

RESEARCH ARTICLE

Open Access



Identifying the vegetation type in Google Earth images using a convolutional neural network: a case study for Japanese bamboo forests

Shuntaro Watanabe^{1,3*} , Kazuaki Sumi² and Takeshi Ise¹

Abstract

Background: Classifying and mapping vegetation are crucial tasks in environmental science and natural resource management. However, these tasks are difficult because conventional methods such as field surveys are highly labor-intensive. Identification of target objects from visual data using computer techniques is one of the most promising techniques to reduce the costs and labor for vegetation mapping. Although deep learning and convolutional neural networks (CNNs) have become a new solution for image recognition and classification recently, in general, detection of ambiguous objects such as vegetation is still difficult. In this study, we investigated the effectiveness of adopting the chopped picture method, a recently described protocol for CNNs, and evaluated the efficiency of CNN for plant community detection from Google Earth images.

Results: We selected bamboo forests as the target and obtained Google Earth images from three regions in Japan. By applying CNN, the best trained model correctly detected over 90% of the targets. Our results showed that the identification accuracy of CNN is higher than that of conventional machine learning methods.

Conclusions: Our results demonstrated that CNN and the chopped picture method are potentially powerful tools for high-accuracy automated detection and mapping of vegetation.

Keywords: Convolutional neural network, Vegetation mapping, Google earth imagery

Background

Classifying and mapping vegetation are essential tasks in environmental science research and natural resource management [1]. Conventional methods (e.g., field surveys, manual interpretation of aerial photographs), however, are not effective for acquiring vegetation data because they are labor-intensive and often economically expensive. Remote sensing technology offers a practical and economical means to acquire information on

vegetation cover, especially over large areas [2]. Because of its ability to perform systematic observations at various scales, remote sensing can potentially enable classification and mapping of vegetation at high temporal resolutions.

Detection of discriminating visual features is one of the most important steps in almost all computer vision problems, including in the field of remote sensing. Because conventional methods such as support vector machines [3] require hand-designed, time-consuming feature extraction, substantial efforts have been dedicated toward the development of methods for the automatic extraction of features. Recently, deep learning has become a new solution for image recognition and classification because

*Correspondence: watanabe@sci.kagoshima-u.ac.jp

¹ Field Science Education and Research Center (FSERC), Kyoto University, Kitashirakawa-iwaki-cho, Sakyo-ku, Kyoto 606-8502, Japan

Full list of author information is available at the end of the article



this new method does not require the manual extraction of features.

Deep learning [4, 5] is a type of machine learning technique that uses algorithms inspired by the structure and function of the human brain, called artificial neural networks. Deep learning involves the learning of features and classifiers simultaneously, and uses training data to categorize image content without a priori specification of image features. Among all deep learning-based networks, the convolutional neural network (CNN) is the most popular for learning visual features in computer vision applications, including remote sensing. Recent research has shown that CNN is effective for diverse applications [4–7]. Given its success, CNN has been used intensively in several distinct tasks in various academic and industrial fields, including plant science. Recent research has shown that the CNN can successfully detect plant diseases and accurately classify plant specimens in an herbarium [8–10].

CNN is a promising technology in the field of remote sensing as well [11, 12]. In recent years, CNNs have started to be used for scene tagging and object detection in remote sensing images. Most previous research employed the supervised learning method and suggested that CNNs can accurately detect or classify the objects in remote sensing images. However, the number of studies that use CNNs for detecting or classifying vegetation in remote sensing images is still limited. Li et al. [13] successfully detected oil palm trees with an accuracy exceeding 96% using CNNs. Recently, Guirado et al. [14] used CNNs to detect a wild shrub (*Ziziphus lotus*) that has a wide range of shapes and sizes, from Google Earth images. They demonstrated that CNNs can successfully detect *Ziziphus lotus* and provide better results than conventional object detection methods. However, the detection targets in the existing works are vegetation distributed in a patch form; therefore, detection or classification of vegetation that has ambiguous and amorphous shapes, such as clonal plants, is still challenging. Semantic segmentation [15] is a possible solution to address this challenge; nevertheless, the application of semantic segmentation for vegetation classification is still limited because this approach requires enormous amounts of labeled data based on pixelwise reference maps [16].

Recently, Ise et al. [17] developed an alternative method (chopped picture method) to conveniently classify ambiguous and amorphous objects. This method dissects the images into numerous small squares and efficiently produces the training images. By using this method, Ise et al. [17] correctly classified three moss species and “non-moss” objects in test images with an accuracy of 95%. However, this method has been applied only to high-resolution images, and its applicability to

low-resolution images, such as remote sensing images, has not yet been investigated.

In this study, we investigated the efficiency of adopting a deep learning model and the chopped picture method for computer-based vegetation detection from Google Earth images. We used bamboo forests as the target for vegetation detection. In recent years, bamboo has become invasive in Japan. The bamboo species moso (*Phyllostachys edulis*) and madake (*Phyllostachys reticulata*) are the two main types of exotic bamboo. Since the 1970s, the bamboo industry in Japan has declined as a result of cheaper bamboo imports and heavy labor costs [18]. Consequently, many bamboo plantations were left unmanaged, eventually leading to the invasion of adjacent native vegetation [19–21].

In this study, we specifically addressed the following research questions: (1) how does the resolution of images affect the accuracy of detection; (2) how does the chopping size of training images affect the accuracy of detection; and (3) can a model that was trained for one geographical location work well for a different location?

Materials and methods

Target area and image acquisition

In this study, we chose three regions (Sanyo-Onoda, Ide, and Isumi) in Japan to perform the analyses (Fig. 1). We used Google Earth as the source of imagery. Google Earth images have various data sources, ranging from medium-resolution Landsat images to high-resolution QuickBird satellite images. Herein, we used high-resolution QuickBird satellite images (spatial resolution: 0.65 m/pixel). From a given sampling location, we obtained the images at zoom levels of 1/500 (spatial resolution: ~0.13 m/pixel), 1/1000 (spatial resolution: ~0.26 m/pixel), and 1/2500 (spatial resolution: ~0.65 m/pixel). Each study site was imaged from Google Earth images during October 2014 (Sanyo-Onoda), May 2017 (Ide), and January 2017 (Isumi).

Methods and background concepts for the neural networks

In this study, we employed CNNs (Fig. 2), which are a special type of feedforward neural networks that consist of several convolutional layers and pooling layers.

A feedforward neural network is an artificial neural network wherein connections between the nodes do not form a cycle. These networks, which perform modeling similar to the neuron activity in the brain, are generally presented as systems of interconnected processing units (artificial neurons) that can compute values from inputs, resulting in an output that may be used on further units. Artificial neurons are basically processing units that compute some operations over several input variables and

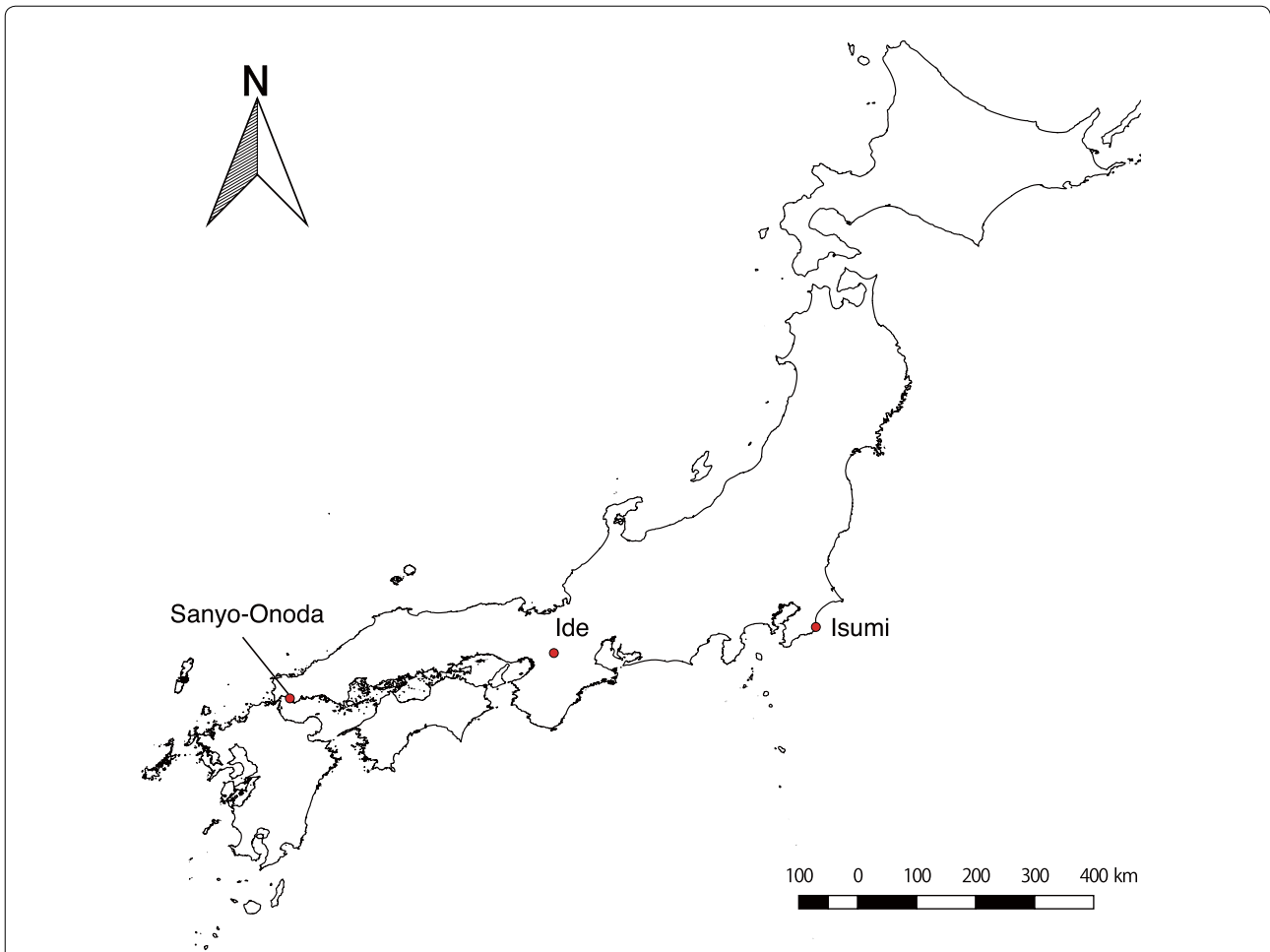


Fig. 1 Target regions of this research. This figure was generated using data Global Map Japan version 2.2 (Geospatial Information Authority of Japan)

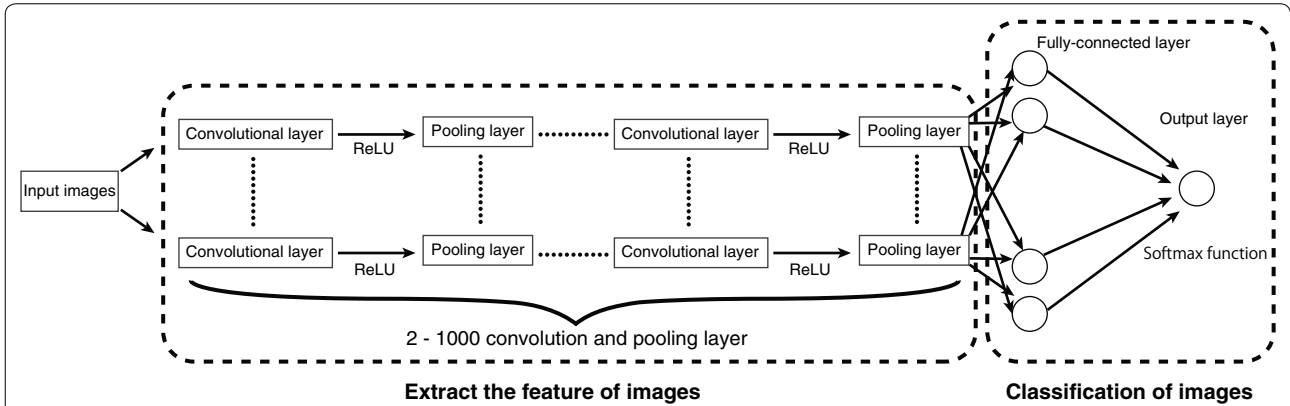


Fig. 2 Schematic of the convolutional neural networks

usually have one output calculated through the activation function. Typically, an artificial neuron has a weight w_i that represents the degree of connection between

artificial neurons, some input variables x_i , and a threshold vector b . Mathematically, the total input and output of artificial neurons can be described as follows:

$$u = \sum_i w_i x_i \quad (1)$$

$$z = f(u + b) = f\left(\sum_i w_i x_i + b\right) \quad (2)$$

where u , z , x , w , and b represent the total input, output, input variables, weights, and bias, respectively. $f(\bullet)$ denotes an activation function: a nonlinear function such as a sigmoid, hyperbolic, or rectified linear function. We employed a rectified linear function as the activation function, and this function is referred to as the Rectified Linear Unit (ReLU). ReLU can be defined as follows:

$$f(u) = \max\{0, u\} = \begin{cases} u (u > 0) \\ 0 (u \leq 0) \end{cases} \quad (3)$$

A CNN consists of a convolutional layer and a pooling layer. The convolutional layer plays a role in capturing the features from the images. In this process, a fixed-size filter runs over the images and extracts the patterns of shades of colors in the images. After each convolutional layer, there are pooling layers that are created to reduce the variance of features; this is accomplished by computing some operations of a particular feature over a region of the image.

The pooling layer has two functions. The first function is to reduce the position sensitivity of the feature that is extracted at the convolution layer so that the output amount of the pooling layer does not change even when the position of the feature amount extracted by the convolution layer is shifted within the image. The second function is to enlarge the receptive field for the following convolutional layers. Two operations may be realized on the pooling layers, namely, max and average operations, in which the maximum and mean values are selected over the feature region, respectively. This process ensures that the same results can be obtained even when image features have small translations or rotations, and this is crucial for object classification and detection. Thus, the pooling layer is responsible for sampling the output of the convolutional layer and preserving the spatial location of the image, as well as selecting the most useful features for the next layers. After several convolutional and pooling layers, there are fully connected layers, which take all neurons in the previous layer and connect them to every single neuron in its layer.

Finally, following all the convolution, pooling, and fully connected layers, a classifier layer may be used to calculate the class probability of each image. We employed the softmax function in this layer. The softmax function calculates the probabilities of each target

class over all possible target classes and is written as follows:

$$y_k = \text{softmax}_k(u_1, u_2, \dots, u_K) = \frac{e^{u_k}}{\sum_{j=1}^K e^{u_j}} \quad (4)$$

where k represents the number of output units and u represents input variables.

To evaluate the performance of the network, a loss function needs to be defined. The loss function evaluates the effectiveness of the network in modeling the training dataset. The objective of training is to minimize the error of the loss function. The cross entropy of the softmax function is defined as follows:

$$E = - \sum_{n=1}^N \sum_{k=1}^K t_{n,k} \log y_k \quad (5)$$

where t denotes the vector for the training data, K represents the possible class, and N represents the total number of instances.

Approach

A schematic of our approach is shown in Fig. 3. We prepared the training data by using the chopped picture method [17]. First, in this method, we collected images that were (a) nearly 100% covered by bamboo and (b) not covered by bamboo. Next, we ‘‘chopped’’ these images into small squares with 50% overlap both vertically and horizontally. Finally, we used the chopped images as training images. Details about the size and number of training images for each study site are presented in Table 1.

We created a model for image classification from a CNN for bamboo forest detection. As opposed to the conventional approaches for training classifiers with hand-designed feature extraction, the CNN learns the feature hierarchy from pixels to classifiers and trains layers jointly. We used the final layer of the CNN model to detect the bamboo coverage from Google Earth images. First, we randomly shuffled all images to avoid overlapping of the training data and validation data. Then, we used 75% of the obtained images as training data and the remaining 25% as validation data.

We used the LeNet network [22], which is a classical deep learning model, because this study uses small-size images as the training and validation data. This network is constructed with two convolutional layers, two pooling layers, and a fully connected layer. The network architecture is discussed in Table 2. The model parameters implemented in this study included the number of training epochs (30), learning rate (0.01), train batch size (64), and validation batch size (32).

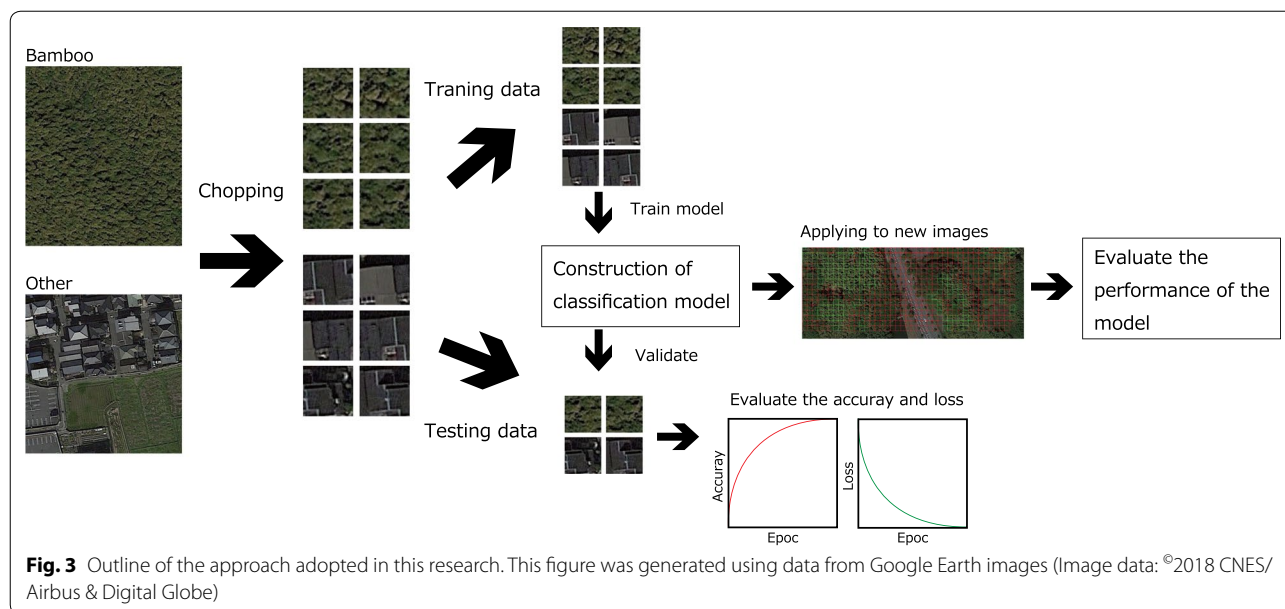


Table 1 Size and number of training images used in each experiment

Zoom level	1/500			1/1000	1/2500
	28	56	84	28	14
Sanyo-Onoda	67398	16207	6866	16000	16000
Ide	79043	18653	2836	18000	18000
Isumi	64106	15083	2837	15000	15000

Evaluation of learning accuracy

The model was validated in each learning epoch by using the accuracy and loss functions. “Accuracy” indicates the accuracy of the model in classifying the validation images, whereas “loss” represents the inaccuracy of prediction by the model. If model learning is successful, loss (val) is low and accuracy is high. However, if loss (val) becomes high during learning, this indicates over fitting.

Evaluation of model performance

To evaluate the performance of the model, we used the confusion matrix. First, we obtained 20 new images, which were uniformly covered by bamboo forest or objects other than bamboo forest, from each study site. In addition, 20 new images were sampled from multiple locations that were at least 5 km from the locations where the training images for each study site were sampled from. Second, we re-sized the images by using the chopped picture method without overlap. Third, we randomly sampled 500 images from the re-sized images. Fourth, we applied the model to the sampled images and

Table 2 Architecture of the network used in this study

Layer name	Filter size	Number of output neurons
Convolution 1	5 × 5	20
Pooling 1	2 × 2	
Convolution 2	5 × 5	50
Pooling 2	2 × 2	
Fully connected 1		500
Fully connected 2		2

evaluated the classification accuracy. Finally, we categorized the classification results into four groups: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Next, we calculated the classification accuracy, recall rate, and precision rate by using the following equations:

$$\text{Classification accuracy} = (TP + TN) / (TP + TN + FP + FN) \tag{6}$$

$$\text{Recall rate} = TP / (TP + FN) \tag{7}$$

$$\text{Precision rate} = TP / (TP + FP) \tag{8}$$

Influence of image resolution on the classification accuracy

To quantify the influence of image resolution on the accuracy of detection, we constructed a model that correspond to image resolution. We obtained images at a

zoom level of 1/500 (spatial resolution: ~ 0.13 m/pixel), 1/1000 (spatial resolution: ~ 0.26 m/pixel), and 1/2500 (spatial resolution: ~ 0.65 m/pixel) from each study site. Next, we applied the chopped picture method. To adjust the spatial extent of each chopped image, we chopped 56, 28, and 14 pixels for the 1/500, 1/1000, and 1/2500 levels, respectively. After constructing the model, we applied it to the 500 new images and calculated the classification accuracy, recall rate, and precision rate.

Influence of chopping grid size on the classification accuracy

To quantify the influence of spatial extent of the chopping grid on the accuracy of detection, we chopped the 1/500 resolution images (spatial resolution: ~ 0.13 m/pixel) at each study site for three types of pixel sizes (84, 56, 28). After constructing the model, we applied it to new images and calculated the classification accuracy, recall rate, and precision rate. Note that our method does not perform dense prediction; therefore, the resolution of the final image will be lower than that of the original image. We considered this to be acceptable for the purpose of this study, because a previous study [23] has reported that more than 90% of the bamboo forests are distributed as patches of size greater than 5000 m² in Japan.

Transferability test

Given the large variations in the visual appearance of bamboo forest across different cities, it is of interest to study to what extent a model trained for one geographical location can be applied to a different geographical location. As such, we conducted experiments in which we trained a model for one (or more) city and applied it to a different set of cities. To test the transferability, we used 10 new 1/500 resolution images, and evaluated the performance of the model based on the classification accuracy, recall rate, and precision rate.

Comparison with existing machine learning method

We compared the classification performance of our method with that of support vector machine (SVM), a supervised nonparametric classification algorithm. The SVM classification algorithm is commonly employed in remote sensing for a range of applications [24]. The SVM classifier tries to find the optimal hyperplane in n -dimensional classification space with the highest margin between classes.

In this experiment, we provided 56 pixel-chopped 1/500 resolution images for each study site as training images. Then, we applied the model to newly sampled 500 images and evaluated the classification accuracy, recall rate, and precision rate. We used RGB value and

object-specific texture measures based on grey-level co-occurrence matrix (GLCM) as an image feature for SVM. To calculate the GLCM texture, we first calculated the luminance value of the image, which is accomplished by defining the weights for R, G, and B. Then, five GLCM texture measures of mean, variance, contrast, homogeneity, and dissimilarity were calculated. We employed RBF kernels for SVM classification and determined the parameters (gamma and cost) by a grid search method.

Robustness assessment

We evaluated the classification performance of our method using different amounts of training data. In this experiment, we provided 15000, 1000, and 100 pixel-chopped images for each study site as training images. Then, we evaluated the classification accuracy, recall rate, and precision rate using 500 images that were randomly sampled from the images used in model performance evaluation.

Results

Fluctuation of accuracy and loss during the learning epochs

The accuracy for classifying the validation data of the final layer ranged from 94 to 99%, with an average of 97.52%. The loss values for the validation data ranged from 0.008 to 0.214, with an average of 0.086 (Fig. 4). The values of accuracy increased and those of loss decreased following the learning epochs (Fig. 4). These results suggest that all the models were not overfitted to the datasets and successfully learned the features of chopped pictures.

Influence of image resolution on the classification accuracy

The classification accuracy ranged from 74 to 93% (Fig. 5a). The recall rate and precision rate for bamboo forest ranged 49% to 90% and 94% to 99%, respectively (Fig. 5b, d), and those for objects other than bamboo forest ranged from 94 to 99% and 66% to 91%, respectively (Fig. 5c, e). The classification accuracy and recall rate for bamboo forest declined following the decline in the image resolution (Fig. 5a, b).

Influence of chopping grid size on the classification accuracy

The classification accuracy ranged from 82 to 93% (Fig. 6a). The recall rate and precision rate for bamboo forest ranged from 72 to 90% and 89% to 99%, respectively (Fig. 6b, d), and those for objects other than bamboo forest ranged from 89 to 99% and 75% to 91%, respectively (Fig. 6c, e). The intermediate-size images (56 pixels) exhibited the highest classification accuracy for all sites (Fig. 6a).

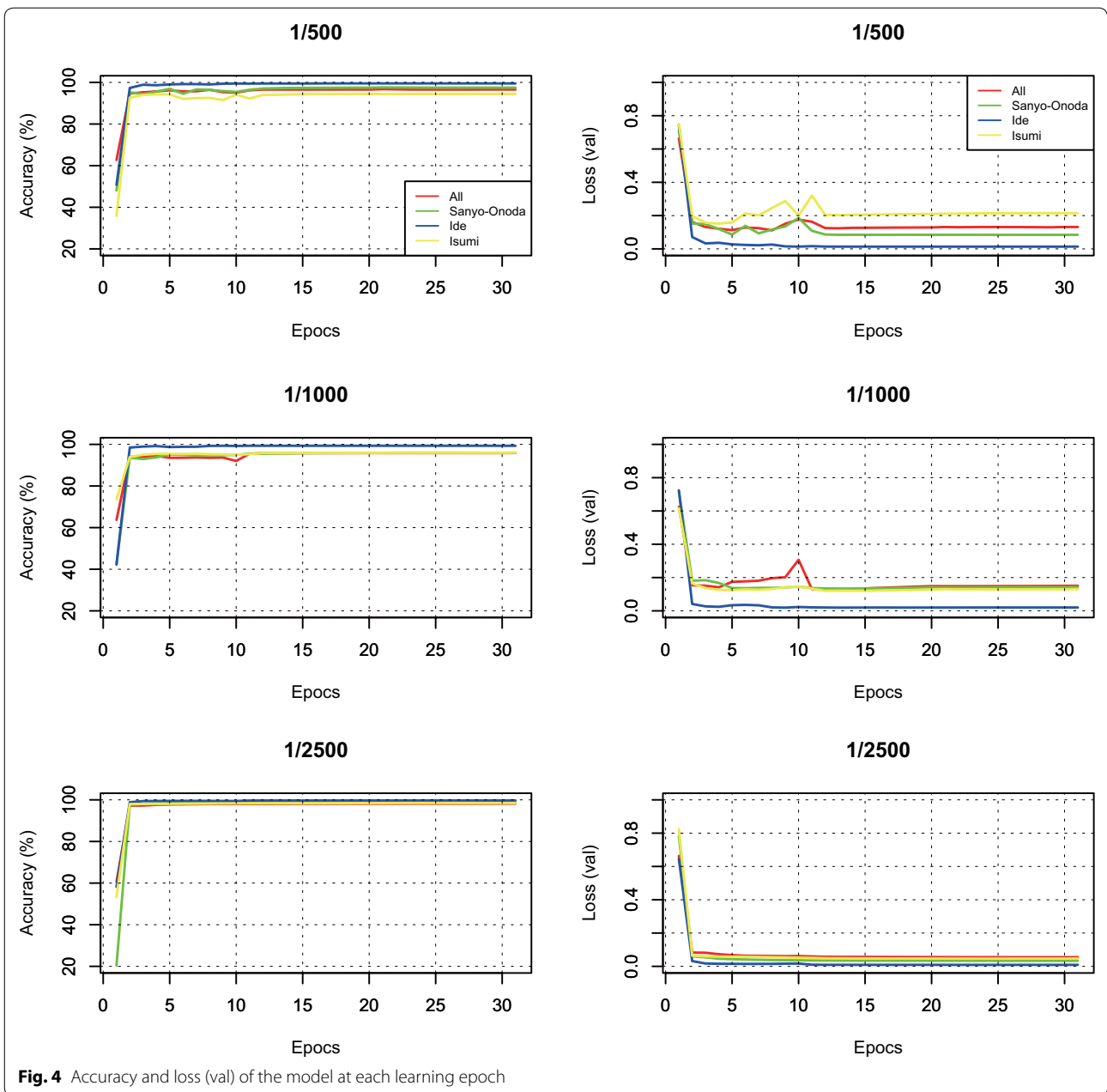


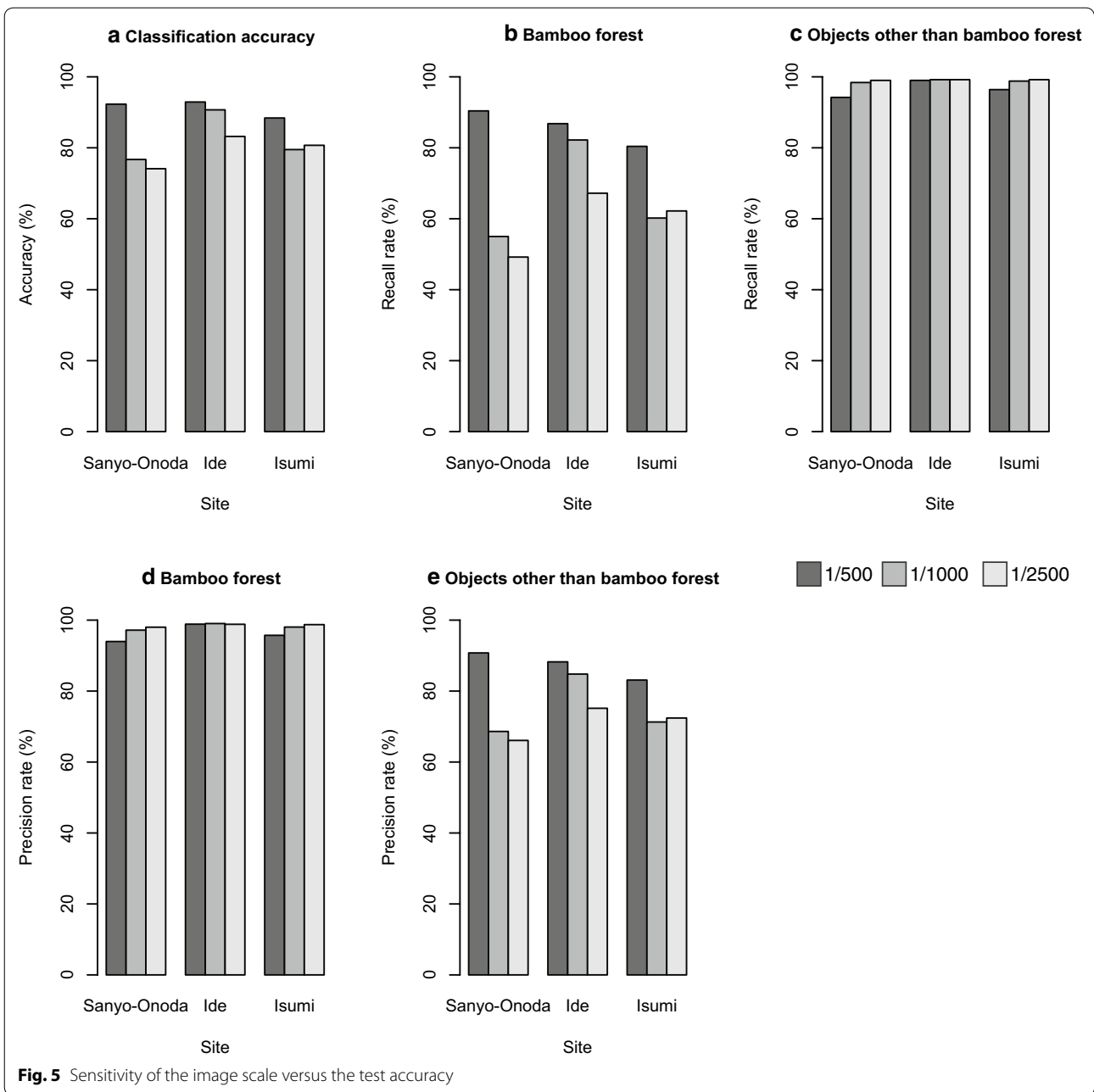
Fig. 4 Accuracy and loss (val) of the model at each learning epoch

Comparison with SVM

Our results show that the CNN classifier outperformed the SVM classifier (Fig. 7). The classification accuracy of CNN ranged from 86 to 88%, whereas that of SVM ranged from 74 to 85% (Fig. 7). The recall rate of CNN ranged from 79 to 92%, whereas that of SVM ranged from 71 to 91% (Fig. 7). The precision rate of CNN ranged from 82 to 97%, and that of SVM ranged from 77 to 93% (Fig. 7).

Robustness assessment

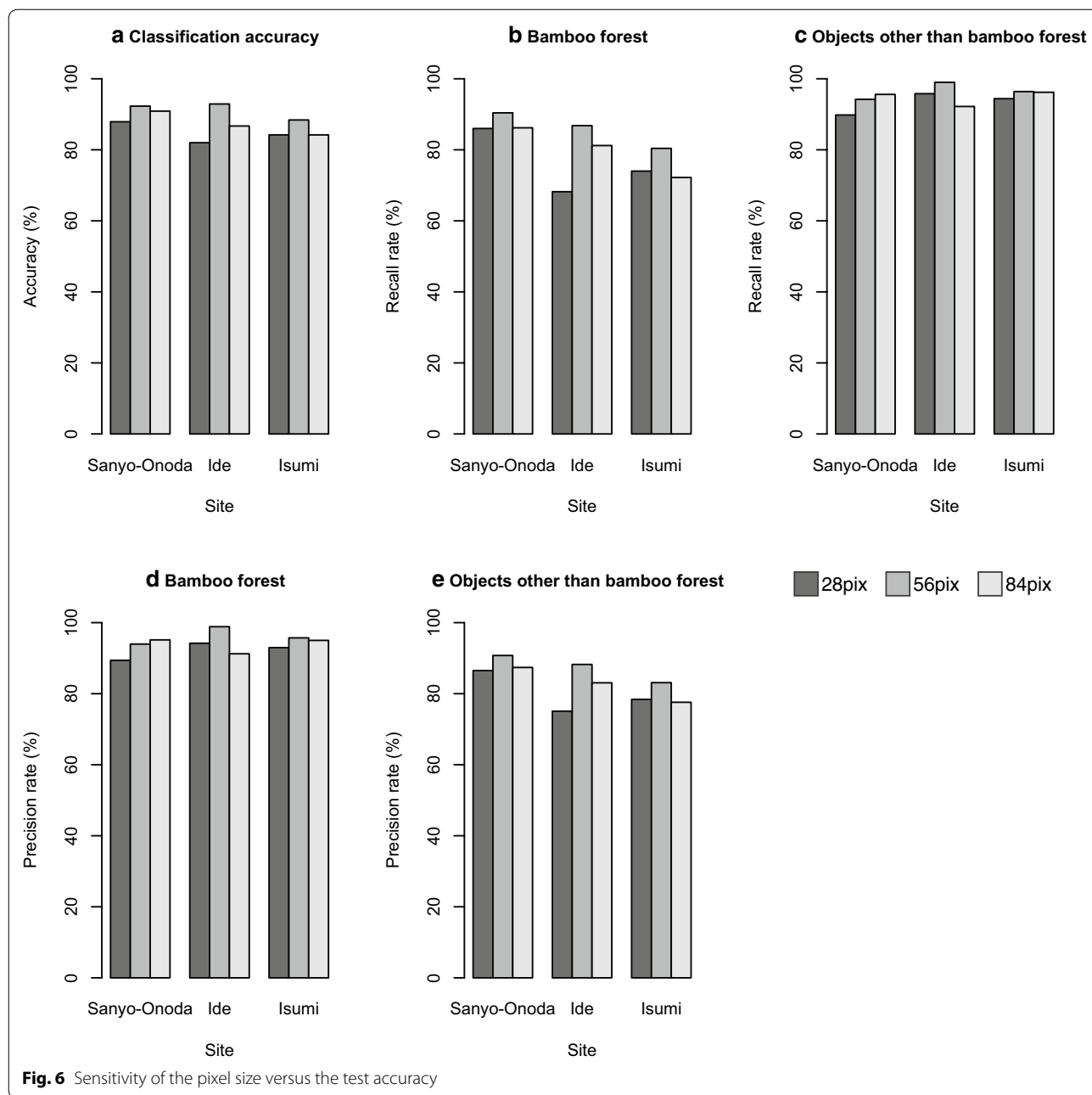
The classification accuracy of the model trained with 15000 images ranged from 88.4% to 92.9% (Fig. 8). The recall rate and precision rate for bamboo forest ranged from 80.4% to 90.4% and 93.9% to 98.8%, respectively (Fig. 8). The classification accuracy of the model trained with 1000 images ranged from 86 to 88% (Fig. 8). In this case, the recall rate and precision rate for bamboo forest ranged from 79% to 92.4% and 82.6% to 97.0%,



respectively (Fig. 8). The classification accuracy of the model trained with 100 images ranged from 76.6% to 82.4% (Fig. 8), in which case the recall rate and precision rate for bamboo forest ranged from 71.2% to 88.8% and 76.4% to 90.1%, respectively (Fig. 8). Note that even when the number of training images was decreased to 100, a classification accuracy of greater than 75% was maintained.

Transferability and classification performance

In general, the performance of the model was poor when it was trained with samples from a given city and tested with samples from a different city (Fig. 9a). When the model trained with the images of Isumi city was applied to the other cities, the recall rate was the worst (Fig. 9b). Conversely, the model trained with the images of Sanyo city had the highest recall rate (Fig. 9b). We noticed that

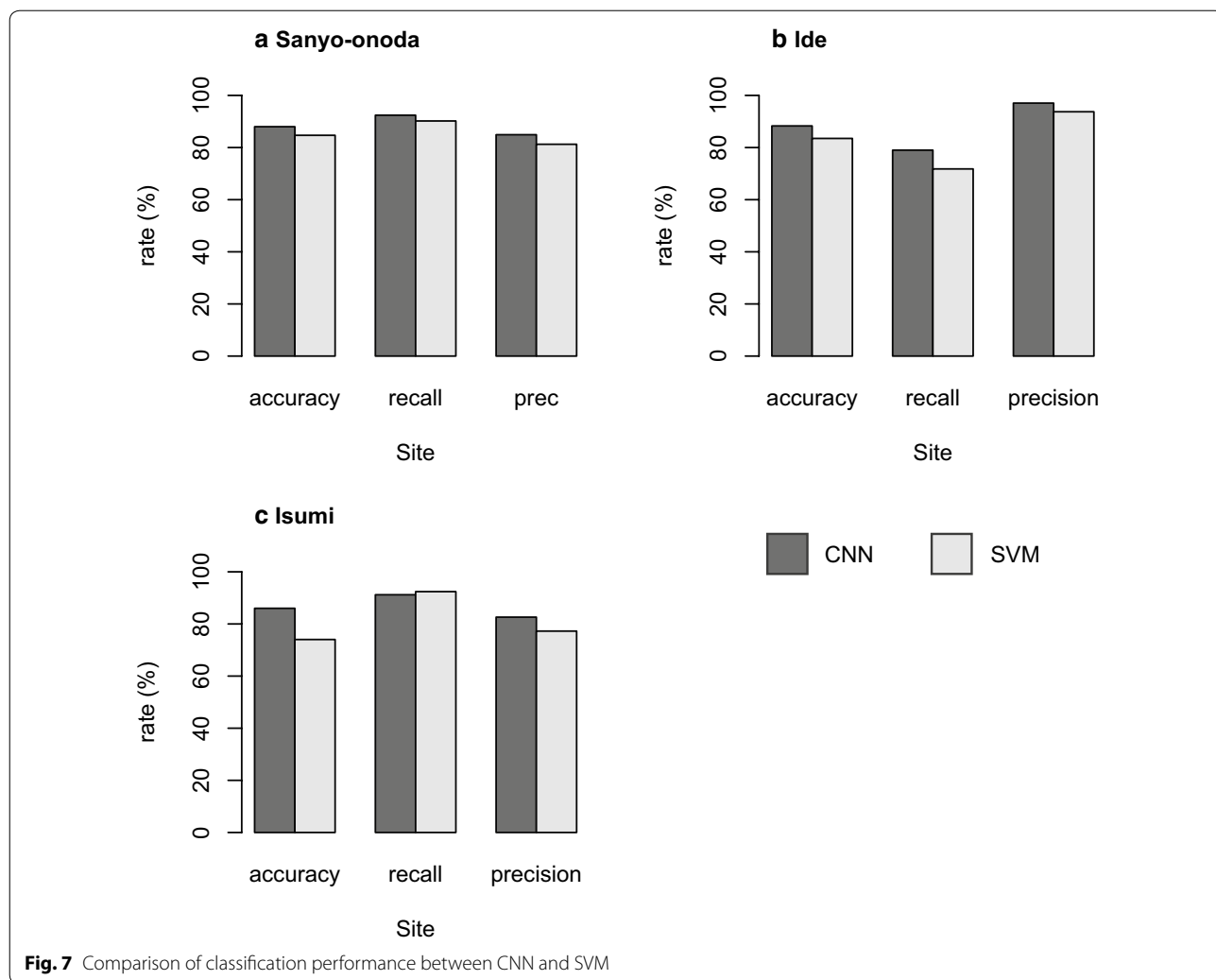


a more diverse set (all) did not yield a better performance when applied at different locations, compared with the models trained on individual cities (Fig. 9).

Discussion

In this paper, we demonstrated that the chopped picture method and CNN could accurately detect bamboo forest in Google Earth imagery. Recent studies [13, 14] have shown that the use of CNN for processing of color (RGB) images of the Earth surface yield high-accuracy results in

the recognition of different tree species. However, most of the existing studies are aimed at identifying tree species at the individual tree level and cannot be applied to the detection of vegetation with ambiguous and amorphous shapes such as clonal plants. In this study, we employed the chopped picture method to detect bamboo forest from Google Earth images, and our results demonstrated good performance even though we employed the most classical CNN (LeNet). Additionally, our method outperformed SVM (Fig. 7), suggesting that the chopped



picture method and CNN would be powerful methods for high-accuracy image-based bamboo forest detection and vegetation mapping (Fig. 10).

Influence of image resolution on the classification accuracy

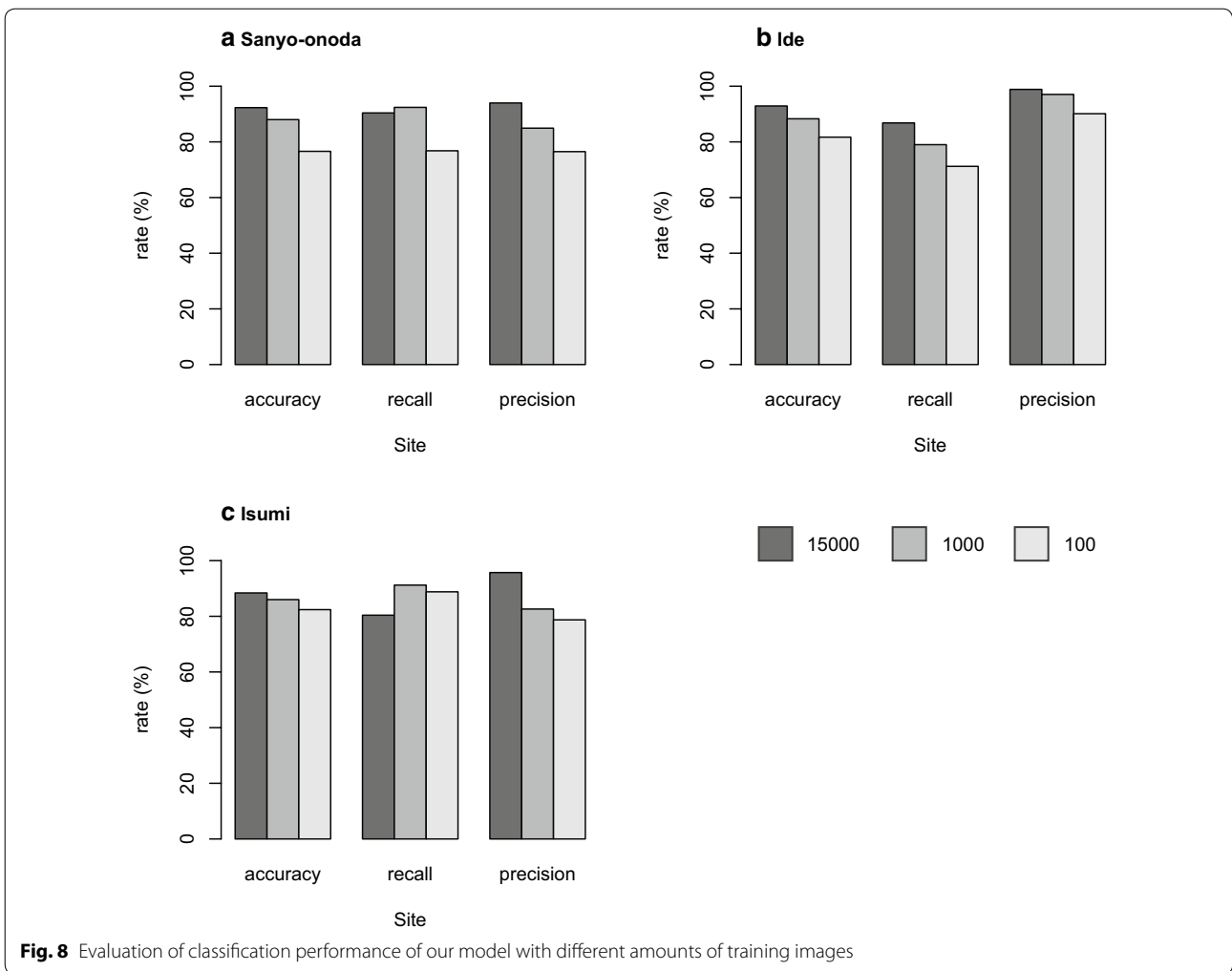
Our results indicate that the image resolution strongly affects the identification accuracy (Fig. 5). As the resolution rate decreased, the performance of the model also declined.

Specifically, in the 1/2500 imagery, the recall rate for bamboo forest of Sanyo-Onoda and Isumi city declined to 59% and 68%, respectively (Fig. 5b). In contrast, the precision rate for bamboo forest increased as the resolution decreased (Fig. 5d); this result indicates that as the resolution decreases, the model overlooks many bamboo forests. Thus, when the image resolution is low,

it is difficult to learn the features of the object, and our approach is critically dependent on high-resolution data, with even the 0.65 m spatial resolution data being insufficient for producing the high-quality results. This result also suggests that in the deep learning model, misidentification due to false negatives is more likely to occur than misidentification due to false positives as the image resolution decreases.

Influence of chopping grid size on classification accuracy

Our results indicate that the chopping grid size also affects the performance of the model. The classification accuracy was the highest at the medium pixel size (56 × 56 pixels; Fig. 6a). In contrast to the effects of image resolution, the recall rate and precision rate for bamboo forest was also the highest at medium pixel

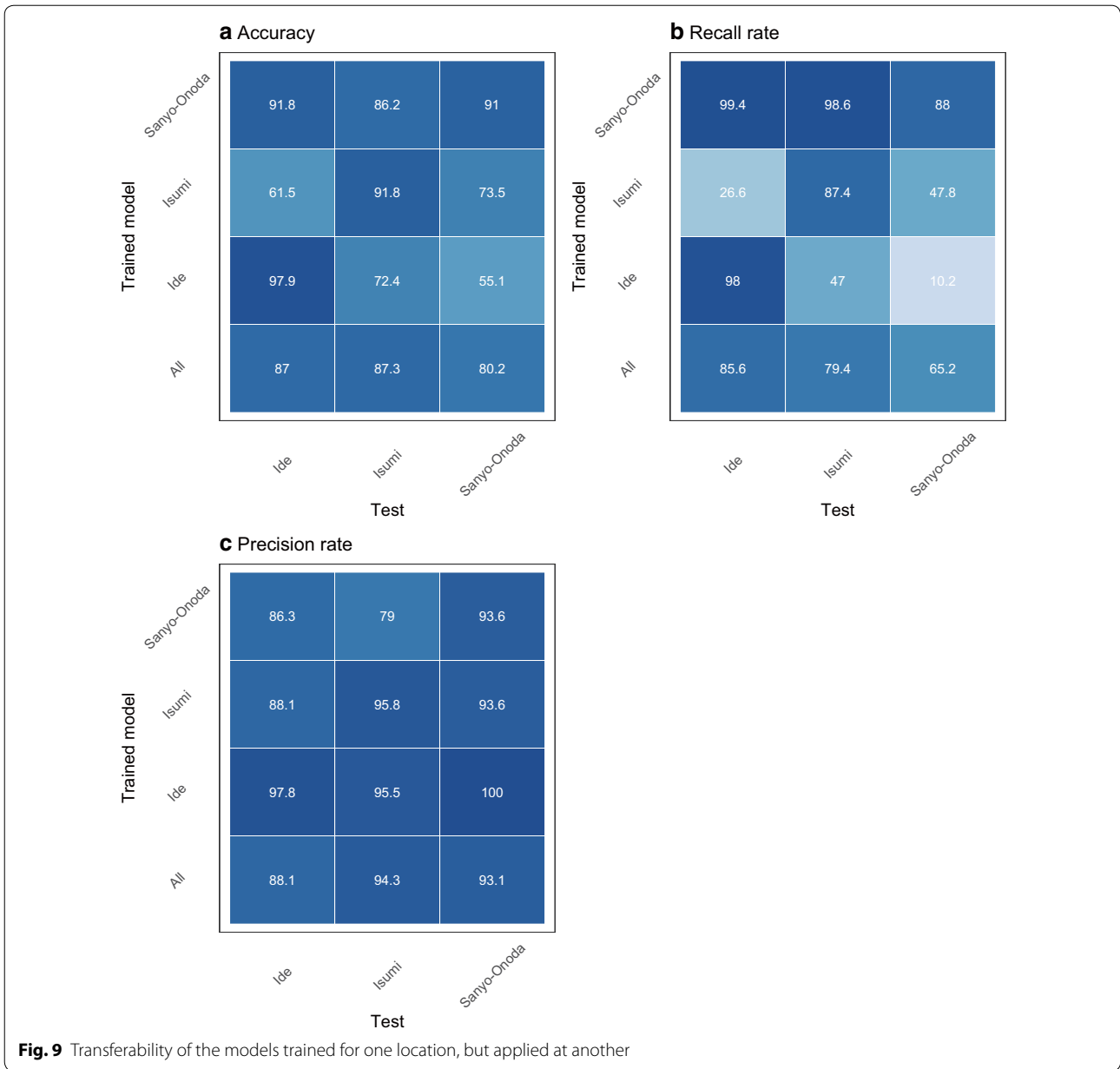


size, except for the recall rate for Ide city (Fig. 6b, d). This result indicates that if the grid size is inappropriate, both false positives and false negatives will increase.

Increasing the chopping grid size results in an increase in the number of chopped pictures, in which bamboo and objects other than bamboo are mixed. In this study, because we evaluated the performance of the model by using images that were uniformly covered by bamboo forest or objects other than bamboo forest, the influence of imagery consisting of mixed objects on the classification accuracy could not be evaluated. We will evaluate the classification accuracy for such images in our future research.

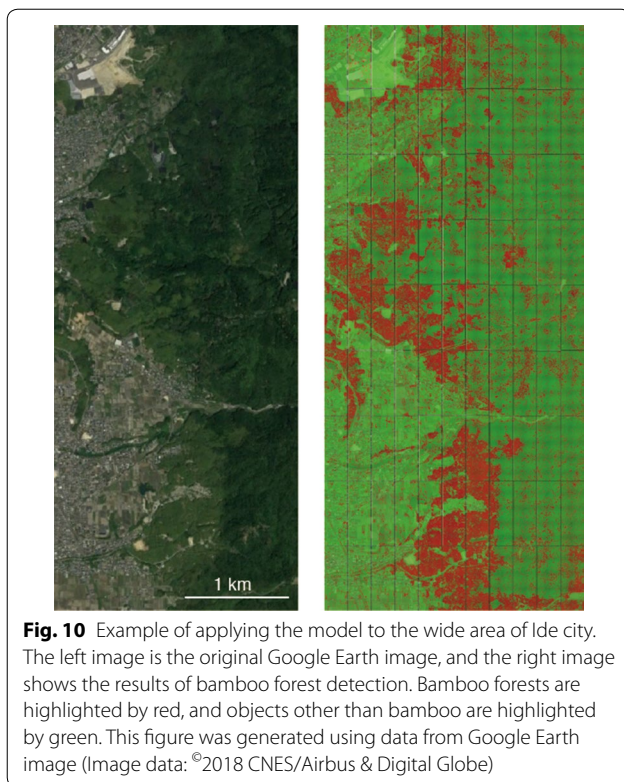
Transferability among the models

The results of the transferability tests show that transferability was generally poor, and they suggest that the spatial extent of acquisition of the training data strongly influences the classification accuracy (Fig. 8). The model trained by Sanyo-Onoda city images yielded high recall rates for the images acquired at all study sites; however, the precision rate was lower than that of the other models (Fig. 8b, c). This means that the model trained by Sanyo-Onoda city images tends to produce false positive errors. Interestingly, transferability was not found to be related to the distance among the study sites (Fig. 8). This result indicates that the classification accuracy across the model reflects the conditions at the local scale, such



as the weather conditions at the time when the image was acquired. Additionally, even when we applied a model that learned from all training images, the performance of the model was not as good as when the training data were obtained from the same city. The same tendencies have

been reported in studies that classified land use by using deep learning [25]. This suggests that increasing the quantity of training data could cause a decrease in the identification accuracy, and it may be difficult to construct an identification model that is applicable to a broad area.



Conclusion

We demonstrated that the deep learning model presented herein can detect bamboo forest from Google Earth images accurately. Our results suggested that the CNN technique and the chopped picture method would be powerful tools for high-accuracy image-based vegetation mapping, and exhibit great potential for reducing the efforts and costs required for vegetation mapping as well as improving the current status of monitoring the distribution of bamboo. Recently, bamboo expansion has become an important problem in Japan because of its invasiveness [18]. Although some studies have analyzed the bamboo forest distribution probability on a national scale [26, 27], monitoring of bamboo expansion is still challenging because it is labor-intensive. Nonetheless, our method has a certain degree of robustness in reducing the amount of training data (see Fig. 8). Therefore, we conclude that our approach could potentially lead to the creation of a semi or even fully automated system for monitoring bamboo expansion. Our results also suggest that the identification accuracy depends on the image resolution and chopping grid size; especially, the spatial resolution of training data strongly affects the model performance. Although satellite-based remote sensing has been widely studied and applied, it still has problems such as insufficient information due to low-resolution

images or inaccurate information due to local weather conditions [28]. A possible way to overcome such difficulties is to use higher resolution satellites such as WorldView-4. It is expected that the classification accuracy of our approach will be further improved by using even deeper neural networks and images obtained by higher resolution satellites such as WorldView-4 in the future.

Abbreviations

CNN: Convolutional neural network; SVM: Support vector machine.

Acknowledgements

We are very grateful to anonymous reviewer for constructive comments.

Authors' contributions

SW designed the research based on discussion with TI. SW and KS conducted research. SW and TI wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by JST PRESTO, Japan (Grant No. JPMJPR1501), JSPS KAKENHI (Grant no. JP18H03357) and the Re:connect Program of the Nippon Foundation—Kyoto University Joint Project.

Availability of data and materials

This study was a re-analysis of existing data that are publicly available. Data on the Google Earth images openly available at <https://earth.google.com/web/>. The Caffe deep learning framework can be accessed at <http://caffe.berkeleyvision.org/>. The glcm packages and e1071 package [29] in R were used for calculating the GLCM texture and SVM implementation. These packages can be accessed at <https://cran.r-project.org/web/packages/glcm/index.html> and <https://cran.r-project.org/web/packages/e1071/index.html>

Ethics approval and consent to participate

This study included no data or analysis requiring special consent to conduct or to publish.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹ Field Science Education and Research Center (FSERC), Kyoto University, Kitashirakawa-iwake-cho, Sakyo-ku, Kyoto 606-8502, Japan. ² Graduate School of Agriculture, Kyoto University, Kitashirakawa-iwake-cho, Sakyo-ku, Kyoto 606-8502, Japan. ³ Graduate School of Science and Engineering, Kagoshima University, 1-21-40 Korimoto, Kagoshima 890-0065, Japan.

Received: 17 July 2019 Accepted: 19 November 2020

Published online: 27 November 2020

References

- Franklin J. Mapping species distribution: spatial inference and prediction. Cambridge: Cambridge University Press; 2009.
- Xie Y, Sha Z, Yu M. Remote sensing imagery in vegetation mapping: a review. *J Plant Ecol*. 2008;1:9–23. <https://doi.org/10.1093/jpe/rtm005>.
- Drucker H, Burges CJ, Kauffman L, Smola A, Vapnik V. Support vector regression machines. *Neural information processing systems 9*. eds Mozer MC, Jordan JI & Petsche T. pp. 155–161, MIT Press 1997.
- Bengio Y. Learning deep architectures for AI. *Foundat Trends Mach Learn*. 2009;2:1–127.
- Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT press. 2016.

6. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Columbus)*, 2014; 1725–1732.
7. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Informat Process Syst*. 2014;27:3320–8.
8. Mohanty SP, Hughes DP, Salathé M. Using deep learning for image-based plant disease detection. *Front Plant Sci*. 2016;7:1419. <https://doi.org/10.3389/fpls.2016.01419>.
9. Ramcharan A, Baranowski K, McCloskey P, Ahamed B, Legg J, Hughes D. Deep learning for image-based cassava disease detection. *Front Plant Sci*. 2017;8:1852.
10. Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. Going deeper in the automated identification of Herbarium specimens. *BMC Evol Biol*. 2017;17:181.
11. Nogueira K, Penatti OAB, Dos Santos JA. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recogn*. 2017;61:539–56.
12. Zhu XX, Tuia D, Mou L, Xia GS, Zhang L, Xu F. Deep Learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci Remote Sens Mag*. 2017;5:8–36.
13. Li W, Fu H, Yu L, Cracknell A. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*. 2017;9:22. <https://doi.org/10.3390/rs9010022>.
14. Guirado E, Tabik S, Alcaraz-Segura D, Cabello J, Herrera F. Deep-learning Versus OBIA for scattered shrub detection with google earth imagery: *Ziziphus lotus* as Case Study. *Remote Sens*. 2017;9:1220. <https://doi.org/10.3390/rs9121220>.
15. Nogueira K, dos Santos JA, Cancian L, Borges BD, Silva TSF, Morellato LP, Torres R. Semantic segmentation of vegetation images acquired by unmanned aerial vehicles using an ensemble of convnets. *Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium*. 2017; 3787–3790.
16. Guo Y, Liu Y, Georgiou T, Lew MS. A review of semantic segmentation using deep neural networks. *Int J Multimedia Informat Retrieval*. 2018;7:87–93.
17. Ise T, Minagawa M, Onishi M. Classifying 3 moss species by deep learning, using the “chopped picture” method. *Open J Ecol*. 2018;8:166–73.
18. Nakashima, A. 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*. 2001; 10: 1–7 (**in Japanese with English abstract**).
19. Nishikawa R, Murakami T, Yoshida S, Mitsuda Y, Nagashima K, Mizoue N. Characteristic of temporal range shifts of bamboo stands according to adjacent landcover type. *J Jpn Forestry Soc*. 2005;87:402–9 (**in Japanese with English abstract**).
20. Okutomi K, Shinoda S, Fukuda H. Causal analysis of the invasion of broad-leaved forest by bamboo in Japan. *J Veg Sci*. 1996;7:723–8.
21. Suzuki S. Chronological location analyses of giant bamboo (*Phyllostachys pubescens*) groves and their invasive expansion in a satoyama landscape area, western Japan. *Plant Species Biol*. 2015;30:63–71.
22. LeCun YL, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278–324. <https://doi.org/10.1109/5.726791>.
23. Fujihara M, Ito K. A preliminary study on the area, distribution, and inclination of bamboo forests on the basis of the actual vegetation map of Awaji Island Hyogo Prefecture. *Hikobia*. 2013;16:393–402 (**in Japanese with English abstract**).
24. Raczko E, Zagajewski B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Eur J Remote Sens*. 2017;50:144–54.
25. Albert A, Kaur J, Gonzalez M. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. *arXiv preprint, arXiv:1704.02965*. 2017.
26. Someya T, Takemura S, Miyamoto S, Kamada M. Predictions of bamboo forest distribution and associated environmental factors using natural environmental information GIS and digital national land information in Japan. *Keikanseitaigaku*. 2010;15:41–54 (**in Japanese with English abstract**).
27. Takano KT, Hibino K, Numata A, Oguro M, Aiba M, Shiogama H, Takayasu I, Nakashizuka T. 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 Evolut*. 2017;7:9848–59.
28. Jones HG, Vaughan RA. *Remote Sensing of Vegetation: Principles Techniques and Applications*. Oxford: Oxford University Press; 2010.
29. Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. e1071: Misc Functions of the Department of Statistics (e1071). Vienna: TU Wien; 2010.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

