

## Article

# Paddy Rice Imagery Dataset for Panicle Segmentation

Hao Wang<sup>1,2,\*</sup>, Suxing Lyu<sup>2</sup> and Yaxin Ren<sup>3</sup>

<sup>1</sup> Beijing Research Center of Intelligent Equipment for Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

<sup>2</sup> Graduate School of Agriculture, Hokkaido University, Sapporo 065-8589, Hokkaido, Japan; johnlyu@bpe.agr.hokudai.ac.jp

<sup>3</sup> Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; renyx@nercita.org.cn

\* Correspondence: wangh@nercita.org.cn; Tel.: +86-10-51503748

**Abstract:** Accurate panicle identification is a key step in rice-field phenotyping. Deep learning methods based on high-spatial-resolution images provide a high-throughput and accurate solution of panicle segmentation. Panicle segmentation tasks require costly annotations to train an accurate and robust deep learning model. However, few public datasets are available for rice-panicle phenotyping. We present a semi-supervised deep learning model training process, which greatly assists the annotation and refinement of training datasets. The model learns the panicle features with limited annotations and localizes more positive samples in the datasets, without further interaction. After the dataset refinement, the number of annotations increased by 40.6%. In addition, we trained and tested modern deep learning models to show how the dataset is beneficial to both detection and segmentation tasks. Results of our comparison experiments can inspire others in dataset preparation and model selection.

**Keywords:** image segmentation; panicle detection; deep learning; smart agriculture; unmanned aerial vehicle platform



**Citation:** Wang, H.; Lyu, S.; Ren, Y. Paddy Rice Imagery Dataset for Panicle Segmentation. *Agronomy* **2021**, *11*, 1542. <https://doi.org/10.3390/agronomy11081542>

Academic Editor:  
Pablo Martín-Ramos

Received: 8 June 2021  
Accepted: 29 July 2021  
Published: 31 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the face of global population growth and climate change, plant breeders and agronomists strive for improved crop yields and quality to ensure regional food security, regulate the market price of the rice grain, and address food-shortage problems [1,2]. The rice-panicle density is one of the important agronomic components in understanding the grain yield and determining the growth period [3,4]. It also plays an important role in nutrition examination and disease detection [5]. Therefore, accurate panicle segmentation is a key step in rice-field phenotyping. However, conventional panicle observation relies to a large degree on the human labor force. It is tedious and labor-intensive. Moreover, the results are prone to unrepresentativeness, due to limited sampling areas [6,7]. Therefore, the panicle detection is desired to be conducted automatically, without harmful impact on the crops. For the application in the area of smart agriculture, the detection method is also required to be accurate, reliable, efficient, and possibly low in cost [6].

Computer-vision methods based on high-spatial-resolution images provide a potential solution to increase the throughput and the accuracy of counting panicles [4,7]. With the development of deep learning methods, convolutional neural networks have been shown to outperform human beings in diverse fields, including phenotyping [1], disease recognition [8], flowering, and ear counting [7,9,10]. Even though many studies have shown success in phenotyping tasks, developing robust and effective models to detect panicles from high-throughput phenotypic data remains a significant challenge, because of varying illumination, diversity of appearance of the panicles, shape deformation, partial occlusion, and complex background [11]. Moreover, requiring large training datasets limits the practical usage of deep learning models in panicle detection [4]. It is a fundamental problem

of most deep learning methods to reach an expected level of accuracy and robustness. Semantic labeling and instance segmentation are two tasks that require particularly costly annotations [12]. However, few public datasets are available for crop phenotyping. Some successful methods for panicle segmentation based on deep learning, such as Panicle-SEG [5], have generally been calibrated and validated on their own limited datasets. Therefore, annotation is still tedious and time-consuming. Data augmentation is a common strategy for handling limited datasets, such as rotation, mirror-reverse, scaling, contrast adjustment, affine transformation, and so on. However, generic data augmentation fails in some cases, and sophisticated choices have to be made for well training the neural network [13].

In recent years, unmanned aerial vehicle (UAV) has been used widely as a high-throughput way of acquiring data. Moreover, deep neural network methods, along with UAV-based photogrammetry, have been proved to be reliable alternatives to the labor-intensive field survey and crop monitoring, such as the phenotyping of wheat plant height and growth rate [14], sorghum head survey [13], crop species identification [15], and so on. However, as the same as the aforementioned difficulty, a large agricultural dataset acquired at the view of UAV is also infrequent. Insufficient data are a difficulty that makes the study go further, especially for specific scenarios. Combining drones and deep learning technologies, we propose a framework for dataset augmentation and refinement. We firstly annotated a part of rice panicles in 400 4K resolution images and trained a standard model, the Mask R-CNN [16], as the basic detector. Furthermore, we refined the whole dataset based on the basic detector. While taking into account the nature of rice panicle distribution and real-world field scenarios, we split the dataset into three subsets of different sizes. We trained and tested modern deep learning models to show how the dataset is beneficial to both detection and segmentation tasks. The contributions of this study can be summarized as follows:

- The proposed dataset argumentation and refinement framework significantly increases the volume of the dataset.
- Cross-experiments were designed to illustrate the usage of the dataset and the principle of model selection.
- The refined dataset is publicly accessible for boosting related research in the agricultural community. We provide the access in Section 4, the Conclusion.

## 2. Materials and Methods

### 2.1. Data Acquisition

The data collection was conducted in August 2018, at the experimental paddy field of Hokkaido University, Sapporo, Japan. The orthoimage of the paddy field is shown in Figure 1. The white dashed line indicates the flight route. There are two rice species, Kitaake and Kokusyokuto-2. The cultivation densities are around 11.46 ~ 16.18 plants/m<sup>2</sup>. We used a commercial UAV to capture the rice field images. It is more efficient than a handheld data-collection system. A larger UAV can load heavier cameras and lenses for capturing higher-quality images. However, the downward wind caused by the rotors will strongly blow rice stems and degrades the image quality. Therefore, based on trade-offs between flight height, image quality, and ease-of-use, the Mavic Pro designed by the DJI Corporation was used in this study. It is a lightweight and compact UAV with an embedded 1/2.3-inch CMOS.

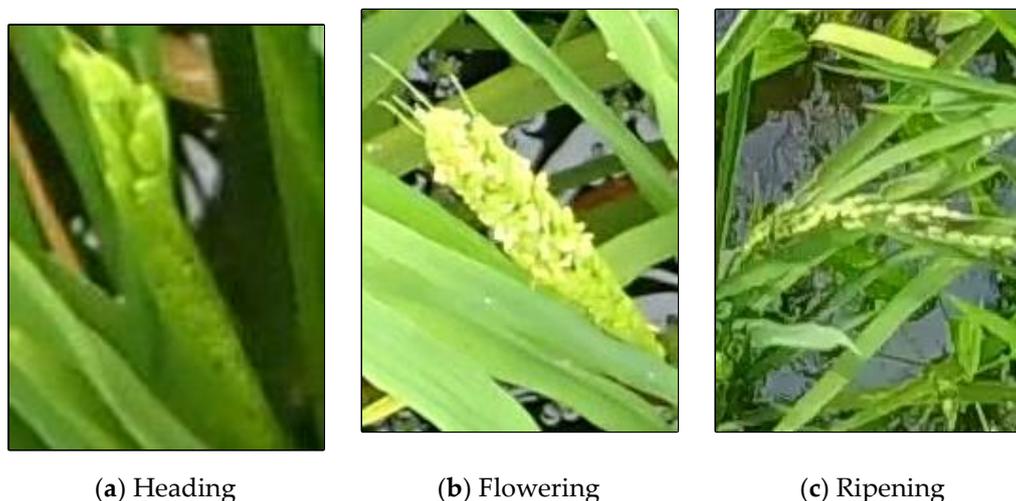


Figure 1. Experimental paddy field and flight route.

The configurations for data acquisition are listed in Table 1. We collected the data by using UAV from 1.2 m altitudes. The setup leads to a 0.04 cm/px ground sampling distance (GSD). The ratio of horizontal overlap is 43%, and the ratio of vertical overlap is 70%. Frames were extracted from acquired videos, while ensuring no duplication and avoiding blurring and overexposure. Each frame is in the resolution of  $4096 \times 2160$ . It covers the phenotype of rice panicles over the entire rice reproductive stage, of which, specifically, rice went through heading, flowering, and ripening. Some sample images are shown in Figure 2. The heading stage is the period that the panicle is fully visible. The flowering stage begins right after heading with the appearance of tiny white spikes on the panicles. The flowering stage lasts about 7 days. The ripening stage starts after flowering and ends before harvest. The texture, color, and shape of the panicle change a lot at this stage.

**Table 1.** Configurations of UAV and the camera.

FOV	Aperture	Gimbal	Resolution	Velocity (m/s)	Altitude (m)	GSD (cm/px)
78.8 (26 mm)	f/2.2	−90	$4096 \times 2160$	0.28	1.2	0.04



**Figure 2.** Sample images at different growth stages. In three subfigures (a–c), rice panicles are at the heading stage, flowering stage and ripening stage, respectively.

## 2.2. Data Processing

### 2.2.1. Basic Dataset Preparation

We manually annotate the panicle's boundary with a polygon, using the tool called Labelme [17]. The annotations of sample images at each stage are analyzed in Figure 3. Images taken at the early stage of rice growth are not informative, because few panicles are observable. As shown in the first panel, the average number of annotations per image taken on August 4 is about 40. The number of annotations in each image increases gradually as the rice matures, from August 4 to August 22. For the same reason, the number of average pixels per annotation accelerates from the flowering period (August 4) to the heading period (August 12). The number of pixels per annotation at the second panel is related to the size of panicles. With the growth of the rice, the panicles cannot be fully observed because of overlapping each other. Therefore, the average pixels per annotation drops from August 12 to August 22. According to the previous observation and analysis, the images taken on August 12 are considered to be representative and informative for panicle identification. The data-collection time (UTC +9) is listed in the third panel. In all, 400 images with 36,089 annotations are dedicatedly selected as a basic dataset. Fifty percent of the total images are taken after the heading period on August 12. We manually label no

more than 200 panicles in each image. Labeling all panicles in one image is a labor-intensive task. The next step was to refine the dataset based on the manual annotations.

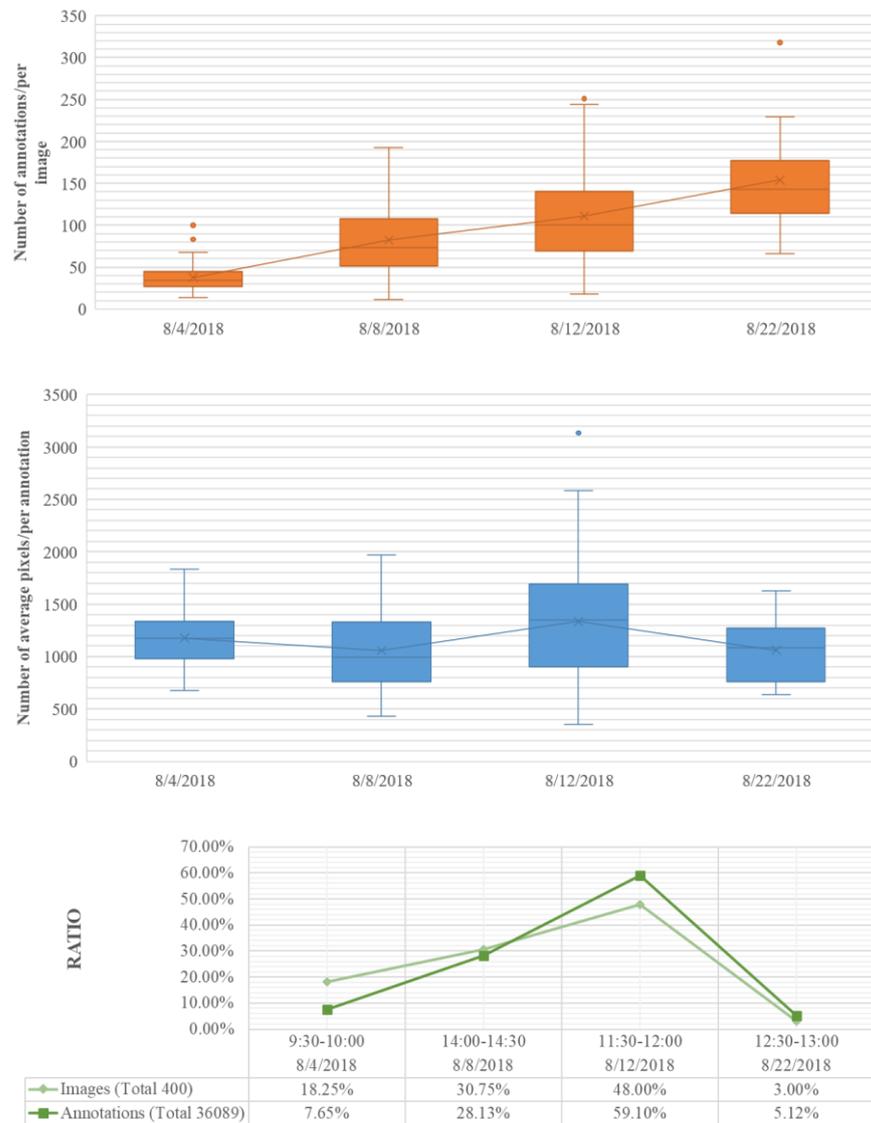
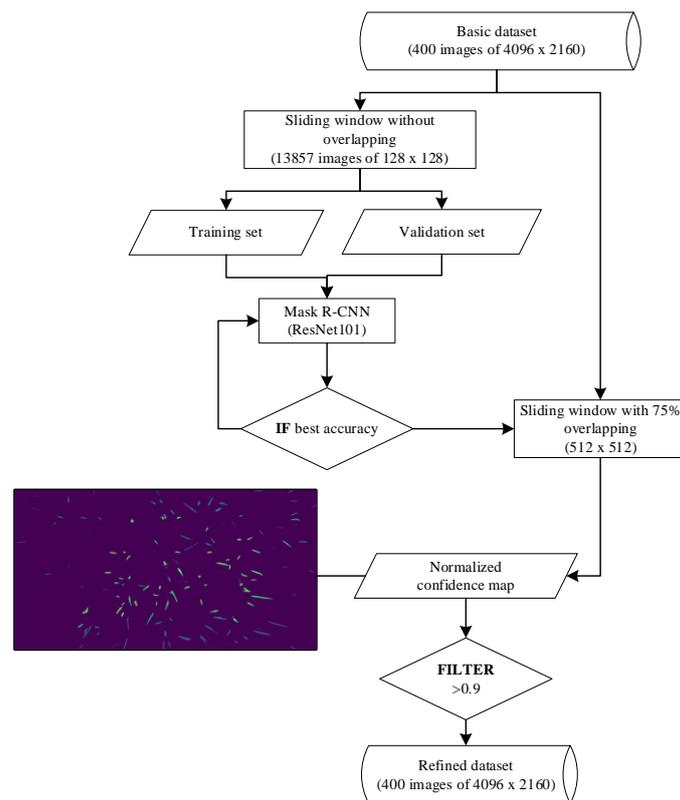


Figure 3. Basic dataset with manual annotations.

### 2.2.2. Dataset Refinement

The dataset refinement process has two steps, as depicted in Figure 4. The first step is to train an instance segmentation model, Mask R-CNN [16], using manual annotations. Processing a full 4K image occupies too much GPU memory, which is not an economical way to training a deep neural network model. Besides, the number of negative labels in an image is much greater than the positive ones. To balance the negative and positive samples and limit the GPU memory occupation, the image with  $4096 \times 2160$  resolution was split into small tiles with  $128 \times 128$  resolution. Tile-to-tile is non-overlapping. We further denoised the dataset by eliminating the tiles with positive labels less than 32 pixels. The dataset contains 13,857 images (with resolution  $128 \times 128$ ) and 38,799 pixel-level annotations. It is divided into a training set (80%) and a validation set (20%) to train the Mask R-CNN model. The backbone of the model is ResNet101 [18].



**Figure 4.** Flowchart of the proposed panicle labeling method.

The next step is to refine the basic dataset by implementing the trained model on it. We scanned the full 4K image, using a  $512 \times 512$  sized sliding window with 75% overlapping in both vertical and horizontal directions. The model gives the confidence of prediction at each pixel. After scanning the whole image, a confidence map is generated. By applying threshold and morphology filtering methods to the confidence map, we can get the refined dataset with manually labeled annotations and newly added annotations created by the model.

### 3. Data Evaluation and Discussions

#### 3.1. Improvement of the Basic Dataset

A sample image from the dataset after re-annotation is shown in Figure 5. It was taken on August 12. The DNN model accurately identified the panicles that are missed in the manual annotations, marked in red circles. As shown in the first panel of Figure 6, the total number of annotations in the refined dataset is 50,730. The improvement of the basic dataset is about 40.6%. The distribution and variation of annotations after refinement are entirely consistent with the results before re-annotation. The number of annotations per image kept increasing from August 4 to August 22. The DNN model can detect panicles all through the growth period. In addition, the annotation generated by the DNN model has proved to be more precise than the manual label; it can be observed from the annotation enlarged in Figure 5. The average pixel per annotation decreased after the refinement process, as shown in the third panel of Figure 6. This tendency also proved the improvement of accuracy for each annotation.

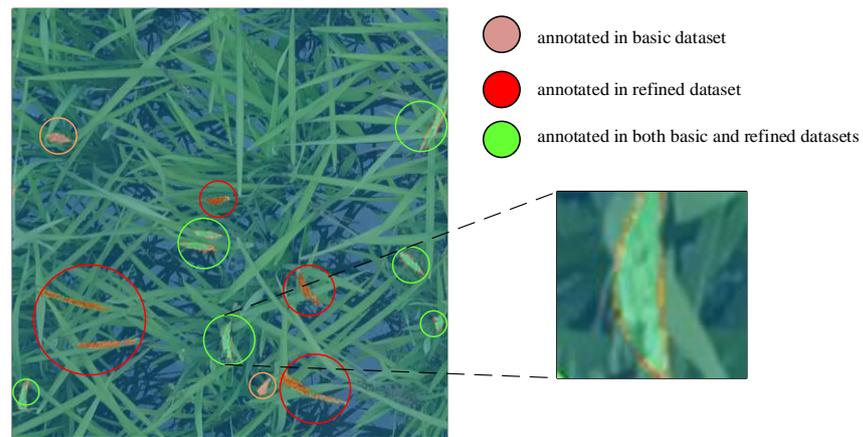


Figure 5. Sample images after refinement.

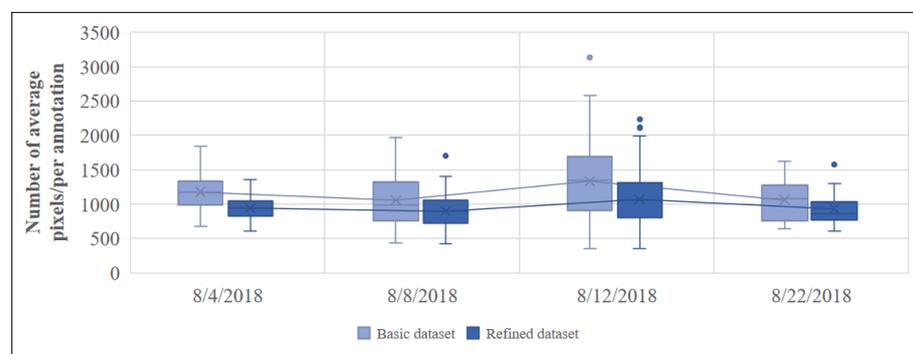
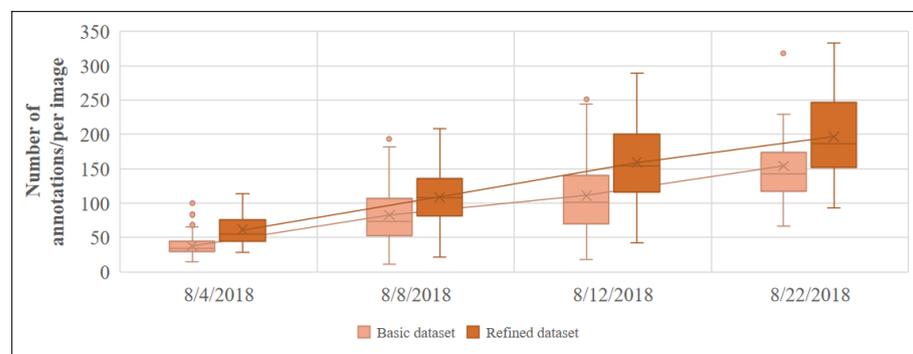
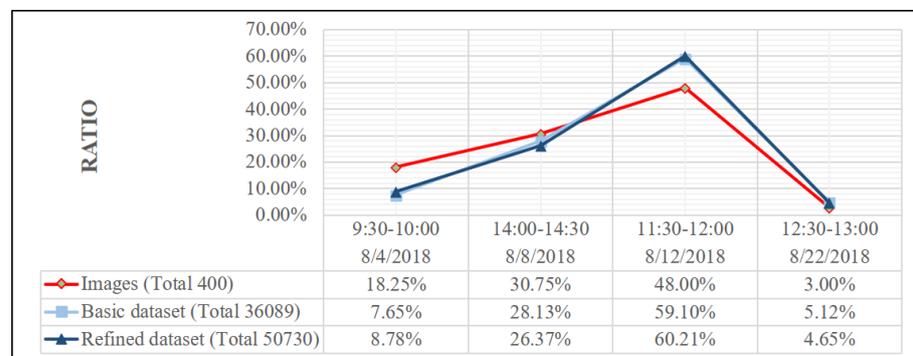


Figure 6. Annotation analysis after refinement.

### 3.2. Application in Detection and Segmentation Tasks

#### 3.2.1. Split into Sub-Datasets for Model Training

In machine learning research, the dataset should be randomly split into training, validation, and testing samples. As shown in Figure 7, the refined dataset was divided into three subsets, namely 70% for training, 10% for validation, and 20% for test. We further split the full 4K images into small tiles, using non-overlapping sliding windows, the size of which ranges from  $128 \times 128$  pixels to  $512 \times 512$  pixels. The tiles with positive labels less than 32 pixels were eliminated from the subsets. The volume of each subset is summarized in Table 2. The small tiles with different sizes are suitable for different computational hardware availability. We do not recommend reducing the sliding window size (less than  $128 \times 128$ ) when the GPU memory is limited. A smaller receptive field of view cannot assure the completeness of a panicle. As shown in the first panel of Figure 8, over half of Subset 1 contains only one annotation. Moreover, the second panel illustrates the proportion of average annotation pixels to the image size. Double peaks can be observed from Subset 1. It indicates the imbalance of the ratio of foreground and background pixels. About half of the images contain incomplete panicles because of the small receptive field. It is consistent with the blue lines in the first panel of Figure 8. There is a high possibility to get one panicle annotation in a receptive field with resolution  $128 \times 128$ . If we further reduce the receptive field, for instance, to  $64 \times 64$ , the ratio of incomplete annotations will increase. Enlarging the receptive field of view with a large sliding window can increase the number of panicles per image. The maximum sliding window size is set as  $512 \times 512$  to keep the balance between positive labels (panicle pixels) and negative labels (background pixels).

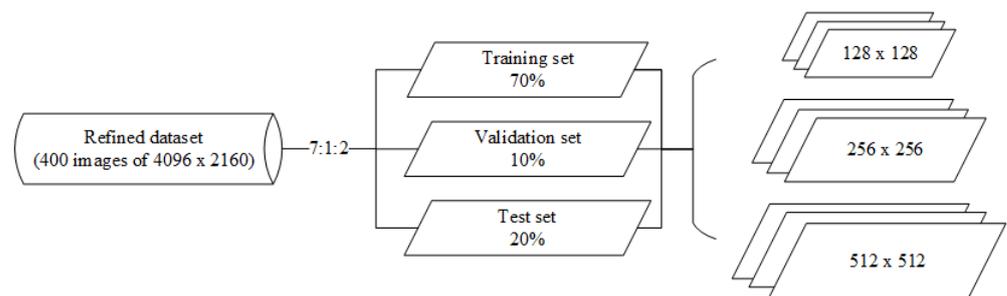
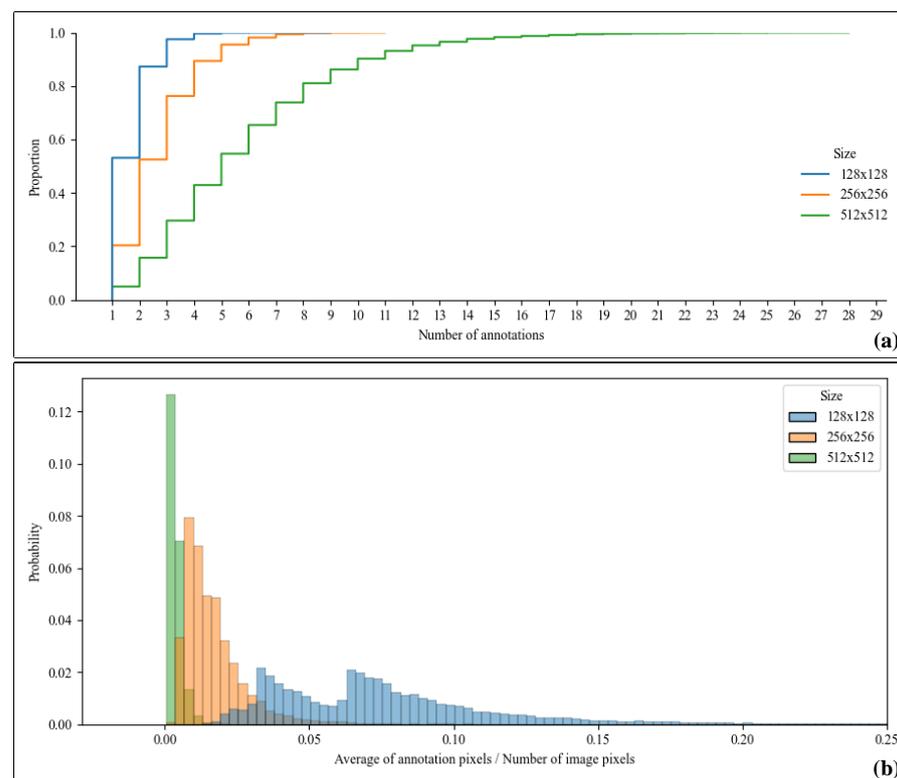


Figure 7. Data preparation for model training.

Table 2. Subsets for deep learning model training.

Order	Subset 1	Subset 2	Subset 3
Image size	$128 \times 128$	$256 \times 256$	$512 \times 512$
Training set	10,857	10,977	5865
Validation set	1022	1086	648
Test set	3944	3922	1086



**Figure 8.** Distribution of panicle number and average pixels per image.

### 3.2.2. Target Applications

Panicle-phenotyping research can be classified as a detection task or a segmentation task. The annotation in this dataset can be used for both applications. We provide results based on modern object detection and segmentation models to set a baseline identification accuracy for the dataset.

Applying the data for panicle detection tasks, we trained two object detection models, the EfficientDet-D7 [19] and the Mask R-CNN, with an Inception-ResNet v2 [20] as the backbone. EfficientDet is a state-of-the-art object detection family which improves both efficiency and accuracy of the model based on EfficientNet [21]. Mask R-CNN evolves from Faster R-CNN [22] by adding a semantic segmentation head for pixel-level masking tasks. The backbone, which is the feature extraction part of Mask R-CNN, is Inception-ResNet v2. The Inception-ResNet v2 backbone consists of a deeper and wider architecture and superiority on detection accuracy, but meanwhile, it is far heavier than any other alternative backbones.

In addition, we trained two segmentation models, U-Net [23] with ResNet 50 as backbone and DeepLab v3+ [24] with Xception as backbone [25]. U-Net is a lightweight model, which is widely implemented in medical-image segmentation [26] and aerial-image segmentation [27]. DeepLab v3+ is an extension based on DeepLab v3 [28], and it currently achieves the best accuracy within DeepLab family.

All the results of panicle detection and segmentation are listed in Supplementary Materials Tables S1 and S2, respectively. In the baseline results, the detection or segmentation effects of a DNN model are evaluated by using four metrics. The standard metrics include Precision, Recall, and F1-score for validating all machine learning tasks. In addition, the mean Average Precision (mAP) and mean Intersection over Union (mIoU) are used for evaluating the performance of the detection model and segmentation model, respectively [29,30]. The difference in the receptive field is reflected in the resolution of the training and test data listed in the second and third columns of Supplementary Materials Tables S1 and S2, respectively. In this research, the results of panicle detection and segmentation derived from EfficientDet-D7 and DeepLab V3+ are better than those derived from Mask R-CNN and U-Net. Adjusting the receptive field is proved to be a good option for

model selection during the training process. The results in Supplementary Materials Table S2 show that the model trained with Subset 2 (with image size  $256 \times 256$ ) performs best, while the model trained with Subset 1 (with image size  $128 \times 128$ ) performs worst. The main reason may be the imbalance of foreground (panicle) and background data in Subset 1, which leads to the instability of the model performance. For the same reason, the low panicle pixels ratio in Subset 3 (with image size  $512 \times 512$ ) also affects the performance of the segmentation model. Therefore, in the application of rice panicle detection, we should pay attention to the matching of the receptive field and spatial resolution of imagery while choosing different models.

### 3.3. Data Expansion

The dataset provided in this research is limited to paddy-rice species. In addition, all the images were captured in the nadir direction. Therefore, the dataset lacks the side view of plants. Users can apply this dataset or the subsets to complement their data. We also invite researchers to complement the dataset with their subsets. The expanded dataset can be used for training panicle-identification models and evaluating their robustness, concerning changes in paddy variety and geographical setup.

Data expansion should follow the Findable, Accessible, Interoperable, and Reusable (FAIR) principles to keep the consistency [31]. For image acquisition, the optimal reference height at which to capture the panicle area, as well as the GSD, is yet to be defined. We observed that the receptive field of view affects the performance of DNN models. Our experience suggests that a near nadir viewing direction and a sub-millimetre resolution is required for efficient panicle identification. We also recommend acquiring panicle images at the stages, between flowering and mature, when the panicle has fully emerged but not matured. These settings can effectively limit the overlap among panicles. Besides the data quality assurance, a minimum set of metadata should be associated with newly added images or datasets. We recommend attaching all the metadata related to the camera and acquisition platform used.

## 4. Conclusions

Localizing and identifying panicles from RGB images are useful for rice breeders and farm managers. This research presents a cost-effective process of panicle-image acquisition and refinement. Our contributions are summarized as follows:

- We presented a solution for paddy-rice-panicle segmentation in RGB images, using deep learning methods. It provides quick and precise panicle phenotyping.
- We provided a pixel-level labeled rice panicle dataset containing 400 images with 50730 pixel-level annotations. The dataset is publicly available at <http://doi.org/10.5281/zenodo.4430186> for developing rice-panicle detection models. The main objective of the dataset is to contribute to solving the problem of rice panicle identification from RGB images. The dataset can be used for both panicle detection and segmentation tasks.
- We proposed guidelines for image acquisition and data expansion. A proper image acquisition method can effectively limit the panicle overlapping.

Panicle identification is elementary for other traits, such as yield estimation, head population density estimation, crop health, and production monitoring. In future work, we hope to develop robust panicle identification methods and expand the variety and volume of the paddy-rice imagery dataset.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/agronomy11081542/s1>. Table S1: Baseline results of panicle detection, measured on the refined dataset. Table S2: Baseline results of panicle segmentation, measured on the refined dataset.

**Author Contributions:** Conceptualization, H.W., S.L., and Y.R.; methodology, H.W. and S.L.; software, H.W. and S.L.; validation, S.L.; formal analysis, H.W. and S.L.; investigation, H.W. and S.L.; writing—original draft preparation, H.W., S.L., and Y.R.; writing—review and editing, H.W., S.L.,

and Y.R.; visualization, H.W., S.L., and Y.R.; supervision, H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Key R&D Program of China (Grant No. [2019YFB1312305]).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset will be publicly available at <http://doi.org/10.5281/zenodo.4430186>.

**Acknowledgments:** The research has been partially funded by the National Key R&D Program of China [2019YFB1312305] and the Outstanding Young Talents Projects of Beijing Academy of Agriculture and Forestry Sciences.

**Conflicts of Interest:** The authors declare that there is no conflict of interest regarding the publication of this article.

## References

- Valipour, M.; Bateni, S.M.; Jun, C. Global Surface Temperature: A New Insight. *Climate* **2021**, *9*, 81. [CrossRef]
- Malambo, L.; Popescu, S.; Ku, N.-W.; Rooney, W.; Zhou, T.; Moore, S. A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. *Remote Sens.* **2019**, *11*, 2939. [CrossRef]
- Chandra, A.L.; Desai, S.V.; Balasubramanian, V.N.; Ninomiya, S.; Guo, W. Active learning with point supervision for cost-effective panicle detection in cereal crops. *Plant Methods* **2020**, *16*, 34. [CrossRef] [PubMed]
- Ikeda, M.; Hirose, Y.; Takashi, T.; Shibata, Y.; Yamamura, T.; Komura, T.; Doi, K.; Ashikari, M.; Matsuoka, M.; Kitano, H. Analysis of rice panicle traits and detection of QTLs using an image analyzing method. *Breed. Sci.* **2010**, *60*, 55–64. [CrossRef]
- Xiong, X.; Duan, L.; Liu, L.; Tu, H.; Yang, P.; Wu, D.; Chen, G.; Xiong, L.; Yang, W.; Liu, Q. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* **2017**, *13*, 104. [CrossRef]
- Xiong, H.; Cao, Z.; Lu, H.; Madec, S.; Liu, L.; Shen, C. TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* **2019**, *15*, 150. [CrossRef]
- Madec, S.; Jin, X.; Lu, H.; De Solan, B.; Liu, S.; Duyme, F.; Heritier, E.; Baret, F. Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.* **2019**, *264*, 225–234. [CrossRef]
- Singh, A.K.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A. Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends Plant Sci.* **2018**, *23*, 883–898. [CrossRef]
- Sadeghi-Tehran, P.; Virlet, N.; Ampe, E.M.; Reyns, P.; Hawkesford, M.J. DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Front. Plant Sci.* **2019**, *10*, 1176. [CrossRef]
- Fernandez-Gallego, J.A.; Kefauver, S.C.; Gutiérrez, N.A.; Nieto-Taladriz, M.T.; Araus, J.L. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. *Plant Methods* **2018**, *14*, 22. [CrossRef]
- Guo, W.; Fukatsu, T.; Ninomiya, S. Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images. *Plant Methods* **2015**, *11*, 7. [CrossRef]
- Khoreva, A.; Benenson, R.; Hosang, J.; Hein, M.; Schiele, B. Simple does It: Weakly supervised instance and semantic segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1665–1674.
- Ghosal, S.; Zheng, B.; Chapman, S.C.; Potgieter, A.B.; Jordan, D.R.; Wang, X.; Singh, A.K.; Singh, A.; Hirafuji, M.; Ninomiya, S.; et al. A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting. *Plant Phenomics* **2019**, *2019*, 1525874. [CrossRef] [PubMed]
- Holman, F.H.; Riche, A.B.; Michalski, A.; Castle, M.; Wooster, M.J.; Hawkesford, M.J. High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sens.* **2016**, *8*, 1031. [CrossRef]
- Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A Review on UAV-Based Applications for Precision Agriculture. *Information* **2019**, *10*, 349. [CrossRef]
- He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
- Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
- Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–19 June 2020; pp. 10778–10787.

20. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
21. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 10–15 June 2019; pp. 10691–10700.
22. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
23. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
24. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision—ECCV 2018, Proceedings of the 15th European Conference, Munich, Germany, 8–14 September 2018*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 801–818.
25. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
26. Siddique, N.; Sidike, P.; Elkin, C.; Devabhaktuni, V. U-Net and its variants for medical image segmentation: Theory and applications. *arXiv* **2020**, arXiv:2011.01118.
27. Ivanovsky, L.; Khryashchev, V.; Pavlov, V.; Ostrovskaya, A. Building Detection on Aerial Images Using U-NET Neural Networks. In Proceedings of the 2019 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 8–12 April 2019; pp. 116–122.
28. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
29. Li, J.; Zhou, H.; Wang, Z.; Jia, Q. Multi-scale detection of stored-grain insects for intelligent monitoring. *Comput. Electron. Agric.* **2020**, *168*, 105114. [[CrossRef](#)]
30. Wang, H.; Ren, Y.; Meng, Z. A Farm Management Information System for Semi-Supervised Path Planning and Autonomous Vehicle Control. *Sustainability* **2021**, *13*, 7497. [[CrossRef](#)]
31. Pommier, C.; Michotey, C.; Cornut, G.; Roumet, P.; Duchêne, E.; Flores, R.; Lebreton, A.; Alaux, M.; Durand, S.; Kimmel, E.; et al. Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS. *Plant Phenomics* **2019**, *2019*, 1–15. [[CrossRef](#)] [[PubMed](#)]