Estimating the Coverage of Multiple Species of Paddy Field Weeds Using Semantic Segmentation

1st Yusuke Sayou
Department of Electrical Engineering and Computer Science
Graduate School of Engineering
Tokyo University of Agriculture and Technology
Tokyo, Japan
s253275s@st.go.tuat.ac.jp

2nd Duc-Anh Pham
Department of Electrical Engineering and Computer Science
Graduate School of Engineering
Tokyo University of Agriculture and Technology
Tokyo, Japan
s241318t@st.go.tuat.ac.jp

3rd Tetsuya Nakamura
NEW GREEN.inc
Tokyo, Japan
tetsuya_nakamura@shonai.inc

4th Keisuke Katsura
Division of Agronomy and Horticultural Science
Graduate School of Agriculture
Kyoto University
Kyoto, Japan
katsura.keisuke.2a@kyoto-u.ac.jp

5th Taiichiro Ookawa
Division of Science of Biological Production
Institute of Agriculture
Tokyo University of Agriculture and Technology
Tokyo, Japan
ookawa@cc.tuat.ac.jp

6th Hironori Nakajo
Division of Advanced Information Technology & Computer Science
Institute of Engineering
Tokyo University of Agriculture and Technology,
Tokyo, Japan
nakajo@cc.tuat.ac.jp

***Abstract*—This study proposes a semantic segmentation-based method for estimating the coverage of multiple types of paddy field weeds. The method classifies weeds into three categories: broadleaf plants and grasses, *Cyperaceae* species, and submerged or floating aquatic plants. Rice and background are also included, resulting in a five-class segmentation task. To evaluate the method, we constructed a new dataset consisting of top-down images of rice paddies captured using a smartphone, with each image covering approximately a 1 m^2 area. Four segmentation approaches were compared: single-stage segmentation; two-stage segmentation with background removal; two-stage segmentation with rice removal; and two-stage segmentation with both background and rice removal. The single-stage segmentation method, which directly classifies all classes in a single pass, produced the best overall results, with a mean IoU of 0.60. These findings suggest that the proposed method is capable of classifying and quantifying multiple weed types from real-world field images. Furthermore, the simplicity and efficiency of the single-stage approach make it a practical and promising tool for weed monitoring and management in precision agriculture, particularly in resource-constrained field environments.**

***Index Terms*—Semantic Segmentation, Rice, Weeds, Multi-stage Segmentation, Multi-class Classification**

I. Introduction

A. Current State of Japanese Agriculture

Japanese agriculture faces serious challenges, including a declining number of core agricultural workers and an aging population. The number of core agricultural workers is decreasing, reaching 1.363 million in 2020 , which is a 22% decrease from 1.757 million in 2015 [1].

To address the challenges faced in Japanese agriculture, it is essential to secure and retain the younger generation of agricultural workers. In the face of labor shortages, there is a growing need to streamline and enhance agricultural operations through technological innovations. Therefore a new approach is required to achieve efficient and sustainable agriculture. [2]

B. Efforts Toward Sustainable Agriculture

In May 2021, the Ministry of Agriculture, Forestry and Fisheries of Japan formulated the “Green Food System Strategy,” setting ambitious goals such as expanding the area of organic farming to 25% of all farmland, reducing the use of chemical pesticides by 50%, and reducing chemical fertilizers by 30% [2]. The promotion of organic agriculture is expected to play a key role in addressing environmental issues currently facing the agricultural sector.

Organic farming faces several challenges. Among these, the most critical is the difficulty in weed management owing to the non-use of herbicides. Inadequate weed control can significantly reduce crop yields, leading to an increased economic burden on producers and potentially diminishing their motivation.

C. Objective

The purpose of this study is to estimate weed coverage in paddy fields by analyzing field images where rice and weeds coexist. Vertically downward images were taken using a smartphone, each covering approximately 1 m².

Traditional methods for estimating weed coverage are often time-consuming and subjective, with results varying by evaluator and field conditions. This makes consistent monitoring difficult. In contrast, image-based methods offer faster and more objective estimation, reducing manual effort and improving consistency.

In this study, weeds are broadly classified into three types based on their growth patterns and shapes. Since weed types differ in structure—such as spreading horizontally or growing upright—treating all weeds as a single category can introduce coverage estimation errors. Estimating the coverage of each type separately helps reduce such errors and improves overall accuracy.

II. Related Works

A. Measurement of Weed Dry Weight

In organic farming, understanding weed growth is essential for determining the necessity of weeding. One common method to assess weed growth is to measure the dry weight of weeds. This involves pulling up all weeds within a fixed area, washing off the mud, and drying them in an oven for more than two days before weighing them [3].

It has been reported that when the weed dry weight is 50 g/m² or less at the panicle initiation stage, additional weeding is generally considered unnecessary. However, because this method requires considerable time and labor, it is difficult to carry out in the field, especially during busy farming periods.

B. MDR Based on the Quadrat Method

The quadrat method is a vegetation survey technique in which a square area is defined, and plant species, number of individuals, and coverage within that area are measured [4]. The Multiplied Dominance Ratio (MDR) based on the quadrat method is an index calculated by multiplying the coverage (%) by plant height (m), providing a simple and quantitative measure of weed abundance, as shown in (1).

$$\mathrm{MDR}\ [\times 0.01\ \mathrm{m}^3/\mathrm{m}^2] = \mathrm{Coverage}\ (\%) \times \mathrm{Plant\ Height}\ (\mathrm{m}) \tag{1}$$

MDR has been shown to be significantly correlated with weed dry weight. In particular, it has been reported that when the MDR at the panicle initiation stage is 6 or less, the weed dry weight remains below 50 g/m², minimizing the impact on yield [5]. Furthermore, when the MDR measured three weeks after transplanting is 1 or less, the MDR at the panicle initiation stage tends to be 6 or less. Therefore, measuring the MDR three weeks after transplantation can be utilized for early decision-making regarding weed control.

However, in practice, the coverage used in MDR calculation is often estimated visually within the quadrat, which introduces subjectivity. As a result, the measured values may vary depending on the evaluator, making it difficult to ensure consistency and reproducibility across different observations.

Fig. 1. Example of a captured image

C. Weed Height Estimation Using Image Recognition

Several methods have been proposed for estimating weed height in paddy fields through image recognition.

The simplest involves manual input by users, but it requires effort and lacks automation. Another approach uses reference markers with known heights in the field; models like YOLOv8 detect both weeds and markers to estimate relative height, though data scarcity can limit accuracy. A third method places known-sized objects near weeds, extracts weed pixels in HSV space, and uses Harris corner detection to find plant endpoints. Background noise is reduced using Local Outlier Factor (LOF) for improved accuracy.

D. Two-Stage Semantic Segmentation

Moazzam et al. proposed a two-stage semantic segmentation approach to improve the accuracy of distinguishing between crops and weeds [6].

In this method, the first stage separates vegetation (including both crops and weeds) from the background by setting the pixel values of the background to zero, thereby simplifying the input data. In the second stage, the image is classified into three classes: crops, weeds, and background. A lightweight U-Net with a Vanilla Mini CNN is employed in the first stage to reduce computational load, while a U-Net with VGG16 as the backbone is used in the second stage to improve classification accuracy.

This approach improved the IoU for crops from 0.67 to 0.85 and for weeds from 0.76 to 0.91 compared to conventional single-stage methods. Furthermore, the use of a lightweight model in the first stage contributed to a reduction in computational cost.

III. Dataset

A. Image Acquisition Method

To calculate the coverage, top-down images of rice paddies were captured to ensure that an area containing a 3-by-3 grid of rice plants was captured. As the row and plant spacing of rice is approximately 30 cm, the photographed area covered approximately 1 m × 1 m. An example of a captured image is shown in Fig. 1. All images were resized to 1920 × 1920 pixels to ensure uniform size.

Fig. 2. Examples of weed classification (Left: broadleaf weeds and Poaceae, Center: *Cyperaceae* [7], Right: submerged and floating-leaved aquatic plants)

B. Weed Classification

Weeds can be broadly classified into three categories. The first category includes broadleaf weeds and Poaceae, which are characterized by horizontally spreading leaves and are typically emergent weeds that grow above the water surface. The second category is *Cyperaceae*, which generally has slender shapes and triangular stems and is classified as an emergent weed. The third category includes submerged and floating-leaved aquatic plants, which grow underwater and float their leaves either on or just below the water surface. Examples from each category are shown in Fig. 2.

C. Details of the Dataset

We used 92 images, each containing one or more weed classes. The number of images in which each class appeared were 68 for broadleaf weeds and Poaceae, 25 for *Cyperaceae*, and 79 for submerged and floating-leaved aquatic plants.

Roboflow [8] was used for image annotation. The dataset was divided into five classes: Weed1 (broadleaf weeds and Poaceae), Weed2 (*Cyperaceae*), Weed3 (submerged and floating-leaved aquatic plants), rice, and background.

The dataset was divided into 66 training, 17 validation, and 9 test images. To avoid a class imbalance within each subset, the data were adjusted to maintain a balanced class distribution.

IV. Proposed Method

In this study, both single- and two-stage semantic segmentation approaches were used to classify images into five classes: Weed1, Weed2, Weed3, rice, and background. Coverage was calculated based on the classification results.

In the two-stage approach, specific classes are predicted in Stage 1, and their corresponding pixels are masked (set to zero). The modified image is then used in Stage 2 for final classification. Four methods were evaluated: one single-stage and three two-stage approaches with different Stage 1 preprocessing strategies.

Images were resized to 1920 ×1920 pixels and divided into nine patches (640× 640) for training and inference. U-Net [9] with a ResNet50 [10] backbone was used. DiceFocalLoss [11] and the Adam optimizer were applied for training.

A. Method 1: Single-Stage Segmentation

Single-stage segmentation is a straightforward approach in which all five classes are predicted simultaneously in a single inference. Because classification is completed in one pass, the main advantage of this method is its low computational cost.

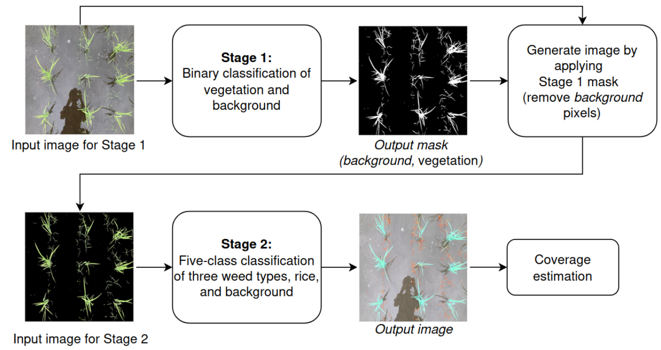

Fig. 3. Flowchart of Method 2: Two-stage segmentation with background removal

B. Method 2: Two-Stage Segmentation with Background Removal

In Stage 1, a binary classification is performed to separate vegetation (rice and weeds) from the background. Then, in Stage 2, the input image is modified by setting the background pixels—identified in Stage 1—to zero. This modified image was then used as input to classify five classes: background, rice, and three types of weeds. A flowchart of this method is shown in Fig. 3.

By removing the background in Stage 1, the influence of the background pixels was eliminated in Stage 2, which was expected to improve the classification accuracy of rice and weeds. Furthermore, reducing the effect of background variability may help maintain performance across different environments, potentially making this method more generalizable.

C. Method 3: Two-Stage Segmentation with Rice Removal

In Stage 1, the rice regions within the image are identified, and a binary classification is performed by treating non-rice pixels as the background. Subsequently, in Stage 2, the input image is modified by setting the background pixels—identified in Stage 1—to zero. This modified image is used to classify four classes: background and the three types of weeds. A flowchart of this method is shown in Fig. 4.

This method aims to improve the accuracy of weed classification by removing rice-related information. In particular, it is expected to reduce misclassification between rice and weeds and mitigate ambiguity around class boundaries.

D. Method 4: Two-Stage Segmentation with Background and Rice Removal

In Stage 1, three-class segmentation is performed to classify the pixels into background, rice, and weeds. In Stage 2, the input image is modified by setting the pixels corresponding to the background and rice—identified in Stage 1—to zero. This modified image is then used to classify four classes:

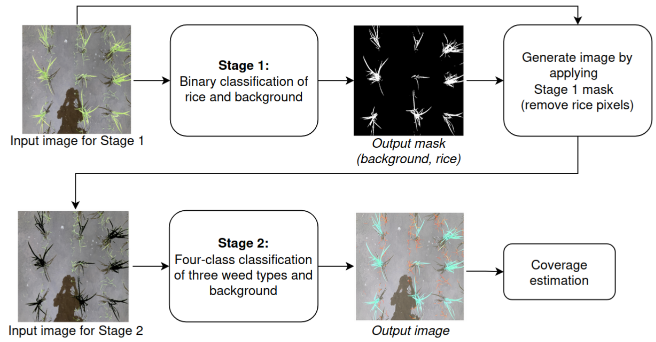

Fig. 4. Flowchart of Method 3: Two-stage segmentation with rice removal

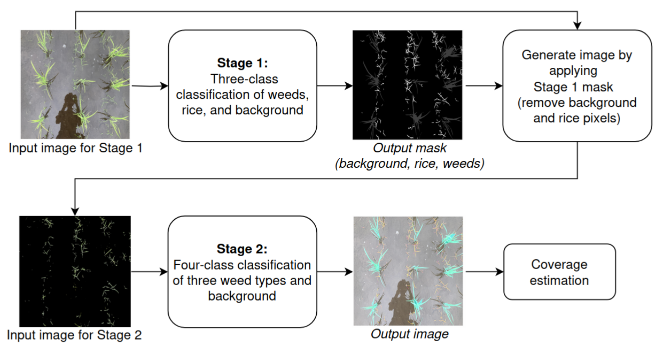

Fig. 5. Flowchart of Method 4: Two-stage segmentation with background and rice removal

background and the three types of weeds. A flowchart of this method is shown in Fig. 5.

By removing both the background and rice regions, weed areas are emphasized, making their distinctive shapes, textures, and color features more apparent. This is expected to improve the learning efficiency of the classification model. Furthermore, eliminating noise from the background and rice may help reduce misclassification, particularly for weed types that visually resemble rice.

V. Evaluation and Discussion

In this section, we evaluate and discuss the four proposed methods from two perspectives: (1) semantic segmentation performance, using metrics such as Precision, Recall, and IoU; and (2) numerical evaluation of the estimated coverage using MAE (Mean Absolute Error). The evaluation includes both quantitative analysis and visualization-based assessment.

A. Quantitative Evaluation

1) IoU: To evaluate the segmentation performance, the IoU for each class was calculated using all four methods. The results are summarized in Table I.

As summarized in Table I, Method 1 achieved the highest mean IoU of 0.59, followed by Method 2 (0.54), Method 3 (0.54), and Method 4 (0.39).

Additionally, Table I shows that for Methods 1–3, the IoU of Weed1 was consistently lower than that of Weed2 and Weed3. Specifically, Method 1 showed an IoU of 0.31 for Weed1, 0.49 for Weed2, and 0.44 for Weed3, indicating that Weed1 was the most difficult to classify. The same trend was observed in Methods 2 and 3, with Weed1 IoUs of 0.26 and 0.22, respectively.

One possible reason for the lower classification accuracy of Weed1 is its morphological similarity to rice. Both Weed1 and rice have elongated, narrow leaves and can appear similar from a top-down view, especially under conditions with strong shadows or overlapping leaves. In addition, Weed1 often grows in close proximity to rice plants, which can make boundary regions more ambiguous. These factors may cause the model to confuse Weed1 with rice, resulting in lower segmentation accuracy.

2) Precision and Recall: To evaluate the classification accuracy, we compared the precision and recall of each method.

As shown in Table I, Method 1 achieved the highest mean IoU. However, as shown in Table II, Methods 2 and 3 outperformed Method 1 in terms of Precision for weed classes. For Methods 2 and 3, Recall for weed classes was consistently lower than Precision, as shown in Table III. This discrepancy suggests that weed regions may have been misclassified as rice or background during Stage 1, potentially leading to the loss of weed information. Method 4, which removes both background and rice in Stage 1, showed a significant drop in both Precision and Recall, especially for Weed3, indicating that excessive filtering may have hindered weed detection.

3) MAE: To evaluate the accuracy of the estimated coverage, the MAE was compared across all methods. The MAE represents the average absolute difference between the predicted and ground truth coverage values, where lower values indicate better estimation accuracy. The results are

TABLE I
Comparison of IoU for each method

Method	Background	Rice	Weed1	Weed2	Weed3	Mean
Method 1	0.98	0.76	0.31	0.49	0.44	0.59
Method 2	0.97	0.75	0.26	0.37	0.37	0.54
Method 3	0.97	0.75	0.22	0.40	0.37	0.54
Method 4	0.98	0.77	0.09	0.11	0.00	0.39

TABLE II
Comparison of Precision

Method	Background	Rice	Weed1	Weed2	Weed3	Mean
Method 1	0.99	0.85	0.46	0.69	0.63	0.72
Method 2	0.98	0.87	0.48	0.76	0.73	0.76
Method 3	0.98	0.86	0.53	0.75	0.81	0.79
Method 4	0.98	0.89	0.13	0.16	0.36	0.51

TABLE III
Comparison of Recall

Method	Background	Rice	Weed1	Weed2	Weed3	Mean
Method 1	0.99	0.87	0.48	0.63	0.59	0.71
Method 2	0.99	0.84	0.36	0.42	0.43	0.61
Method 3	0.99	0.86	0.27	0.46	0.40	0.59
Method 4	0.99	0.85	0.21	0.25	0.00	0.46

TABLE IV
COMPARISON OF MAE FOR WEED CLASSES

Method	Weed1	Weed2	Weed3	Mean
Method 1	0.45	0.11	0.30	0.29
Method 2	0.52	0.09	0.30	0.30
Method 3	0.50	0.11	0.29	0.30
Method 4	0.75	0.56	1.33	0.88

summarized Table IV.

From Table IV, Methods 1 to 3 maintained stable accuracy, with mean MAE values around 0.30. In contrast, Method 4 produced notably large errors across all weed classes, particularly for Weed3, which reached a MAE of 1.33.

Among the weed types, Weed1 consistently showed higher MAE values (0.45–0.52), while Weed2 had the lowest (0.09–0.11), and Weed3 fell in between (0.29–0.30). These results indicate that coverage estimation was most accurate for Weed2 and least accurate for Weed1.

The overall trend shows that higher segmentation accuracy corresponds to lower MAE, suggesting that improving classification performance leads to better coverage estimation.

B. Visual comparison of single-stage segmentation and three two-stage segmentation methods

To compare the outputs of each method, predicted masks were overlaid on input images using transparency-based visualization. Class colors were set as follows: cyan for rice, pink for Weed1, yellow for Weed2, and orange for Weed3. This section highlights differences in weed recognition between the single-stage method (Method 1) and the three two-stage methods (Methods 2–4).

1) Comparison Between Method 1 (Single-Stage) and Method 2 (Two-Stage with Background Removal): The visualization shows that removing the background in Method 2 increased confusion between rice and Weed1. With background pixels removed, their similar shapes and colors became more prominent, making classification harder. As shown in Fig. 6, misclassification occurred especially in overlapping leaves and varying growth stages.

2) Comparison Between Method 1 (Single-Stage) and Method 3 (Two-Stage with Rice Removal): As shown in the visualization results in Fig. 7, the removal of rice led to an increase of correctly classified pixels for Weed1. However, Weed3 was misclassified more frequently than Weed1. This result suggests that rice may have served as a contextual clue for distinguishing Weed1 and Weed3. With the removal of rice, the boundary between the features of Weed1 and Weed3 became more ambiguous, making it more difficult for the model to correctly classify the two.

3) Comparison between Method 1 (Single-stage Segmentation) and Method 4 (Two-stage Segmentation with Background and Rice Removal): In the visualization shown in Fig. 8, Weed3 was misclassified as Weed1 owing to the removal of both background and rice. This suggests that the model relied on the presence of rice and the background as contextual cues for weed classification. By eliminating both, the model lacked sufficient contextual information to distinguish between the weed classes, which likely led to an increase in misclassification among the weed types.

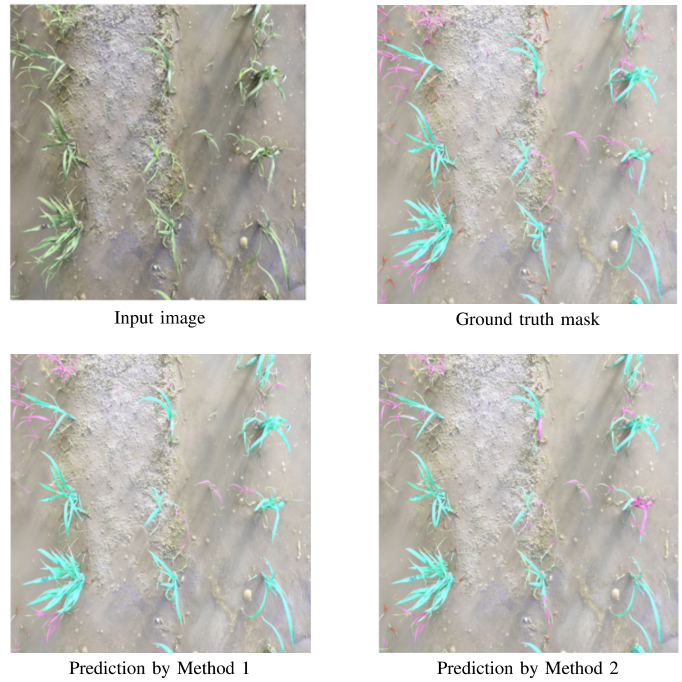

Fig. 6. Comparison of output results by Method 1 and Method 2 (example of Weed1)

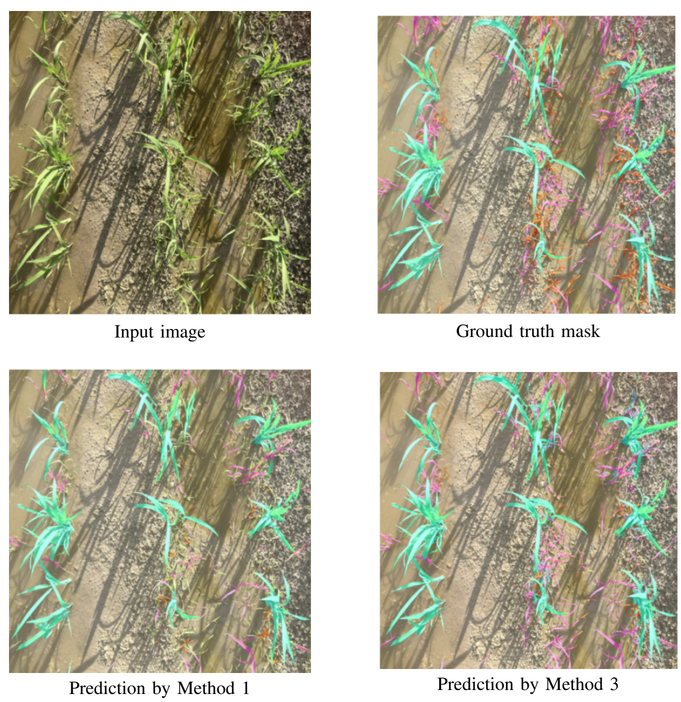

Fig. 7. Comparison of output results by Method 1 and Method 3 (example of Weed1 and Weed3)

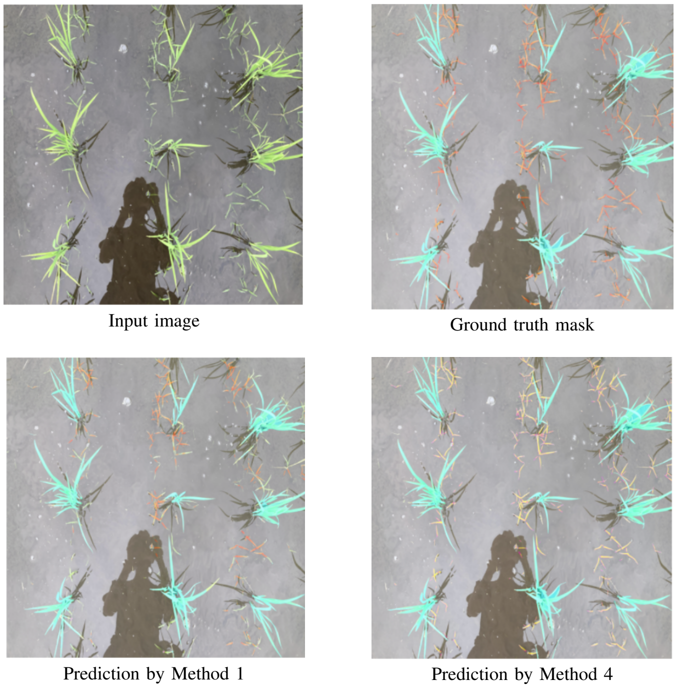

Fig. 8. Comparison of output results by Method 1 and Method 4 (example of Weed3)

VI. Conclusion and Future Work

A. Conclusion

In this study, we proposed and evaluated a method for classifying multiple weed types and estimating their coverage using top-down field images containing both rice plants and various weeds. Four methods were compared: one single-stage segmentation approach and three two-stage segmentation approaches. The results showed that the single-stage segmentation method (Method 1), which directly classifies all classes—including rice and weeds—in a single step, achieved the highest accuracy.

The lower performance of the two-stage methods likely stems from two factors. First, rice and background regions provide useful context for distinguishing weed types, and removing them in Stage 1 may have discarded valuable information. Second, errors in Stage 1 can propagate to Stage 2, reducing overall accuracy. In contrast, the single-stage method not only achieved higher accuracy but also has practical benefits. It requires only one forward pass, lowering computational cost and system complexity, making it more suitable for real-time or low-resource agricultural settings.

The evaluation showed that Weed1 was the most difficult class to distinguish, mainly due to its morphological similarity to rice—such as narrow and elongated leaves—which makes visual separation challenging in top-down images. Its tendency to grow close to rice further increases misclassification risks due to overlapping vegetation.

B. Future Work

To improve the classification performance of Weed1, which showed low accuracy, we plan to expand the dataset with a particular focus on this class and enhance data diversity through methods such as Copy-Paste Augmentation. On the modeling side, we will explore incorporating instance-level features—such as attention mechanisms, spatial relationships, and shape-based cues—to better distinguish visually similar classes like Weed1 and rice. In addition, we will evaluate inference time, model size, and memory usage to assess the feasibility of deployment on smartphones.

Currently, the height estimation process does not incorporate weed classification. As a next step, we aim to integrate weed type classification into the height estimation pipeline, enabling the estimation of both coverage and height for each class. Ultimately, our goal is to develop an application that calculates the Multiplied Dominance Ratio (MDR) for each weed type and uses the total MDR to assess the necessity of weeding.

Acknowledgment

This work is partially supported by JSPS Grant-in-Aid for Scientific Research (C) 24K14874.

References

[1] Ministry of Agriculture, Forestry and Fisheries (MAFF), "2020 Census of Agriculture and Forestry," 2020. [Online]. Available: https://www.maff.go.jp/ [in Japanese]

[2] Ministry of Agriculture, Forestry and Fisheries (MAFF), "Green Food System Strategy," May 2021. [Online]. Available: https://www.env.go.jp/content/900449080.pdf [in Japanese]

[3] Niigata Agricultural Research Institute, Basic Research Division, "Criterion for additional weeding three weeks after rice planting in organic cultivation," *FY2023 Research Achievements*, p. 0603995, 2023. [Online]. Available: https://www.pref.niigata.lg.jp/sec/0603995.html [in Japanese]

[4] J. E. Weaver, "The Quadrat Method in Teaching Ecology," *The Plant World*, vol. 21, no. 11, pp. 267–283, November 1920. [Online]. Available: https://www.jstor.org/stable/2480223?seq=5

[5] Niigata Agricultural Research Institute, Basic Research Division, "Simple evaluation of weed emergence in organic rice cultivation using the quadrat method," *FY2022 Report*, pp. 3–4, 2022. [Online]. Available: https://www.pref.niigata.lg.jp/uploaded/attachment/376670.pdf [in Japanese]

[6] S. I. Moazzam, U. S. Khan, W. S. Qureshi, T. Nawaz, and F. Kunwar, "Towards automated weed detection through two-stage semantic segmentation of tobacco and weed pixels in aerial imagery," *Smart Agricultural Technology*, vol. 4, p. 100142, 2023. doi: 10.1016/j.atech.2022.100142. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S277237552200106X

[7] Yukai na Nakamatachi, "Kuwai (Arrowhead)," *livedoor Blog*, Dec. 2007. [Online]. Available: http://blog.livedoor.jp/varietyus/archives/64934476.html [in Japanese]

[8] Roboflow, Inc., "Roboflow," *Roboflow*, 2024. [Online]. Available: https://roboflow.com/

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, vol. 9351, pp. 234–241, 2015. Springer. doi: 10.1007/978-3-319-24574-4_28

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90

[11] Y. Mu, D. Thomas, J. Hattne, A. Rohou, Z. Ke, Y. Kwon, Q. Zhang, S. J. Ludtke, and W. Chiu, "The combined focal loss and dice loss function improves the segmentation of beta-sheets in medium-resolution cryo-electron-microscopy density maps," *Bioinformatics Advances*, vol. 4, no. 1, p. vbae169, 2024. Oxford University Press. doi: 10.1093/bioadv/vbae169. [Online]. Available: https://academic.oup.com/bioinformaticsadvances/article/4/1/vbae169/7907198