Semi-supervised learning for accurate weed mapping of UAV imagery

Huasheng Huang[1], Yubin Lan[2,3*], Jizhong Deng[3,4], Yali Zhang[3,4], Aqing Yang[1*]

(1. *College of Computer Sciences, Guangdong Polytechnic University, Guangzhou 510665, China*;
2. *College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China*;
3. *National Center for International Collaboration Research on Precision Agricultural Aviation Pesticide Spraying Technology, Guangzhou 510642, China*;
4. *College of Engineering, South China Agricultural University, Guangzhou 510642, China*)

Abstract: Weed mapping is essential for Site Specific Weed Management (SSWM) applications. Semantic segmentation is the mainstream algorithm to perform the weed mapping at pixel level, which is proven to be superiors than the traditional Object Based Image Analysis (OBIA) approaches. However, the semantic segmentation requires large amount of annotated data for parameter updating, and the development of such methods are currently limited by the shortage of annotated data in the SSWM community. To address this problem, this paper proposed a semi-supervised learning method for accurate weed mapping of UAV imagery. Firstly, we applied limited training data with annotation to train the classifiers of the OBIA method. Secondly, we used the trained OBIA model to produce the pseudo labels for other training data without annotations. Finally, we applied both the manual annotations and the generated pseudo labels to train the semantic segmentation models. The proposed method is compared with the mainstream semantic segmentation at different training sizes. Experimental results showed that the proposed semi-supervised learning method significantly improved the prediction precision at different training sizes. Furthermore, the proposed algorithm well addressed the overfitting problem of supervised learning at extremely small training set. The proposed semi-supervised method is expected to reduce the manual annotation efforts and enhance the weed mapping researches in the context of SSWM applications.

Keywords: weed mapping; UAV imagery; OBIA; semantic segmentation; semi-supervised learning
DOI: 10.33440/j.ijpaa.20220501.190

Citation: Huang H H, Lan Y B, Deng J Z, Zhang Y L, Yang A Q. Semi-supervised learning for accurate weed mapping of UAV imagery. Int J Precis Agric Aviat, 2022; 5(1): 29–34.

1 Introduction

Weed mapping is an important step in the context of Specific Site Weed Management (SSWM). With the instruction of weed distribution map, the spraying machines can accurately target on the weeds, which is of great significance to reduce the use of herbicides while maintaining the chemical effects. In recent years, it was widely accepted that the UAV remote sensing offers a feasible platform on efficient data acquisition for weed mapping. Compared with the traditional satellite remote sensing and piloted aircraft remote sensing, UAV can efficiently obtain large scale imagery in high spatial resolution, which builds a good foundation for the following data interpretation[1,2]. Stroppiana et al.[3] proposed an automatic procedure for classification of UAV imagery to map weed presence in rice paddies at early stages of the growing cycle. Experimental results showed that best results are provided by a set of spectral indices, where the overall accuracy was up to 95%. Later, the weed map was aggregated to a grid layer of 5×5 m to simulate variable rate management. López-Granados et al.[4] generate georeferenced weed seedling infestation maps in two sunflower fields by analyzing overlapping aerial images of the visible and near-infrared spectrum (using visible or multi-spectral cameras) collected by an unmanned aerial vehicle (UAV) flying at 30 and 60 m altitudes. Regarding weed discrimination, high accuracies were observed using the multi-spectral camera at any flight altitude, with the highest (approximately 100%) precision recorded for the 15% weed threshold. Gašparović et al.[5] applied the UAV for data collection using a low-cost RGB camera. Classification algorithms were based on the random forest machine learning algorithm for weed and bare soil extraction, following an unsupervised classification with the K-means algorithm for further estimation of weeds and bare soil presence in non-weed and non-soil areas. Experimental results showed that proposed methods yields an overall accuracy of 89.0% for subset A and 87.1% for subset B.

After data acquisition, data interpretation plays an important role in obtaining the weed cover information. Object based Image Analysis (OBIA) and semantic segmentation are two main branches to perform the dense classification in weed mapping[6]. OBIA method generally segments the imagery into several homogeneous objects and then performs the classification for each object[7-9]. Compared with the per-pixel classification mode, OBIA performs the classification in the scale of objects, which may reduce the noises and increase the accuracy. De Castro et al.[10] developed a robust and innovative automatic OBIA algorithm on Unmanned Aerial Vehicle (UAV) images to design early post-emergence prescription maps. Specifically, the OBIA-based plant heights were accurately estimated and used as a feature in the

Received date: 2022-11-02 **Accepted date:** 2022-12-12
Biographies: Huasheng Huang, associate professor, research interests: UAV remote sensing in smart agriculture. Email: huanghsheng@gpnu.edu.cn; **Yubin Lan**, PhD, professor, Director, research interests: precision agricultural aviation application, Email: ylan@scau.edu.cn; **Jizhong Deng**, professor, research interests: Unmanned aerial remote sensing, Email: jz-deng@scau.edu.cn; **Yali Zhang**, associate professor, research interests: smart agriculture and sensors, Email: ylzhang@scau.edu.cn; **Aqing Yang**, lecturer, research interests: smart agriculture and image processing, Email: 512289446@qq.com.
*** Corresponding author: Aqing Yang**, lecturer, research interests: smart agriculture and image processing, Mailing Address: Guangdong Polytechnic Normal University. Email: 512289446@qq.com.

automatic sample selection by the RF classifier. Gao et al.[11] developed a semi-automatic Object-Based Image Analysis (OBIA) procedure combined with feature selection techniques to classify soil, weeds and maize. The developed approach was evaluated by 5-fold cross validation, and it obtained an overall accuracy of 0.945, and Kappa value of 0.912. Compared with OBIA, the semantic segmentation performs the weed mapping in an end to end mode, which only involves one processing stage[12,13]. Zou et al.[14] proposed a simplified U-net architecture for weed mapping. The proposed network was trained by a two-stage training method composed of pre-training and fine-tuning. After training, the intersection over union (IoU) of this method was 92.91% and the average segmentation time of a single image (ST) was 51.71 ms. Sa et al.[15] developed a novel crop/weed segmentation and mapping framework that processes multispectral images obtained from an unmanned aerial vehicle (UAV) using a deep neural network (DNN). Furthermore, the authors release a large sugar beet/weed aerial dataset with expertly guided annotations for further research in the fields of remote sensing, precision agriculture, and agricultural robotics. Generally speaking, the semantic segmentation algorithms significantly outperform the OBIA method in terms of accuracy and efficiency, which have more potential in the SSWM applications.

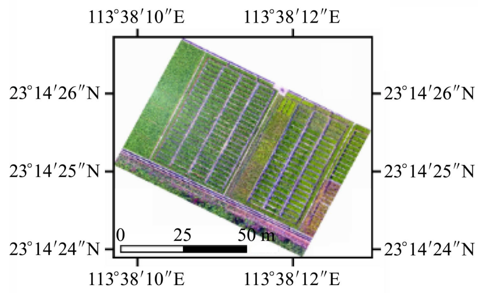
Despite its superiority, the development of the deep learning is limited by its demand for large amount of annotated data. Especially in the SSWM context, the natural rice fields are complicated and it is hard to annotate the UAV imagery at pixel level. Current solutions mainly contain the semi-supervised and unsupervised learning approaches. He et al.[16] presented Momentum Contrast (MoCo) for unsupervised visual representation learning. From a perspective on contrastive learning as dictionary look-up, the authors build a dynamic dictionary with a queue and a moving-averaged encoder. MoCo can outperform its supervised pre-training counterpart in 7 detection/segmentation tasks on PASCAL VOC, COCO, and other datasets, sometimes surpassing it by large margins. However, the concept of contrast learning is based on the public dataset with large amount of categories and unlabeled data, which is not appropriate in the context of SSWM researches. Ouali et al.[17] proposed cross-consistency training, where an invariance of the predictions is enforced over different perturbations applied to the outputs of the encoder. Li et al.[18] proposed a generative adversarial network that captures the joint image-label distribution and is trained efficiently using a large set of unlabeled images supplemented with only few labeled ones. However, this semi-supervised method is designed for out-of-domain image generalization, which is also not suitable in our application scenarios.

To address the shortage of annotated data in the SSWM community, we proposed a semi-supervised learning algorithms for accurate weed mapping of UAV imagery. The proposed method is based on the current mainstream supervised weed mapping algorithms, and is expected to reduce the manual annotation and promote the weed mapping researches.

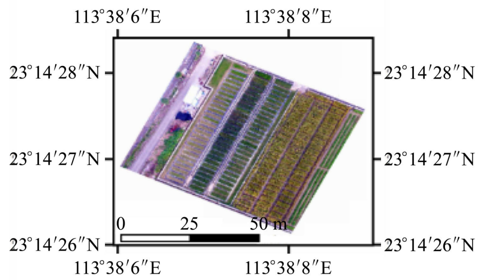
2 Materials and Methods

2.1 Materials

Experiments were conducted in two rice fields located in Guangdong Provinces, Southern China, where the rice was in its seedling and tillering stages. The general location of two studied fields are shown in Figure 1, where two rice fields are named as F1 and F2. The weed control managements are usually taken in the early growing stages of rice, thus the weed monitoring in theses stages are meaningful for real applications.



a. Rice field F1



b. Rice field F2.

Figure 1 The general location of the studied sties

In our experiments, the employed UAV is Phantom 4 (DJI company, Shenzhen, China). During data collection, the UAV was kept 10 meters from the ground, where the spatial resolution is 0.10 cm. The UAV imagery were taken with constant overlapping rate, where the forward lap and side lap were set to 70% and 60%. The image sequences in each field were ortho-mosaicked and split into 1000×1000 sub images to avoid the exhaustion of computational resources during image processing. The image sequences in one UAV campaign consist one dataset, as shown in Table 2.

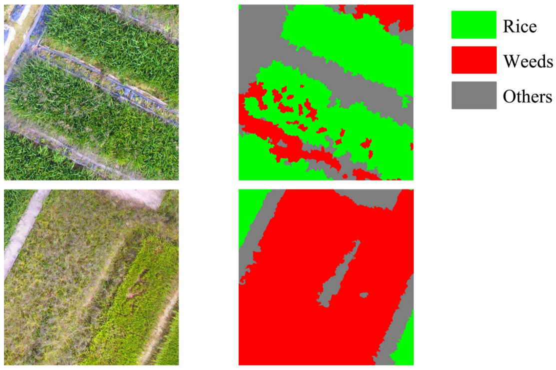
Table 2 Specification of the dataset

Name	Number of imagery	Growth Stage	Description
D1	182	tillering stage	Captured on 02 October 2017 in field F1
D2	182	tillering stage	Captured on 10 October 2017 in field F1
D3	120	seedling stage	Captured on 10 November 2017 in field F2
D4	120	seedling stage	Captured on 18 November 2017 in field F2

The UAV images were manually labeled at pixel level under the instruction of agronomic experts. The demonstration of UAV images and its corresponding labels are shown in Figure 2. The labels are used as the ground truth for training and validation in this paper. The dataset of D1 and D3 are used as the training set, and the dataset of D2 and D4 are used as the testing set. The reason for this dataset split is that the training set and testing set are collected from different dates, which may evaluate the generalization capability of the analysis models.

2.2 Methods

The general methodology of this paper is shown in Figure 3. In the weed mapping domain, it is well accepted that the semantic segmentation is accurate and efficient with large amount of labeled data, and the OBIA is relatively slow which is not appropriate for real applications. However, the OBIA method required less labeled data. In this case, we proposed the connection bridge to combine the both advantages of OBIA and deep learning. Firstly, we trained the classifier of the OBIA with limited training samples. Secondly, we applied the trained OBIA models to generate the pseudo labels for the images without labels. Finally, we applied both the manual annotations and the pseudo labels to train the semantic segmentation models.



a. UAV images b. Label images

Figure 2 The demonstration of two samples

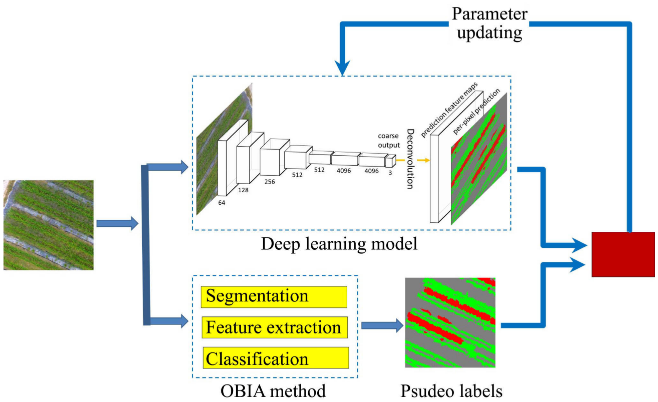


Figure 3 The general methodology of this paper

2.2.1 OBIA methods

The OBIA methods mainly consist of three stages: segmentation, feature extraction, and classification. Among them, the unsupervised segmentation is the key step which may influence the accuracy of the following steps. In this work, we applied the multiresolution segmentation algorithm to generate homogeneous objects. During the segmentation process in multiresolution segmentation, the clustering for the pixels is iteratively optimized until no significant improvement is found on the heterogeneity.

To the manual experience of weed identification, the difference of the rice and weeds involve color and texture characteristics. Therefore, we extracted the color and texture features for the discrimination of different categories. Specifically, we computed the mean value of different channels to represent the color feature, and used the local binary pattern (LBP) to express the texture distribution. The LBP descriptor use each pixel as center, and compares its values with its 3×3 neighborhood. The comparison results are stored in binary number and transformed in decimal form and later used to compute the histograms. Specifically, the mean value and the LBP feature features are selected based on our previous work on the feature representation for weed mapping[6]. For the concatenated feature vector, some traditional methods were applied for classification, such as BP network, SVM, and random forests. In this research, the classifiers (i.e. BP network, SVM, and random forests) are trained with limited annotated samples and then used to generate pseudo labels for other samples.

2.2.2 Semantic segmentation

In recent years, the deep learning researches developed semantic segmentation for pixel level classification tasks. Generally, the semantic segmentation model extracted the deep representation of the input image via convolutional and down sampling operations. Different from the classification problems, semantic segmentation applied deconvolutional methods to restore the full spatial resolution in the end to end mode. In this fashion, the spatial details of the input images are retained, which brings signification performance boosts in both accuracy and efficiency.

It was noticeable that the down sampling results in the loss of spatial details, and the deconvolutional operation cannot well address this problem. In this case, Long et al.[12] proposed to combine the deep layers and the shallow layers of the network to achieve better precision. Similar with the multiscale algorithms in object detection[19-21], the skip architecture builds a baseline in semantic segmentation to increase the awareness of the objects in different scales.

3 Results and discussions

In this work, all the experiments were conducted on a computer with the i7 6700 CPU and RTX 2080 TI GPU. Similar with most semantic segmentation researches, we applied the pixel accuracy, mean accuracy, mean IU and f.w. IU as the evaluation metrics. Among them, the mean IU was generally accepted as the main metrics of semantic segmentation.

3.1 Implementation details

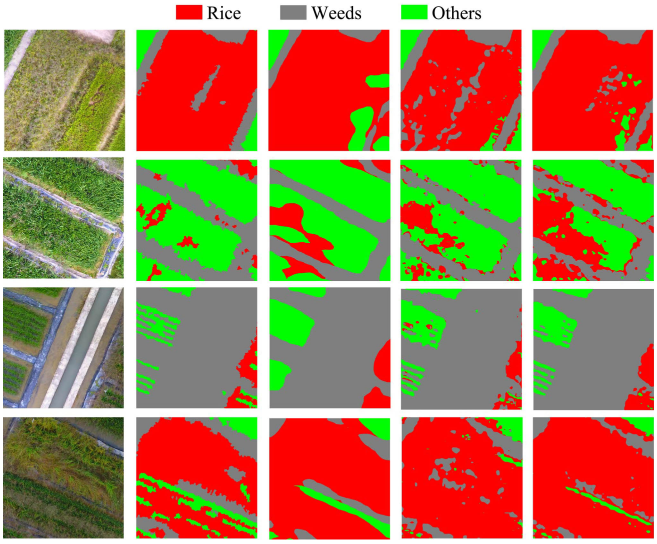
For the semantic segmentation, the VGG-19 network was used as the backbone architecture. The fully connected layers of the backbone network were removed and replaced with convolutional layers. After the deep representation extraction, the deconvolutional layers were appended after the network to restore the full spatial resolution. During training, we used stochastic gradient descent (SGD) for optimization. We used the fixed learning rate of 10^{-5}. The momentum was set to 0.99, and the weight decay was not employed. The ImageNet pretrained weights were adapted and transferred to our training set, and the training was repeated until the model converged on our training set. For the OBIA method, the scale, shape parameter, compactness and smoothness of the multiresolution segmentation were set to 100, 0.1, 0.5 and 0.5. The mean values of three color channels were used as the color feature, and the LBP descriptor was applied to represent the texture features. The BP neural network was applied for classification of each segment. We established two hidden layers for the BP network, and the number of neurons of each layer were set to 10 and 5, respectively.

3.2 Experiments on skip architecture

Similar with most semantic segmentation researches, we applied the skip architecture to address the spatial loss caused by the down sampling operation. The original framework of semantic segmentation is 32 stride upsampling, which is denoted as FCN-32s. We add a 1×1 convolution layer on top of pool4 and fused that with the final predictions of FCN-32s, which is denoted as FCN-16s. We continue this fashion and fusion the representation form pool3 layer, which is denoted as FCN-8s. It can be seen from Table 2 that adding skip architecture brings no noticeable boosts in accuracy. Though related literatures revealed that skip architecture brings slight improvement on performance, our research scenario demonstrate similar accuracy with different skip strategies. One possible reason is that the spatial resolution of our UAV imagery was not high as the public dataset like ImageNet or coco, thus the shallow layers like pool3 and pool4 cannot capture the details. Figure 4 gives some prediction results of different skip architecture. It can be seen from Table 7 that the prediction results from FCN-32s, FCN-16s, and FCN-8s are all close to the ground truth, demonstrating no significant difference.

Table 2 Experiments on skip architecture

	pixel acc.	mean acc.	mean IU	f.w. IU
FCN-32s	86.3%	87.3%	74.9%	76.2%
FCN-16s	86.1%	84.9%	73.2%	76.1%
FCN-8s	86.2%	87.1%	73.6%	76.7%



a. Input images b. Label images c. Outputs by FCN-32s d. Outputs by FCN-16s e. Outputs by FCN-8s

Figure 4 The experiments on skip architecture

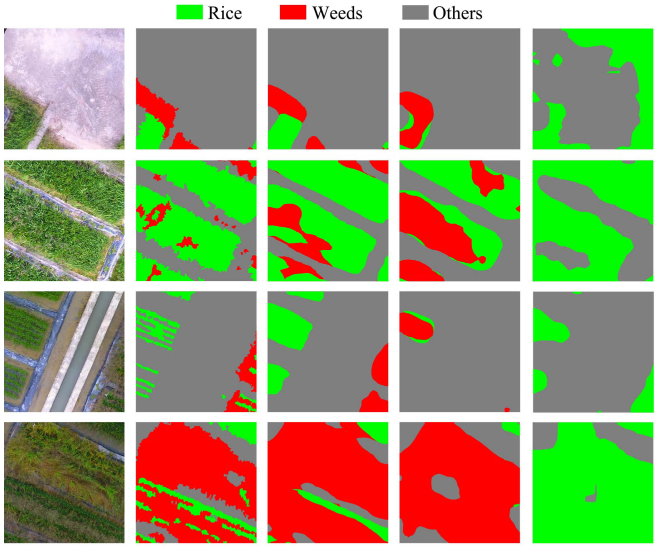
In our research, we did not consider the inference speed, since related literatures have proven that different skip architectures shared similar efficiency[1]. Therefore, we chose the FCN-32s as the backbone network for the following experiment, since it achieved close performance with simplified architecture.

3.2 Experiments on different training sizes

In this section, we trained the FCN-32s with different training sizes, where the employed training sizes were set to 302, 160, 80, 40, 20 and 10. The total training samples of this study is 302, thus the training size of 302 represent full training size. After training with different training sizes, the performance of the model was reported on all the validation set. Table 3 gives the accuracy of the model with different training set. It can be seen from Table 3 that the performance of the semantic segmentation model decrease significantly when the training size becomes smaller. The experimental results confirmed that the deep learning model requires large amount of training data for parameter optimization, and the performance decreases with small training set. With 10 training samples, the mean IU of FCN-32s is only 36.6%, which is only half of the full training size. Figure 5c shows that the FCN-32s obtained reasonable accuracy on the testing set with full training size. The rice and weeds can be correctly distinguished, and the predicted contours are close to the labels. When the training size is reduced to 80 samples, the classification error increase significantly, and the predicted contours are different from the ground truth, as shown in Figure 5d. We continue this fashion, and the training size was reduced to 10 training samples. It can be seen from Figure 5e that the model misclassified all the weeds into rice. This result demonstrated that too much training on the limited training samples caused the typical overfitting problem, and the optimizer fall into the local optimization point since the category of weeds are in small portion among all samples.

Table 3 Experiments of supervised learning on different training sizes

training sizes	pixel acc.	mean acc.	mean IU	f.w. IU
302	86.3%	87.3%	74.9%	76.2%
160	84.7%	84.7%	72.4%	73.6%
80	75.6%	74.4%	57.2%	60.1%
40	75.6%	75.9%	58.4%	61.4%
20	68.1%	56.0%	40.3%	50.4%
10	63.6%	52.8%	36.6%	45.8%

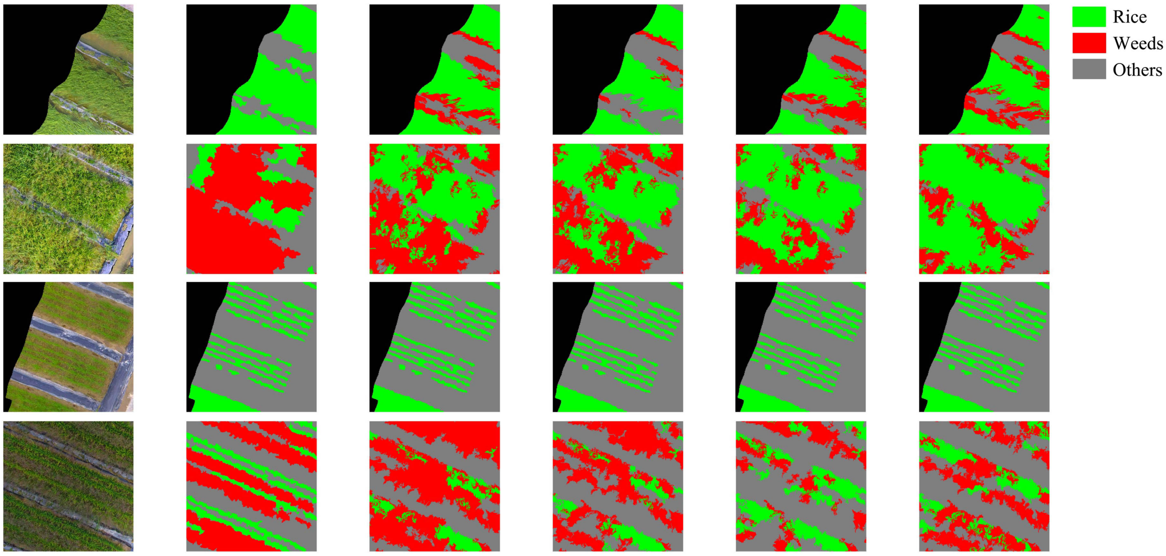


a. Input images b. Label images c. Outputs with full training size d. Outputs with 80 training samples e. Outputs with 10 training samples

Figure 5 The experiments of supervised learning on different training set

3.3 Experiments on semi-supervised learning

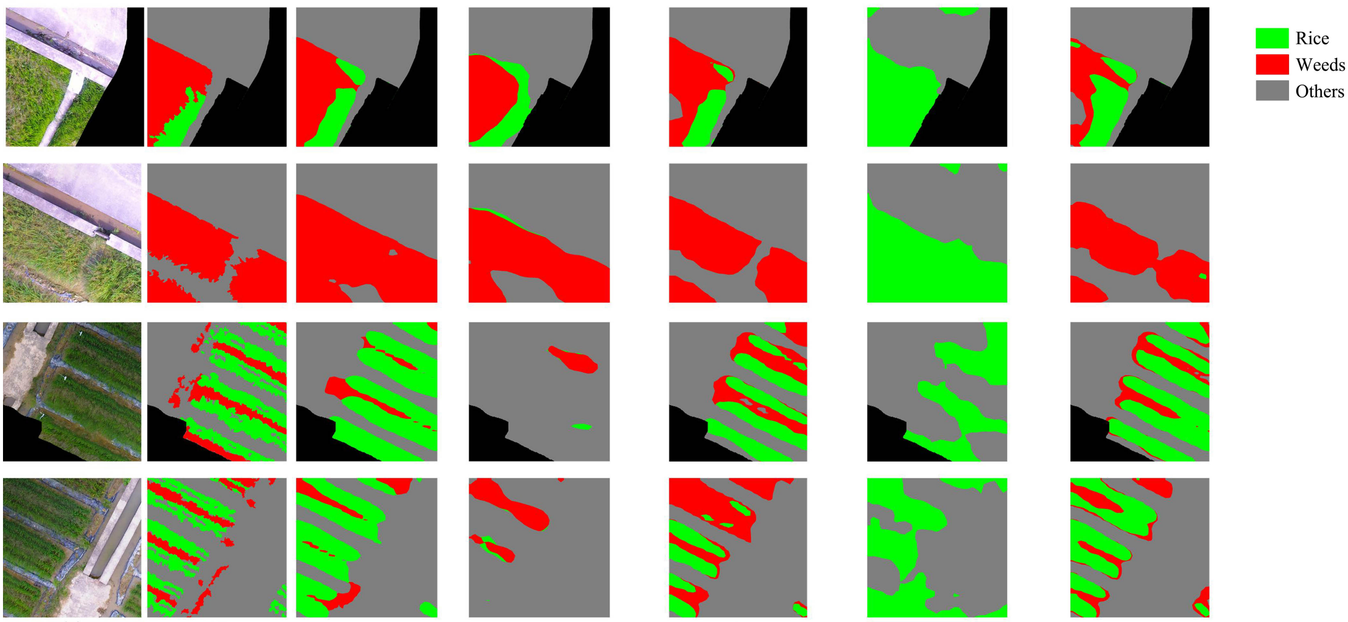
In this section, we chose part of the training samples as our training set with varying training sizes, similar with section 3.2. Besides that, we applied OBIA to generate pseudo labels for other training samples that were not selected. We applied the multiresolution segmentation to generate homogeneous segments, extracted the color and texture features to form the feature vector for each segment, and used the BP neural network for classification. Specifically, our previous work has conducted extensive work to explore the performance of different classifiers (i.e. BP neural network, SVM, and random forests), and it was proven that the BP neural network achieve the state of art over other classifiers under this application scene[6]. The classifier in the OBIA method was trained on the limited annotated samples of the training set, and later used to generate the pseudo labels for other samples that were not annotated. Figure 6 gives the illustration on the pseudo labels generated by OBIA method under different training sizes. It can be seen from Figure 6 that the accuracy of the pseudo labels decreases when the training sizes becomes smaller, which is similar with the deep learning models. However, the accuracy of OBIA is still acceptable when the training size is extremely small, reflecting that the OBIA method does not require that much labeled data compared with the deep learning algorithms. Therefore, the pseudo labels generated by the OBIA method is used to provide extra training for the deep learning model, which is expected to address the overfitting problem of the semantic segmentation models.



a. Input images b. Label images c. Pseudo labels with 80 training samples d. Pseudo labels with 40 training samples e. pseudo labels with 20 training samples f. Pseudo labels with 10 training samples

Figure 6 Pseudo labels generated by OBIA method under different training sizes

Table 4 gives the comparison of the traditional supervised learning and the proposed semi-supervised learning algorithms. It is obvious from Table 4 that the semi-supervised learning significantly outperforms the supervised in accuracy especially with extremely small training sizes. Though the performance of the semi-supervised models also decreases when the training sizes become smaller, the precision is still maintained at an acceptable margin. Figure 7 gives some prediction results by the supervised and semi-supervised learning under different training sizes. It can be seen from Figure 7 that when the training size is reduced to 80 samples, the accuracy dramatically drops with much misclassification, as showed in Figure 7d. However, the proposed semi-supervised method still achieves reasonable precision, as shown in Figure 7e. When the training size is further reduced to 10 samples, the supervised learning was stuck with minimum optimization and cause the overfitting problem. It can be seen from Figure 7f that FCN misclassifies all the weeds into rice, since the rice has the largest amount of samples and classifying all pixels into rice can lead to a smaller loss. This overfitting problem is also reflected in the quantitative results. It can be seen from Table 4 that most supervised models achieve a relatively high pixel accuracy with significantly low mean IU value, and this is caused by misclassifying other categories into the category that appears most. In contrast, our proposed semi-supervised learning algorithm well address this problem, and the weeds can still be identified despite that this category is in small portion. From the qualitative and quantitative results, it is clear that the proposed semi-supervised learning can well address the overfitting problem for the deep learning model caused by the shortage of training data, which may improve the generalization ability of the deep learning model with limited annotated data.



a. Input images b. Label images c. Outputs by FCN with full training size d. Outputs by supervised learning with 80 training samples e. Outputs by semi-supervised learning with 80 training samples f. Outputs by supervised learning with 10 training samples g. Outputs by semi-supervised learning with 10 training samples

Figure 7 Experiments on semi-supervised learning

Table 4 Experiments on semi-supervised learning

training strategies	pixel acc.	mean acc.	mean IU	f.w. IU
training with full training size	86.3%	87.3%	74.9%	76.2%
supervised training with 160 samples	84.7%	84.7%	72.4%	73.6%
semi-supervised training with 160 samples	87.0%	85.4%	75.6%	77.2%
supervised training with 80 samples	75.6%	74.4%	57.2%	60.1%
semi-supervised training with 80 samples	75.6%	72.2%	58.2%	63.5%
supervised training with 40 samples	75.6%	75.9%	58.4%	61.4%
semi-supervised training with 40 samples	77.9%	71.6%	59.1%	64.5%
supervised training with 20 samples	68.1%	56.0%	40.3%	50.4%
semi-supervised training with 20 samples	76.6%	67.7%	55.2%	61.7%
supervised training with 10 samples	63.6%	52.8%	36.6%	45.8%
semi-supervised training with 10 samples	75.3%	68.1%	55.4%	61.7%

4 Conclusions

This paper collected UAV imagery from the two rice fields in its early growing stages at four different dates and constructed a weed mapping dataset. After that, we proposed a semi-supervised learning algorithm to address the shortage of annotated data in weed mapping with deep learning methods. Firstly, we established an OBIA model, and trained the classifier using limited training samples. Secondly, we used the OBIA model to generate the pseudo labels for the samples without ground truth. Finally, we used both the manual annotation and the generated pseudo labels for the training of the deep learning models. Experimental results showed that the proposed semi-supervised method significantly outperformed the supervised algorithms, and can well address the overfitting problem of traditional supervised learning at extremely small training sizes. The proposed method has potential in weed mapping research with limited annotated data, which may reduce the efforts of manual labeling.

In the future work, we plan to collect more UAV imagery in the natural rice field to enhance the classifier both in training and validation. Also, more sophisticated methods for small training set, like few-shot learning or Transformer should be investigated to extend the methodology of this paper.

Acknowledgements

This study was supported by the Basic and Applied Basic Research Fund in Guangdong Province, China (Grant No. 2021A1515110756), and the Project of Educational Commission of Guangdong Province, China (Grant No. 2021KQNCX044).

[References]

[1] H Huang, J Deng, Y Lan, et al. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. PLOS ONE, 2018, 13(4): e196302. DOI: 10.1371/journal.pone.0196302.

[2] H Huang, Y Lan, J Deng, et al. A Semantic Labeling Approach for Accurate Weed Mapping of High Resolution UAV Imagery. Sensors, 2018, 18(7): 2113. DOI: 10.3390/s18072113.

[3] D Stroppiana, P Villa, G Sona, et al. Early season weed apping in rice crops using multi-spectral UAV data. International Journal of remote sensing, 2018, 39(15-16): 5432–452. DOI: 10.1080/ 01431161.2018.1441569.

[4] F López-Granados, J Torres-Sánchez, A Serrano-Pérez, et al. Early season weed mapping in sunflower using UAV technology: variability of herbicide treatment maps against weed thresholds. Precision Agriculture, 2016, 17(2): 183–199. DOI: 10.1007/s11119-015-9415-8.

[5] M Gašparović, M Zrinjski, Barković, et al. An automatic method for weed mapping in oat fields based on UAV imagery. Computers and Electronics in Agriculture, 2020, 173: 105385. DOI: 10.1016/ j.compag.2020.105385.

[6] H Huang, Y Lan, A Yang, et al. Deep learning versus Object-based Image Analysis (OBIA) in weed mapping of UAV imagery. International Journal of Remote Sensing, 2020, 41(9): 3446–3479.

[7] M D Hossain, D Chen. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 150: 115–134. DOI: 10.1016/j.isprsjprs.2019.02.009.

[8] L Janowski, K Tylmann, K Trzcinska, et al. Exploration of glacial landforms by object-based image analysis and spectral parameters of digital elevation model. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1–17. DOI: 10.1109/TGRS.2021.3091771.

[9] D Ventura, A Bonifazi, M F Gravina, et al. Mapping and classification of ecologically sensitive marine habitats using unmanned aerial vehicle (UAV) imagery and object-based image analysis (OBIA). Remote Sensing, 2018, 10(9): 1331. DOI: 10.3390/rs10091331.

[10] A I De Castro, J Torres-Sánchez, J M Peña, et al. An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. Remote Sensing, 2018, 10(2): 285. DOI: 10.3390/rs10020285.

[11] J Gao, W Liao, D Nuyttens, et al. Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. International journal of applied earth observation and geoinformation, 2018, 67: 43-53. DOI: 10.1016/j.jag.2017.12.012.

[12] J Long, E Shelhamer, T Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp: 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

[13] L Chen, G Papandreou, I Kokkinos, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834–848. DOI: 10.1109/TPAMI.2017.2699184.

[14] K Zou, X Chen, Y Wang, et al. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. Computers and Electronics in Agriculture, 2021, 187: 106242. DOI: https://doi.org/10.1016/j.compag.2021.106242.

[15] I Sa, M Popović, R Khanna, et al. WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. Remote Sensing, 2018, 10(9): 1423. DOI: 10.3390/rs10091423.

[16] K He, H Fan, Y Wu, et al. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp: 9729–9738. DOI: 10.1109/cvpr42600.2020.00975.

[17] Y Ouali, C Hudelot, M Tami. Semi-supervised semantic segmentation with cross-consistency training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp: 12674–12684. DOI: 10.1109/CVPR42600.2020.01269.

[18] D Li, J Yang, K Kreis, et al. Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp: 8300–8311. DOI: 10.1109/ cvpr46437.2021.00820.

[19] T Lin, P Dollár, R Girshick, et al. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp: 2117–2125. DOI: 10.1109/CVPR.2017.106.

[20] S Liu, L Qi, H Qin, et al. Path aggregation network for instance segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp: 8759–8768. DOI: 10.1109/ CVPR.2018.00913.

[21] W Liu, D Anguelov, D Erhan, et al. Ssd: Single shot multibox detector. In European conference on computer vision, 2016, pp: 21–37. DOI: 10.1007/978-3-319-46448-0_2.