# Building Data Sets Using Clustering and Evaluation of Learning Accuracy in Convolutional Neural Networks

Yuichi Miyata, Yoko Uwate and Yoshifumi Nishio
Depertment of Electrical and Electronic Engineering, Tokushima University
2-1 Minami-Josanjima, Tokushima, 770-8506, Japan
E-mail: {y.miyata, uwate, nishio}@ee.tokushima-u.ac.jp

## SUMMARY

In recent years, aerial photography became easier than before by using the camera loaded in the drone. Also, convolutional neural network (CNN) is one of deep learning and is a network often used for image recognition. With the development of CNN, drones are being researched for applications in various fields such as agriculture [1]. Actually, wild animals such as deer and boars are rapidly increasing in Japanese forests. Agricultural crops damaged in nearby farms. By using a drone, we expected to wild animals management. However, the drone camera's battery and memory are limited. It is important to use the collected data effectively.

In this study, we used clustering to make more efficient data sets from the collected image data. Clustering is used to classify similar data and can also be used to classify images [2]. This time, we used k-means method to classify the image into two clusters and constructed data sets. This data set is trained using CNN. We examine differences in learning accuracy and computational cost on the data set before and after clustering. We compare the differences in learning accuracy and computational cost with the data set before and after clustering. We also change the assignment of cluster images.

## REFERENCES

[1] J. Rebetez, H.F. Satizbal, M. Mota, D. Noll, L. Bchi. "Augmenting a convolutional neural network with local histograms - A case study in crop classification from high-resolution UAV imagery" ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016.

[2] V. Birodkar, H. Mobahi, and S. Bengio. "Semantic redundancies in image-classification datasets: The 10 % you don't need" ArXiv, 1901.11409, 2019.