2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2019)
Honolulu, Hawaii, USA, March 4-7, 2019

Evaluation of Training Accuracy in Color Cut Out Images Using Convolutional Neural Network

Yuichi Miyata, Ryuta Yoshimura, Yoko Uwate, and Yoshifumi Nishio

Dept. of Electrical and Electronic Engineering, Tokushima University
2-1 Minami-Josanjima, Tokushima 770-8506, Japan
Email: {y.miyata, yoshimura, uwate, nishio}@ee.tokushima-u.ac.jp

Abstract

In this paper, we investigate the accuracy of the digitally processed images by using convolutional neural network (CNN). In order to examine the effect of digital processing, we compared with digitally processed images which are similar color cut out images and original images for deep learning. The procedure includes: 1) collecting the original image and creating a data set: 2) digital processing of image used by threshold processing: 3) deep learning network learning on a data set of 200 images. We investigate the influence of the digitally processing for deep learning. We applied the training and test accuracy at simulation result.

1. Introduction

In recent years, wild animals such as deers and boars are rapidly increasing in Japanese forests. Therefore, agricultural crops damaged in nearby farms. Covering around a broader area agricultural field, with nets or wires needs enomous costs for installation, materials, and maintenance. It needs to prevent damage over wider areas where wild animals appear[1]. As that countermeasure, humans frequently enter the mountains and hunt wild animals. However, effective control is limited. By using the camera loaded in the drone, it is effective for the detection and monitoring of wild animals in the forests.

In order to realize this, it is important to recognize animals in images captured in the forests. In recent years, CNN entered the stage in the field of image recognition. CNN is one of artificial intelligence(AI) technology that learning methods for image recognition. AI is focused on industry of all over the world. We are expecting AI to be active in every field. The CNN we use is no exception. CNN is a family of multi layer neural network particularly designed for use on two-dimensional data, such as images and videos[2]. Using by CNN, recognition accuracy has dramatically improved. The research on object recognition using by CNN are actively conducted. However, no studies on the recognition and detection of specific animals such as animals that eat agricultural crops have been reported.

In this study, we recognize images by using CNN. We compare with digitally processed images which are similar color cut out images and original images for deep learning. In addition, we also investigate whether the accuracy is affected when changing the color of the object in the image.

2. Max pooling and Spatial Pyramid Pooling

CNN used for image recognition convolutional layers, max pooling layers and fully connected layers. The pooling layer has the function of compressing the information in order to transform the input data into a more manageable form. The pooling layer has three features. First, it does not have parameters to learn. Second, number of channels does not change. Third, it is robust against a small gap changed. Fig. 1 shows a process of applying max pooling configuration. Max pooling is usually used in the pooling layer of CNN. This is an operation to leave the largest one from small area in the input area. Even if there is a small gap in the input data, the pooling have a similar value. Therefore, by entering the pooling layer, it becomes a strong network against the small gap of input data. Furthermore, we introduce another pooling. Figure 2 shows a process of applying Spatial Pyramid Pooling(SPP) configuration. SPP is a kind of pooling. SPP is a method mainly used in the field of image processing using by CNN. The difference from ordinary pooling is that the target area can be divided at the time of pooling and processed hierarchically. SPP can obtain fixed length output regardless of input image size. In addition, it is possible to improve the accuracy of CNN of various structures by using SPP in the pooling layer [3][4]. Input can be changed to variable size instead of fixed.

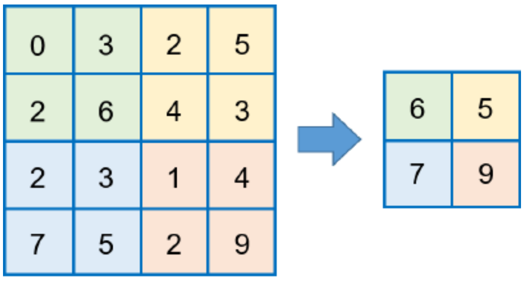

Figure 1: Max pooling.

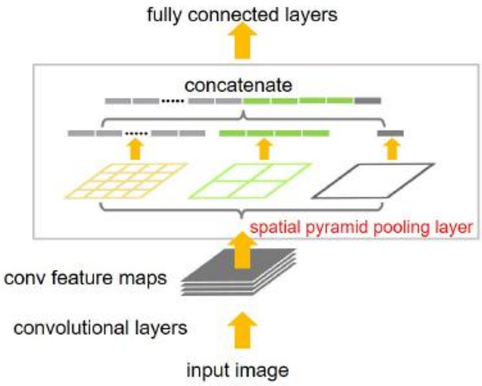

Figure 2: A network structure with a spatial pyramid pooling layer.

3. Image data set creation

All image data we used this time were the original images. We prepare data set like Fig. 3 to Fig. 6 for training and test. As the object of recognition, we used a brown Bear's stuffed toy and a green flog's stuffed toy. The reason is that wild animals such as deer and wild boar are brown. This object was placed in various places and take a picture. The other reason is that most of the ground in the mountain is covered with grasses and trees. The background of the object is green and brown. This time, we focused on the case where the background of the object is green or brown. Therefore, we also prepare the image that include green and brown in the buckgrond. We took a picture both when we placed the object and did not put it at the same point. By using CNN, we classify when the object exists and when it is not. In order to compare the training accuracy, we prepared each of the original data set and digitally processed data set. Training accuracy is the accuracy of 200 data for learning correctly learned. Test accuracy is the accuracy of whether 10 newly prepared test data are correctly classified. The type of digital processing is threshold processing from OpenCV. First, we calculated the histogram of the color of the object from the image of only the object. Next, we remove the color of the area deviating from the histogram of the object in the image to process. Therefore, we can obtain the image that cut out the background. We process by the same for image without the object.

After digital processed, labels of the data were created. The training image data were named as 1-200.jpg, of which 1-100.jpg were the images that object exists (labeled as 1) and 101-200.jpg were the images that object doesn't exist (labeled as 0). The test image data were named as 1-10.jpg. We prepared for green objects as well.

Figure 3: Original image (Object color is brown).

Figure 4: Original image (no object).

Figure 5: Original image (Object color is brown).

Figure 6: Original image (no object).

4. Proposed system

Figure 7 shows a process of applying CNN configuration. This CNN used for image recognition consists of input layers, convolutional layers, max pooling layers, fully connected layers and output layers. There are two layers in convolutional layers, max pooling layers, fully connected layers for each one. Its function is to progressively reduce the spatial size and length output from convolutional layers and reduce the amount of parameters and calculation in the network. When introducing SPP, it is optimal just before the entire bonding layer. Others have the same network structure as before, simply changing max pooling to Spatial Pyramid Pooling improves accuracy. Training and test images are compressed 28×28 pixels. We define as the training steps = 200. We learned these by using CNN and investigate the accuracy of color cut out images and original images.

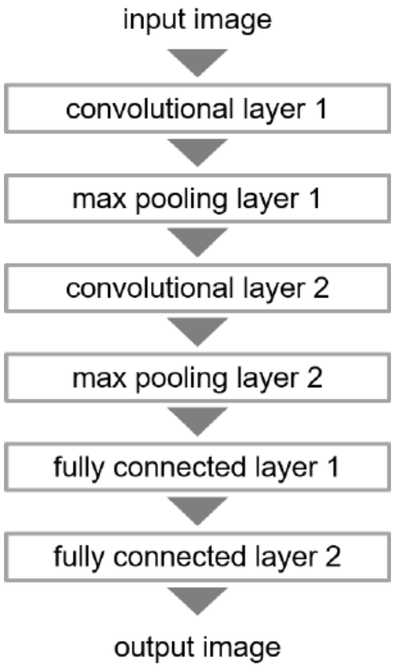

Figure 7: CNN Configuration.

5. Simulation result

In this section, we show some simulation results by using the CNN. First, we digitally processed the original images using by threshold processing. Fig. 8 to Fig. 11 shows results of digitally processed. From Figs. 8 and 9, background of image is black, that is no background. However, when the object is brown, fallen leaves in the background are similar in color to the object. Therefore, we could not completely remove the background. when the object is green, grasses are similar in color to the object like Figs. 10 and 11. We consider the reason why brown components that remained is the color similar to an object.

Figure 8: Digitally processed image (brown).

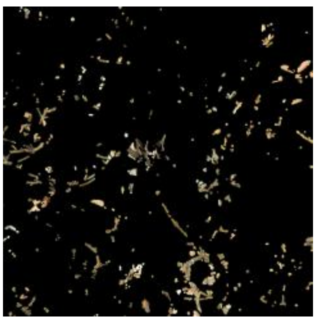

Figure 9: Digitally processed image (no object).

Figure 10: Digitally processed image (green).

Figure 11: Digitally processed image (no object).

We define as the number of learning data sets = 200. Learning data contain 100 images with the object and 100 images without the object. Fig. 12 shows the training accuracy of original images (brown) and digitally processed images (brown). The horizontal axis is the number of epochs. When epoch is 0 to 100, the training accuracy of digitally processed images reached 1 earlier than original images. In addition, the training accuracy of digitally processed images reached 1 without decreasing accuracy in the steps.

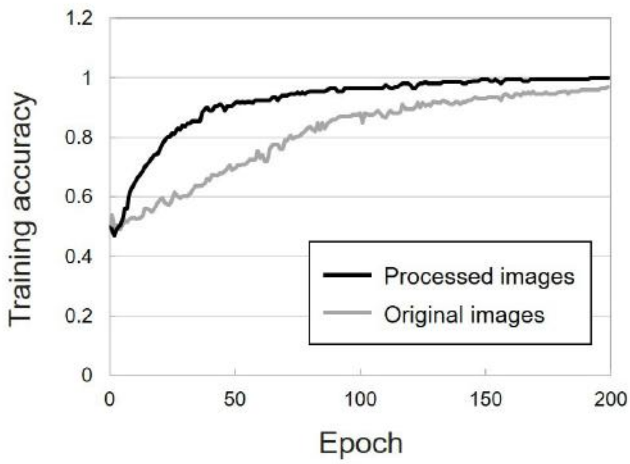

Figure 12: Training accuracy of two pattern (brown).

Figure 13 shows the training accuracy of original images (green) and digitally processed images (green). The horizontal axis is the number of epochs. Similar to Fig. 12, training accuracy of digitally processed images reached 1 earlier than original images. However, there is not much difference between two training accuracies. We consider that improving accuracy results from removing background. The image prepared as a data set contained a lot of grasses. Terefore, the same green to the object couldn't remove as a background.

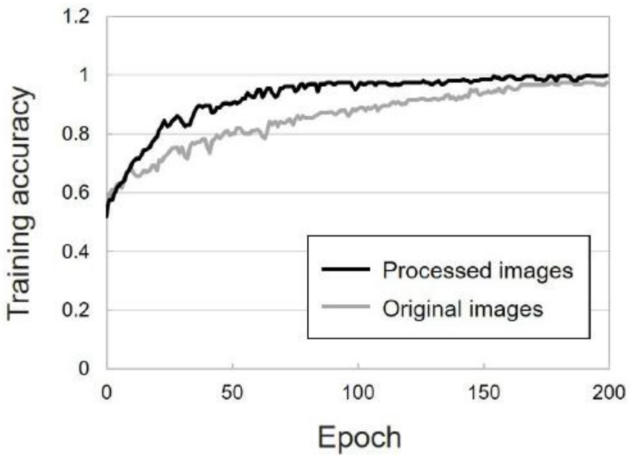

Figure 13: Training accuracy of two pattern (green).

In Table 1, digitally processed images obtains a little bit better test accuracy than original image when we learned images of two patterns. However, test accuracy is not good.

Table 1: Average of training and test accuracies (brown).

	Original images	Processed images
Training accuracy	0.968	0.998
Test accuracy	0.5	0.56

In Table 2, although the training accuracy was improved, test accuracy was the same. As we mentioned earlier, this factor was that, the background as green color couldn't be completely remove.

Table 2: Average of training and test accuracies (green).

	Original images	Processed images
Training accuracy	0.961	0.990
Test accuracy	0.53	0.53

6. Conclusion

In this study, we succeeded in cutting the green background in images used for threshold processing. However, green parts similar to object in the background could not cut. It is necessary to consider about a method that can remove background color similar to the object. Moreover, we compared with digitally processed images which are similar color cut out images and original images for deep learning. We investigated the training accuracy of color cut out images and original images. We considered that the object becomes clear and stands out becouse. Therefore, the test accuracy of the digitally processed image was improved. As our future works, we improve a pooling layer in CNN using SPP method and would like to obtain better test accuracy. We will try to recognize images taken of animals in the forest using drone.

References

[1] Kouichi Nakajima, Katsuhiko Seo, Shinji Kitagami, Tetsuo Shiotsuki Hisao Koizumi," Development of Unmanned Aerial Vehicle Systems for Monitoring and Prevention of Bird and Animal Damage and its evaluation ", Information Processing Society of Japan SIG technical Report, 2015

[2] Itamar Arel, Derek C. Rose, and Thomas P. Kamowski, " Deep Machine Learning-A New Frontier in Artificial Intelligence Research ", IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE, 2010

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun," Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition " , Computer Vision ECCV, 13th European Conference, Zurich, Switzerland, September 6-12, 2014, proceedings, Part, pp.346-361.

[4] Makoto Okada, Hidekazu Yanagimoto, Kiyota Hashimoto, " Sentiment Classification with Gated CNN using Spatial Pyramid Pooling ", The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018