

Dataset Article

Global rice multiclass segmentation dataset (RiceSEG): comprehensive and diverse high-resolution RGB-annotated images for the development and benchmarking of rice segmentation algorithms



Junchi Zhou^a, Haozhou Wang^b, Yoichiro Kato^b, Tejasri Nampally^c, P. Rajalakshmi^d, M. Balram^e, Keisuke Katsura^f, Hao Lu^g, Yue Mu^a, Wanneng Yang^h, Yangmingrui Gao^a, Feng Xiao^a, Hongtao Chen^a, Yuhao Chen^a, Wenjuan Liⁱ, Jingwen Wang^j, Fenghua Yu^k, Jian Zhou^l, Wensheng Wang^m, Xiaochun Huⁿ, Yuanzhu Yangⁿ, Yanfeng Ding^a, Wei Guo^{b,*}, Shouyang Liu^{a,**}

^a Engineering Research Center of Plant Phenotyping, Ministry of Education, Jiangsu Collaborative Innovation Center for Modern Crop Production, Sanya Institute of Nanjing Agricultural University, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing, China

^b Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

^c Department of Artificial Intelligence, Indian Institute of Technology, Hyderabad, India

^d Department of Electrical Engineering, Indian Institute of Technology, Hyderabad, India

^e Institute of Biotechnology, Professor Jayashankar Telangana Agricultural State University, Hyderabad, India

^f Graduate School of Agriculture, Kyoto University, Kyoto, Japan

^g Key Laboratory of Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

^h National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, and Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agricultural University, Wuhan, China

ⁱ State Key Laboratory of Efficient Utilization of Arid and Semiarid Arable Land in Northern China, The Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China

^j Center for Geospatial Information, Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shenzhen, China

^k School of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang, China

^l Rice Research Institute, Jilin Academy of Agricultural Sciences, Changchun, China

^m Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China

ⁿ Yuan Long Ping High-Tech Agriculture Co., Ltd., Changsha, China

ARTICLE INFO

ABSTRACT

Keywords:

RiceSEG dataset
Rice phenotyping
Semantic segmentation
Deep learning
Crop monitoring

The development of computer vision-based rice phenotyping techniques is crucial for precision field management and accelerated breeding, which facilitate continuously advancing rice production. Among phenotyping tasks, distinguishing image components is a key prerequisite for characterizing plant growth and development at the organ scale, enabling deeper insights into ecophysiological processes. However, owing to the fine structure of rice organs and complex illumination within the canopy, this task remains highly challenging, underscoring the need for a high-quality training dataset. Such datasets are scarce, both because of a lack of large, representative collections of rice field images and because of the time-intensive nature of the annotation. To address this gap, we created the first comprehensive multiclass rice semantic segmentation dataset, RiceSEG. We gathered nearly 50,000 high-resolution, ground-based images from five major rice-growing countries (China, Japan, India, the Philippines, and Tanzania), encompassing more than 6000 genotypes across all growth stages. From these original images, 3078 representative samples were selected and annotated with six classes (background, green vegetation, senescent vegetation, panicle, weeds, and duckweed) to form the RiceSEG dataset. Notably, the subdataset from China spans all major genotypes and rice-growing environments from northeastern to southern regions. Both state-of-the-art convolutional neural networks and transformer-based semantic segmentation models were used as baselines. While these models perform reasonably well in segmenting background and green

* Corresponding author.

** Corresponding author.

E-mail addresses: guowei@g.ecc.u-tokyo.ac.jp (W. Guo), shouyang.liu@njau.edu.cn (S. Liu).

vegetation, they face difficulties during the reproductive stage, when canopy structures are more complex and when multiple classes are involved. These findings highlight the importance of our dataset for developing specialized segmentation models for rice and other crops. The RicesEG dataset is publicly available at www.global-rice.com.

1. Introduction

As a core pillar of global agricultural production, rice is widely cultivated worldwide and feeds more than half of the global population [1]. However, facing global warming, the variability and uncertainty in rice-growing environments pose severe challenges for the sustainability of rice production [2]. To leverage unfavorable growth conditions, great efforts have been made to improve both cultivar and cultivation practices according to the adaptation of phenotypic traits [3]. Hence, the success of these efforts strongly relies on the precision and throughput of the plant phenotyping techniques. Unfortunately, the measurement of plant phenotypic traits is accomplished mainly manually, which is very time-consuming and labor-intensive [4,5,6]. Therefore, the development of high-throughput phenotyping techniques is crucial for overcoming these limitations and consequently ensuring rice production [7,8,9].

Compared with traditional human observation, computer vision techniques have greatly advanced plant phenotyping because they provide higher throughput and accuracy [10]. A key step in this domain is image segmentation, which underpins the extraction of critical traits such as canopy structure [11], light interception [12], and stress status [13]. With respect to single-class segmentation, when green vegetation is distinguished from the background, deep learning-based models have demonstrated robust performance across various crops [14,15], maintaining consistent accuracy under diverse environments, genotypes, and spatial resolutions [16]. However, there is a growing need for more detailed segmentation that distinguishes multiple plant organs (e.g., panicles and both green and senescent leaves), as this enables deeper insights into organ development and the source–sink relationship [17]. Moreover, because weeds commonly appear in rice fields, simultaneously segmenting weeds alongside crop organs both reduces misclassification and informs weed management strategies. Although recent deep learning segmentation models, such as the SAM [18], show promise, none have successfully addressed the multiclass segmentation of rice canopies, encompassing both organs and weeds, across diverse genotypes and environmental conditions. This is primarily due to the unique challenges posed by rice canopies, which feature fine leaves, thin stems, and substantial genotype-dependent variations. Fluctuating field illumination further complicates segmentation by creating mutual shading within the canopy, while reflective water surfaces in paddy fields produce mirror-like reflections and glare, distorting certain image regions and reducing clarity. As with other complex computer vision tasks, improving current models or developing specialized approaches hinges on the availability of comprehensive training datasets that capture the full complexity of rice field conditions.

High-quality training datasets are critical for adapting state-of-the-art computer vision (CV) models to plant phenotyping [19]. In recent years, numerous phenotyping datasets have emerged for various crops, both indoors and in the field, focusing primarily on plant counting [20], organ detection [21,22], and disease or pest classification [23,24]. However, few datasets target semantic segmentation because of the labor-intensive nature of pixel-level annotation. This issue is especially pronounced in rice, where fine leaves and dense canopies complicate the annotation process, leading to a shortage of publicly available datasets [25,26]. Table 1 provides an overview of representative plant semantic segmentation datasets, which for rice crops are largely confined to single classes—such as panicle segmentation [27] for basic green segmentation from the background [28]. In summary, no existing rice segmentation dataset jointly encompasses multiple genotypes, diverse field conditions, multiple organs (leaf, stem, and panicle), and weeds.

Table 1
Representative semantic segmentation datasets.

Dataset	Crop type	Class	# Images	Image Size
CVPPP [29]	Rosette plants	2	1311	2048 × 2448
CWFID [30]	Carrot	3	60	1291 × 966
Oil Radish Growth [31]	Oil radish	7	129	1601 × 1601
PhenoBench [32]	Sugar beet	3	2872	1024 × 1024
Paddy Rice Imagery [27]	Rice	2	400	4096 × 2160
VegAnn [28]	Rice, wheat, etc.	2	466	512 × 512
RicesEG	Rice	6	3078	512 × 512

The main objectives of this work are to construct a broad, multiclass, high-resolution semantic segmentation dataset for rice crops. This dataset includes 3078 ground-based RGB images collected from 5 countries and 12 different institutions, taken throughout the entire growth cycle, and covering a wide range of genotype–environment–management combinations. Pixels in all the images are finely annotated into six categories: background, green vegetation, senescent vegetation, panicle, weed, and duckweed. Furthermore, to assess the dataset, we report baseline results for most classic and cutting-edge semantic segmentation algorithms. The main contributions of this study are twofold:

- To the best of our knowledge, we present the largest global rice semantic segmentation dataset, offering precise pixel-level annotations across multiple detailed classes in real rice fields.
- We conducted extensive experiments with various segmentation models on this dataset to create benchmark performance, thereby facilitating the development of more effective rice segmentation algorithms.

2. Materials and methods

2.1. Dataset collection

To maximize the representativeness and diversity of the dataset, we collected approximately 50,000 images in total, contributed by 12 institutions between 2012 and 2024, from 14 sites located in 5 countries, including China, Japan, India, the Philippines, and Tanzania (Table 2). They were taken by different types of cameras, such as digital single-lens reflex cameras, portable action cameras, or smartphones. The configuration of the cameras was set 1–2 m above the canopy with different orientations (0°–90°) toward the canopy. This ensured the high resolution of the images with ground sampling distances (GSD) ranging from 0.1 to 1.8 mm/pixel.

- 1) **Dataset from China.** The dataset originated from various sites across China, encompassing all major rice production regions from the northeasternmost to the southernmost areas where rice is cultivated. This extensive coverage includes more than 6000 rice varieties, resulting in a large collection of diverse images. Specifically, images provided by Nanjing Agricultural University (JS_1, JS_2, JS_3, JS_4, and HN) were meticulously gathered from experimental fields in Jiangsu and Hainan Provinces, featuring more than one thousand rice varieties. These images highlight the challenges associated with segregating plant organs due to the high variability in canopy structures among genotypes under diverse field light conditions, as well as the presence of weeds or duckweed in the background. Additionally, images from Changsha were captured in the

Table 2
Metadata of the subdatasets comprising the RiceSEG dataset.

Name	Institute	Site	Images	Lat (°)	Long (°)	Year	Growth stage ^a	Genotypes	Platform	Camera	Image size (pixels)	GSD (mm/px)
CHINA	JS_1	NJAU	Jiangsu	4000 4000	31.5 N 8000	119.3 E 2020	Vegetative, Transition Vegetative	1000	Handheld rod	SONY RX0	4800*3200	0.1–0.3
	JS_2						Vegetative, Transition					
	JS_3						Reproductive					
	JS_4											
HN	HUST	Hainan	2000	18.2 N	109.5 E	2023	Vegetative, Transition	20	Fixed rod	Canon EOS 1100D	4272*2848	0.3–0.5
GX	CAS	Guangxi	280	24.3 N	109.4 E	2012	Vegetative	35	Fixed rod	OLYMPUS E-450	3648*2736	0.3&1.2
JX	SYAU	Jiangxi	355	28.7 N	115.9 E	2013	Vegetative	104	Tripod	NIKON D7100	6000*4000	1.8
HB	LPHT	Hubei	104	30.5 N	114.3 E	2016	Transition		Handheld rod	NIKON D7100	2000*2000	0.3
HLJ	JAAS	CAAS	40	45.7 N	126.6 E	2016	Vegetative	40	Handheld rod	iphone11	2048*1536	0.1–0.3
GD		Guangdong	90	22.6 N	113.1 E	2022	Reproductive	60	Handheld rod	SONY RX0	4032*3024	0.1–0.3
LN		Shenyang	154	41.8 N	123.4 E	2024	Vegetative, Transition	50	Handheld rod	SONY RX0	4800*3200	0.1–0.3
HUN		Changsha	14994	28.2 N	112.9 E	2024	Transition	5000	Handheld rod	SONY RX0	4800*3200	0.1–0.3
JL		Changchun	2642	43.8 N	125.3 E	2024	Reproductive	700	Handheld rod	SONY RX0	4800*3200	0.1–0.3
JAPAN	TKO_1	UTokyo	645	35.4 N	139.3 E	2013	Vegetative	5	Fixed rod	Canon EOS Kiss X5	5184*3456	0.1
	TKO_2		142			2014	All stage					
	TKO_3											
INDIA	TG	IITB	768			2015	Transition	50	Handheld rod	Sony RX100	5472*3648	0.3–0.5
TANZANIA	Kil	KATC	271	17.3 N	78.4 E	2018	Vegetative	4	Handheld rod	RICOH WG-4	3072*2304	0.2–0.4
PHILIPPINES	Lag	IRRI	126	3.45 S	37.4 E	2019	Reproductive	1596	Handheld rod	OLYMPUS TG-620	1600*1200	0.3–0.5

^a Growth stages are categorized into three main phases: (a) vegetative: seedling, tillering, and jointing; (b) transition: shooting, heading, and flowering; and (c) productive: filling and maturity.

rice experimental fields of Yuan Long Ping High-Tech Agriculture Co., Ltd. (<https://lpht.com.cn/>), a leading firm in rice breeding renowned for its hybrid rice varieties. This collection includes images of nearly 5000 rice genotypes at various growth stages (transition and reproductive stages), encompassing both domestic and international varieties.

The northeastern region significantly contributes to China's production of rice, particularly high-quality japonica rice adapted to cold climates. Images were collected from each of the northeastern provinces, including Heilongjiang (HL), Jilin (JL), and Liaoning (LN). The 'HL' dataset was captured by the Institute of Agricultural Resources and Regional Planning using a fisheye camera, providing a unique wide-angle perspective of the rice canopy across several varieties. The 'JL' dataset comprises images from more than 700 rice varieties obtained from the Rice Research Institute of the Jilin Academy of Agricultural Sciences, whereas the 'LN' dataset was provided by Shenyang Agricultural University. The 'JX' and 'GX' subdatasets, contributed by Huazhong University of Science and Technology, document images from various growth stages ranging from seedling to jointing across more than 40 genotypes in Jiangxi and Guangxi Provinces, respectively. The 'HB' subdataset, provided by Huazhong Agricultural University, includes data from 104 varieties, and the 'GD' dataset, supplied by the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, encompasses images from more than 60 genotypes.

- 2) **Dataset from Japan.** This dataset encompasses a broad spectrum of rice genotypes in Japan. Notably, the dataset sourced from the University of Tokyo (TKO_1, TKO_2, TKO_3) comprises time series images of rice captured by field-fixed cameras. The UTokyo dataset was collected from paddy phenotyping field trials at the Institute for Sustainable Agro-ecosystem Services (ISAS) (35°44'20.3"N, 139°32'29.8"E) in Tokyo, Japan, during the 2014 season. A field server system cite{utokyo_2015} collected images of five genotypes throughout the entire growth stage. The camera module of the system is based on a digital single-lens reflex (DSLR) camera, the Canon EOS Kiss X5 camera, with an EF-S18-55 mm lens (Canon, Inc., Tokyo) that provides high-quality and high-resolution (18 megapixels) image data. A preprogrammed microcontroller board automatically controls the power and shutter of the camera.
- 3) **Dataset from India.** The dataset was obtained from the Institute of Biotechnology of Professor Jayashankar Telangana State Agriculture University, which is located in Hyderabad, Telangana, India. The study area covers an area of 15.3 m × 34.8 m and includes two repetitions of 203 plots, each representing a different variety/genotype of aerobic paddy, resulting in a total of 406 plots. Each plot covers an area of 1.26 square meters and contains 42 crop strands. The dataset provides a collection of images of upland rice, which are unique because of the presence of many weeds in complex backgrounds. The images were captured by a team from the Indian Institute of Technology Hyderabad using a high-resolution Sony RX 100 camera. Each image has a resolution of 3456 × 2592 pixels.
- 4) **Dataset from the Philippines.** The dataset was collected from the International Rice Research Institute (IRRI) farm located in Los Baños, Philippines at 14°11' N, 121°15' E and an elevation of 21 m above sea level. The study encompassed three distinct paddy fields containing a comprehensive collection of rice varieties with varying experimental conditions. In total, the dataset comprises 1596 rice varieties/lines distributed across 2172 plots, with some overlap in varieties between fields. All the fields maintained a consistent planting density of 20 cm × 20 cm between plants, creating a uniform growing environment for comparative analysis. The experimental design allows for the systematic evaluation of rice phenotypes under different field management strategies. All the images were captured during the vegetative stage of rice growth, specifically 3–4 weeks after transplanting.

5) **Dataset from Tanzania.** Field experiments were conducted at the irrigated lowland field in the Kilimanjaro Agricultural Training Centre in the Republic of Tanzania ($3^{\circ}45'08''$ S, $37^{\circ}39'68''$ E, 720 m above sea level) in 2019. Four rice varieties, NERICA 1, IR64, TXD 306 and Wahiwahi, were subjected to four different water management practices with three replications: continuous flooded conditions and three alternate wetting and drying conditions. Irrigation was repeated until the water depth reached 10 cm when the surface water level decreased to 0 cm, 15 cm and 30 cm, respectively. At maturity, images of the rice canopy were taken vertically downward from 80 cm above the rice canopy using a digital camera (WG-4; Ricoh, Japan). Twenty-four rice hills (4 hills \times 6 hills, 1.2 m \times 0.9 m)

that were captured in the images were then harvested to ground level, and the yield and yield components were investigated.

2.2. Construction of the RiceSEG dataset

Considering the substantial variation in the number of images collected from China and other countries (Table 2 and Fig. 1), we employed distinct selection strategies to maximize the dataset's representativeness (Fig. 2 and Table 3). In China, compared with other countries, collaborations across all major rice-growing regions enabled the largest overall collection of images. From each Chinese site, 60–100 images were randomly chosen to capture diverse growth stages, varieties, and environmental conditions. In contrast, acquiring high-resolution rice images from other countries proved more challenging; hence, for the remaining five countries, we utilized nearly all the originally collected data.

Cropping can increase the relative size of small but important targets (e.g., panicles, senescent leaves, and weeds) and capture phenotypic variation within the same scene, thereby enhancing organ-scale feature learning and dataset diversity. Most datasets adopted this approach, such as GWHD [21,22], VegAnn [28], and PhenoBench [32]. After finalizing the image selection across all the sites, a cropping procedure was adopted. With continual advancements in computational resources, larger models can leverage higher-resolution images for potentially enhanced performance [33]. Nonetheless, balancing the annotation costs with the demand for high-resolution imagery led us to fix the final cropping size to 512×512 pixels. For the Chinese dataset, we selected a representative 512×512 subimage from each original image, whereas for images from other countries, a 1024×1024 region was first cropped from the center and then subdivided into one to four subimages using a sliding-window approach, with each subimage carefully inspected for quality.

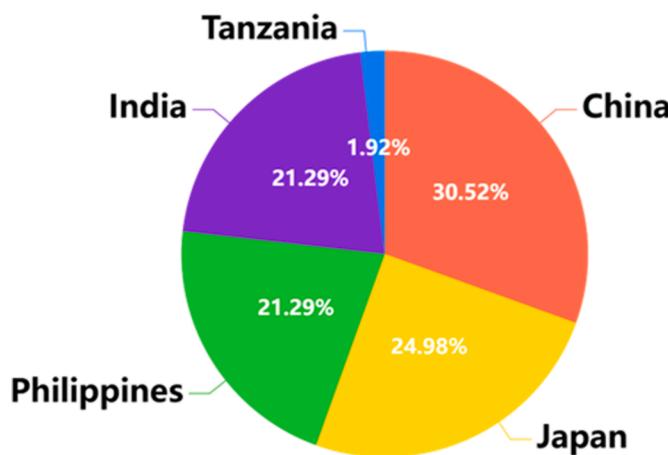


Fig. 1. Global range distribution and composition of the datasets (a detailed station distribution map is shown in Fig. S1).

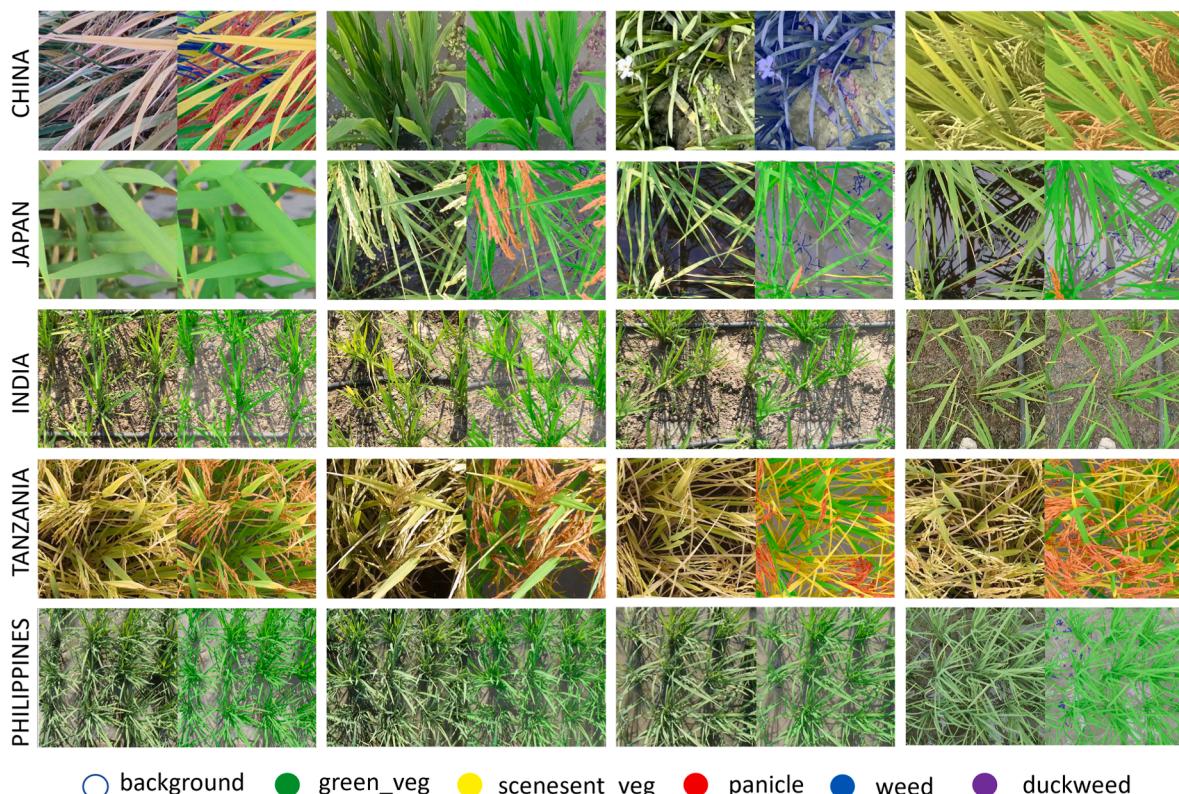


Fig. 2. Image annotation process.

Table 3
Statistics of the RiceSEG dataset.

Name	Images	No. of classes	Category Proportions (%)					
			background	green_veg	senescent_veg	panicle	weed	duckweed
JS_1	100	5	47.1	50.6	0.2	0.5		1.6
JS_2	100	4	53.9	44.4	0.2			1.6
JS_3	100	6	25.3	22.9	0.4	0.4	36.2	14.9
JS_4	80	5	11.4	37.1	32.7	16.4	2.5	
HN	100	5	56.8	42.6	0.2		0.1	0.3
GX	60	3	79.0	19.0	1.9			
JX	60	2	76.8	23.2				
HB	100	4	24.9	67.2	5.6	2.4		
HLJ	100	2	48.6	51.5				
GD	100	4	8.3	72.5	4	15.1		
LN	60	6	21.5	70.7	3.6	1.6	0.2	2.5
HUN	100	5	3.0	66.4	4.5	26.1	0.1	
JL	60	5	6.4	63.2	1.2	26.4		2.8
TKO_1	100	5	49.6	47.6	2.1	2.2	0.7	
TKO_2	504	6	67.4	28	0.8	2.3	0.8	0.8
TKO_3	100	4	15.4	82.3	2.1	0.2		
TG	600	5	58.9	39.1	1.2	0.2	0.6	
Kilimanjaro	54	4	15.7	27.5	25.8	30.9		
Laguna	600	4	55.7	43.5	0.3		0.3	
Summary 3078	3078	6	48.3	43.4	2.5	3.4	1.6	0.8

2.3. Data annotation

We engaged specially trained volunteers, primarily graduate students studying agronomy at Nanjing Agricultural University, to manually annotate the images. In total, the annotation process involved 11 volunteers, with the time cost for each image ranging from 0.5 to 1.5 h, depending on its complexity. Collectively, the annotators dedicated 2440 h to data annotation and an additional 800 h to verification and refinement, culminating in a total of 3240 h.

The training program included fundamental knowledge of rice growth physiology, equipping annotators to identify diverse characteristics and morphological traits of rice at various growth stages. Participants were further trained to categorize each pixel into one of six predefined classes: background, green vegetation, senescent vegetation, panicle, weed, and duckweed, which were labeled as numbers from 0 to 5, respectively. Detailed explanations and annotation samples of each category were provided to ensure consistent classification criteria (Fig. 2). Moreover, annotators were trained to use a JavaScript-based image annotation tool (<https://github.com/kyamagu/js-segment-annotator>) [34]. This tool was selected because it was developed on the basis of the superpixel annotation method. This approach significantly enhances annotation efficiency while ensuring precise alignment with natural boundaries. Note that annotators were required to adjust the superpixel resolution carefully to capture fine details and textures in the rice images.

To ensure annotation quality and consistency, a strict protocol was followed throughout the process. After the initial round of annotation, approximately 10 % of the labeled images from each annotator were randomly selected for double-checking by a second annotator. During this process, common misclassifications were identified and corrected, with documentation provided by the project leader. Feedback was then promptly given to the annotators to improve their practices. In summarizing the lessons learned from this iterative annotation process, we found that among the six categories, distinguishing senescent leaves, particularly those at the bottom of the canopy with substantial shadows, was often challenging. Additionally, residual plant matter from previous crop rotations sometimes resembled senescent rice, further complicating the labeling task. To minimize subjectivity, each annotation was cross-verified by at least three individuals to ensure reliability. Finally, weeds such as water onions, which structurally resemble rice at certain growth stages, were sometimes misclassified as green vegetation. Extra care was taken to maintain precision in the annotations.

Owing to the nature of agricultural ecosystems, the labels in the RiceSEG dataset are not evenly distributed across categories, as expected (Table 3). The background category is the most dominant, accounting for nearly 50 % of all labels. Following this, the green vegetation category ('green_veg') ranks second, comprising more than 40 %, as green plants cover a significant portion of the rice fields and are the primary visual component throughout the growth cycle. In contrast, categories such as senescent vegetation ('senescent_veg') and rice panicle ('panicle') appear only during the reproductive stage and thus represent a relatively small proportion of the dataset. Additionally, owing to the use of herbicides across all the experimental sites, the presence of weeds and duckweed is minimal. We released an additional file, class_pixel_counts.csv, which lists the per-image pixel count of all six classes to facilitate customized subset creation.

2.4. Baseline test

2.4.1. Baseline models

To create the baseline accuracy for the RiceSEG dataset, we determined baseline results for six semantic segmentation models divided into two major categories (Table 4): Convolutional neural networks (CNN) and transformer-based models. We chose the FCN [35,36], PSPNet [37], and DeepLabV3+ [38], which are three methods based on the CNN backbone. With respect to the Transformer architecture, we adopted SegFormer [39], KNet [40] and Mask2Former [41]. These models represent the classic and cutting-edge technologies in the field of semantic segmentation. Our RiceSEG dataset was randomly split 8:2 for the training and test datasets. All six models selected were trained and tested accordingly.

2.4.2. Evaluation metrics

At the pixel scale, to evaluate the baseline models, we determined the intersection over union (IoU) and accuracy for each class while using the mean intersection over union (mIoU) and mean accuracy (mAcc) as performance metrics across all classes. Afterward, at the image scale, we calculated the proportion of each class in the entire image and compared it with the corresponding proportions from the manually labeled images. Furthermore, we calculated R² and RMSE values to assess the model's performance.

Table 4
Baseline model for semantic segmentation.

Model		Backbone	Venue	Key Features
CNN backbone	FCN	Resnet50	2015-CVPR/2017-TPAMI [35,36]	Fully convolutional network for semantic segmentation
	PSPNet		2017-CVPR [37]	Employing pyramid pooling to capture multiscale contextual information.
	DeepLabV3+		2018-ECCV [38]	Combining atrous convolutions with a new decoder for enhanced boundary delineation.
Transformer backbone	SegFormer	Mit b0	2021-NeurIPS [39]	Efficient transformer-based model with a lightweight MLP decoder.
	KNet	SwinT	2021-NeurIPS [40]	Uses kernel-based convolution for multiscale feature extraction.
	Mask2Former		2022-CVPR [41]	Unifies semantic and instance segmentation with dynamic mask prediction.

3. Results

3.1. Dataset diversity analysis

The UMAP projections of the RiceSEG samples from the five contributing countries are shown in Fig. 3. Overall, the data from China exhibit a relatively larger distribution area because of the broad variation in genotype–environment–management factors in the collected rice images. In contrast, the distribution of samples from the other four countries largely overlaps with that of the Chinese dataset but within a narrower domain. Nevertheless, the datasets from these four countries

demonstrate distinct distribution patterns. Ultimately, the combined samples from all five countries contribute to expanding the dataset's distribution and improving its representation of the diverse range of high-resolution field images of rice.

3.2. Baseline results at the pixel scale

With respect to the average performance across all the classes, the transformer-based segmentation models outperformed their CNN counterparts (Table 5). Specifically, all the baseline models generally performed well in segmenting background, green vegetation, and panicles. However, significant differences were observed in more challenging categories, such as senescent vegetation, weeds, and duckweed. With respect to senescent vegetation, none of the models yielded satisfactory results; the best-performing model, Mask2Former, achieved an IoU of only 52.98, and SegFormer achieved an ACC of 66.47. For weeds, although the classification accuracy of the top-performing model reached 77.06, the IoU remained low at 65.73 (Fig. 4).

The segmentation performance of all the baseline models on the test set is shown in Fig. 5. During the vegetative growth stages, the majority of the images consist of green vegetation and background, which are highly contrasted and easily distinguishable. As a result, only minor differences in segmentation performance were observed among the models during this phase. However, during the transition phase, segmentation became more challenging because of the emergence of weeds and duckweed. The high morphological similarity between these and rice parts leads to misidentification as rice, resulting in less accurate segmentation and an increased occurrence of false positives. In the reproductive stage, the canopy begins to saturate, leaving only a small portion of the background visible. This leads to the misclassification of yellow leaves, which are predominantly classified as green vegetation or background. Achieving reliable recognition performance remains difficult for both traditional CNN models and state-of-the-art transformer models.

3.3. Baseline results at the imaging scale

At the image scale, for green vegetation and panicles, the models generally performed well. However, for more complex categories, such as weeds and senescent vegetation, the CNN models performed poorly.

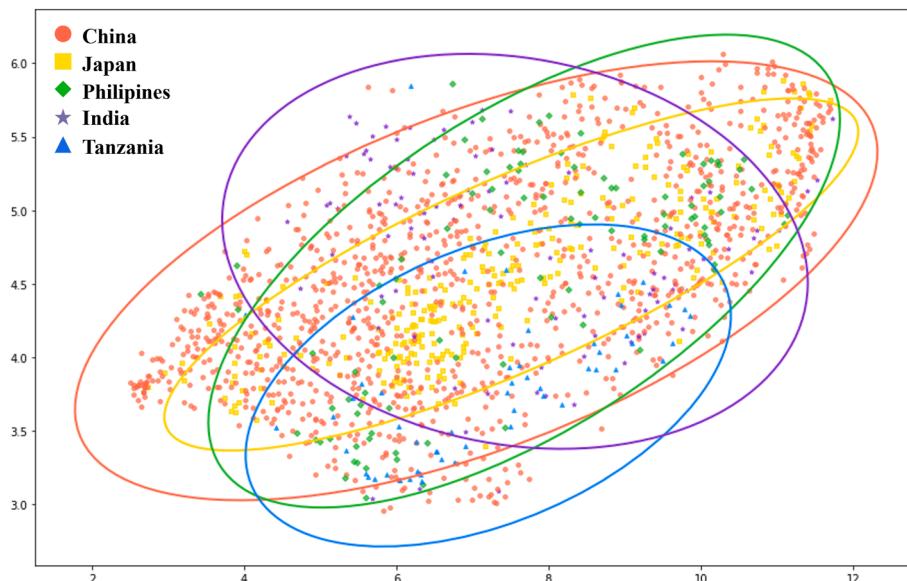
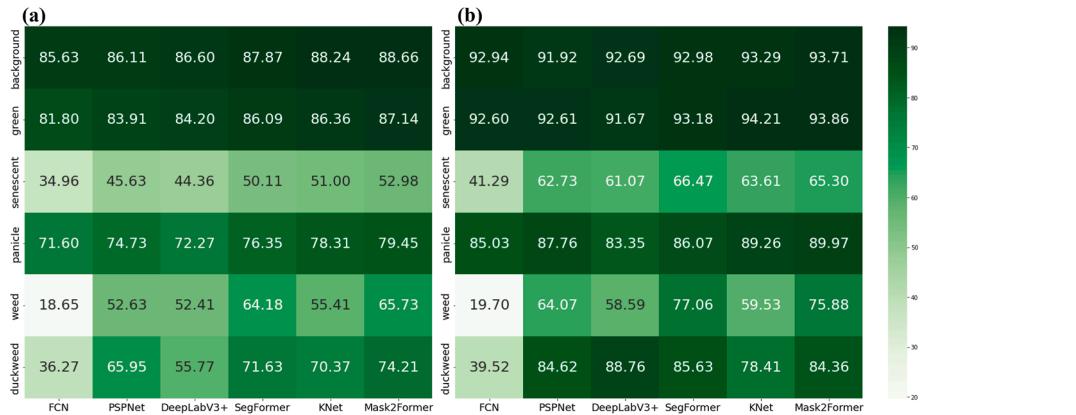


Fig. 3. Distribution of subdatasets from different contributions. The images were projected into two dimensions using uniform manifold approximation and projection (UMAP). Each colored domain represents the confidence ellipse (90 %) of the country's dataset. The color of each ellipse matches the points of its respective country.

Table 5

Performances of different models on the RiceSEG dataset.

Metrics	CNN backbone						Transformer backbone			
	FCN		PSPNet		DeepLabV3+		SegFormer		KNet	Mask2Former
	mIoU	54.82	68.16	65.93	72.70	71.87	74.69			
mAcc	61.85	80.48	79.35	83.57	80.50	83.85				

**Fig. 4.** Segmentation performance at the pixel scale. Six classical and state-of-the-art semantic segmentation models were compared in terms of the IOU (a) and ACC (b). The test dataset includes 601 images in 6 classes.

In contrast, transformer-based models significantly improved performance (Fig. 6). Furthermore, we demonstrated the dynamics of the rice canopy from the seedling stage to the maturity stage on the basis of the best-performing Mask2Former model (Fig. 7). This further indicates that dispersals of the segmentation at the reproductive stage are consistent with those at the pixel scale.

4. Discussion

4.1. Potential contribution of the RiceSEG dataset

To the best of our knowledge, we created the first comprehensive multiclass rice semantic segmentation dataset, RiceSEG. We gathered nearly 50,000 high-resolution, ground-based images from five major rice-growing countries (China, Japan, India, the Philippines, and Tanzania), encompassing more than 6000 genotypes across all growth stages. From these original images, 3078 representative samples were selected to form the RiceSEG dataset. Notably, the subdataset from China spans all major genotypes and rice-growing environments from northeastern to southern regions. The RiceSEG uniquely captures key rice crop organs, including the primary source organs—leaves (classified as green and senescent)—and the sink organ, the panicle. Unlike previous rice segmentation datasets, which were limited to binary segmentation of vegetation and background [14], our dataset enables the development of advanced segmentation models to track the dynamics of these critical organs throughout the entire rice growth cycle (Fig. 6). Providing detailed time series data on organ development offers insights that are unattainable through manual measurements, potentially revealing new ecophysiological processes underlying crop adaptation to local environments and yield formation [42]. Additionally, the dataset incorporates both aquatic and nonaquatic weeds, enabling simultaneous segmentation of weeds and rice crops. By facilitating accurate weed and crop differentiation, datasets may play a crucial role in the development of advanced computer vision models for automated weed control, addressing the growing demand for precision agriculture solutions such as field robots [43]. However, existing models encounter difficulties during the reproductive stage, when canopy structures become more complex and multiple classes are involved. These findings highlight the importance of our dataset for developing specialized segmentation

models for rice and other crops. Finally, through collaboration with international partners, we expanded the dataset to include samples from 5 countries, representing diverse genotype–environment–management combinations. This broad representation ensures the robustness and scalability of the resulting segmentation models, enabling precise differentiation of fine phenotypic traits among hundreds or even thousands of genotypes for breeding programs.

4.2. Challenges in rice image segmentation

Compared with other computer vision tasks, semantic segmentation in agriculture—particularly for rice—presents unique challenges. In the broader computer vision field, widely used datasets such as COCO and ADE20K typically encompass a larger number of categories and significantly more images than RiceSEG. However, these general-purpose datasets predominantly feature large objects with relatively planar surfaces, whereas crop images often contain dense, finely detailed structures—primarily leaves—characterized by numerous edges and complex spatial arrangements. This inherent complexity is further compounded by varying illumination within the canopy, where mutual shading and reduced light transmittance at greater canopy depths make it particularly difficult to segment leaves located near the bottom. In paddy rice fields, water surfaces introduce additional complications, including reflections and mirror-like effects that resemble vegetation, whereas submerged or partially submerged weeds add yet another layer of segmentation difficulty. Although a few existing datasets address crop image segmentation, their limited scope and categories do not fully capture the complexity of real-world agricultural settings. Consequently, our RiceSEG dataset offers distinct value for developing and validating specialized segmentation models tailored to rice and other plant species.

Because crop image segmentation datasets are both scarce and unique, current state-of-the-art methods are not fully optimized because of the complexities inherent in rice imagery. Nonetheless, owing to the robust feature-extraction capabilities of deep learning models, most tested architectures accurately classify dominant image components (e.g., background and green vegetation) at the pixel level. Beyond pixel-level performance, we also evaluated segmentation accuracy at the image scale, as many phenotypic trait estimations (e.g., the green

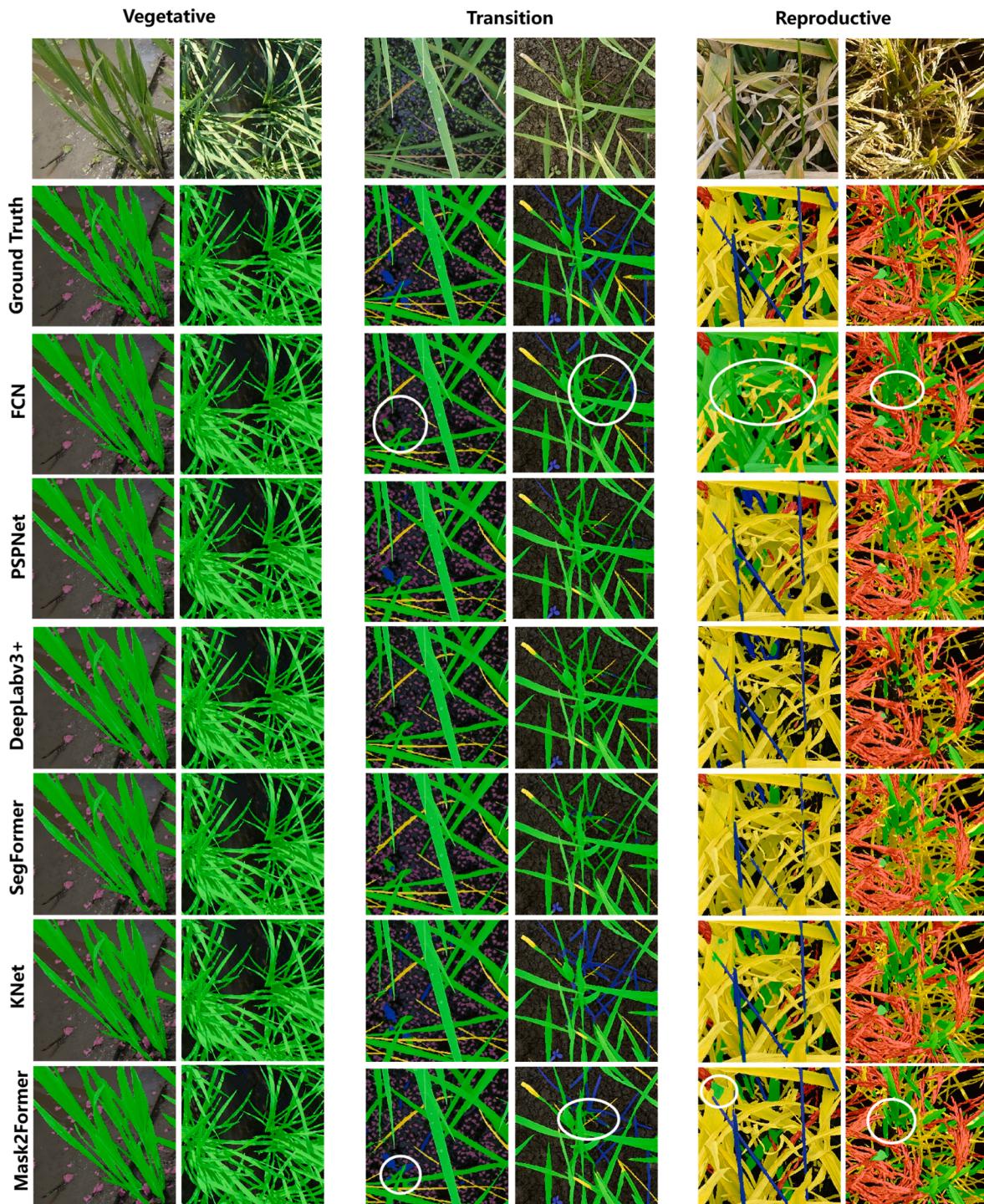


Fig. 5. Visualization of the segmentation results on the test set. Six classical and state-of-the-art semantic segmentation models were tested on three growth stages of rice. The white circles represent key misclassified regions.

vegetation fraction for the green area index, [11]) depend on organ-specific pixel fractions. Overall, image-level evaluations largely parallel pixel-level results but exhibit slight improvements, potentially because of compositional effects across each image. However, pronounced performance gaps remain in more challenging categories, such as senescent vegetation, weeds, and duckweed. Transformer-based models (e.g., SegFormer and Mask2Former) demonstrate superiority in these domains, likely because their self-attention mechanisms capture long-distance dependencies and effectively handle intricate visual patterns [44]. In contrast, CNN-based architectures, which primarily

extract local features, struggle to recognize fine structures that require a more global contextual understanding [45]. Moving forward, research could focus on further refining transformer-based models to enhance segmentation performance in these nuanced categories.

4.3. Limitations of the dataset

We made significant efforts to collect rice images from the most representative rice-growing conditions. Nevertheless, our dataset still has limitations regarding its overall representativeness. For instance, in

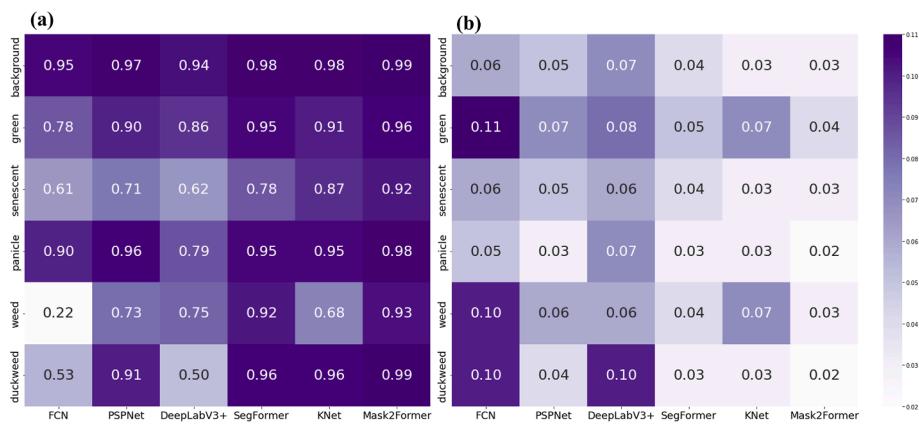


Fig. 6. Segmentation performance at the image scale. Six classical and state-of-the-art semantic segmentation models were compared in terms of R2 (a) and RMSE (b) values. The test dataset includes 601 images in 6 classes. The vertical axis corresponds to the proportion of each class's pixels relative to the total pixels in the entire image.

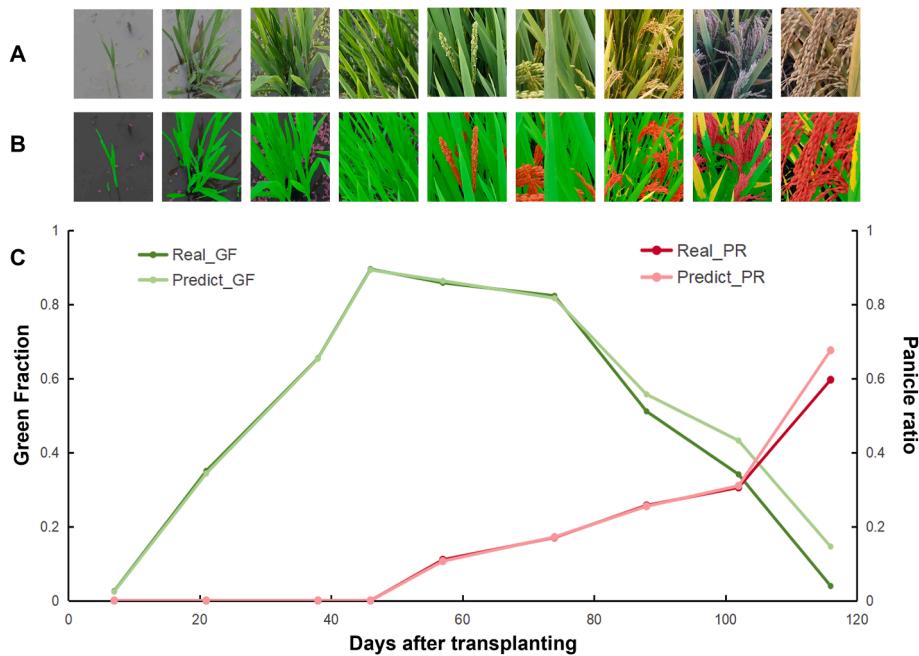


Fig. 7. Dynamics of the GF and leaf-to-panicle ratio generated by time series images. (A) RGB images and (B) estimated (C) dynamics of the GF and panicle ratio.

China, we gathered images from nearly all major rice-producing regions, capturing a wide range of genotype–environment–management combinations. In contrast, although we obtained an almost equivalent number of images in Japan, the Philippines, and India, their geographic and genotypic diversity is much narrower, potentially biasing the model toward Chinese conditions and reducing its generalizability elsewhere. Additionally, in assembling each site's dataset, we included images spanning all growth stages to improve the model's ability to handle the entire crop cycle. Despite this, the annotated pixel counts across categories are imbalanced, particularly for senescent leaves, which constitute only 2.8 % of the annotated pixels. This imbalance may partly account for the relatively low segmentation accuracy observed for senescent leaves (Fig. 5). However, for images collected in natural environments, such a pixel distribution is a normal representation of the natural world. Another factor could be the inherent ambiguity of annotating senescent leaves, especially those in lower canopy layers where shading is more pronounced. Furthermore, our current dataset does not include a detailed classification of weeds. To achieve more precise in-field weed management, a broader range of weed species is

essential. Therefore, we are considering both collecting more field data and employing data generation techniques to further diversify the dataset. Previous studies [46–50] have demonstrated the potential of synthetic data: by supplementing a small set of real images and labels with high-quality virtual samples, it is possible to approach the desired accuracy while greatly reducing annotation costs. Additionally, foundation models [51] trained on large-scale unsupervised data have shown strong performance across a wide range of downstream tasks—even when only small amounts of labeled data are available.

To facilitate distribution and track updates, we provide detailed descriptions of the dataset at <http://www.global-rice.com> and <http://www.phenix-lab.com>. Unfortunately, open-access datasets remain scarce in plant phenotyping research. In contrast, the computer vision community has achieved rapid progress largely through shared resources that reduce redundant efforts and enhance efficiency. We encourage more researchers in plant phenotyping and digital agriculture to collectively foster an open-access culture. Such collaboration will expedite the development of robust deep learning algorithms for

agricultural applications, ultimately having a greater impact on crop breeding and smart farming.

Authors' contributions

Conceptualization, S.L., J.Z., and W.G.; Methodology, S.L., and W.G.; Software, J.Z., and H.W.; Validation, J.Z.; Formal analysis, S.L., J.Z., and W.G.; Investigation, J.Z., Y.G. and F.X.; Data curation, W.G., Y.K., T.N., M.B., P.R., K.K., H.L., Y.M., H.C., Y.C., W.Y., W.L., J.W., F.Y., J.Z., X.H., Y.Y. and W.W.; Writing original draft preparation, J.Z.; Writing review and editing, J.Z., S.L., and W.G.; and Supervision, S.L. and W.G.

Funding

This work was supported by the National Key R&D Program of China (No. 2022YFE0116200 and No. 2022YFD2300700), the Young Scientists Fund of the National Natural Science Foundation of China (No. 42201437 and No. 32201893), the PhD Scientific Research and Innovation Foundation of The Education Department of Hainan Province Joint Project of Sanya Yazhou Bay Science and Technology City (No. HSPHDSRF-2024-09-001), the Hainan Provincial Natural Science Foundation of China (No. 325QN370), the “JBGS” Project of Seed Industry Revitalization in Jiangsu Province (JBGS [2021] 007), the Japan Society for the Promotion of Science (No. 22KK0083 and No. JP25H01110), and the Sarabetsu Village “Endowed Chair for Field Phenomics” project in Hokkaido, Japan.

Data availability

The RiceSEG dataset is publicly available at <http://www.global-rice.com/>.

Conflicts of interest

The authors declare that there are no conflicts of interest relevant to the content of this article.

Acknowledgment

We thank Professor Shirong Zhou from the State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University and Professor Lizhong Xiong from the National Key Laboratory of Crop Genetic Improvement at Huazhong Agricultural University for providing data support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.plaph.2025.100099>.

References

- [1] Z. Jin, T. Shah, L. Zhang, H. Liu, S. Peng, L. Nie, Effect of straw returning on soil organic carbon in rice–wheat rotation system: a review, *Food Energy Secur.* 9 (2) (2020) e2002, <https://doi.org/10.1002/fes.200>.
- [2] H.C.J. Godfray, J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, C. Toulmin, Food security: the challenge of feeding 9 billion people, *Science* 327 (5967) (2010) 812–818, <https://doi.org/10.1126/science.1185383>.
- [3] K.G. Cassman, R.R. Harwood, The nature of agricultural systems: food security and environmental balance, *Food Policy* 20 (5) (1995) 439–454, [https://doi.org/10.1016/0306-9192\(95\)00037-F](https://doi.org/10.1016/0306-9192(95)00037-F).
- [4] C. Chen, H. Mcnairn, A neural network integrated approach for rice crop monitoring, *Int. J. Rem. Sens.* 27 (7) (2006) 1367–1393, <https://doi.org/10.1080/01431160500421507>.
- [5] S. Madec, X. Jin, H. Lu, B. De Solan, S. Liu, F. Duyme, E. Heritier, F. Baret, Ear density estimation from high resolution RGB imagery using deep learning technique, *Agric. For. Meteorol.* 264 (2019) 225–234, <https://doi.org/10.1016/j.agrformet.2018.10.013>.
- [6] D. Mandal, V. Kumar, A. Bhattacharya, Y.S. Rao, P. Siqueira, S. Bera, Sen4Rice: a processing chain for differentiating early and late transplanted rice using time-series Sentinel-1 SAR data with google Earth engine, *IEEE Geosci. Remote Sens. Lett.* 15 (12) (2018) 1947–1951, <https://doi.org/10.1109/LGRS.2018.2865816>.
- [7] W. Maohua, Possible adoption of precision agriculture for developing countries at the threshold of the new millennium, *Comput. Electron. Agric.* 30 (1) (2001) 45–50, [https://doi.org/10.1016/S0168-1699\(00\)00154-X](https://doi.org/10.1016/S0168-1699(00)00154-X).
- [8] A.R. Mermut, H. Eswaran, Some major developments in soil science since the mid-1960s, *Geoderma* 100 (3) (2001) 403–426, [https://doi.org/10.1016/S0016-7061\(01\)00030-1](https://doi.org/10.1016/S0016-7061(01)00030-1).
- [9] F. Yan dun Narvaez, G. Reina, M. Torres-Torriti, G. Kantor, F.A. Cheein, A survey of ranging and imaging techniques for precision agriculture phenotyping, *IEEE ASME Trans. Mechatron.* 22 (6) (2017) 2428–2439, <https://doi.org/10.1109/TMECH.2017.2760866>.
- [10] Z. Li, R. Guo, M. Li, Y. Chen, G. Li, A review of computer vision technologies for plant phenotyping, *Comput. Electron. Agric.* 176 (2020) 105672, <https://doi.org/10.1016/j.compag.2020.105672>.
- [11] J. Wang, R. Lopez-Lozano, M. Weiss, S. Buis, W. Li, J. Zhang, F. Baret, Estimating Green Area Index (GAI) from Radiative Transfer Model: Application of the Bayesian Theory to Account for crop-specificities, *2020*, 2020, B009-10.
- [12] L. Shouyang, J. Shichao, G. Qinghua, Z. Yan, B. Fred, An algorithm for estimating field wheat canopy light interception based on digital plant phenotyping Platform, *Smart Agric.* 2 (1) (2020) 87, <https://doi.org/10.12133/j.smartag.2020.2.1.202002-SA004>.
- [13] F. Baret, S. Madec, K. Irfan, J. Lopez, A. Comar, M. Hemmerlé, D. Dutartre, S. Praud, M.H. Tixer, Leaf-rolling in maize crops: from leaf scoring to canopy-level measurements for phenotyping, *J. Exp. Bot.* 69 (10) (2018) 2705–2716, <https://doi.org/10.1093/jxb/ery071>.
- [14] Y. Gao, Y. Li, R. Jiang, X. Zhan, H. Lu, W. Guo, W. Yang, Y. Ding, S. Liu, Enhancing green fraction estimation in rice and wheat crops: a self-supervised deep learning semantic segmentation approach, *Plant Phenomics* 5 (2023) 64, <https://doi.org/10.34133/plantphenomics.0064>.
- [15] M. Serouart, S. Madec, E. David, K. Velumani, R. Lopez Lozano, M. Weiss, F. Baret, SegVeg: segmenting RGB images into green and senescent vegetation by combining deep and shallow methods, *Plant Phenomics* 2022 (2022), <https://doi.org/10.34133/2022/9803570>.
- [16] Y. Gao, L. Li, M. Weiss, W. Guo, M. Shi, H. Lu, R. Jiang, Y. Ding, T. Nampally, P. Rajalakshmi, F. Baret, S. Liu, Bridging real and simulated data for cross-spatial-resolution vegetation segmentation with application to rice crops, *ISPRS J. Photogrammetry Remote Sens.* 218 (2024) 133–150, <https://doi.org/10.1016/j.isprsjprs.2024.10.007>.
- [17] Z. Zhao, C. Wang, X. Yu, Y. Tian, W. Wang, Y. Zhang, W. Bai, N. Yang, T. Zhang, H. Zheng, Q. Wang, J. Lu, D. Lei, X. He, K. Chen, J. Gao, X. Liu, S. Liu, L. Jiang, J. Wan, Auxin regulates source-sink carbohydrate partitioning and reproductive organ development in rice, *Proc. Natl. Acad. Sci.* 119 (36) (2022) e2121671119, <https://doi.org/10.1073/pnas.2121671119>.
- [18] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, P. Dollár, R. Girshick, *Segment anything* (arXiv: 2304.02643), arXiv (2023). Retrieved from, <http://arxiv.org/abs/2304.02643>.
- [19] A. Garcia-Garcia, S. Orts-Escalona, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A review on deep learning techniques applied to semantic segmentation, *arXiv.Org* (2017, April 22). Retrieved from, <https://arxiv.org/abs/1704.06857v1>.
- [20] X. Bai, P. Liu, Z. Cao, H. Lu, H. Xiong, A. Yang, Z. Cai, J. Wang, J. Yao, Rice plant counting, locating, and sizing method based on high-throughput UAV RGB images, *Plant Phenomics* 5 (2023) 20, <https://doi.org/10.34133/plantphenomics.0020>.
- [21] E. David, S. Madec, P. Sadeghi-Tehran, H. Aasen, B. Zheng, S. Liu, N. Kirchgessner, G. Ishikawa, K. Nagasawa, M.A. Badhon, C. Pozniak, B. de Solan, A. Hund, S. Chapman, F. Baret, I. Stavness, W. Guo, Global Wheat Head Detection (GWHD) dataset: a large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods, *Plant Phenomics* 2020 (2020), <https://doi.org/10.34133/2020/3521852>.
- [22] E. David, M. Serouart, D. Smith, S. Madec, K. Velumani, S. Liu, X. Wang, F. Pinto, S. Shafiee, I.S.A. Tahir, H. Tsujimoto, S. Nasuda, B. Zheng, N. Kirchgessner, H. Aasen, A. Hund, P. Sadeghi-Tehran, K. Nagasawa, G. Ishikawa, W. Guo, Global wheat head detection 2021: an improved dataset for benchmarking wheat head detection methods, *Plant Phenomics* (2021), <https://doi.org/10.34133/2021/9846158>, 2021.
- [23] H.B. Prajapati, J.P. Shah, V.K. Dabhi, Detection and classification of rice plant diseases, *Intell. Decis. Technol.* 11 (3) (2017) 357–373, <https://doi.org/10.3233/IDT-170301>.
- [24] X. Wu, C. Zhan, Y.-K. Lai, M.-M. Cheng, J. Yang, IP102: a large-scale benchmark dataset for insect Pest recognition, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 8779–8788, <https://doi.org/10.1109/CVPR.2019.00899>.
- [25] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic urban scene understanding, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3213–3223, <https://doi.org/10.1109/CVPR.2016.350>.
- [26] B.C. Russell, A. Torralba, K.P. Murphy, W.T. Freeman, LabelMe: a database and web-based tool for image annotation, *Int. J. Comput. Vis.* 77 (1) (2008) 157–173, <https://doi.org/10.1007/s11263-007-0090-8>.
- [27] H. Wang, S. Lyu, Y. Ren, Paddy rice imagery dataset for panicle segmentation, *Agronomy* 11 (8) (2021) 1542, <https://doi.org/10.3390/agronomy11081542>.
- [28] S. Madec, K. Irfan, K. Velumani, F. Baret, E. David, G. Daubige, L.B. Samatan, M. Serouart, D. Smith, C. James, F. Camacho, W. Guo, B. De Solan, S.C. Chapman,

- M. Weiss, VegAnn, vegetation annotation of multi-crop RGB images acquired under diverse conditions for segmentation, *Sci. Data* 10 (1) (2023) 302, <https://doi.org/10.1038/s41597-023-02098-y>.
- [29] H. Scharr, M. Minervini, A. Fischbach, S. Tsafaris, *Annotated Image Datasets of Rosette Plants*, 2014.
- [30] S. Haug, J. Ostermann, A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks, in: L. Agapito, M.M. Bronstein, C. Rother (Eds.), *Computer Vision - ECCV 2014 Workshops*, Springer International Publishing, Cham, 2015, pp. 105–116, https://doi.org/10.1007/978-3-319-16220-1_8.
- [31] A.K. Mortensen, S. Skovsen, H. Karstoft, R. Gislum, The oil radish growth dataset for semantic segmentation and yield estimation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 2703–2710, <https://doi.org/10.1109/CVPRW.2019.00328>.
- [32] J. Weyler, F. Magistri, E. Marks, Y.L. Chong, M. Sodaño, G. Roggiolani, N. Chebrolu, C. Stachniss, J. Behley, PhenoBench – A large dataset and benchmarks for semantic image interpretation in the agricultural domain (arXiv:2306.04557), arXiv (2023), <https://doi.org/10.48550/arXiv.2306.04557>.
- [33] Z. Jia, J. Chen, X. Xu, J. Kheir, J. Hu, H. Xiao, S. Peng, X.S. Hu, D. Chen, Y. Shi, The importance of resource awareness in artificial intelligence for healthcare, *Nat. Mach. Intell.* 5 (7) (2023) 687–698, <https://doi.org/10.1038/s42256-023-00670-0>.
- [34] P. Tangseng, Z. Wu, K. Yamaguchi, *Looking at outfit to parse clothing* (arXiv: 1703.01386), arXiv (2017), <https://doi.org/10.48550/arXiv.1703.01386>.
- [35] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation (arXiv:1411.4038), arXiv (2015), <https://doi.org/10.48550/arXiv.1411.4038>.
- [36] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (4) (2017) 640–651, <https://doi.org/10.1109/TPAMI.2016.2572683>.
- [37] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid Scene Parsing Network, 2017, pp. 6230–6239, <https://doi.org/10.1109/CVPR.2017.660>.
- [38] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), *Computer Vision – ECCV 2018*, Springer International Publishing, Cham, 2018, pp. 833–851, https://doi.org/10.1007/978-3-030-01234-2_49.
- [39] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: simple and efficient design for semantic segmentation with transformers, *Adv. Neural Inf. Process. Syst.* 34 (2021) 12077–12090. Retrieved from, <https://proceedings.neurips.cc/paper/2021/hash/64f127bf1b4ec22924fd0acb550c235-Abstract.html>.
- [40] W. Zhang, J. Pang, K. Chen, C.C. Loy, K-Net: towards unified image segmentation, in: *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2024, pp. 10326–10338.
- [41] B. Cheng, I. Misra, A.G. Schwing, A. Kirillov, R. Girdhar, Masked-Attention Mask Transformer for Universal Image Segmentation (arXiv:2112.01527), arXiv, 2022, <https://doi.org/10.48550/arXiv.2112.01527>.
- [42] T.-G. Chang, Z.-W. Wei, Z. Shi, Y. Xiao, H. Zhao, S.-Q. Chang, M. Qu, Q. Song, F. Chen, F. Miao, X.-G. Zhu, Bridging photosynthesis and crop yield formation with a mechanistic model of whole-plant carbon–nitrogen interaction, *Silico Plants* 5 (2) (2023) diad011, <https://doi.org/10.1093/insilicoplants/diad011>.
- [43] A. Wang, W. Zhang, X. Wei, A review on weed detection using ground-based machine vision and image processing techniques, *Comput. Electron. Agric.* 158 (2019) 226–240, <https://doi.org/10.1016/j.compag.2019.02.005>.
- [44] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: transformers for image recognition at scale, in: *International Conference on Learning Representations*, 2020, October 2. Retrieved from, <https://openreview.net/forum?id==YicbFdNTTy>.
- [45] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [46] Y. Gao, Y. Li, R. Jiang, X. Zhan, H. Lu, W. Guo, W. Yang, Y. Ding, S. Liu, Enhancing green fraction estimation in rice and wheat crops: a self-supervised deep learning semantic segmentation approach, *Plant Phenomics* 5 (2023) 64, <https://doi.org/10.34133/plantphenomics.0064>.
- [47] Y. Gao, L. Li, M. Weiss, W. Guo, M. Shi, H. Lu, R. Jiang, Y. Ding, T. Nampally, P. Rajalakshmi, F. Baret, S. Liu, Bridging real and simulated data for cross-spatial-resolution vegetation segmentation with application to rice crops, *ISPRS J. Photogrammetry Remote Sens.* 218 (2024) 133–150, <https://doi.org/10.1016/j.isprjprs.2024.10.007>.
- [48] Y. Li, X. Zhan, S. Liu, H. Lu, R. Jiang, W. Guo, S. Chapman, Y. Ge, B. Solan, Y. Ding, F. Baret, Self-supervised plant phenotyping by combining domain adaptation with 3D plant model simulations: application to wheat leaf counting at seedling stage, *Plant Phenomics* 5 (2023) 41, <https://doi.org/10.34133/plantphenomics.0041>.
- [49] S. Liu, P. Martre, S. Buis, M. Abichou, B. Andrieu, F. Baret, Estimation of plant and canopy architectural traits using the digital plant phenotyping Platform1 [OPEN], *Plant Physiol.* 181 (3) (2019) 881–890, <https://doi.org/10.1104/pp.19.00554>.
- [50] S. Liu, F. Baret, M. Abichou, L. Manceau, B. Andrieu, M. Weiss, P. Martre, Importance of the description of light interception in crop growth models, *Plant Physiol.* 186 (2) (2021) 977–997, <https://doi.org/10.1093/plphys/kiab113>.
- [51] M. Oquab, T. Darcret, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, P. Bojanowski, DINov2: learning robust visual features without supervision (arXiv:2304.07193), arXiv (2024). Retrieved from, <http://arxiv.org/abs/2304.07193>.