CNN-Based Tree Species Identification from Bark Image

Junya Ido, Takeshi Saitoh
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan
saitoh@ces.kyutech.ac.jp

ABSTRACT

This paper proposes a convolutional neural network (CNN)-based tree species identification method from bark image. The proposed method uses the well-known CNN model. The difficulty of our problem is to use a special tree image in which a colorful tag is stick on the bark. The tag is irrelevant to the species. In order to recognize with CNN, it is necessary to extract a region (ROI) excluding the tag. Thus, this paper proposes a ROI extraction method. Extracted ROI is fed to CNN. We evaluated the proposed method with six tree species. We carried out the evaluation experiment with various conditions, and found an optimal condition for our problem.

Keywords: tree species identification, bark image, convolutional neural network.

1. INTRODUCTION

In order to assess the land price, there is a task to identify trees. However, there are few experts on trees, and amateur workers use tree books to identify trees in the field. This task is time-consuming. It is useful to have a system or application that can automatically identify from tree image taken with a smart device. The final objective of this research is to develop this system, this paper proposes an image-based tree identification method.

Recently, natural object recognition from the image has attracted attention [1]-[6]. The most important item for recognition is the choice of organs. To recognize the tree, we can use flowers, leaves, barks, fruits, tree shape, and branching. Flowers are more attractive than leaves and barks which have simple shapes and boring colors. However, flowers can be taken only at the specific time because the blooming period is restricted. There are deciduous trees in the trees and there are times when leaves cannot be taken. There is also a time when fruits cannot be taken. Analysis of tree shape is difficult because trees are rarely inhabited isolated. Analysis of branches is also a difficult task. Thus, we focus on the bark. Bark texture is easy to take. Even though it is the same species, the thickness of the trunk varies depending on the growth condition. Moreover, the thickness of the trunk in the image is different depending on the shooting distance. One way of characterizing the bark texture is the description of uniform, scaly, crack-wise, grainy etc. Another way is to use structural characteristics such as deep or shallow crack, rough or fine grain, lengthwise and side crack.

This paper proposes a convolutional neural network (CNN)-based tree species identification method using a bark image. CNN is a deep artificial neural network that is used primarily to classify images, and has attracted much attention due to its high performance in many tasks in the field of image recognition. Our method uses the existing CNN model. However, since our method will be used not for hobby but for work, the difficulty of this research is to use a special tree image instead of a general tree image. Thus, the main contribution of this paper is to extract a bark region of interest (ROI) from the tree image. Evaluation experiments are carried out with six tree species, and this paper shows an optimal condition to obtain high recognition accuracy for our problem.

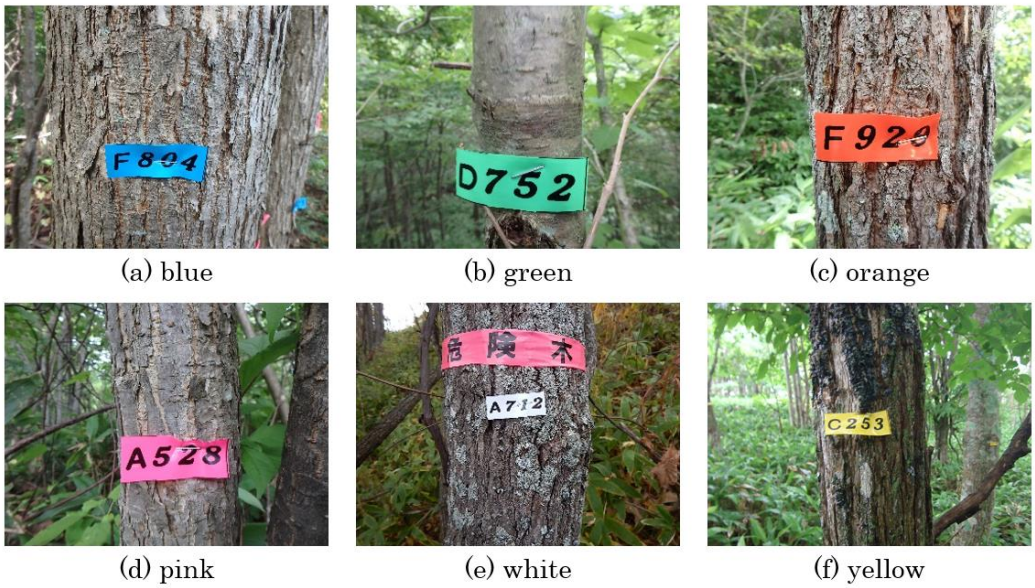
2. PROPOSED METHOD

2.1 Input Data

Input data for identifying the tree species of this research is a bark image as shown in Fig. 1. These images are taken by some workers. The target tree is a native tree, and its background is various. The background may be the color of soil and leaves as shown in Fig. 1(b)(c)(d)(f), however, different trees may be overlapped as shown in Fig. 1(a)(e). Moreover, some barks are growing moss as shown in Fig. 1(f).

Tenth International Conference on Graphics and Image Processing (ICGIP 2018), edited by
Chunming Li, Hui Yu, Zhigeng Pan, Yifei Pu, Proc. of SPIE Vol. 11069, 110693G
© 2019 SPIE · CCC code: 0277-786X/19/$18 · doi: 10.1117/12.2524213

Nicholson Manufacturing Ltd.
EX-1005, Page 1



(a) blue (b) green (c) orange

(d) pink (e) white (f) yellow

Figure 1. Target bark images.

At present, the tree identification task is done by the operator visually. This task is not done on the shooting place, but it is done after collecting some tree images back to the company office. Therefore, in order to be able to confirm individuals of trees, all trees are put tag by the worker. Thus, the tag is shown in all tree images. After the image-based identification technology is established, the tag is no longer used. There are six colors of tags; blue, green, orange, pink, white, and yellow, as shown in Fig. 1. However, the color of the tag is irrelevant to the species. Moreover, a string which is written in the tag such as "F804" in Fig. 1(a), is irrelevant to the species. The size of the real tag is almost the same, however, the shooting distance of tree varies depending on individuals and photographers, the size of the tag in the image is not the same.

2.2 Tag Region Extraction

As mentioned above, the tree has the tag, however, this is not irrelevant to the species. Thus, it should analyze the bark area without the tag region. Therefore, first, the tag region is extracted. The following the proposed tag region extraction method is simple, however, the proposed method is sophisticated for our real problem.

The color of the tag is one of six colors, this is a different color from the bark. Thus, the tree image is converted from RGB color space into HSV color space. Next, the thresholding method is applied to extract the tag region. Thresholding values are determined empirically. Figure 2 shows binary images after applying the thresholding method. The left side of Fig. 2 is the result of Fig. 1(e), and the right side of Fig. 2 is the result of Fig. 1(f).



Figure 2. Binary images.

Nicholson Manufacturing Ltd.
EX-1005, Page 2

It is clear from Fig. 2 that many small white pixels are detected. To remove these pixels, an erosion process for white pixels is applied three times. The size of erosion filter is 3×3 [pixel]. The filtered results to the images of Fig. 2 are shown in Fig. 3.

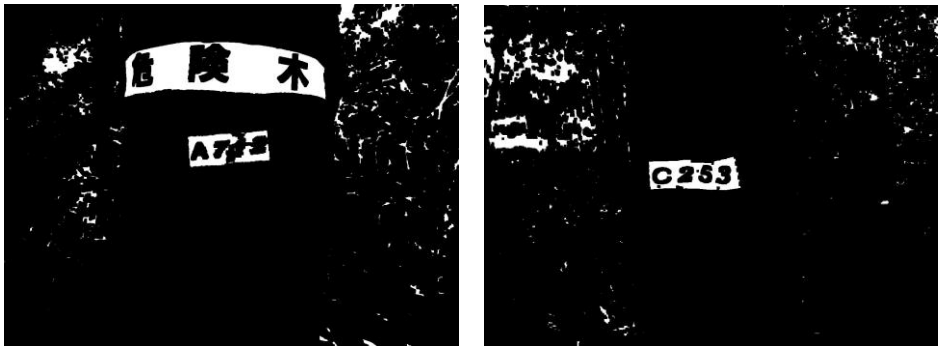


Figure 3. Filtered binary images.

2.3 ROI Extraction

Next process is a region of interest (ROI) extraction. Here, depending on the shooting condition, the tree image may be inclined. In order to obtain high recognition accuracy, it is necessary to remove the inclination problem. Thus, before extracting ROI, the rotation normalization process is applied. Here, the next problem is how to compute the inclination of the tree. Since the background of the tree is complicated and it is difficult to automatically extract a tree boundary. In this paper, the boundary information is given manually. Then, the affine transformation is applied to the tree image using the given boundary information. Figure 4 shows the rotational normalized binary images of Fig. 3. In these images, the yellow lines are the reference lines for the normalization process.

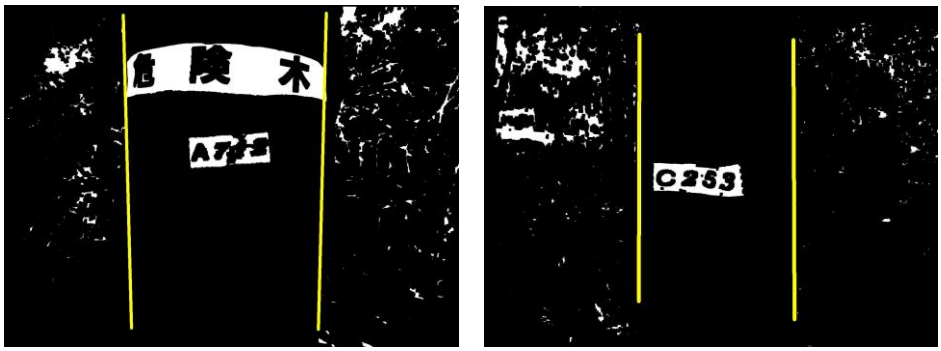


Figure 4. Normalized binary images.

As mentioned above, the tag information of color and string are not irrelevant to the species, and the ROI not including tag region is extracted. However, there are various positions and sizes of tags. Then, the ROI is extracted at a position where the tag region is the minimum.

This paper defines a scale ratio r to the width of the bark region for extracting the ROI. The size s of ROI is decided by using r, that is, $s = w \times r$, where w is the bark width in the image space. Next, a squared ROI of $s \times s$ [pixels] which the number of tag pixel is the minimum is extracted. Figure 5 shows the extraction process. In the figure, the yellow lines are the reference lines as in Fig. 4, the right graph of each image indicates the number of tag pixel at each y-position. The red square on the image is the ROI at the minimum position which $r = 0.9$. Extracted ROI is shown in Fig.6. In the proposed method, these extracted ROIs are fed to CNN.

Nicholson Manufacturing Ltd.
EX-1005, Page 3

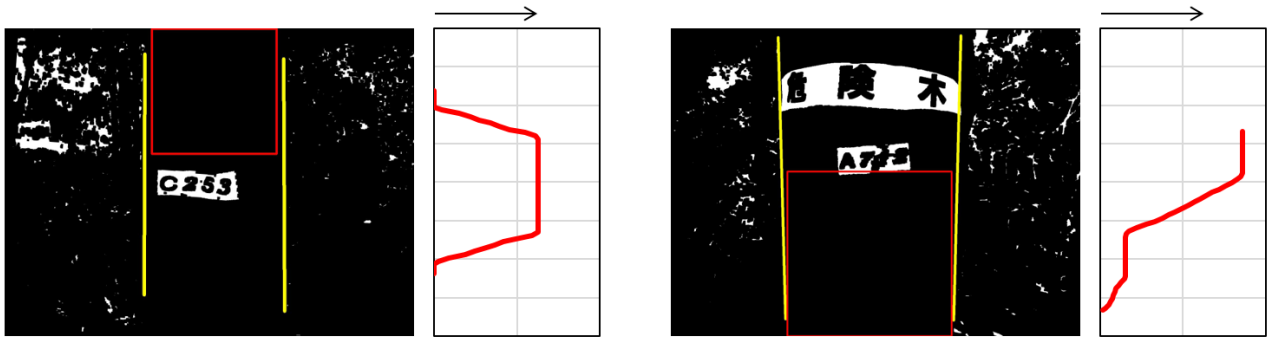


Figure 5. ROI Extraction.



Figure 6. Extracted ROI images.

2.4 CNN Architecture

Recently, deep learning technologies have greatly advanced the performance of the state-of-the-art of many visual recognition tasks. In particular, CNN has been established as a powerful class of models for image recognition tasks.

This paper uses four well-known CNN architectures of AlexNet [7], Inception-v3 [8], VGG-16 [9], and VGG-19 [9]. As for the training of CNN, there are several parameters or methods, such as dropout rate, optimizer, fine-tuning, data augmentation (DA). These are discussed in the experiment.

The input data of each CNN model is an extracted ROI image. The size of the ROI image is resized to 256×256 [pixels]. ROI image has three-channels.

3. EVALUATION EXPERIMENT

3.1 Dataset

Six tree species inhabiting the Hokkaido area of Japan are targeted for recognition task. Table 1 shows the names of each species and the number of collected samples. Sample tree images are shown in Fig.7. Collecting task is cooperated by several company workers. In our dataset, there are five types of tree image sizes; 1280×960 [pixels], 1440×1920 [pixels], 1600×1200 [pixels], 1920×1440 [pixels], 2048×1536 [pixels].

Table 1. Six species and number of samples.

No.	Scientific name	Number of samples
1	Ulmus	241
2	Quercus	630
3	Salix	216
4	Betula	223
5	Itaya amicorum	305
6	Alnus japonica	318

Nicholson Manufacturing Ltd.
EX-1005, Page 4

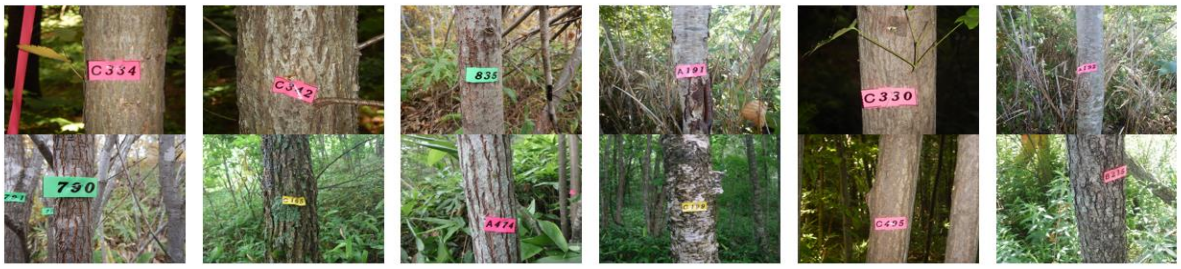


Figure 7. Sample tree images.

In our dataset, the sample number of each species is different. We randomly selected 100 samples for each species as the test data, and the remaining samples were used as the training data. That is, when we did not apply data augmentation (DA), the number of training data and test data is 1333 and 600, respectively.

3.2 Recognition Results

As for the CNN, there are several parameters and methods. In the experiment, we focused on one of the parameters or methods, and, such as optimizer, DA, fine-tuning, and scale ratio. These are discussed as below:

3.2.1 Optimizer

We applied six methods of optimizer; Adam, Adagrad, Gradient Descent, Momentum, Proximal Adagrad, and RMSProp, where the CNN model is VGG-19, r = 0.9, with DA, and applying fine-tuning. The experimental result is shown in Table 2. It is confirmed that Adam obtained the best accuracy among other optimizers.

Table 2. Recognition result (Optimizer).

Optimizer	Adam	Adagrad	Gradient Descent	Momentum	Proximal Adagrad	RMSPorp
Accuracy	**0.747**	0.443	0.677	0.687	0.717	0.437

3.2.2 Data augmentation and fine-tuning

It is generally said that CNN can obtain good performance when a large-scale dataset is provided. However, there are cases where such a dataset cannot be prepared. To overcome this problem, we generate our own data with the existing data which we have. This methodology of generating our own data is known as data augmentation (DA). The popular DA techniques are scaling, translation, rotation, flipping, adding salt and pepper noise. In this experiment, we applied both scaling and rotation and generate 15 patterns. In this case, the number of training data is 19,995.

Another approach is to use a pre-trained model. A basic approach is to train the network from scratch. Fine-tuning is a procedure based on the concept of transfer learning. We start training a CNN to learn features for a broad domain with a classification function targeted at minimizing error in that domain. Next, we replace the classification function and optimize the network again to minimize error in another, more specific domain. Under this setting, we are transferring the features and the parameters of the network from the broad domain to the specific one. In this experiment, we used existing networks that are trained on ImageNet [10].

In this experiment, we used four CNN models and two initial training approaches: random values and pre-trained value, without DA and with DA. Here, we used Adam optimizer, r = 0.9. The experimental result is shown in Table 3. Interestingly, applying DA decreased accuracy in all CNN models. From the results, it was confirmed that fine-tuning with DA is effective for our task.

Table 3. Recognition result (DA and fine-tuning).

CNN model	Random value		Fine-tuning	
	Without DA	With DA	Without DA	With DA
AlexNet	0.587	0.620	–	–
Inception-v3	0.580	0.647	0.630	0.638
VGG-16	0.330	0.328	0.733	0.722
VGG-19	0.287	0.295	0.718	**0.747**

Nicholson Manufacturing Ltd.
EX-1005, Page 5

3.2.3 Scale ratio *r*

The final experiment is to search for the optimal scale ratio *r*. Experiments were carried out with *r* is changed to 0.7, 0.8, 0.9, respectively. Here, we used two CNN models, fine-tuning model, Adam optimizer and with DA. The experimental result is shown in Table 4. This result indicates that large ROI obtains higher recognition accuracy than small ROI.

Table 4. Recognition result (scale ratio r).

r	0.7	0.8	0.9
Inception-v3	0.620	0.657	0.638
VGG-19	0.697	0.725	**0.747**

4. CONCLUSION

This paper proposes the CNN-based tree species identification method from bark image. The input bark image is not a general tree image but a special tree image in which a colorful tag is stick on the bark. The tag is irrelevant to the species. In order to recognize with CNN, it is necessary to extract a ROI excluding the tag. This paper proposes a ROI extraction method. The extracted ROI is input to well-known CNN model. We evaluated the proposed method with six tree species. We carried out the evaluation experiment with various conditions, and obtained the highest recognition accuracy of 74.7%.

The target number of species was only six. The future task is to increase the number of species. The final goal of this research is to develop the tree identification system that can automatically identify from tree image taken with a smart device. Then, it is also the future task to develop the system that implements the proposed method.

REFERENCES

[1] Wäldchen, J., Mäder, P.: Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. Archives of Computational Methods in Engineering, Vol.**25**, Issue 2, pp. 507–543 (2018).

[2] Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., Soares, J. V. B.: Leafsnap: A Computer Vision System for Automatic Plant Species Identification. ECCV2012, pp. 502–516 (2012).

[3] Saitoh, T., Iwata, T., Wakisaka, K.: OKIRAKU search: Leaf images based visual tree search system. IAPR International Conference on Machine Vision Applications (MVA2015), pp. 242–245 (2015).

[4] Berg, T., Liu, J., Lee, S. W., Alexander, M. L., Jacobs, D. W., Belhumeur, P. N.: Birdsnap: Large-Scale Fine-Grained Visual Categorization of Birds. CVPR2014 (2014).

[5] Saitoh, T., Shibata, T., Miyazono, T.: Feature points based fish image recognition. International Journal of Computer Information Systems and Industrial Management Applications, Vol. **8**, pp. 12–22 (2016).

[6] Miyazono, T., Saitoh, T.: Fish species recognition based on CNN using annotated image. 7th iCatse International Conference on IT Convergence and Security (ICITCS2017), LNEE **449**, pp. 153–163 (2017).

[7] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems (NIPS2012), pp. 1097–1105 (2012).

[8] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. CVPR2016, pp. 2818–2826 (2016).

[9] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.

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

Nicholson Manufacturing Ltd.
EX-1005, Page 6