Efficient Parameters for Rotation Processing of Data Augmentation

Naoki Nakamura
School of Information and
Telecommunication Engineering
Tokai University
Tokyo, Japan
e-mail: nakamura.naoki@star.tokai-u.jp

Duke Maeda
School of Information and
Telecommunication Engineering
Tokai University
Tokyo, Japan
e-mail: maedaduke@gmail.com

Kenta Morita
Graduate School of Engineering
Mie University
Mie, Japan
e-mail: k-morita@ip.elec.mie-u.ac.jp

Naoki Morita
School of Information and
Telecommunication Engineering
Tokai University
Tokyo, Japan
e-mail: morita@tokai.ac.jp

***Abstract*— Deep learning typically requires a large amount of training data. However, in some cases, it is not possible to prepare enough data to achieve the desired recognition accuracy. A number of approaches to training models with a limited amount of data are available, such as data augmentation and fine-tuning. In the present study, we focus on rotation processing, which has the capacity to augment image data more easily than other augmentation methods. With this method, for example, we could produce 360 images from a single image by rotating the image a full 360° by 1° increments. However, if the rotation angle is not chosen appropriately, essential features of the rotated object may be lost. No clear standards have been previously determined for setting appropriate rotation parameters. This study presents an approach to efficient rotary processing for cases in which the key features of the object either does or does not distort, depending on the angle of rotation. The approach should make it easier for general users to set proper parameters when using rotation processing for data augmentation.**

Keywords- Convolutional neural network; Training data augmentation; Rotation angle; Augmentation rate.

I. INTRODUCTION

A Convolutional Neural Network (CNN) [1] is a machine learning algorithm frequently used for image recognition. For training, a CNN requires a dataset consisting of many pairs of images and labels. Training conducted with insufficient data impairs the ability of the CNN to accurately recognize objects.

Large labeled datasets, such as CIFAR-10 [2] and MNIST [3], are available for training. However, such published datasets cover limited categories, so it is not uncommon for practitioners to create their own datasets. Unfortunately, creating an original dataset normally requires a substantial investment in time and human resources, since more than 1,000 images per object are typically needed. In such cases, data augmentation [4] offers an attractive solution. Data augmentation is a technique that enables practitioners to significantly increase the diversity of the data available for training without actually collecting new data. Data augmentation takes various forms, including rotation, horizontal flipping, and color jittering. In this study, we focus on rotation.

The paper is structured as follows: Section II defines a rotation operation; Section III describes the purpose of the study; Section IV describes the experiments conducted, presents results and considerations; and Section V provides a summary and offers conclusions.

II. ROTATION

For the purposes of data augmentation, a rotation has two parameters: Rotation Angle (RA) and Augmentation Rate (AR). Augmentation Rate is defined as the number of augmented images produced from a single image. For example, if an image is rotated 360° in 1° increments, 360 images will be created. The corresponding values of the RA and the AR would be 1° and 360, respectively.

One notable disadvantage of rotation processing is that key features of the training object can be lost or distorted depending on the angle of rotation, leading to incorrect object recognition. Letters and numbers are one such example. For instance, when the number "6" is rotated by 180°, it becomes "9". As shown in Figure 1, as the rotation angle increases, recognition accuracy decreases.

Kitakaze et al. found rotation to be an effective data augmentation approach for recognizing harmful birds. In their study, each image was rotated over a range of -15° to +15° in 5° increments, resulting in an AR of 7 [5]. In a binary classification test, they reported a recognition accuracy rate of 85% using 536 images per object (for a total of 2 x 536 = 1,072 images). When the dataset was augmented with the rotated images, the accuracy rate increased to 90%. In another study, L. Taylor et al. used Caltech101 to show that the recognition accuracy could be

 ISBN: 978-1-61208-765-8

improved by using images that were rotated over a range of -30° to +30° [6]. Hu et al. rotated CIFAR-10 and MNIST images to determine the effects of rotation. They reported that, when the RA was between -30° and +30°, an AR of two or three times was sufficient for training [7].

In this paper, we call an object whose core appearance is unchanged by the angle of rotation an "unaffected object." For such an object, RA × AR = 360°, as the essential meaning of the object is not altered even if rotated by any possible angle.

Kawasaki et al. used a leaf dataset consisting of 800 images composed of three objects to identify plant diseases. The AR was 36 and RA was 10°. They reported that recognition accuracy improved from 77.0% to 92.5% after the rotated images were included [8].

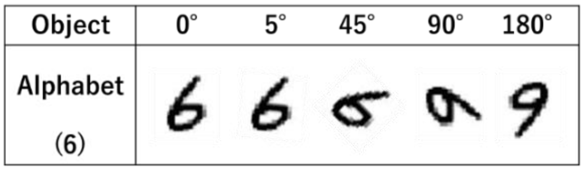


Figure 1. Rotation processing (Example)

III. Purpose

The present study has two purposes. First, we attempt to determine a method for efficiently rotating objects in a dataset consisting of images that are not affected by the rotation processing. Finding the optimal RA and AR values in such cases would clearly be desirable. As described above, Kawasaki et al. reported that recognition accuracy increased from 77% to 92.5%. when the RA was set to 10° and the resulting AR was 36. It is possible that the recognition rate may be further improved by setting the rotation angle to 5° and AR to 72.

One might assume that there is a straightforward relationship between recognition accuracy and the augmentation rate, and that recognition accuracy will continue to improve as the number of images increases. However, it seems at least as plausible that at some point the recognition rate will no longer improve, since smaller angular increments will produce many similar images. If a CNN can accurately recognize objects that are rotated with a lower AR, then increasing the AR becomes unnecessary. We need to be able to determine at what point the accuracy stops increasing as RA is reduced.

The second objective is to identify the effective way to apply rotation processing to images that are affected by rotation. Images that are affected by rotation may be distorted depending on the angle of rotation, so care must be taken when performing rotation processing.

Our overall intent is to identify a simple way to set proper RA and AR parameters to ensure a high degree of recognition accuracy for any object.

IV. Experiment I: Effective Parameters For Unaffected Objects

Experiment 1 was conducted in order to determine the most efficient value for AR for cases in which the images are not affected by the rotation. The datasets, experimental method, and results are described below.

A. Dataset

In this experiment, we used three datasets which consist of unaffected objects: a HEp-2 cell dataset, a Malaria-infected cell dataset [9], and a Branches dataset, as shown in Table I. The HEp-2 cell dataset was provided by the 22nd International Conference on Pattern Recognition (ICPR 2014). We rotated the objects to augment the three datasets with an AR of 2 to 10. The RA of the augmented datasets conforms to the equation RA×AR = 360°.

TABLE I. Datasets of unaffected objects

	For learning	*For validation*
HEp-2 cell (4 classes)	500 images per object (2000 images in total)	100 images per object (400 images in total)
Malaria-infected cell (2 classes)	200 images per object (400 images in total)	50 images per object (100 images in total)
Branches (4 classes)	300 images per object (1200 images in total)	200 images per object (800 images in total)

B. Experimental methods

We used 100 epochs for training with the HEp-2 and Branches datasets and 300 epochs for the Malaria-infected cell dataset. The experimental conditions are shown in Table II.

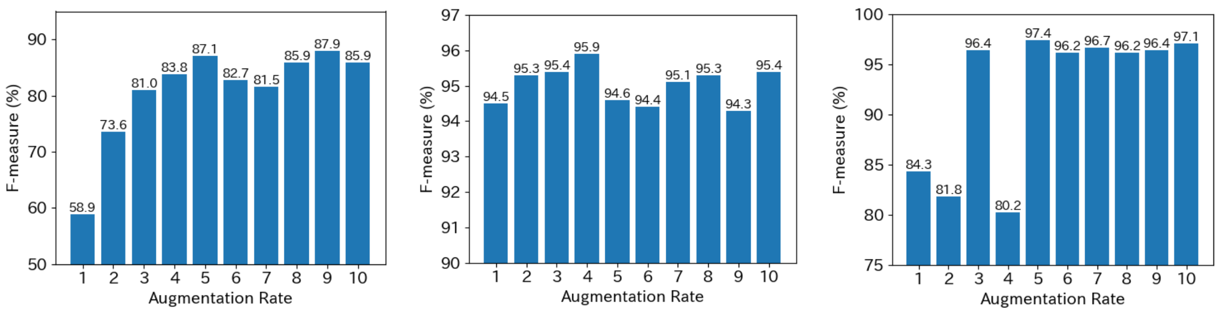
TABLE II. Experiment Conditions

OS	Ubuntu 18.04 LTS
CUDA	10.0.130
cuDNN	7.6.4.38
Python	2.7.15+
OpenCV	3.4.0
Framework	Caffe [10]
Network	GoogleNet

For testing, we used 200 test images per object. We evaluated the training models with the F-measure [11], which is the harmonic mean of precision and recall, as defined below.

$$\text{F-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

The experimental results are shown in the next section. The results represent the average value when the experiment was performed 3 times.



(a) (b) (c)

Figure 2. Results for unaffected objects datasets: (a) HEp-2 cell, (b) Malaria-infected cell, (c) Branches

C. Results and Discussion

Figure 2 shows the training results for unaffected objects datasets. According to the results for the HEp-2 cell and branch datasets, rotation processing in which the AR values were from 1 to 5 the performance improved. An AR of more than 6 produced no significant improvement.
For the malaria-infected cell dataset, rotation processing in which the AR values were from 1 to 4 the performance improved. For an AR of more than 5, no further improvement occurred.

These results suggest that there are limits to the capacity of rotation processing to improve the performance of CNNs. It is assumed that once the AR exceeds 4 or 5, the training data includes many similar images. Thus, simply increasing the number of similar images has little effect on increasing accuracy, as the features of the object do not increase. The learning time, however, does increase. This suggests that efficient rotation processing for unaffected objects should use an AR of 4 to 5.

It should be noted that for the branches dataset, when AR was set to either 2 or 4, performance decreased. In these two cases, it is assumed that the decrease was related to the object's specific characteristics, as suggested in Figure 3. When AR was 2, for example, the training data consisted exclusively of images rotated by 0° and 180°. These images are very similar, since branches are essentially linear objects. Results also show a decrease in performance for the HEp-2 cell dataset when AR is 6 or 7. Investigation of the reason for this will be conducted in future studies.

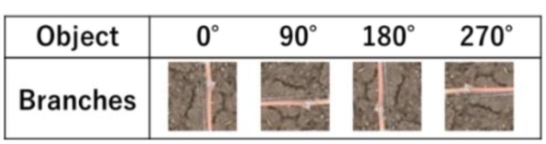


Figure 3. Rotation processing for Branches

V. EXPERIMENT II: EFFECTIVE PARAMETERS FOR OBJECTS AFFECTED BY THE ROTATION

The purpose of Experiment 2 was to determine the most efficient rotation procedure—that is, the best combination of the parameters AR and RA—for images affected by the rotation.

A. Datasets

Two datasets which consist of objects affected by the rotation were used: ImageNet [12] and MNIST (Table III). We augmented the datasets with AR = 2 and AR = 3.

To accomplish the targeted augmentations, we used the following rotation scheme: For an AR of 3, we rotated the original image plus x° and minus x°, where x is set to a specific value. (For example, we could triple the size of the dataset by using the original image plus the image rotated 5° clockwise and 5° counterclockwise, or 10° clockwise and 10° counterclockwise, etc.) For an AR of 2, we simply removed the base dataset from the AR = 3 augmented data. We used seven possible angles for x, from 5° to 35°, in 5° increments.

TABLE III. DATASETS OF OBJECTS AFFECTED BY THE ROTATION

	For learning	*For validation*
ImageNet (10 classes)	450 images per object (4500 images in total)	50 images per object (500 images in total)
MNIST (10 classes)	180 images per object (1800 images in total)	20 images per object (200 images in total)

B. Experimental methods

We trained 100 epochs using the training data described in Table III, with 100 test images per object. The experimental results using the F-measure are shown in the next section. The experimental conditions are the same as in Table II. The results are given as the average value when the experiment was performed 3 times.

C. Results and Discussion

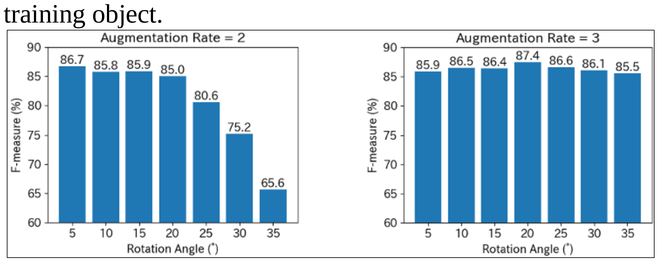
Figure 4 shows the results for ImageNet and MNIST when the size of the dataset was doubled and tripled, respectively.

From the experimental results, it seems that the most efficient rotation procedure for objects affected by rotation is defined by RA = 15° or 20° and AR = 3.

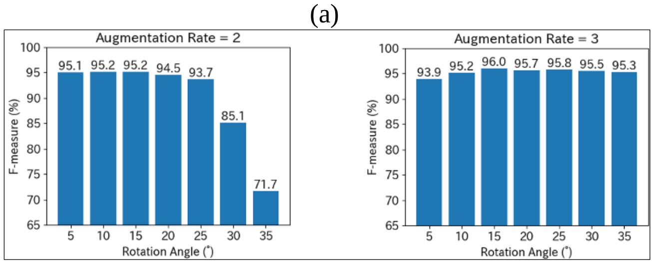
The result for ImageNet with AR = 3 showed that the performance improved with the increase in RA. For RA of more than 25°, the performance dropped. For the result of AR = 2, the performance degraded with an increase in RA.

The results for the MNIST dataset showed the same tendency as the results for the ImageNet dataset. For the result of AR = 3, performance improved with the increase in RA. For RA of more than 20°, the performance dropped.

From these results, the performance when the dataset was tripled was better than when the dataset was doubled in most cases from 5° to 35°. From this result, it is reasonable to believe that the tripled dataset contains more information about the objects in the dataset than the doubled dataset. These results suggest that adjusting an RA more than a certain RA value started to distort the key features of the training object.



(a)



(b)

Figure 4. Results for datasets of objects affected by the rotation: (a) ImageNet, (b) MNIST

VI. CONCLUSION

In the present study, we clarified an efficient rotation processing procedure for augmenting a dataset for training a CNN. Our results provide a reference when performing rotation processing.

When rotating images of objects whose meaning is affected by their rotation, such as those of alphanumeric characters, cars, and dogs, the efficient rotation procedure is to augment the number of images tripled and tilt images to -15°, 0°, 15° or -20°. 0°, 20°. When rotating images of objects whose essential meaning is unchanged when rotated, such as those of cells and branches, the most efficient rotation procedure is to augment the number of images from 4 to 5 times over the full 360° range.

Rotation is only one of many data augmentation techniques. In the future, we plan to clarify optimum parameters for other augmentation approaches.

REFERENCES

[1] Y. LeCun et al., “Handwritten Digit Recognition with a Back-Propagation Network.” Advances in Neural Information Processing Systems, pp. 396-404, 1990.

[2] A. Krizhevsky, “Learning Multiple Layers of Features from Tiny Images” Tech. Rep., University of Toronto, 2009.

[3] Y. LeCun and C. Cortes, “The MNIST database of handwritten digits.”, http://yann.lecun.com/exdb/mnist/, [retrieved: February 2020]

[4] European Space Agency. ESA: Missions, Earth Observation: ENVISAT. [Online]. Available from. https://bair.berkeley.edu/blog/2019/06/07/data_aug/, [retrieved: February 2020]

[5] H. Kitakaze, S. Okabe, R. Yoshihara, and R. Matsumura, “Recognition Rate-Improvement of Injurious Bird Recognition System by Increasing CNN Learning Image using Data Augmentation” The Japanese Journal of the Institute of Industrial Applications Engineers, Vol. 7. No. 2, pp. 69-76, 2019.

[6] L. Taylor and G. Nitschke, “Improving Deep Learning using Generic Data Augmentation” IEEE Symposium Series on Computational Intelligence, pp. 1542-1547, 2018.

[7] B. Hu, C. Lei, D. Wang, S. Zhang, and Z. Chen, “A Preliminary Study on Data Augmentation of Deep Learning for Image Classification” arXiv:1906.11887v1, 2019.

[8] Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, “Basic Study of Automated Diagnosis for Viral Plant Diseases with Convolutional Neural Networks” 31st Fuzzy System Symposium, pp.391-394, 2015.

[9] S. Rajaraman et al., “Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images” PeerJ, Vol. 6, p. e4568, 2018.

[10] Y. Jia et al., “Caffe: Convolutional Architecture for Fast Feature Embedding" 22nd ACM international conference on Multimedia, pp. 675-678, 2014.

[11] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation” 19th Australian Joint Conference on Artificial Intelligence, pp. 1015-1021, 2006.

[12] J. Deng et al., “Imagenet: A large-scale hierarchical image database” IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009.