Using deep learning for bamboo forest detection from Google Earth images.

Shuntaro Watanabe[1], Kazuaki Sumi[2], Takeshi Ise[1,3]

1. Field Science Education and Research Center (FSERC), Kyoto university, Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto, 606-8502, Japan
2. Graduate School of Agriculture, Kyoto University, Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto, 606-8502, Japan
3. PRESTO, Japan Science and Technology Agency, 7 Goban-cho, Chiyoda-ku, Tokyo, 102-0076, Japa

Author for correspondance: SW
E-mail: sh.watanabe1006@gmail.com

Abstract

Classifying and mapping vegetation are very important in environmental science or natural resource management. However, these tasks are not easy because conventional methods such as field survey are highly labor intensive. Automatic identification of the objects from visual data is one of the most promising way to reduce the cost for vegetation mapping. Although deep learning has become the new solution for image recognition and classification recently, in general, detection of ambiguous objects such as vegetation still has been considered difficult. In this paper, we investigated the potential for adapting the chopped picture method, a recently described protocol of deep learning, to detect plant community in Google Earth images. We selected bamboo forests as the target. We obtained Google Earth images from 3 regions in Japan using Google Earth. Applying deep convolutional neural network, the model successfully learned the features of bamboo forests in Google Earth images and the best trained model successfully detected 97 % targets. Our results also show that identification accuracy is strongly depends on the image resolution and the quality of training data. Our results highlight that deep learning and chopped picture method potentially become a powerful tool for high accuracy automated detection and mapping of vegetation.

Key Words

Deep learning, vegetation mapping, Google Earth image

Introduction

Classifying and mapping vegetation are essential for environmental science or natural resources management (Franklin, 2009). Traditional methods (e.g. field surveys, literature reviews, map interpretation), however, are not effective to acquire vegetation data because they are labor intensive and often economically expensive. The technology of remote sensing offers a practical and economical means to acquire information on vegetation cover, especially over large areas (reviewed by Xie et al., 2008). Because of its systematic observations at various scales, remote sensing technology possibly enable classification and mapping of vegetation at high temporal resolution.

Detecting the discriminating visual features are one of the most important steps in almost any computer vision problem, including in the remote sensing. Since conventional methods such as support vector machine (Hearst et al., 1998) requires hand-designed, time-consuming feature extraction, substantial efforts have been dedicated to develop the method for automatic extraction of features. Recently, deep learning has become the new solution for image recognition and classification because the new method does not need manual extraction of features.

Deep learning (Bengio et al., 2009; Goodfellow et al., 2016) is one of the types of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. Deep learning learn features and classifiers at once and it uses training data to categorize image content without a priori specification of image features. Among all deep learning-based networks, a specific type, called Convolutional (Neural) Networks (Bengio et al., 2009; Goodfellow et al., 2016), is the most popular for learning visual features in computer vision applications, including remote sensing. Recent researches have shown that CNN is effective for diverse applications (Karpathy et al., 2014; Yosinski et al., 2014).

Given its success, deep learning has been intensively used in several distinct tasks of different academic and industrial fields, including plant science. Recent research show that deep learning technique successfully detect plant disease or correctly classify the plant specimens in herbarium (Mohanty et al., 2016; Ramcharan et al., 2017; Carranza-Rojas et al., 2017). Deep learning is also a promising technology in the field of remote sensing because it has a natural ability to effectively encode spectral and spatial information (Yue et al., 2015; Nogueira et al., 2017), but application is not

sufficient yet because automatic object identification including deep learning tends not to work well on ambiguous, amorphous objects. Thus, description of vegetation cover is still considered as a challenging task. Recently, Ise et al., (2018) developed a method to extract the characteristics from ambiguous and amorphous objects. This method dissected the images into numerous small squares and efficiently produces the training images. Using this method, Ise et al., (2018) correctly classified 3 moss species and "non-moss" objects in test images with accuracy more than 90%.

In this paper, we investigated the potential for adapting a deep learning model and chopped picture method to vegetation detection in Google Earth images, especially for bamboo forest. Recent years, bamboo become invasive in Japan. Japanese people have mainly introduced and used two exotic bamboos (Poaceae), moso (*Phyllostachys edulis*) and madake (*P. bambusoides Siebold*), for a long time but, since 1970s bamboo industry in Japan was decline due to cheaper bamboo imports and heavy labor costs (Nakashima, 2001). Consequently, many bamboo plantations became unmanaged and eventually invading the adjacent native vegetation (Nishikawa et al., 2005; Okutomi et al., 1996; Suzuki, 2015).

We specifically focused on following questions. 1) How the resolution of images affects the accuracy of detection? 2) How the chopping sizes of training images affects the accuracy of detection? 3) Can a model learned on one geographical location work well for different location?

Materials and methods

Target area and image acquisition

In this study, we choose three regions (Sanyo-Onoda, Ide and Isumi) in Japan (Fig.1). We used the Google Earth as the source of imagery. From a given sampling location, we obtained the images at a zoom level 1/500 (around 0.13m/pix spatial resolution), 1/1000 (around 0.26m/pix spatial resolution) and 1/2500 (around 0.65m/pix spatial resolution) spatial scale.

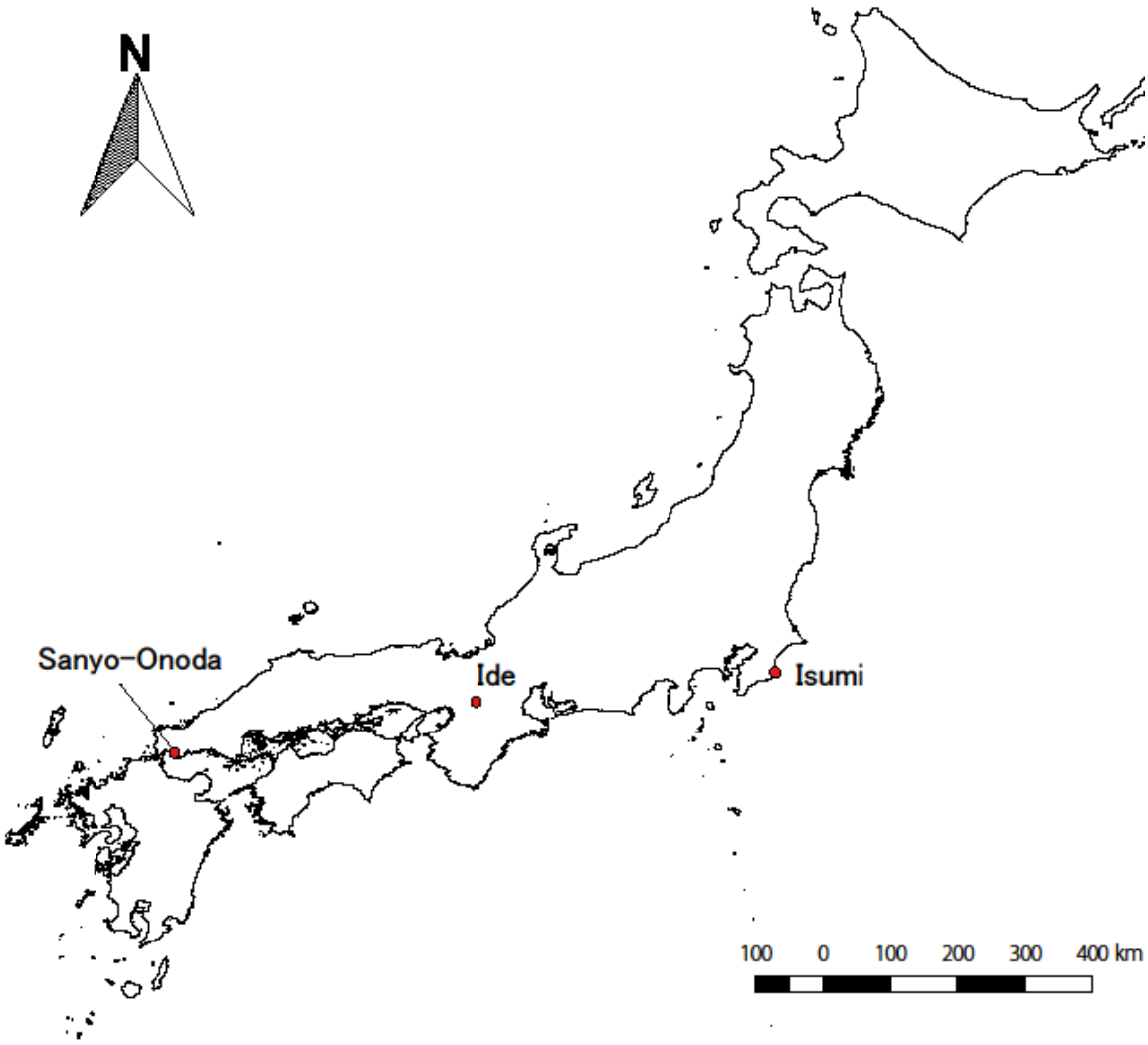

Fig.1 Target regions of this research

Approach

The schematic diagram of our approach was shown in Fig. 2. We prepared the training data using the chopped picture method (Ise et al., 2018). First, in this method, we collect the images that are (1) nearly 100% covered by bamboo and (2) not covered by bamboo. Next, we "chopped" this picture into small squares with 50% overlap both vertically and horizontally.

We made a model for image classification from a deep convolutional neural network model (CNN) for the bamboo forest detection. As opposed to traditional approaches of training classifiers with hand-designed feature extraction, CNN learn feature hierarchy from pixels to classifier and train layers jointly. We use the final layer of the CNN model for detecting the bamboo coverage from Google Earth images. To make a model for object identification, we used the deep learning framework of NVIDIA DIGITS (NVIDIA 2016). We used 75% of the obtained images as training data and the remaining 25% as the validation data. We used the LeNet network model (LeCun et al., 1998). The model parameters implemented in this study included the number of training epochs (30), the learning rate (0.01), train batch size (64), and the validation batch size (32).

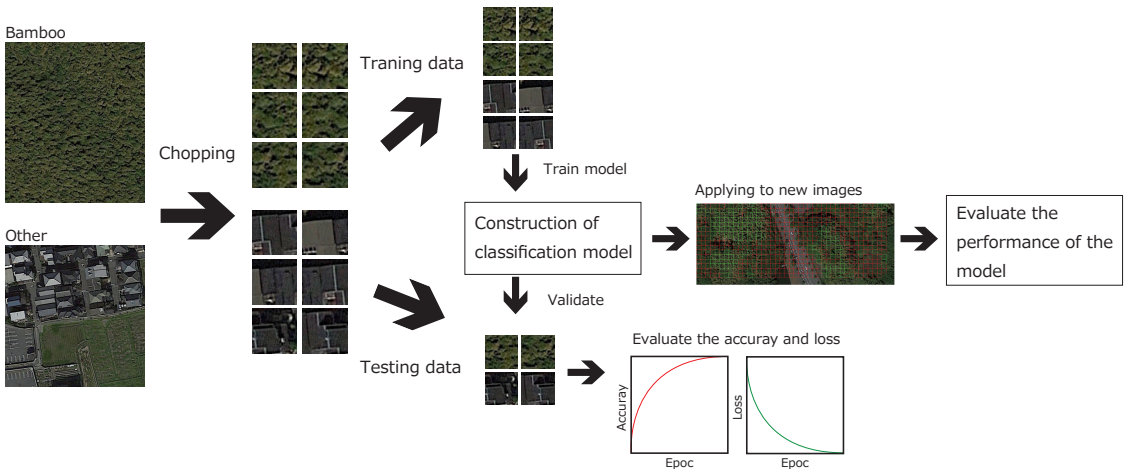

Fig.2 Schematic diagram of research approach

Model validation

Evaluation of learning accuracy

Validation of model in each learning epoch was conducted using accuracy and loss function obtained from cross validation images. The accuracy indicates how accurately the model can classify the validation images. Loss represents that the inaccuracy of prediction of the model. If model learning is successful, loss (val) is low and accuracy is high. However, when loss (val) becomes high during learning, it indicates that over fitting is occurring.

Evaluation of performance of model

We obtained 10 new images, which are uniformly covered by a bamboo forest only or other than bamboo forest only from each study sites. Next, we re-size the images using chopped picture method. Third, we randomly sampled 500 images from re-sized images. Finally, we applied the model to sampled images and evaluate the classification accuracy. To evaluate the performance of model, we classified the classification results into following four categories, true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Next we calculated the classification accuracy, recall rate and precision rate using following equation.

$$Classification\ accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Recall\ rate = TP/(TP + FN) \quad (2)$$

$$Precision\ rate = TP/(TP + FP) \quad (3)$$

Testing the effects of image resolution on classification accuracy

To quantify the effects of image resolution on the accuracy of detection, we obtained images at a zoom level 1/500 (~0.13m/pixel), 1/1000 (around 0.26m/px spatial resolution) and 1/2500 (around 0.65m/px spatial resolution) spatial scale from each study site. Next, we applied chopped picture method. To adjust the spatial extent of each chopped image, we chopped 56 pix for 1/500, 28 pix for 1/1000 and 14 pix for

1/2500 image. After construct the model, we applied the model for new images and calculated the classification accuracy, recall rate and precision rate.

Testing the effects of chopping grid size on classification accuracy

To quantify the effects of spatial extent of chopping grid on the accuracy of detection, we chopped 1/500 images of each study site for 3 type pixel size (84, 56, 28). After construct the model, we applied the model for new images and calculated the classification accuracy, recall rate and precision rate.

Transferability test

Given the large amount of variation in the visual appearance of bamboo forest across different cities, it is of interest to study to what extent a model learned on one geographical location can be applied to a different geographical location. As such, we perform experiments in which we train a model for one (or more) cities, then apply the model to a different set of cities. Performance of the model was evaluated by classification accuracy, recall rate and precision rate.

Results

Fluctuation of accuracy and loss during the learning epochs.

The accuracy in classifying the validation data of final layer was ranged from 94% to 99%. Loss values of validation data was ranged from 0.008 to 0.214 (Fig.3). Values of accuracy was increase and loss was decrease following the learning epochs (Fig.3). Results suggest the all models were not overfit to the datasets and successfully learned the features of chopped pictures.

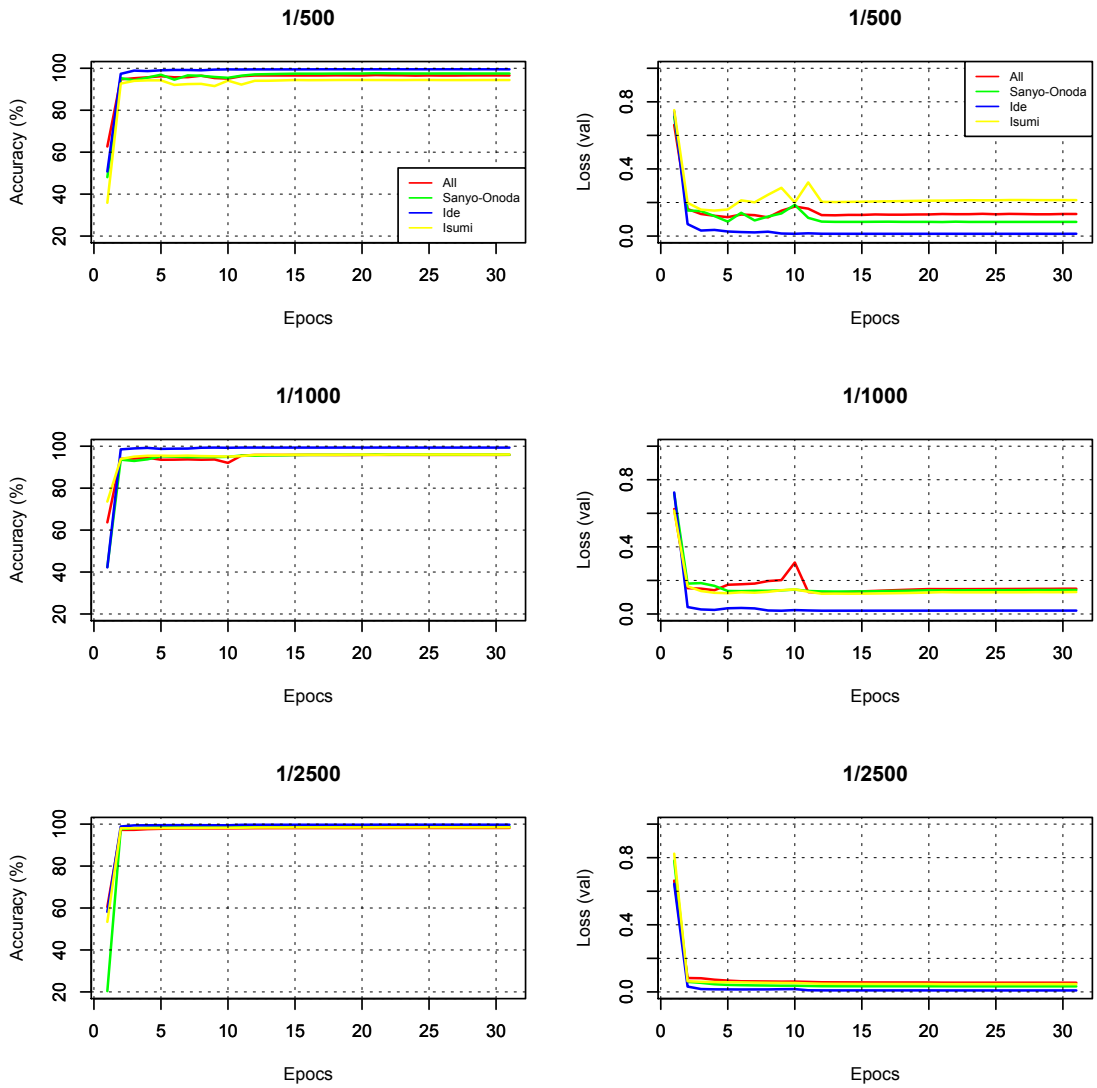

Fig.3 Accuracy and loss (val) of model at each learning epoch.

Effects of image resolution on classification accuracy

The classification accuracy was ranged 76% to 97% (Fig.4a). The recall rate and precision rate of bamboo forest was ranged 52 % to 96 % and 91 % to 99 %, respectively (Fig.4 b d). The recall rate and precision rate of objects other than bamboo forest was ranged 92 % to 99 % and 67% to 96%, respectively (Fig.4 c e). The recall rate of bamboo forest was decline following the image resolution rate was declined and it was dramatically declined when we use 1/2500 (around 0.65m/pix spatial resolution) images (Fig.4 a).

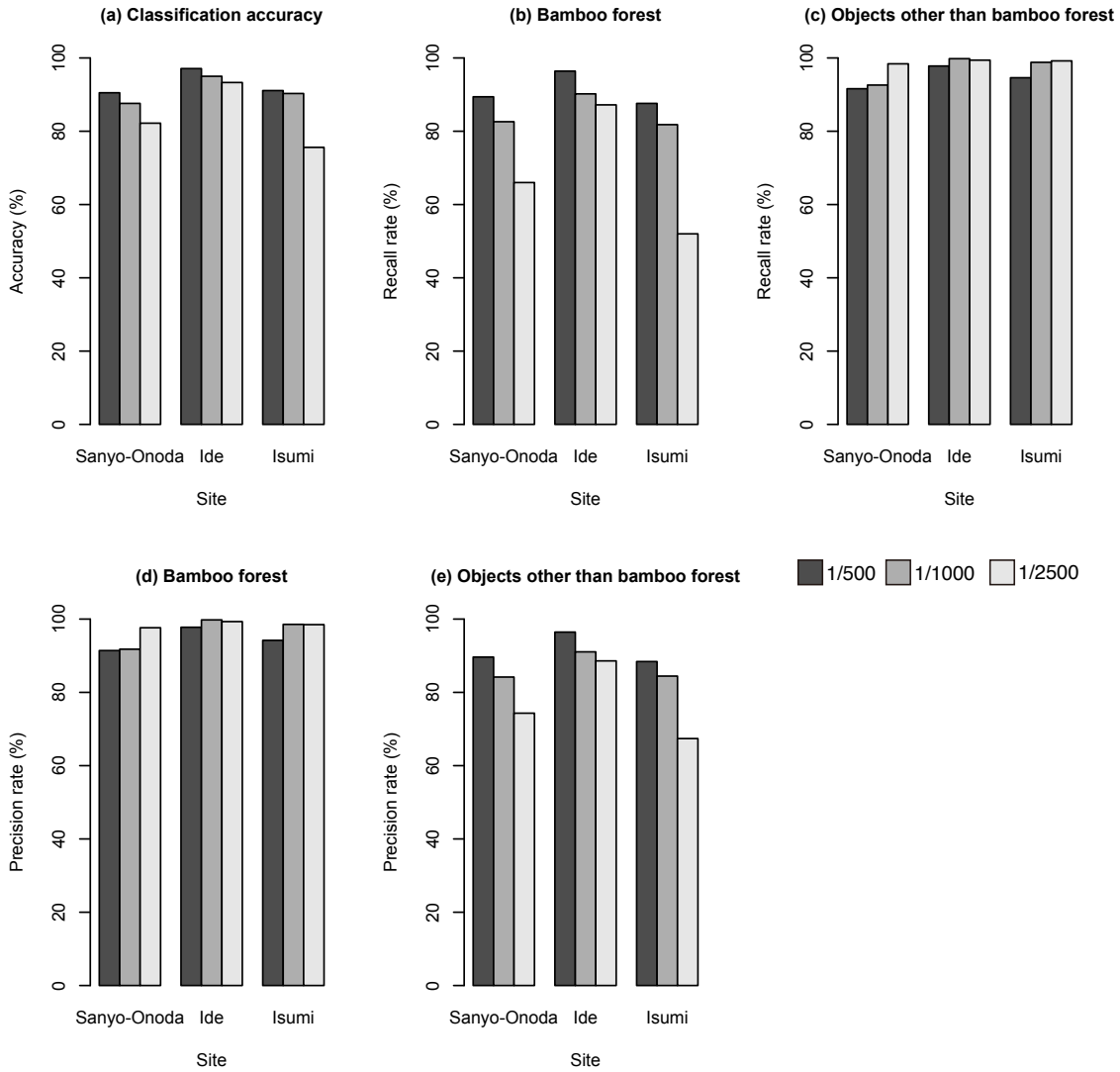

Fig.4 Sensitivity of image scale vs test accuracy.

Effects of chopping grid size on classification accuracy

The classification accuracy was ranged 85 % to 96 % (Fig.5 a). The recall rate and precision rate of bamboo forest was ranged 79 % to 99 % and 89 % to 98 %, respectively (Fig.5 b d). The recall rate and precision rate of objects other than bamboo forest was ranged 88 % to 98 % and 79 % to 99 %, respectively (Fig.5 c e). The intermediate size images (56pix) shows highest classification accuracy in all sites (Fig.5 a). The example of classification image was shown in Fig.6.

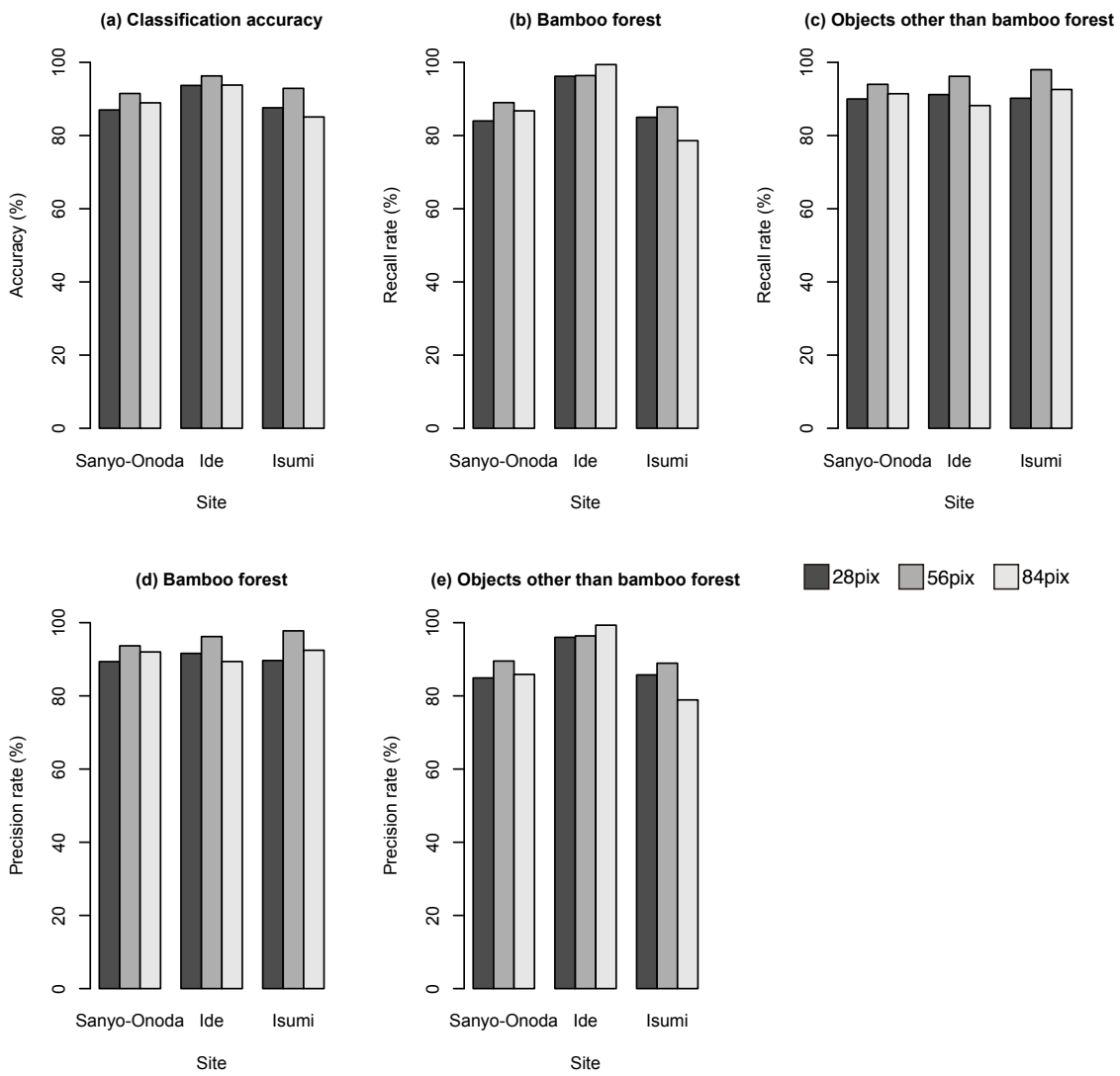

Fig.5 Sensitivity of pixel size vs test accuracy.

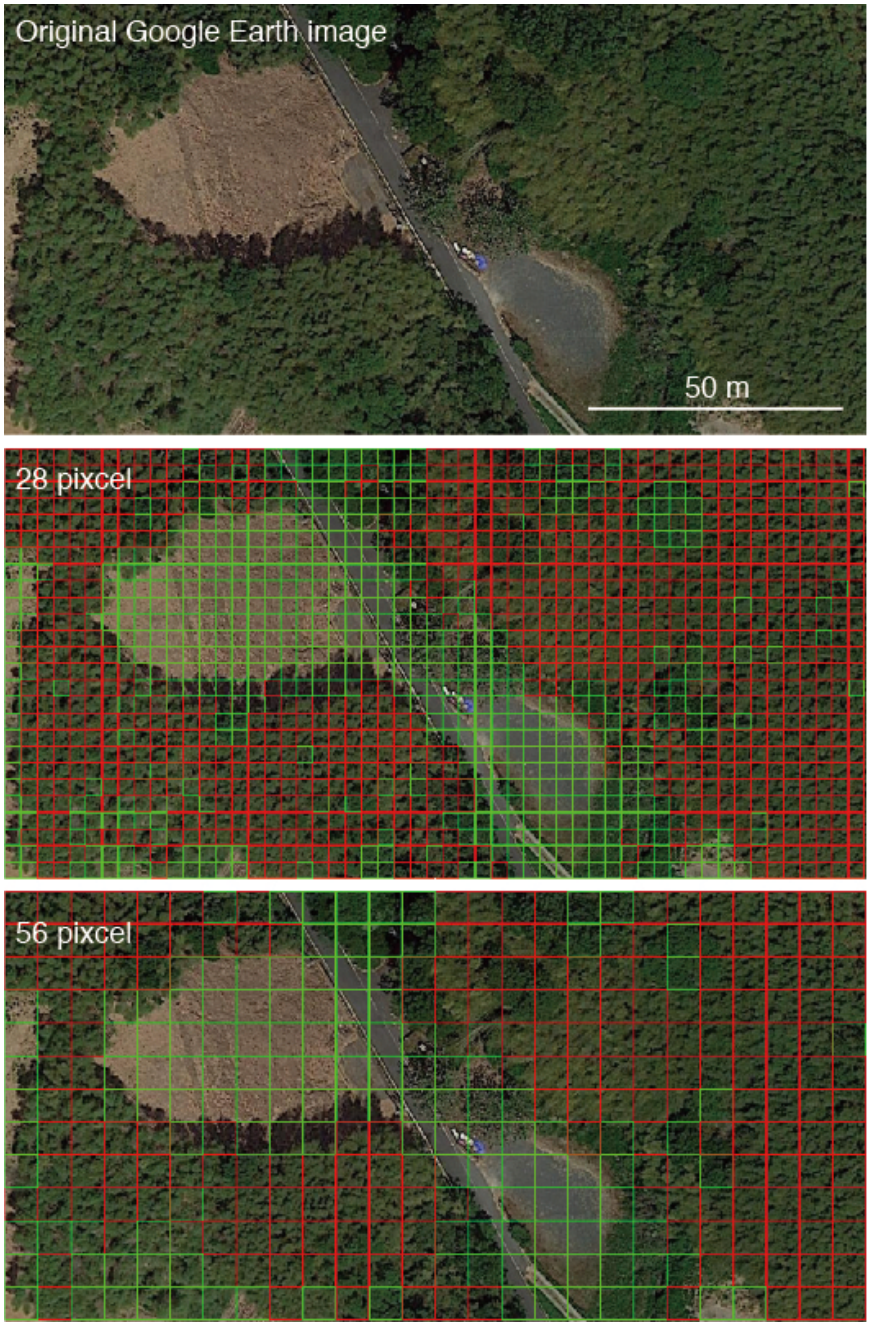

Fig.6 The example of classification image.

Bamboo forests are highlight by red and objects other than bamboo are highlighted by green.

Transferability and classification performance

In general, performance is poor when training on samples from a given city and testing on samples from a different city (Fig.7 a). When the model which trained by the images of Isumi city applied other cities, the recall rate was worst (Fig.7 b). Contrastingly, the model which trained by the images of Sanyo city shows high recall rate (Fig.7 b). We notice that a more diverse set (all) yields not better performance when applied at different locations than models trained on individual cities (Fig.7).

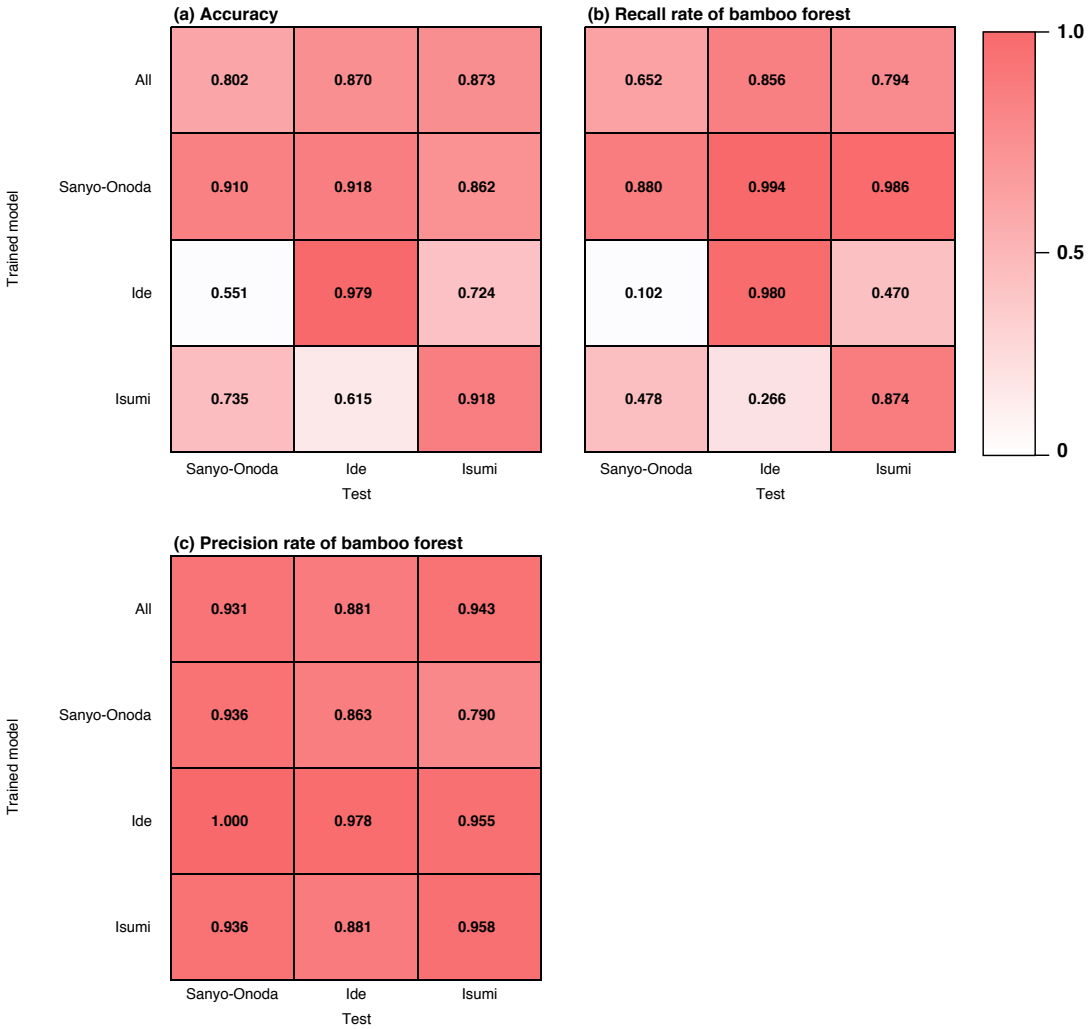

Fig.7 Transferability of models learned at one location and applied at another.

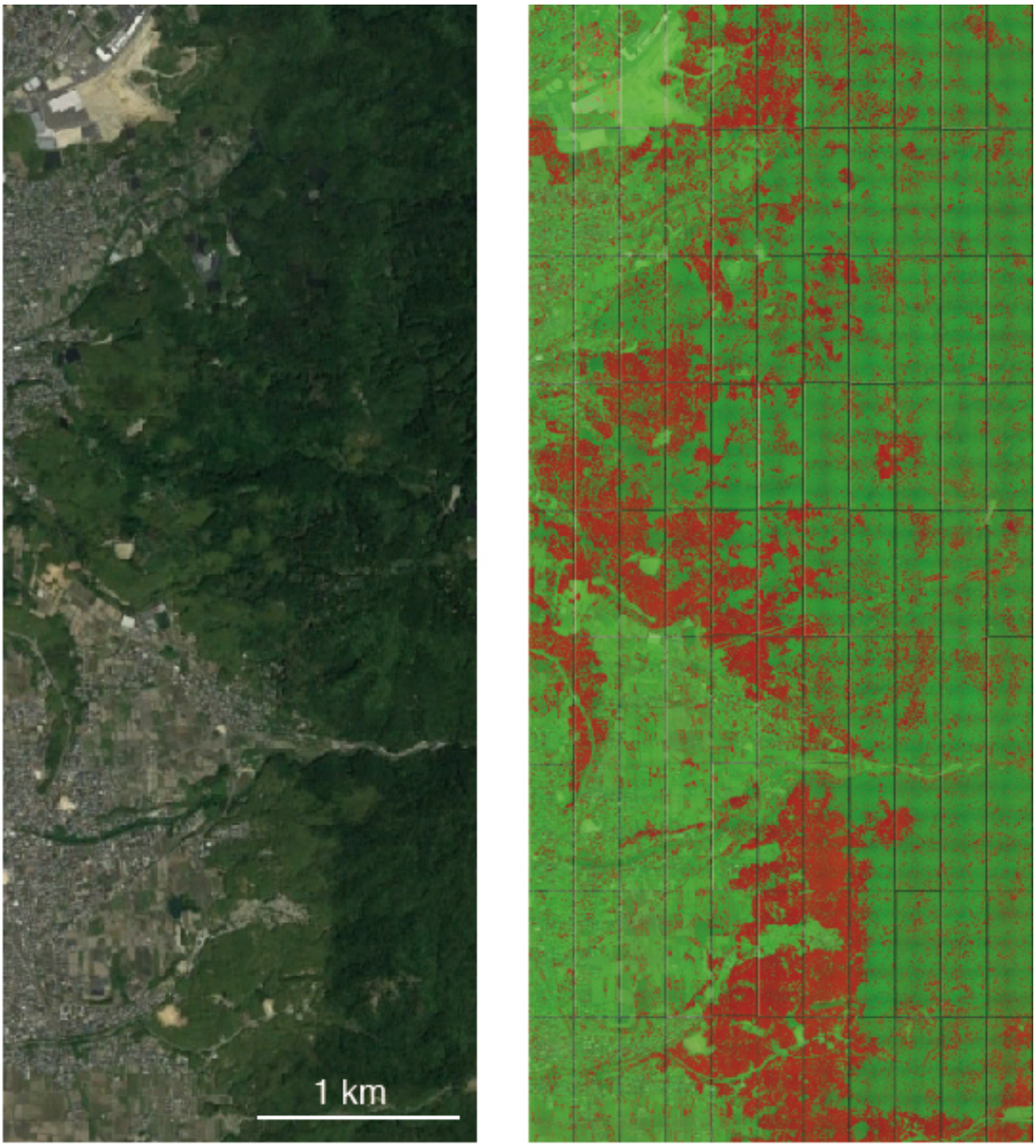

Fig.8 An example of applying the model to the wide area of Ide city.
The left image is an original google earth image and the right image is a results of bamboo forest detection. Bamboo forests are highlight by red and objects other than bamboo are highlighted by green.

Discussion

In this paper we demonstrated that the deep learning technique accurately detect bamboo forest in the Google Earth image. Although we employed most classical network (LeNet), the model can detect the bamboo forest accurately. In general, performance of model was good when training on images from a same city. So far, it is difficult to detect the ambiguous object such as vegetation but our results show good performance to detect bamboo forest from Google Earth image using chopped picture method. Our results highlight deep learning and suggests that deep learning would be a powerful method for high accuracy automated bamboo forest detection and vegetation mapping (see Fig.7).

Effects of image resolution on classification accuracy

Our result shows image resolution rate strongly affect the identification accuracy (Fig.4). As the resolution rate decreased, performance of model also declined (Fig. 4). Especially in 1/2500 image, recall rate of bamboo forest of Sanyo-Onoda and Isumi city decline to 53 % and 64 % respectively (Fig.4b). Contrastingly, precision rate of bamboo forest was increase as the the resolution rate decreased (Fig. 4d). This result means that as the resolution decreases, the model overlooks many bamboo forests and indicates that when the image resolution rate is low, it is difficult to learn the features of the object. This result also suggests that in the deep learning model, the misidentification due to false negatives was more likely occur than misidentification due to false positive as the image resolution rate decline.

Effects of chopping grid size on classification accuracy

Our result indicates that chopping grid size also affects the performance of model. Classification accuracy was highest at medium pixel size (56x56 pixels; Fig. 5a). In contrast to the effects of image resolution, recall rate and precision rate of bamboo forest was also highest at medium pixel size except recall rate at Ide city (Fig. 5 b, d). This result means that if the grid size is inappropriate, both false positives and false negatives will increase.

Increases of the chopping grid size will cause an increase in the number of chopped pictures in which objects other than bamboo and bamboo are mixed. In this

paper, as we evaluated the performance of model using the picture that is uniformly covered by bamboo forest or objects other than bamboo forest, effects of the picture that consisted with mixed objects on the classification accuracy could not evaluated. Evaluation of classification accuracy of such images is a future task.

Transferability among the models

Results of transferability test show that transferability was generally poor and suggests that the spatial extent of acquisition of training data strongly influence the classification accuracy (Fig.7). The model trained by Sanyo-Onoda city images showed high recall rate for images of any study site but the precision rate was lower than the other models (Fig.7b c). It means that the model trained by Sanyo-Onoda city images tend to occur false positive mistake. Interestingly, transferability did not relate to the distance among the study site (Fig.7). This result indicates that classification accuracy across the model reflects the conditions as local scale such as the climate at the timing when the image was taken. Additionally, even when we applied a model that learned all traning images (all), the performance of model was not as good as when traning data was obtained within the same city. The same tendencies are reported in studies that classified land use using deep learning (Albert et al., 2017). This may suggest that increasing the number of training data may also lead to a decrease in identification accuracy and it is difficult to construct an identification model applicable to a broad area.

Conclusions and future directions

Our results show deep learning model can detect bamboo forest from Google Earth images accurately. Our results also suggest that deep learning and chopped picture method would be a powerful tool for high accuracy automated vegetation mapping and may offer great potential for reducing the effort and cost for vegetation mapping as well as improving monitoring of distribution. Recently, bamboo expansion is important social problem in Japan due to its invasiveness (Okutomi et al., 1996). Some research analyzed bamboo forest distribution probability on a national scale (Someya et al., 2010; Takano et al., 2016) but monitoring of bamboo expansion still challenging

problem due to its labor. Our approach could potentially lead to the creation of a semi, or even fully automated system to monitoring of expansion.

Our result also suggest that identification accuracy depends on the image resolution rate and chopping grid size. Especially, resolution rate of training data strongly affects model performance. Generally, satellite based remote sensing has been widely studied and applied but suffers from insufficient information due to low resolution images, inaccurate information due to local weather conditions (Jones and Vaughan, 2010). Our result also shows that the performance of the model is greatly influenced by the spatial extent of acquisition of training data and the model learned on one geographical location is difficult to applied to a different geographical location. It is a future task to develop a model that can be applied to wide spatial scale.

Acknowledgements

This work was supported by JST PRESTO, Japan (Grant No. JPMJPR15O1).

References

Albert, A., Kaur, J., and Gonzalez, M. (2017). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. arXiv preprint arXiv:1704.02965.

Bengio, Y. (2009). Learning deep architectures for AI. Found. Trends Mach. Learn. 2, 1–127.

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A. (2017). Going deeper in the automated identification of Herbarium specimens. BMC Evolutionary Biology. 17:181.

Franklin, J. (2009). Mapping Species Distributions: Spatial Inference and Prediction. New York, NY: Cambridge University Press.

Goodfellow, I.; Courville, A.; Bengio, Y. Deep Learning. Book in Preparation for MIT Press; The MIT Press: Cambridge, MA, USA, 2016.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Schölkopf, B. (1998). Support vector machines. IEEE Intell. Syst. Appl. 13, 18–28. doi: 10.1109/5254. 708428

Ise, T., Minagawa, M., and Onishi, M. (2018) Classifying 3 Moss Species by Deep Learning, Using the "Chopped Picture" Method. Open Journal of Ecology. 8, 166–173.

Jones, H. G., and R. A. Vaughan. (2010). Remote sensing of vegetation: principles techniques and applications. Oxford University Press, Oxford, UK.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). "Large-scale video classification with convolutional neural networks," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (Columbus), 1725–1732.

LeCun, Y. L., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. doi: 10.1109/5.726791

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. Front. Plant Sci. 7:1419. doi: 10.3389/fpls.2016.01419

Nakashima, A. (2001). The present situation of the bamboo forests management in the suburbs: A case study of the bamboo shoot producing districts in the suburbs of Kyoto City. Applied Forest Science, 10, 1–7. (in Japanese with English abstract).

Nishikawa, R., Murakami, T., Yoshida, S., Mitsuda, Y., Nagashima, K., and Mizoue, N. (2005). Characteristic of temporal range shifts of bamboo stands according to adjacent landcover type. Journal of the Japanese Forestry Society, 87, 402–409. (in Japanese with English abstract).

Nogueira, K., Penatti, O.A.B., and Dos Santos, J.A. (2017). Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification. Pattern Recognit. 61, 539–556.

NVIDIA, (2016). NVIDIA deep learning gpu training system. https://developer.nvidia.com/digits. Accessed: 2018-06-01.

Okutomi, K., Shinoda, S., and Fukuda, H. (1996). Causal analysis of the invasion of broad-leaved forest by bamboo in Japan. Journal of Vegetation Science, 7, 723–728.

Ramcharan, A., Baranowski, K., McCloskey, P., Ahamed, B., Legg, J., Hughes, D. (2017). Deep learning for image-based cassava disease detection. Front. Plant Sci. 8, 1852.

Someya, T., Takemura, S., Miyamoto, S., and Kamada, M. (2010). Predictionsof bamboo forest distribution and associated environmental factors using natural environmental information GIS and digital national land information in Japan. Keikanseitaigaku [Landscape Ecology], 15, 41–54. (in Japanese with English abstract).

Suzuki, S. (2015). Chronological location analyses of giant bamboo (Phyllostachys pubescens) groves and their invasive expansion in a satoyama landscape area, western Japan. Plant Species Biology, 30, 63–71.

Takano KT, Hibino K, Numata A, Oguro, M., Aiba, M., Shiogama, H., Takayasu, I., and Nakashizuka, T. (2017). Detecting latitudinal and altitudinal expansion of invasive bamboo *Phyllostachys edulis* and *Phyllostachys bambusoides* (Poaceae) in Japan to project potential habitats under 1.5°C–4.0°C global warming. Ecol Evol, 7, 9848–9859.

Xie, Y., Sha, Z., and Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. J. Plant Ecol. 1, 9–23. doi: 10.1093/jpe/rtm005

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). “How transferable are features in deep neural networks?,” in Advances in Neural Information Processing Systems (Montreal), 3320–3328.

Yue, J., Zhao, W., Mao, S., and Liu, H. (2015) Spectral—Spatial classification o hyperspectral images using deep convolutional neural networks. Remote Sens. Lett. 6, 468–477.