied to the pixel values, scaling them into a standard range, and enhancing the efficiency of the training process.

2.3 Introduction to Deep Learning Models Used

2.3.1 SegNet

SegNet is a semantic segmentation model based on a Fully Convolutional Network (FCN) that adopts an encoder-decoder architecture [8]. The encoder utilizes the first 13 convolutional layers of VGG16 to extract the image features, whereas the decoder progressively reconstructs the image from these features. The key innovation of SegNet is the use of max-pooling indices during the upsampling process, which helps preserve edge information and improves segmentation accuracy. As an earlier foundational model, SegNet has been widely recognized and extensively applied in the field of semantic segmentation. Its relatively simple architecture makes it easy to implement and effective across various image segmentation tasks. Therefore, SegNet was selected as the baseline model to facilitate the evaluation of performance improvements in the other models.

2.3.2 EfficientNet

In resource-constrained environments, lightweight models are of significant importance. EfficientNet was designed to reduce computational complexity and improve runtime efficiency, thereby meeting the demands of real-world applications [9].

EfficientNet comprises a family of convolutional neural networks optimized through an automated architecture search and a compound scaling strategy, achieving high efficiency in terms of both parameter count and computational cost. By employing EfficientNet as the encoder in SegNet, it is possible to retain the model performance while significantly reducing the number of parameters and the demand for computational resources.

2.3.3 MobileNet

In agricultural automation and drone-based applications, computational resources are often limited. MobileNet is a lightweight convolutional neural network designed for mobile and embedded devices. It reduces the number of parameters and computational load using depthwise separable convolutions [10]. Employing MobileNet as the encoder in SegNet can further reduce the resource requirements, making the model suitable for deployment on devices with limited computational capabilities.

2.3.4 DeepLabv3_ResNet50 and ResNet101

DeepLabv3 is an advanced semantic segmentation model that employs atrous convolution to expand the receptive field of the convolutional kernels, thereby capturing more contextual information [11]. ResNet50 and ResNet101 serve as the backbone networks for DeepLabv3, utilizing residual connections to alleviate the vanishing gradient problem in deep networks and enhance the learning capability of the model. The combination of DeepLabv3 with ResNet's deep feature extraction capabilities enables a superior performance in complex semantic segmentation tasks. These models are particularly useful for evaluating the effectiveness of deep networks in distinguishing rice seedlings from weeds, particularly in scenarios with complex backgrounds and detailed patterns.

2.3.5 SegFormer

In recent years, transformers have achieved remarkable success in natural language processing, and their application in image processing has garnered increasing attention. Introducing SegFormer allows for the exploration of the potential of transformer architectures in rice seedling and weed recognition and enables comparison with traditional convolutional models.

SegFormer is a semantic segmentation model based on transformer architecture, which leverages the powerful global feature modeling capability of transformers to capture long-range dependencies in images [12].

This study selected a diverse range of representative deep learning models, ranging from foundational architectures to lightweight designs, and from deep convolutional networks to transformer-based frameworks, to comprehensively evaluate their effectiveness in rice seedling and weed recognition and to provide a reference for future applications in agricultural automation.

3. Experimental Results and Discussion

3.1 Overall Model Performance Comparison

The following section presents a comparison of the performance of each model on the test set, including Pixel Accuracy (ACC) and Mean Intersection over Union (mIoU).

Table 1. Overall performance of each model on the test set

Model	**ACC (mean ± std)**	**mIoU (mean ± std)**
SegNet	**0.9486 ± 0.0238**	**0.7204 ± 0.0560**
EfficientNetSegNet	0.9368 ± 0.0258	0.6898 ± 0.0543
MobileNetSegNet	0.9379 ± 0.0251	0.6914 ± 0.0516
DeepLabV3-ResNet50	0.8932 ± 0.0445	0.5479 ± 0.0894

DeepLabV3-ResNet101	0.8534 ± 0.0531	0.3283 ± 0.0424
SegFormer	0.9218 ± 0.0347	0.6471 ± 0.0707
CED-Net [6]	-	0.7105

In terms of the overall performance, SegNet demonstrated the most stable and superior results in the segmentation of rice seedlings and weeds, achieving a pixel accuracy of 0.9486 and a mean IoU of 0.7204. These results suggest that, with appropriate adjustments, this architecture can effectively balance fine boundary preservation with the overall recognition accuracy. In comparison, although EfficientNet-SegNet and MobileNet-SegNet performed slightly lower on both metrics (achieving 0.9368 / 0.6898 and 0.9379 / 0.6914, respectively), they still maintained a satisfactory prediction level. Furthermore, their lightweight characteristics render them particularly suitable for deployment in agricultural environments with limited computational resources.

On the other hand, SegFormer achieved a pixel accuracy of 0.9218 and a mean IoU of 0.6471, indicating that the transformer architecture has significant potential for capturing long-range dependencies in images. With further optimization of the training strategies and loss function weighting, there remains considerable room for performance improvement.

Regarding the DeepLabV3 models, both ResNet50 and ResNet101 variants underperformed significantly in comparison to the other models. Specifically, DeepLabV3-ResNet50 achieved a pixel accuracy of 0.8932 and a mean IoU of 0.5479, whereas DeepLabV3-ResNet101 showed even poorer performance with an mIoU of only 0.3283. One possible reason for this underperformance could be the high computational complexity and model size associated with these architectures, especially in comparison to more lightweight alternatives, such as MobileNet-SegNet. The depth and complexity of ResNet50 and ResNet101 may lead to difficulties in convergence when trained on relatively small datasets, as the model may overfit or fail to effectively learn relevant features from paddy field images. Additionally, the original DeepLabV3 architecture, despite being powerful for general segmentation tasks, might not be well suited to the specific task of paddy field segmentation, where fine-grained differences between rice seedlings and weeds need to be captured. This indicates the need for further architectural adjustments or training strategies tailored to the nuances of agricultural imaging.

To further illustrate the prediction differences among the models on actual images, Figures 1 and 2 present the original images, ground truth labels, and predicted outputs under scenarios where rice seedlings and weeds coexist. Figure 1 includes a variety of weed species alongside rice seedlings, providing a comprehensive test of model performance, whereas Figure 2 depicts a simpler and more field-realistic scenario.

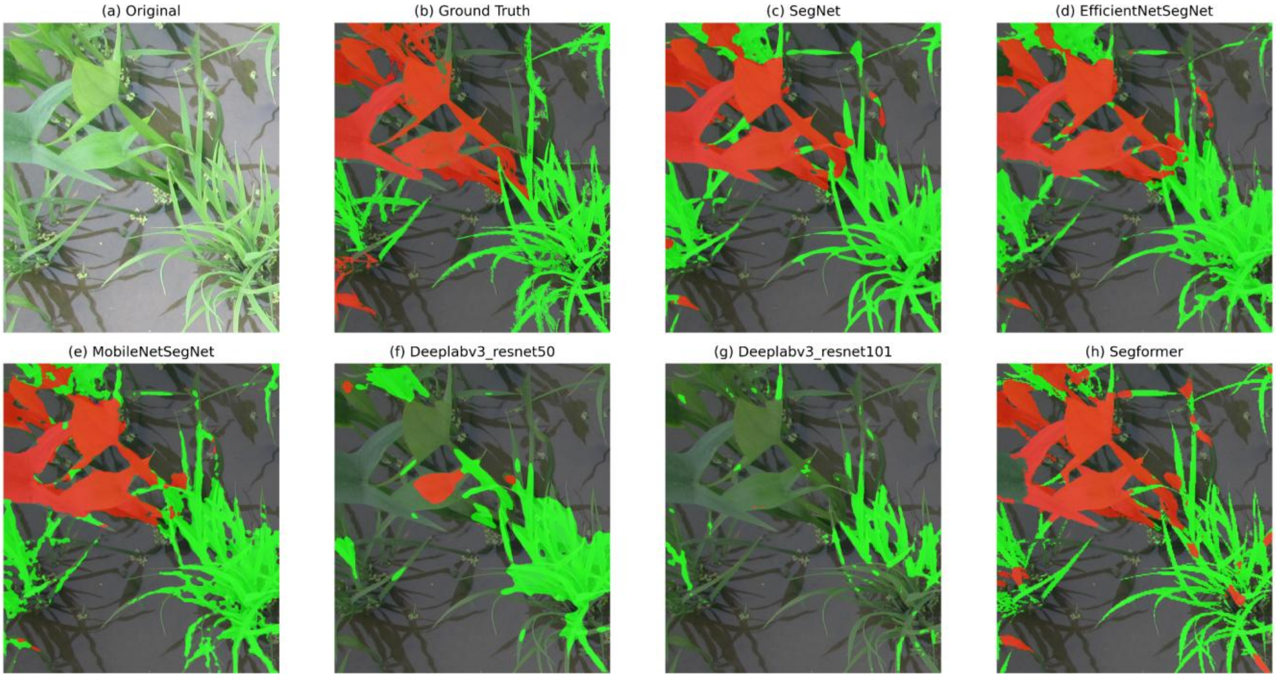

Figure 1. A scenario where rice seedlings and weeds coexist, showing segmentation results of different models for rice and weed identification (green indicates rice seedlings, red indicates weeds). (a) Original image; (b) Ground truth annotation. (c)–(h) show the prediction results of SegNet, EfficientNet-SegNet, MobileNet-SegNet, DeepLabv3 ResNet50, DeepLabv3 ResNet101, and SegFormer, respectively.

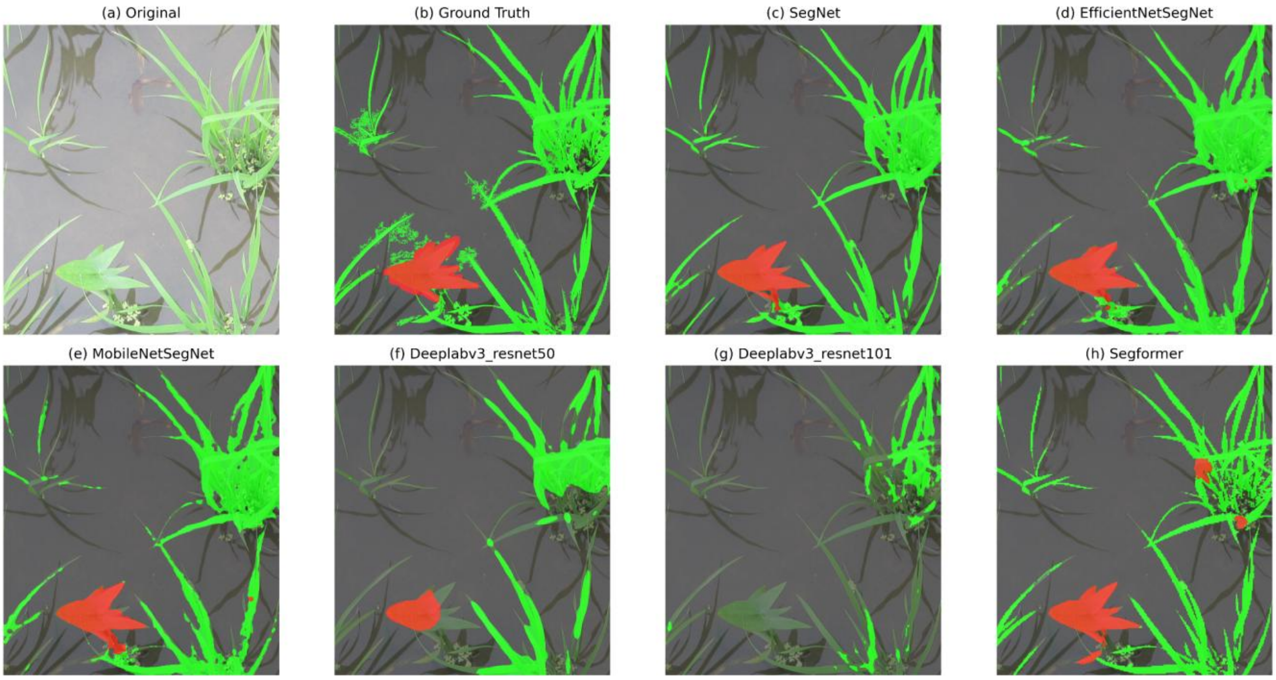

Figure 2. Another scenario where rice seedlings and weeds coexist, showing segmentation results of different models for rice and weed identification (green indicates rice seedlings, red indicates weeds). (a) Original image; (b) Ground truth annotation. (c)–(h) show the prediction results of SegNet, EfficientNet-SegNet, MobileNet-SegNet, DeepLabv3 ResNet50, DeepLabv3 ResNet101, and SegFormer, respectively.

It is worth noting that the number of pixels in Class 1 (rice seedlings), Class 2 (background), and

Class 3 (weeds) are 98,077, 727,848, and 45,610, respectively. The background class (Class 2) accounts for 83.5% of all pixels, while rice seedlings and weeds account for only 11.2% and 5.2%, respectively. Under such a severely imbalanced distribution, a model predicting the entire image as background would still achieve approximately 83.5% pixel accuracy. In this context, the result of DeepLabv3 ResNet101 reaches only 85.34%, which is nearly equivalent to predicting all pixels as background. A similar outcome is observed for DeepLabv3 ResNet50.This suggests that due to the overwhelming presence of background pixels in paddy field images (around 83.5%), the model tends to overfit toward predicting background during training. This overfitting occurs because the model's optimization process gravitates toward minimizing the loss caused by the abundant background class, neglecting the minority classes (rice seedlings and weeds). Moreover, if the hyperparameter settings of atrous convolutions are not well adapted to such highly imbalanced and boundary-sensitive scenarios, the model may fail to effectively learn minority classes, such as weeds, which require more precise boundary delineation. The observed behavior is further validated by visual inspection of the predicted outputs in Figures 1 and 2. In Figure 1, where multiple weed species coexist with rice seedlings, DeepLabv3 models fail to properly segment the weeds, often classifying them as background, which aligns with the statistical data. The segmentation boundaries between rice seedlings and weeds are poorly defined, indicating that the model's inability to properly handle class imbalances led to poor performance in segmenting these smaller, less frequent classes. Similarly, in Figure 2, a simpler scenario with fewer weed species shows that even in this less complex setting, the model still struggles with the boundary delineation, further emphasizing the difficulty of handling highly imbalanced pixel distributions in paddy field images.

3.2 IoU Analysis by Class

Table 2. IoU Performance of Each Model for Rice Seedlings, Background, and Weeds

Model	**Class 1(Seedlings)IoU**	**Class 2 (Background) IoU**	**Class 3 (Weeds)IoU**
SegNet	**0.5871**	0.9006	**0.6735**
EfficientNet-SegNet	0.5353	0.8887	0.6453
MobileNet-SegNet	0.5424	0.8893	0.6424
Deeplabv3-Resnet50	0.4090	0.8556	0.3790
Deeplabv3-Resnet101	0.1656	0.7950	0.0243
Segformer	0.5143	**0.9075**	0.5195

Due to the high proportion of Class 2 (background) pixels in the dataset, the segmentation of background is relatively easier for the models, with most achieving IoU scores above 0.79, and some models even surpassing 0.90. In contrast, Class 1 (rice seedlings) and Class 3 (weeds) exhibit moderate to low IoU values, highlighting the difficulty models face in distinguishing these smaller and morphologically similar classes. This is particularly evident in the low performance on weeds, which are often indistinguishable from rice seedlings due to their fine structures and similarity in appearance.

Notably, DeepLabv3 ResNet101 achieves an extremely low IoU of 0.0243 for weeds, suggesting that it largely failed to learn discriminative features for this class. This aligns with its overall poor performance in terms of mIoU, where it struggles significantly with class imbalance. Its tendency to predict the background class predominates, leaving the minority classes underrepresented.

While SegFormer shows a higher overall performance with an IoU of 0.5195 for weeds and a strong background IoU of 0.9075, its performance in distinguishing weeds from rice seedlings in complex scenarios remains suboptimal. Despite its ability to handle background segmentation well, SegFormer struggles in regions where rice seedlings and weeds are densely mixed. This is particularly evident in scenarios like the one presented in Figure 3, where rice seedlings and weeds coexist in close proximity. The model's prediction results show that while SegFormer can segment background effectively and handle large areas of rice seedlings, it faces challenges when attempting to segment weeds in such intricate settings.

The case in Figure 3 illustrates these difficulties more clearly. In this high-density situation, where weeds and rice seedlings are intermixed, the models, including SegFormer, tend to blur the boundaries between these classes. The fine-grained structures of weeds are particularly difficult to delineate from rice seedlings, leading to reduced accuracy in segmentation. Thus, while SegFormer's transformer-based architecture holds promise in handling global dependencies, it still faces limitations in distinguishing fine details, particularly when the two classes share highly similar features in complex, dense environments.

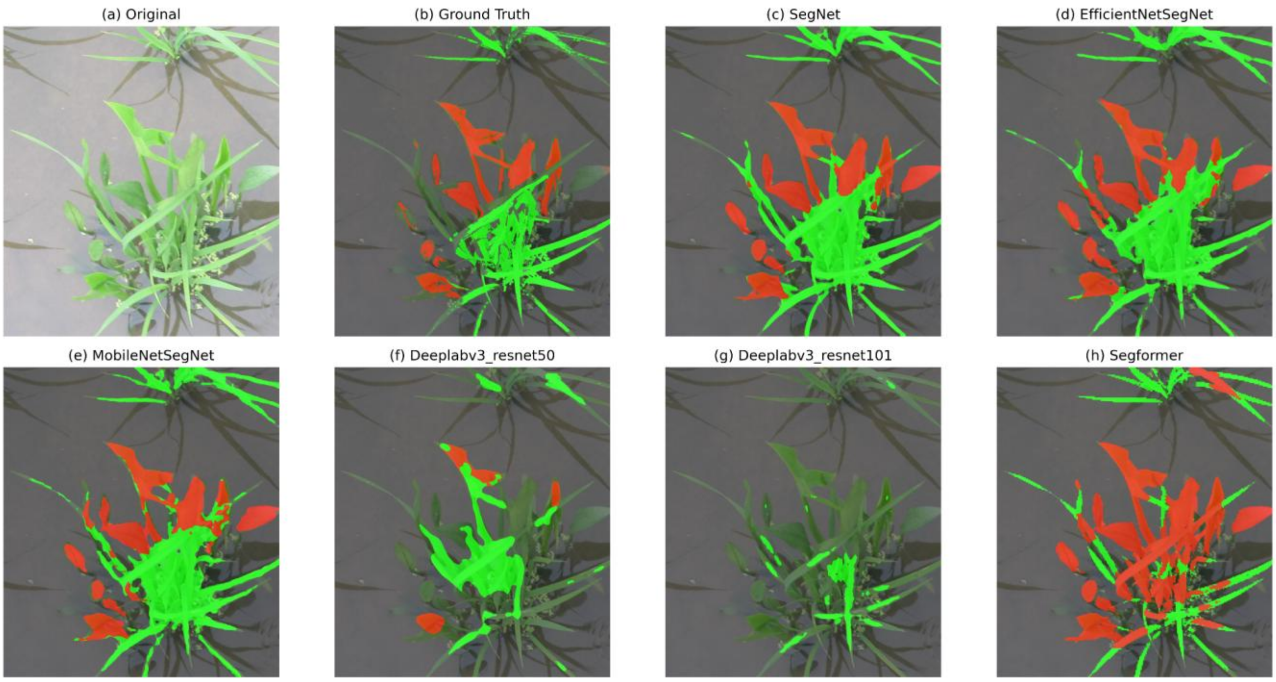

Figure 3. High-density mixing between rice seedlings and weeds, illustrating the difficulty in

distinguishing fine weed structures within a shared region. (a) Original image; (b) Ground truth annotation. (c)–(h) show the prediction results of SegNet, EfficientNet-SegNet, MobileNet-SegNet, DeepLabv3 ResNet50, DeepLabv3 ResNet101, and SegFormer, respectively.

In summary, the models, while capable of segmenting the background efficiently, still struggle with the finer details of rice seedlings and weeds. The performance of SegFormer highlights both its strengths and weaknesses, particularly in complex segmentation tasks where fine-grained class differentiation is required. Further improvements in model architectures and training strategies are needed to enhance performance in these challenging scenarios.

3.3 Model Efficiency and Lightweight Design

In addition to accuracy and segmentation performance, the practicality of deploying deep learning models in real-world agricultural settings greatly depends on computational efficiency and model size. To this end, we evaluated each model based on the following indicators: parameter count, floating-point operations (FLOPs), model size, and average inference time per image. These metrics provide critical insights into the trade-off between accuracy and efficiency. Table 3 presents a summary of the efficiency evaluations of the model.

Table 3. Model Complexity and Inference Efficiency

Model	**Params**	**FLOPs**	**Size (MB)**	**Inference Time (ms)**
SegNet	16.29M	356.57G	62.21	80.41
EfficientNetSegNet	11.46M	82.72G	44.00	20.38
MobileNetSegNet	8.95M	73.83G	34.33	**13.96**
DeepLabV3-ResNet50	42.00M	619.20G	160.56	134.03
DeepLabV3-ResNet101	60.99M	897.10G	233.32	217.84
SegFormer	3.72M	24.17G	14.26	42.39

The results show that while SegNet achieved the highest accuracy (ACC = 0.9486, mIoU = 0.7204), its model size and inference time were comparatively larger (62.21MB, 80.41ms). In contrast, MobileNetSegNet demonstrated a compelling balance between efficiency and accuracy, achieving 93.79% ACC and 0.6914 mIoU with a significantly smaller model size (34.33MB) and faster inference speed (13.96ms). SegFormer stands out as the most compact model, with only 3.72M parameters and a model size of 14.26MB, while maintaining a respectable 0.6471 mIoU. These characteristics make it suitable for edge deployment, especially in UAVs or portable agricultural devices. In contrast, DeepLabV3 models exhibit high computational cost and memory requirements, with DeepLabV3-ResNet101 requiring over 897G FLOPs and 233MB storage, making them less suitable for real-time or embedded applications.

3.4 Interpretation of t-test Results

Table 4. Paired t-test Results (SegNet vs Others)

Compared Model	ACC (t / p)	mIoU (t / p)
EfficientNetSegNet	t = 12.76, p < 0.0001	t = 13.48, p < 0.0001
MobileNetSegNet	t = 11.20, p < 0.0001	t = 11.04, p < 0.0001
DeepLabV3-ResNet50	t = 12.15, p < 0.0001	t = 13.56, p < 0.0001
DeepLabV3-ResNet101	t = 14.59, p < 0.0001	t = 39.56, p < 0.0001
SegFormer	t = 9.34, p < 0.0001	t = 7.89, p < 0.0001

The paired t-test results presented in Table 4 statistically validate the performance differences observed in Table 1. All comparisons against SegNet yielded extremely high t-values and corresponding p-values of less than 0.0001, indicating that the differences in both pixel accuracy and mean IoU are statistically significant.

Among the models, DeepLabV3-ResNet101 exhibited the largest performance gap from SegNet, particularly in the mean IoU (t = 39.56), reflecting its inability to generalize well in this domain. Conversely, SegFormer showed the smallest t-values (t = 9.34 ACC, t = 7.89 for mIoU), suggesting that it is the most competitive among the compared models, despite its lightweight design.

These results reinforce the conclusion that SegNet achieves superior segmentation performance across metrics and that its advantage is not due to random variation but is statistically robust. Therefore, although lightweight models offer deployment benefits, they may incur trade-offs in segmentation precision, which must be carefully balanced depending on the application scenario.

4. Conclusion and Future Work

The results of this study demonstrated the feasibility of applying deep learning to weed segmentation in paddy fields. SegNet achieved the highest segmentation accuracy, excelling in both pixel accuracy (ACC) and mean Intersection over Union (mIoU). However, lightweight models, such as MobileNet-SegNet and SegFormer, offer significant advantages in terms of model size and inference efficiency, making them well-suited for deployment on resource-constrained platforms, such as agricultural drones, mobile robotic systems, and portable edge devices.

In particular, SegFormer delivered competitive accuracy despite having the lowest parameter count and FLOPs, making it a strong candidate for real-time agricultural applications where computational efficiency is critical. Similarly, MobileNet-SegNet struck a favorable balance between speed and accuracy, further supporting its deployment in low-power embedded systems. SegNet's superior performance underscores the value of hyperparameter optimization in maintaining boundary fidelity, which is vital in fine-grained vegetation segmentation.

Several specific deployment scenarios can be envisioned to enhance real-world applicability. Drone-based weed management systems can utilize real-time segmentation for targeted herbicide spraying, thereby reducing chemical use. Autonomous ground vehicles equipped with on-device weed-recognition models can conduct real-time weed detection and spot treatment. Edge-computing modules embedded in smart sprayers or field robots can further enable responsive weeding operations without the need for cloud connectivity.

However, this study had several limitations. The dataset was collected from a single geographic region under stable environmental conditions, which may limit its generalizability. The impact of domain shift arising from variations in climate, soil background, or rice cultivars was not investigated. Additionally, smallholder farms may face barriers to adopting such systems because of hardware costs, integration complexity, and limited technical support.

Future work should address these challenges by expanding datasets across different regions and growing seasons, investigating domain adaptation strategies, and exploring hybrid CNN–Transformer architectures to enhance generalization. Incorporating semi-supervised or weakly supervised learning can also reduce the high cost of manual labeling. Ultimately, integrating segmentation models into automated systems for precision spraying, field surveillance, and yield forecasting can advance smart agriculture, optimize resource usage, and support sustainable farming practices.

References

[1] Wang, N.-W. Sustainability? Development? Discourse and practice in the changing of farming craft: An analysis focused on the rice of Yuli [doctoral thesis]. National Taiwan University, 2023. https://hdl.handle.net/11296/grx2xp

[2] Yang, C.-M. and Lin, J.-Y. Research on the precision agriculture system for rice. Ministry of Agriculture, Taiwan, 2005. https://www.moa.gov.tw/ws.php?id=5093

[3] Council of Agriculture, Executive Yuan. Section 5: Weeds. In: Plant Protection Picture Book Series No. 8: Rice, pp. 212–214. Retrieved July 15, 2025. https://www.aphia.gov.tw/Publish/plant_protect_pic_8/ricePDF/05-0.pdf

[4] Council of Agriculture, Executive Yuan. Plant Protection Picture Book Series No. 8: Rice. Retrieved July 15, 2025. https://www.intelligentagri.com.tw/xmdoc/cont?sid=0K311369563868891650&xsmsid=0K303431216608203659

[5] Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y. and others. Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. PLOS ONE, 2019.

[6] Khan, A., Ilyas, T., Umraiz, M., Mannan, Z.I. and Kim, H. Ced-net: Crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture. Electronics, 2020, 9(10), 1602.

[7] Deng, X., Ma, X., Qi, L., Jiang, Y., Li, H. and Xing, X. The dataset of the manuscript Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields (Version 5). figshare, 2019. https://doi.org/10.6084/m9.figshare.7488830.v5

[8] Badrinarayanan, V., Kendall, A. and Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12), 2481–2495.

[9] Tan, M. and Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, 2019, pp. 6105–6114. PMLR.

[10] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint, 2017. arXiv:1704.04861.

[11] Chen, L.C., Papandreou, G., Schroff, F. and Adam, H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint, 2017. arXiv:1706.05587.

[12] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M. and Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems, 2021, 34, 12077–12090.