Two-step Transfer Learning for Semantic Plant Segmentation

Shunsuke Sakurai[1], Hideaki Uchiyama[1], Atsushi Shimada[1], Daisaku Arita[2] and Rin-ichiro Taniguchi[1]
[1]*Kyushu University, Fukuoka, Japan*
[2]*University of Nagasaki, Sesebo, Japan*

Keywords: Semantic Segmentation, Transfer Learning, Deep Learning, CNN, Plant Segmentation.

Abstract: We discuss the applicability of a fully convolutional network (FCN), which provides promising performance in semantic segmentation tasks, to plant segmentation tasks. The challenge lies in training the network with a small dataset because there are not many samples in plant image datasets, as compared to object image datasets such as ImageNet and PASCAL VOC datasets. The proposed method is inspired by transfer learning, but involves a two-step adaptation. In the first step, we apply transfer learning from a source domain that contains many objects with a large amount of labeled data to a major category in the plant domain. Then, in the second step, category adaptation is performed from the major category to a minor category with a few samples within the plant domain. With leaf segmentation challenge (LSC) dataset, the experimental results confirm the effectiveness of the proposed method such that F-measure criterion was, for instance, 0.953 for the A2 dataset, which was 0.355 higher than that of direct adaptation, and 0.527 higher than that of non-adaptation.

1 INTRODUCTION

Segmentation of plant leaves is a fundamental issue in plant phenotyping aiming to capture and analyze leaf shape, size, color and growth. Image-based automatic segmentation plays an important role in reducing the cost of phenotype analysis. Challenges have been organized on phenotyping based on computer vision techniques [1] [2]. In the challenges, Arabidopsis and young tobacco were focused as the most common rosette model plants, and several methods were applied to leaf segmentation tasks (Pape and Klukas, 2014; Scharr et al., 2015). Although these methods performed well in case of the particular plant leaves, the performance was supported by explicitly designed methodologies. In terms of broad applicability, there are various limitations to design methodologies for many types of leaves.

In this paper, we discuss the applicability of the fully convolutional networks(FCN) (Long et al., 2015) to plant segmentation tasks as a basic stage of leaf segmentation. The most important and challenging issue is how to train the FCN with plant features from a small image dataset. Compared with large-scale image datasets with hundreds of images for each category, plant image datasets provided for the leaf segmentation challenges (LSC) contained fewer images, as described in Section 3.1. To solve this problem, we first introduce the idea of transfer learning. Transfer learning is a research problem in machine learning that focuses on storing knowledge gained while solving one problem in a source domain and applying it to a different but related problem in a target domain (Pan and Yang, 2010). In our study, knowledge of object segmentation in a source domain (i.e., the FCN trained by a large amount of object images) is transferred to plant segmentation tasks.

We define two words, “domain” and “category”, to clarify the explanation in the rest of the paper. We apply transfer learning to realize adaptation from a source domain to a target domain. A domain can include several categories, and the target domain is limited to plant categories in our study. In other words, only plant categories belong to the target domain, namely “plant domain”. The goal of our study is to realize precise plant segmentation even for a category having a small number of training samples. One of the simplest method of transfer learning is to perform the adaptation from the source domain to a target category directly. However, limited training samples may cause insufficient adaptation to the target category .

In this paper, we propose a two-step transfer learn-

[1]www.plant-phenotyping.org/CVPPP2014
[2]www.plant-phenotyping.org/CVPPP2015

Sakurai, S., Uchiyama, H., Shimada, A., Arita, D. and Taniguchi, R-i.
Two-step Transfer Learning for Semantic Plant Segmentation.
DOI: 10.5220/0006576303320339
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 332-339
ISBN: 978-989-758-276-9

ing. In the first step, we apply transfer learning for domain adaptation from the source domain to the plant domain by using a major category in the plant domain. Then, in the second step, category adaptation is performed from the major category to minor categories within the plant domain. With the LSC datasets, the experimental results confirm the effectiveness of the proposed method.

2 RELATED WORK

Image-processing based approaches have been proposed for the LSC. The details can be found in the collation study report (Scharr et al., 2015). In this paper, we review the approaches by focusing on segmentation procedures.

In IPK Gatersleben (Pape and Klukas, 2014), 3-D histogram cubes with three color channels are used instead of multiple one-dimensional color components. The histogram values are then encoded as probabilities of foreground/background, therefore, the pixels are assigned to foreground if the corresponding cube contains a higher value than the cell of the background cube. After morphological operations are applied to smooth the object borders, the remaining large greenish objects are regarded as leaf regions.

In Nottingham (Scharr et al., 2015), a superpixel-based approach is proposed. Simple linear iterative clustering (SLIC) (Achanta et al., 2012) is applied in the Lab color space, to obtain superpixels. Then, the foreground (plant) is extracted from the background using simple seeded region growing in the superpixel space. Although this approach does not require any training, parameter tuning is required by using the training dataset.

In MSU (Scharr et al., 2015), a multi-leaf alignment and tracking framework is modified in order to adapt to the LSC. The original method performed well due to the clean background (Yin et al., 2014). The MSU introduces a more advanced background segmentation process.

In Wageningen (Scharr et al., 2015), supervised classification with a neural network is used for plant segmentation from the background. To separate the plants from the background, four color features and two texture features are used: R, G, B and the excessive green value (2G-R-B) for color features, the pixel values of the variance filtered green channel, and the pixel values of the gradient magnitude filtered green channel for texture features. A multi layer perceptron (MLP) with one hidden layer is used for the feature training.

3 TWO-STEP TRANSFER LEARNING

3.1 Dataset

The dataset used in this study is the one provided in the LSC of CVPPP 2014 (Minervini et al., 2014; Scharr et al., 2014). Three categories are available in the LSC dataset: A1, A2 and A3, which correspond to Arabidopsis thaliana, Arabidopsis thaliana variant, and tobacco plant images, respectively. These categories are regarded as the target domain.

Each dataset contains RGB plant images taken from the top view and corresponding label (leaf or background) images. There are 128 images of A1, 31 images of A2, and 27 images of A3 in the dataset. The resolution of A1, A2 and A3 are 480×512, 512×544, and 2176×1792, respectively. In order to use the leaf instance labels as the ground truth of the plant semantic segmentation, we created foreground label images by combining all instances.

3.2 Overview

The overview of our proposed framework is shown in Figure 1. First, training of the FCN is performed in the source domain, where a large amount of labeled data is available such as ImageNet (Russakovsky et al., 2015) and PASCAL VOC. We use the ImageNet dataset for the training in the source domain. Then, the model obtained from the source domain is transferred to a target domain (plant domain). We use the LSC dataset for the transfer learning in the target domain. There are three categories in the LSC dataset as mentioned above, and the number of labeled samples in A2 and A3 is much smaller than that in A1. In this study, we call A1 as a major category and A2 and A3 as minor categories in terms of available data size. In the first step, the transfer learning is performed by using a major category in the target domain for domain adaptation. We assume that the transferred network will grasp abstract features of the target domain. Subsequently, category adaptation from the major category to a minor category is performed in the second step.

3.3 Network Architecture

The architecture is the FCN with 3 deconvolution (transposed convolution) layers (FCN-8s) as shown in Figure 2. The FCN has 2 channels of output and represents each class label(plant or background). Rectified Linear Unit (ReLU) is used as a network activation function.

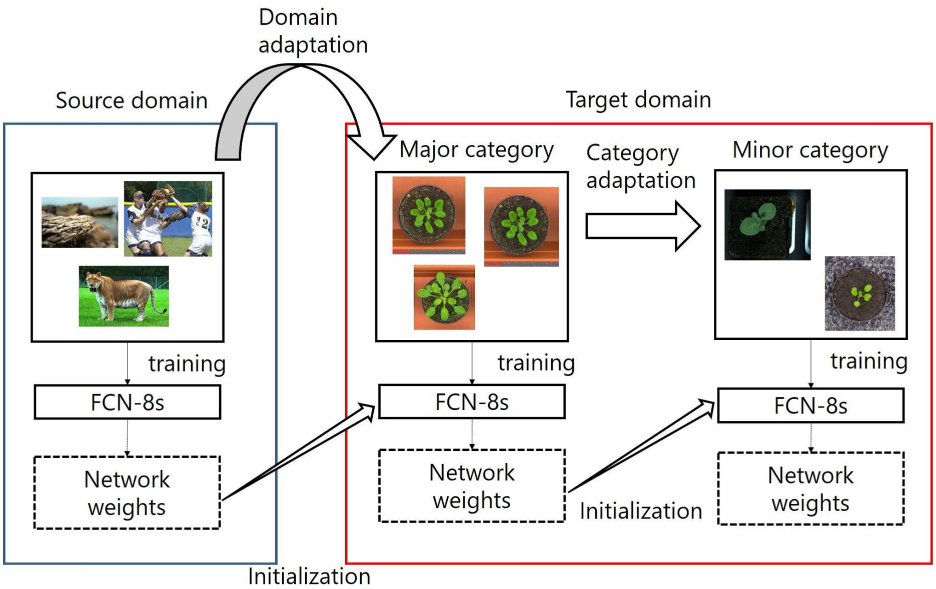


Figure 1: Overview of two-step transfer learning.

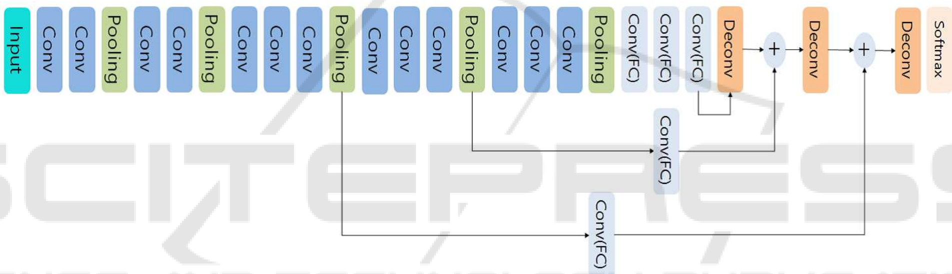


Figure 2: Network architecture.

The FCN is initialized with parameters of the model trained using the ImageNet dataset in the source domain, as done in the original FCN (Long et al., 2015). Specifically, the network weights obtained from the source domain are used as the initial weights in the transfer learning with the same network architecture. In this step, the weights of deconvolution and convolution layers in skip connection are not transferred because of the difference of network architecture between the source domain and the target domain. The weights of these layers are always randomly initialized. In the first step, the A1 dataset is used for the training of the domain adaptation, in which the initial weights come from the source domain. In the second step, the A2 or A3 dataset is used for training of the category adaptation. It should be noted that the initial weights at the second step come from the results of the transfer learning in the first step.

During the training, the order of input images is randomly shuffled with a batch size of 1, for each epoch. The FCN outputs the plant label or the background label for each pixel. The estimated segmentation result is compared with the ground truth, and the error between them is calculated by the cross entropy. Thereafter, the network parameters are updated by a back propagation method. Adam (Kingma and Ba, 2014) was used as a method to update network parameters. The learning rate was set to 10^{-5}.

4 EXPERIMENT

4.1 Evaluation Criteria

Precision, recall and F-measure were used as the evaluation criteria. The scores are calculated in terms of true positive (TP), false positive (FP) and false negative (FN) for each pixel.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\mathrm{F-measure} = \frac{2Recall \times Precision}{Recall + Precision}$$

Precision is the relevance ratio. The higher the value, the less is the false detection of the foreground. Recall signifies the ratio of foreground that is not detected. F-measure is a harmonic mean of precision and recall, providing an evaluation of both false detection and detection of foreground. We calculate the precision, recall and F-measure for each test image. The final scores are acquired by averaging over all the test images.

4.2 Effectiveness of Two-step Transfer Learning

We investigated the effectiveness of two-step transfer learning for plant segmentation. The LSC dataset consists of the training dataset and the test dataset. However, the ground truth is not given for the test dataset. For this reason, we decided to filter out the test dataset. Instead, we divided the original training dataset into two subsets; 104 images in A1, 7 images in A2, and 5 images in A3 for the training, and the rest of the images for the testing.

We compared the proposed two-step transfer learning with other conditions of direct adaptation from the source domain to the target category and/or skipping the training the source domain. In each condition, we fixed the number of training epochs as 150, even if domain adaptation was not applied. We named each condition in terms of the dataset used for the training. The details of each condition are summarized as follows.

- **Random_A1.** The FCN parameters were initialized with random values. Then the FCN is trained by the major category (A1) in the target domain. In this condition, no transfer learning is performed.
- **ImgNet_A1.** The FCN parameters were initialized with the parameters trained ImageNet dataset. Then, the domain adaptation is performed by the major category (A1) in the target domain.
- **Random_direct_A2, Random_direct_A3.** The FCN parameters were initialized with random values. In both the steps of domain adaptation and category adaptation, we transferred no knowledge from the source domain to the target category. In other words, no transfer learning is conducted. The training data in the target category with fewer samples (A2 or A3) are directly used for the FCN training.
- **Random_A1_A2, Random_A1_A3.** The FCN parameters were initialized with random values. In the step of domain adaptation, we transferred no knowledge from the source domain to the major category. In the step of category adaptation, we apply transfer learning from the major category (A1) to the minor category (A2 or A3).
- **ImgNet_direct_A2, ImgNet_direct_A3.** The FCN parameters were initialized with the parameters trained ImageNet dataset. Then, the domain adaptation is skipped and A2 or A3 samples are directly used for one-step transfer learning.
- **ImgNet_A1_A2, ImgNet_A1_A3.** The following is the proposed approach. The FCN parameters were initialized with the parameters trained ImageNet dataset. In the step of domain adaptation, we apply transfer learning from the source domain to the major category (A1) in the target domain. In the step of category adaptation, we apply transfer learning from the major category (A1) to the minor category (A2 or A3).

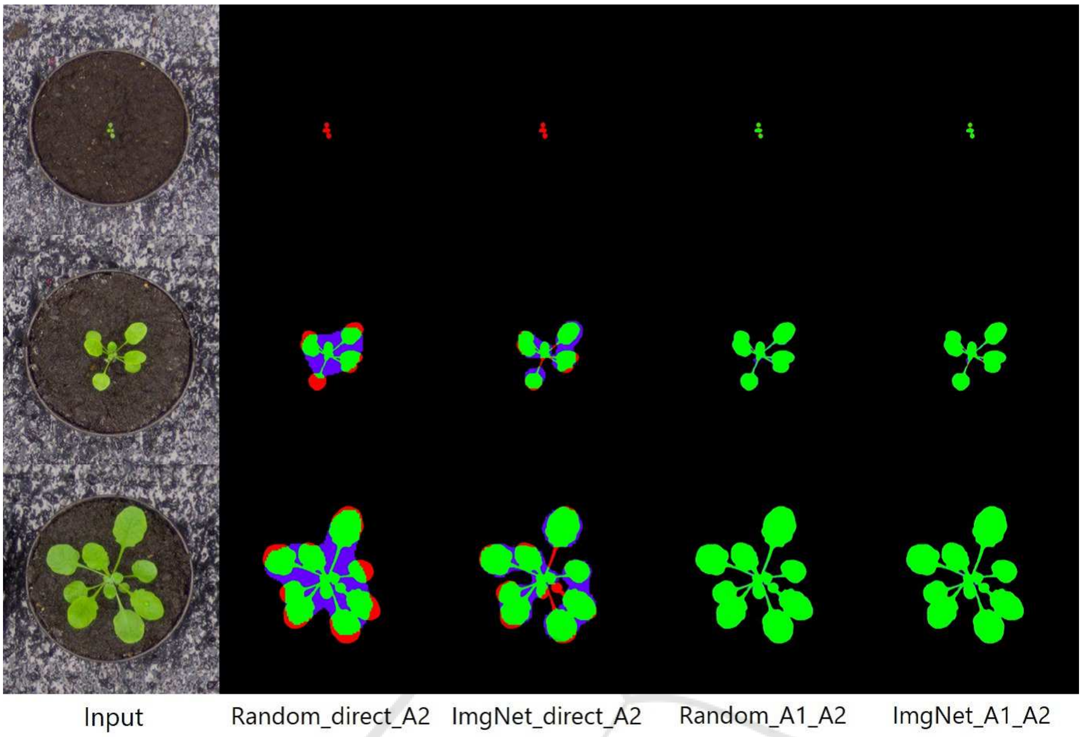
The segmentation results are illustrated in Figure 3. The green, red and purple color pixels denote true positive, false negative and false positive pixels, respectively. Overall, the proposed two-step transfer learning approach generated accurate segmentation results, as shown in the 5th column in the figure. The evaluation results are shown in Table 1.

First, with regards to Table 1a, when we used the ImageNet dataset for the initial training of the FCN, the trained network provided higher precision, recall, and F-measure. This result confirms the effectiveness of the domain adaptation.

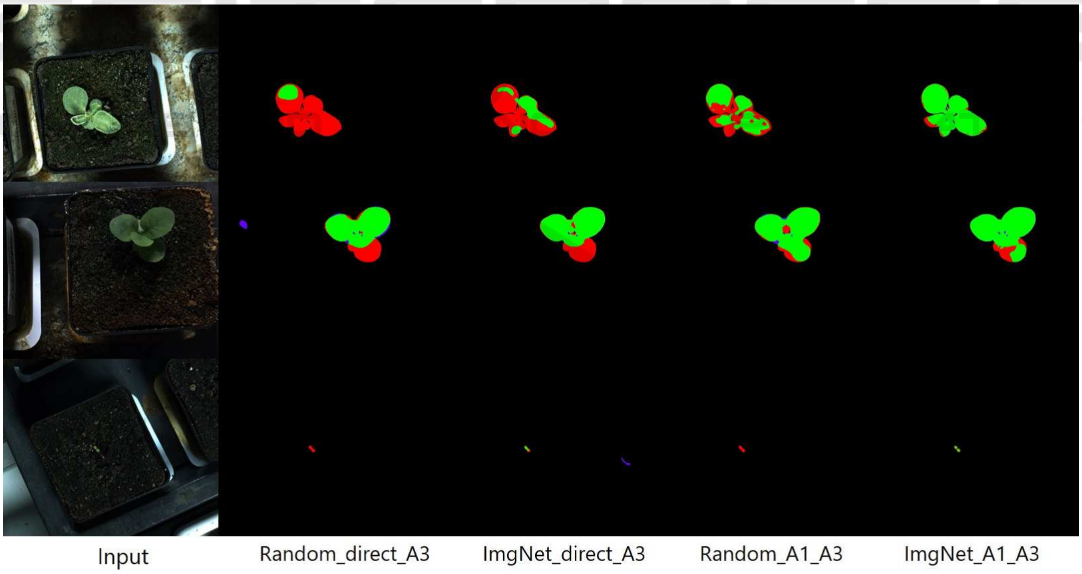
Second, in the case in which we did not use transfer learning, i.e., Random_direct_A2 , Random_direct_A3, the evaluation scores were much worse than the other conditions. The scores were improved when we used the ImageNet dataset for the initial training of the FCN and the adaptation to the target category (A2 or A3) was performed (see ImgNet_direct_A2in Table 1b and ImgNet_direct_A3 inTable 1c).

Third, in the case of category adaptation only without domain adaptation, i.e., Random_A1_A2 and Random_A1_A3, better results were obtained for both A2 and A3. These results indicate that training with the major category is effective to capture the abstract features of plants, and the transfer learning worked well to adapt to the minor categories.

Finally, the proposed approach, namely ImgNet_A1_A2 and ImgNet_A1_A3, outperformed the other conditions in every category. For instance, for the A2 dataset, the F-measure criterion of proposed approach was 0.953, which was 0.355 higher than that of direct adaptation and 0.527 higher than



(a) A2 result



(b) A3 result

Figure 3: Learning result. The green areas represent true positive. The red areas represent false negative. The purple areas represent false positive.

that of non-adaptation. Comparing the results in the 3rd and 4th columns, we found that the domain adaptation contributes to improving the evaluation scores of segmentation.

4.3 Data Size for Domain Adaptation

In Section 4.2, we found that the major category in the target domain played an important role in acquiring features of the target domain, and to bridge the

Table 1: Evaluation result of plant segmentation in three categories.

(a) A1 result

	Random_A1	ImgNet_A1
Precision	0.931	**0.968**
Recall	0.912	**0.983**
F-measure	0.921	**0.975**

(b) A2 result

	Random_direct_A2	Random_A1_A2	ImgNet_direct_A2	ImgNet_A1_A2
Precision	0.430	0.938	0.616	**0.953**
Recall	0.454	**0.958**	0.592	0.955
F-measure	0.426	0.946	0.598	**0.953**

(c) A3 result

	Random_direct_A3	Random_A1_A3	ImgNet_direct_A3	ImgNet_A1_A3
Precision	0.631	0.948	0.865	**0.963**
Recall	0.409	0.731	0.771	**0.941**
F-measure	0.444	0.808	0.794	**0.948**

Table 2: Segmentation accuracy when reducing the number of training samples for domain adaptation. The first row of the table is the number of samples in A1 used for training. The highest score in each category is marked in bold.

(a) Domain adaptation

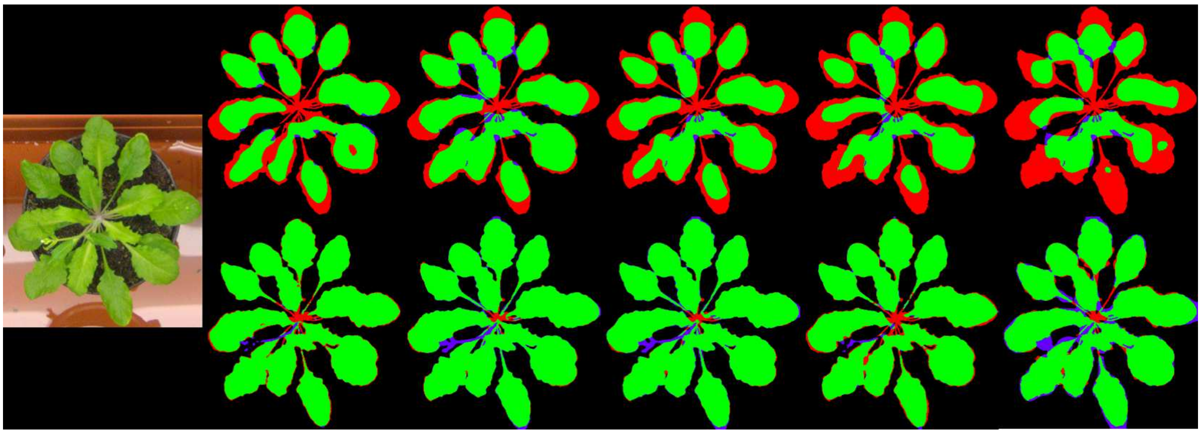
		90 images	70 images	50 images	30 images	10 images
Random_A1	Precision	0.921	0.884	0.880	0.864	0.813
	Recall	0.862	0.882	0.838	0.839	0.784
	F-measure	0.889	0.881	0.856	0.848	0.790
ImgNet_A1	Precision	**0.982**	0.957	0.980	0.949	0.927
	Recall	0.968	**0.984**	0.950	0.967	0.931
	F-measure	**0.975**	0.970	0.965	0.958	0.928

(b) Category adaptation

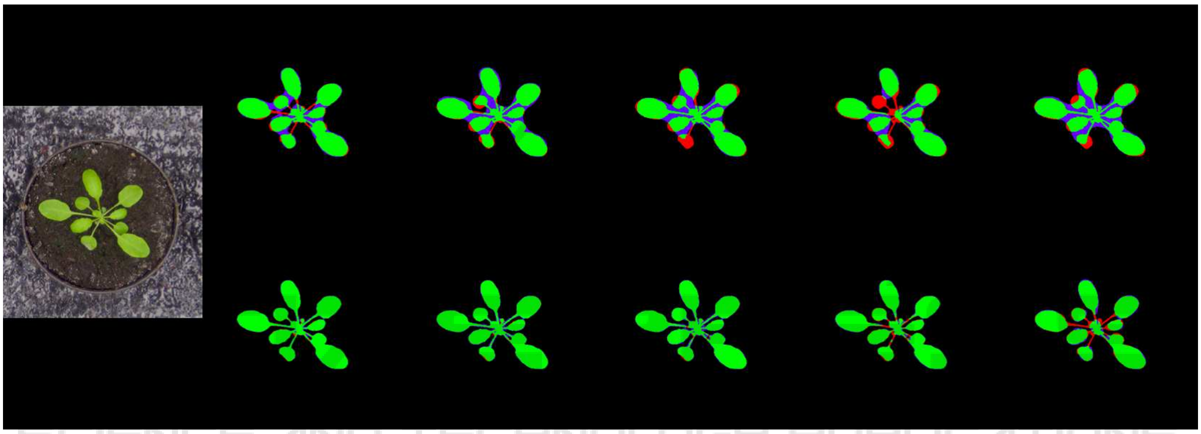
		90 images	70 images	50 images	30 images	10 images
Random_A1_A2	Precision	0.728	0.683	0.630	0.666	0.571
	Recall	0.776	0.759	0.682	0.616	0.670
	F-measure	0.727	0.716	0.653	0.630	0.613
ImgNet_A1_A2	Precision	**0.947**	0.910	0.877	0.930	0.790
	Recall	0.961	0.977	**0.980**	0.921	0.834
	F-measure	**0.954**	0.942	0.924	0.925	0.810
Random_A1_A3	Precision	0.920	0.905	0.815	0.865	0.735
	Recall	0.740	0.696	0.817	0.659	0.740
	F-measure	0.809	0.763	0.791	0.709	0.709
ImgNet_A1_A3	Precision	**0.980**	0.979	0.961	0.921	0.872
	Recall	0.834	0.851	**0.922**	0.883	0.823
	F-measure	0.883	0.906	**0.928**	0.899	0.835

features from the source domain to target minor categories. In this section, we investigate how many samples are required for adequate training of the major category A1. We reduced the number of training samples from 104 images to 90, 70, 50, 30 and 10 images gradually. Then, domain adaptation and/or category adaptation was performed.

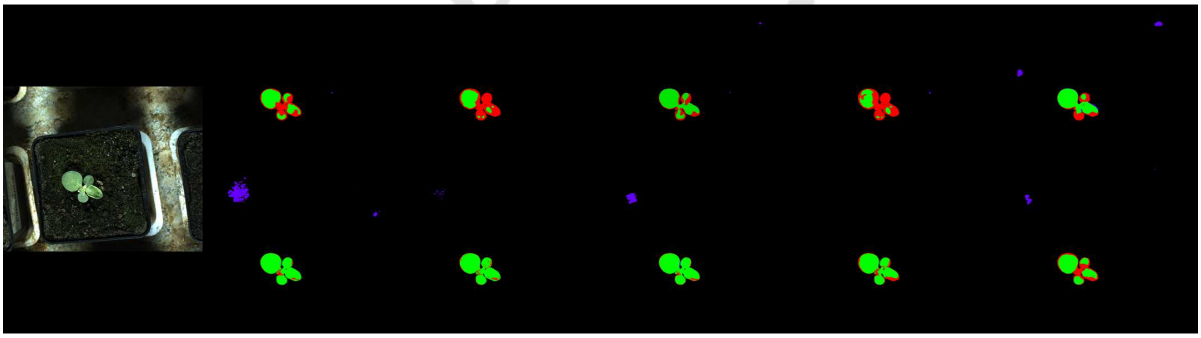
The results of domain adaptation and category adaptation are shown in Table 2a and Table 2b, respectively. The segmentation results are shown in Figure 4. As the number of training samples of the major category decreased, the evaluation scores wors-



(a) The upper is random_A1 result and the lower is ImgNet_A1 result.



(b) The upper is random_A1_A2 result and the lower is ImgNet_A1_A2 result.



(c) The upper is random_A1_A3 result and the lower is ImgNet_A1_A3 result.

Figure 4: Leaf segmentation result. Start from the left, input image, 90 images, 70 images, 50 images, 30 images, 10 images in A1 used for training.

ened. Comparing random_A1 and ImgNet_A1, when the FCN was initially trained by the ImageNet dataset, the evaluation scores of the A1 segmentation were not seriously decreased. The segmentation results of ImgNet_A1 were clearly better than those of Random_A1 as shown in Figure 4a. In the cases of category adaptation, the evaluation scores were decreased as well as in the cases of domain adaptation. Comparing the results of “Random_” and “ImgNet_”, we can see that the domain adaptation is necessary to maintain satisfactory evaluation scores even if the number of training samples in the major category decreased. In the LSC dataset, we found out that approximately 70 images are required for promising results.

5 CONCLUSION

In this study, we investigated the effectiveness of transfer learning in plant segmentation tasks. In our proposed approach, we applied two-step transfer learning; domain adaptation from a the source (object) domain to target (plant) domain, and category adaptation from the major category to a minor category. We used a FCN for transfer learning and segmentation of whole leaf regions. In our experiments, we used the LSC dataset for the evaluation, and found that the two-step transfer learning yielded much higher accuracy of segmentation. In our future work, we will evaluate our approach using other datasets provided by CVPPP2015 and CVPPP2017.

ACKNOWLEDGMENT

This work is supported by JSPS KAKENHI Grant Number JP15H01695 and JP17H01768, and grants from the Project of the NARO Bio-oriented Technology Research Advancement Institution (the special scheme project on regional developing strategy)

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2274–2282.

Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440.

Minervini, M., Abdelsamea, M. M., and Tsaftaris, S. A. (2014). Image-based plant phenotyping with incremental learning and active contours. *Ecological Informatics*, 23:35–48.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. on Knowl. and Data Eng.*, 22(10):1345–1359.

Pape, J. and Klukas, C. (2014). 3-d histogram-based segmentation and leaf detection for rosette plants. In *Computer Vision - ECCV 2014 Workshops - Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part IV*, pages 61–74.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Scharr, H., Minervini, M., Fischbach, A., and Tsaftaris, S. A. (2014). Annotated image datasets of rosette plants. In *European Conference on Computer Vision. Zürich, Suisse*, pages 6–12.

Scharr, H., Minervini, M., French, A., Klukas, C., Kramer, D., Liu, X., Luengo, I., Pape, J., Polder, G., Vukadinovic, D., Yin, X., and Tsaftaris, S. (2015). Leaf segmentation in plant phenotyping: a collation study. *Machine Vision and Applications*, pages 1–22.

van Opbroek, A., Ikram, M., Vernooij, M., and de Bruijne, M. (2015). Transfer learning improves supervised image segmentation across imaging protocols. *I E E E Transactions on Medical Imaging*, 34(5):1018–1030.

Yin, X., Liu, X., Chen, J., and Kramer, D. M. (2014). Multi-leaf tracking from fluorescence plant videos. In *in Proceedings of IEEE Conference on Image Processing*.