

The Effect of Data Augmentation in Deep Learning Approach for Peripheral Blood Leukocyte Recognition

Hiroyuki Nozaka^a, Miku Oda^b, Ami Sasaki^b, Honoka Harako^b, Mae Miyazaki^b, Suzuka Kaga^b, Niina Sakaiya^b, Kyouka Kudo^b, Shou Kimura^b, Manabu Nakano^a, Miyuki Fujioka^a, Kazufumi Yamagata^a

^a Hirosaki university Graduate school of health sciences, JAPAN

^b Hirosaki university School of health sciences, JAPAN

Abstract

Data augmentation is reported as a useful technique to generate a large amount of image datasets from a small image dataset. The aim of this study is to clarify the effect of data augmentation for leukocyte recognition with deep learning. We performed three different data augmentation methods (rotation, scaling, and distortion) as pretreatment on the original images. The subjects of clinical assessment were 51 healthy persons. The thin-layer blood smears were prepared from peripheral blood and stained with MG. The effect of data augmentation with rotation was the only significant effective technique in AI model generation for leukocyte recognition. On contrast, the effect of data augmentation with image distortion or image scaling was poor, and accuracy improvement was limited to specific leukocyte categories. Although data augmentation is one effective method for high accuracy in AI training, we consider that a highly effective method should be selected.

Keywords:

Artificial intelligence, Convolutional neural network, Data augmentation, Leukocyte, Hematology

Introduction

Artificial Intelligence (AI) technology is rapidly developing in recent years as a new medical technology that enables judgments that reflect human thought processes. AI technology is characterized by learning with large amounts of patient data diagnosed by medical specialists based on years of experience. Therefore, AI is also expected to become a next-generation medical technology that enables the provision of a diagnosis based on the experience of medical specialists [1, 2]. Deep learning as one of AI technology is a multi-layered neural network that imitates the human cranial nerve circuit. Deep learning automatically extracts the common features of each training image, and it realizes accurate and efficient judgment. Although it is expected to automate leukocyte morphology analysis with deep learning in hematology fields, a lot of training images are required for high accuracy AI model creation in leukocyte recognition. Data augmentation is reported as one useful technique to generate a large amount of image dataset from a small image dataset [3-5]. However, there are few reports on generation techniques of optimal image dataset for AI training in hematology fields. In addition, data augmentation limits of recognition improvement and optimal data augmentation methods are not reported in hematology field. The aim of this study is to clarify the effect of data augmentation for

leukocyte recognition with deep learning. We evaluated AI models generated by data augmentation in leukocyte recognition on peripheral blood smears.

Materials and Methods

Materials

1. Original images for supervised learning

The subjects were 40 healthy persons. The thin-layer blood smears were prepared from peripheral blood supplemented with EDTA-2Na and stained with May-Grünwald-Giemsa (MG). The MG-stained smears were observed under a microscope using an objective 100x oil immersion lens, and typical normal leukocyte images were captured with a microscope color camera (Axiocam ERc5s, Carl Zeiss). Leukocyte images were labeled and classified into six categories by three medical specialists: rod-shaped neutrophils (Band), segmental nucleus neutrophils (Segment), eosinophils (Eosino), basophils (Baso), monocytes (Mono), and lymphocytes (Lymph). The image of each cell group was trimmed at 480×480 pixels. A database of labeled 1632 typical mature leukocyte images were generated.

2. Clinical assessment images

The subjects were 51 healthy persons. The thin-layer blood smears were prepared from peripheral blood supplemented with EDTA-2Na and stained with MG. The MG-stained smears were observed under a microscope using an objective 100x oil immersion lens, and leukocyte color images (150 to 200 cells per smear slide) were captured with a microscope color camera (Axiocam ERc5s, Carl Zeiss). Leukocyte images were labeled and classified into six categories (Band, Segment, Eosino, Baso, Mono and Lymph) by three medical specialists. The image of each cell group was trimmed at 320×320 pixels.

Data augmentation

In this study, we performed three different data augmentation methods (rotation, scaling, and distortion) as pretreatment on the original images (Fig 1).

1. Image rotation

Data augmentation processing was performed to increase the number of original data. Total 5340 images were generated from 1632 original images by rotation. Two types of dataset were generated with original images, and three types of

dataset were generated with mixture of original and rotated images (Fig 2).

In any type, 80% of the randomly extracted images (480×480 pixels) were used as the training data. The remaining 20% of the images were converted to low resolution images (320×320 pixels), and it was used as the validation data.

Table 1 The effect of image rotation

Recall	Original		Mix (Original and Rotation)		
	1064 images	1302 images	2385 images	3284 images	4106 images
Baso	0.989	1.000	1.000	1.000	1.000
Eosino	1.000	1.000	1.000	1.000	1.000
Lymph	0.990	0.982	0.987	0.984	0.988
Mono	0.932	0.943	0.941	0.918	0.905
Neutrophil (Band)	0.826	0.631	0.933	0.910	0.924
Neutrophil (Segment)	0.618	0.775	0.724	0.772	0.766
Total Accuracy	0.768	0.844	0.833	0.857	0.855

F-measure	Original		Mix (Original and Rotation)		
	1064 images	1302 images	2385 images	3284 images	4106 images
Baso	0.956	0.880	0.951	0.951	0.967
Eosino	0.986	0.998	0.993	0.979	0.993
Lymph	0.988	0.986	0.988	0.984	0.985
Mono	0.945	0.938	0.939	0.925	0.921
Neutrophil (Band)	0.287	0.320	0.391	0.428	0.427
Neutrophil (Segment)	0.756	0.856	0.837	0.867	0.864
Average F-measures	0.820	0.830	0.850	0.856	0.859

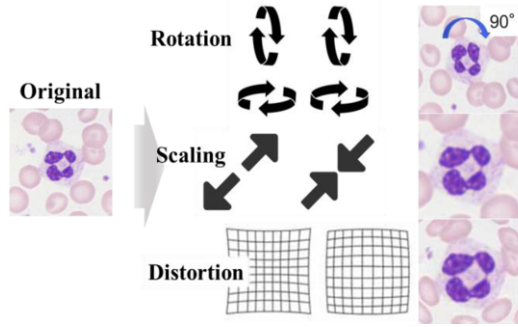


Figure 1. Data augmentation methods

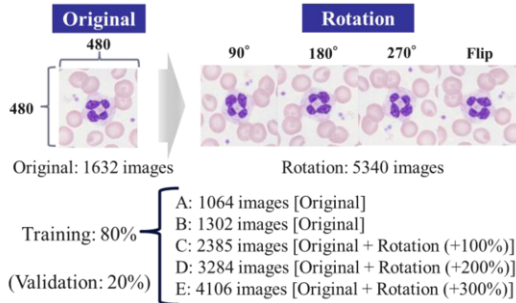


Figure 2. Training images augmentation with rotation processing

2. Image scaling

Image scaling processing was performed to two types of training datasets ("Original" or "Original + Rotation"). In both types, the scaling conditions were -20% to +20% or -5% to +5%, and data augmentation process was applied to randomly selected images.

3. Image distortion

Image distortion processing was performed to two types of training datasets ("Original" or "Original + Rotation"). In both types, the distortion conditions were -20% to +20% or no distortion, and data augmentation process was applied to randomly selected images.

pixels), and it was used as the validation data.

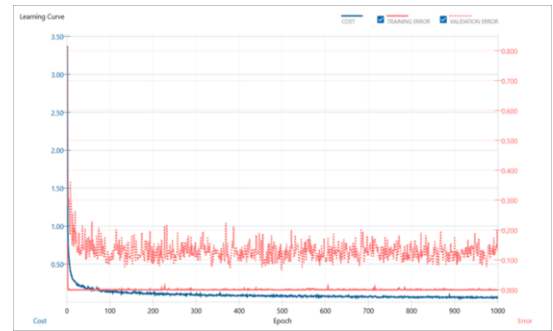


Figure 3. Training curve and loss curve

Hardware and software for deep learning

The hardware consisted of Intel(R) Core (TM) i7-8700 3.2GHz for CPU, NVIDIA GeForce RTX 2070 Super with 8GB memory for GPU (8.2 TFLOPS). Nnabla (SONY. INC) was used as Neural Network Libraries[6], and Anaconda3.0 and Python3.5 were used as the development environment. In this study, Convolutional Neural Network (CNN) based object detection algorithm, ResNet-18 was used for AI model generation of leukocyte recognition. [7]. We performed transfer learning and fine tuning in 1000 epoch with each training dataset, and then we evaluated the accuracy, precision and F-measure of the generated AI models using clinical assessment images. Statistical analysis was carried out with paired t-test. P-value less than 0.05 was considered statistically significant.

Results

The supervised training with ResNet-18 is shown in Fig 3. In any datasets, the validation error converged within 100 epochs, and the loss curve reached a plateau within 300 epochs.

1. Image rotation

Accuracy and F-measure value in data augmentation with image rotation was shown in Table 1. It reached the best accuracy at 3284 mixed images dataset (Dataset D). Statistical analysis showed a significant difference ($p < .05$) in accuracy

Table 2. The effect of image scaling (A) Original images (1064 images), (B) Original and rotated images (3284 images)

A-1 Ratio: $\pm 5\%$							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	85	0	0	0	2	1	0.966
Eosino	0	213	0	0	0	0	1.000
Lymph	0	2	2795	61	0	0	0.978
Mono	1	0	26	533	0	0	0.952
Band	0	0	0	0	465	71	0.868
Segment	0	1	0	1	2204	3335	0.602
Precision	0.988	0.986	0.991	0.896	0.174	0.979	
F-measure	0.977	0.993	0.984	0.923	0.290	0.745	
B-1 Ratio: $\pm 5\%$							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	88	0	0	0	0	0	1.000
Eosino	0	213	0	0	0	0	1.000
Lymph	0	0	2823	35	0	0	0.988
Mono	9	0	27	523	0	1	0.934
Band	0	0	0	0	432	104	0.806
Segment	2	1	0	0	1434	4104	0.741
Precision	0.889	0.995	0.991	0.937	0.232	0.975	
F-measure	0.941	0.998	0.989	0.936	0.360	0.842	
A-2 Ratio: $\pm 20\%$							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	87	0	0	0	0	1	0.989
Eosino	0	213	0	0	0	0	1.000
Lymph	0	1	2810	47	0	0	0.983
Mono	5	0	32	523	0	0	0.934
Band	0	0	1	0	446	89	0.832
Segment	4	1	1	1	1862	3672	0.663
Precision	0.906	0.991	0.988	0.916	0.193	0.976	
F-measure	0.946	0.995	0.986	0.925	0.314	0.789	
B-2 Ratio: $\pm 20\%$							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	88	0	0	0	0	0	1.000
Eosino	0	213	0	0	0	0	1.000
Lymph	1	7	2811	39	0	0	0.984
Mono	2	0	43	515	0	0	0.920
Band	0	0	0	0	489	47	0.912
Segment	5	1	0	1	1272	4262	0.769
Precision	0.917	0.964	0.985	0.928	0.278	0.989	
F-measure	0.956	0.982	0.984	0.924	0.426	0.865	

Table 3. The effect of image distortion (A) Original images (1064 images), (B) Original and rotated images (3284 images)

A-1 Distortion: 0							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	85	0	0	0	2	1	0.966
Eosino	0	213	0	0	0	0	1.000
Lymph	0	2	2795	61	0	0	0.978
Mono	1	0	26	533	0	0	0.952
Band	0	0	0	0	465	71	0.868
Segment	0	1	0	1	2204	3335	0.602
Precision	0.988	0.986	0.991	0.896	0.174	0.979	
F-measure	0.977	0.993	0.984	0.923	0.290	0.745	
B-1 Distortion: 0							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	88	0	0	0	0	0	1.000
Eosino	0	213	0	0	0	0	1.000
Lymph	0	0	2823	35	0	0	0.988
Mono	9	0	27	523	0	1	0.934
Band	0	0	0	0	432	104	0.806
Segment	2	1	0	0	1434	4104	0.741
Precision	0.889	0.995	0.991	0.937	0.232	0.975	
F-measure	0.941	0.998	0.989	0.936	0.360	0.842	
A-2 Distortion: $\pm 20\%$							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	87	0	0	0	0	1	0.989
Eosino	0	213	0	0	0	0	1.000
Lymph	3	0	2819	34	0	2	0.986
Mono	2	0	42	515	0	1	0.920
Band	0	0	0	0	453	83	0.845
Segment	6	2	0	1	1891	3641	0.657
Precision	0.888	0.991	0.985	0.936	0.193	0.977	
F-measure	0.935	0.995	0.986	0.928	0.314	0.786	
B-2 Distortion: $\pm 20\%$							
Expert/AI	Baso	Eosino	Lymph	Mono	Band	Segment	Recall
Baso	87	0	0	0	1	0	0.989
Eosino	0	213	0	0	0	0	1.000
Lymph	0	0	2843	15	0	0	0.995
Mono	0	0	75	485	0	0	0.866
Band	1	0	0	0	480	55	0.896
Segment	3	0	0	1	1735	3802	0.686
Precision	0.956	1.000	0.974	0.968	0.217	0.986	
F-measure	0.972	1.000	0.984	0.914	0.349	0.809	

between original image dataset A (1064 images) and all rotated image datasets (Dataset C/D/E). Whereas statistical analysis showed no significant difference in accuracy between original image dataset B (1302 images) and all rotated image datasets (Dataset C/D/E). Statistical analysis showed no significant difference in accuracy between all rotated image datasets (Dataset C/D/E). The effect by data augmentation with rotation was equivalent to about a 25% increase in the number of original images. Statistical analysis in Average F-measure also showed similar results to accuracy. By cell

category, a significant rise in recognition rate was shown in neutrophils.

2. Image scaling

Confusion matrix of total 9756 images (n=51 cases) in data augmentation with image scaling is shown in Table 2. The AI model generated using the original image dataset showed an accuracy of 0.758 in the $\pm 5\%$ image scaling condition and 0.791 in the $\pm 20\%$ image scaling condition. On the contrary, the AI model generated using the mixture of original image and the rotated image dataset showed an accuracy of 0.835

under $\pm 5\%$ image scaling and 0.855 under $\pm 20\%$ image scaling. Image scaling showed an improvement in accuracy of 0.03 for the original image dataset and 0.02 for the mixed image dataset. Neutrophils showed improved accuracy by data augmentation with image scaling, while other cell groups did not show improved accuracy. Data augmentation with scaling showed slight improvement in leukocyte recognition.

3. Image distortion

Confusion matrix of total 9756 images ($n=51$ cases) in data augmentation with image distortion is shown in Table 3. The AI model generated using the original image dataset showed an accuracy of 0.758 in the $\pm 5\%$ image scaling condition and 0.789 in the $\pm 20\%$ image scaling condition. In contrast, the AI model generated using the mixture of original image and the rotated image dataset showed an accuracy of 0.835 under $\pm 5\%$ image scaling and 0.807 under $\pm 20\%$ image scaling. The original image dataset showed an improvement in accuracy of 0.03, while the mixed image dataset showed a reduction in accuracy of 0.03. Data augmentation with distortion did not show effective improvement in leukocyte recognition.

Discussion

1. Image rotation

In this study, statistical analysis showed a significant difference in accuracy between original image dataset and all mixed image datasets. The +200% augmented data set showed the highest value. This result suggests that data augmentation with image rotation can contribute to higher accuracy in generating AI models for leukocyte morphology recognition. However, the upper limit of the effect of data augmentation was three times that of the original data, it was demonstrated that the necessity of limiting the application of data augmentation with this method to less than three times.

2. Image scaling

Both the original and mixed image datasets showed an improvement in accuracy by data augmentation with image scaling. However, the effect was small and limited to improving the recognition of segmented neutrophils. Therefore, it is suggested that image scaling is an effective method to improve the accuracy in recognition of leukocytes with complex nuclear structures. However, some segmented neutrophils were recognized as basophils or eosinophils when the impact of image scaling was expanded. It is suggested that the size of intracytoplasmic granules changed by scaling and matched the characteristics of intracytoplasmic granules of basophils or eosinophils. These results demonstrated that data augmentation with image scaling was a useful method for the recognition of leukocytes, which had complex nuclear structures and few intracytoplasmic granules. On the contrary, this method was not considered suitable for data augmentation for leukocyte recognition among different multiple cell categories including mononuclear cells.

3. Image distortion

It was suggested that data augmentation with distortion does not contribute to the diversity of leukocyte images. This reason may be that the image distortion effect was weak. However, the image distortion effect has a different impact on the center and the periphery of the image. Strong image distortions are leukocyte morphologies that do not exist, and the AI may learn wrong morphological features. These results

demonstrated that data augmentation with image distortion was not a useful method for the recognition of leukocytes.

Although the diversity of images in the training data is effective in generating highly accurate AI models [8, 9], our results of the three approaches demonstrated that it was necessary to selectively apply an effective technique for each AI model in data augmentation for leukocyte recognition.

Previous studies have reported the highly accurate AI models for leukocyte recognition were generated with comparing the number of layers CNN models or different algorithm CNN models [10, 11]. However, in all of them, their approaches performed the number of labeled images for training was less than 20,000 images, and a lot of studies performed image augmentation for AI generation. This suggests a lot of studies were not able to collect sufficient leukocyte images or the upper limit of GPU performance has been reached for each AI model generation. Furthermore, it is reported that up to 100 hours or more of calculation was performed to generate one AI model. However, extracting more morphological features of leukocyte and generating more accurate AI models for hematological diagnosis requires more computing performance and large memory.

Kimura et al. [12] studied an AI model that combined CNN and XGBoost, and they reported that it showed more than 93.5% accuracy in discriminating between myelodysplastic syndromes (MDS) and Aplastic anemia (AA). It required the extraction of 17 cell types and 97 morphological features for the diagnosis of these two extremely difficult to distinguish diseases. In addition, 695030 labeled images were used for AI training. Their study shows that the generation of AI models for immature or reactive leukocyte recognition identification or diagnosis requires extremely high computational performance and a large number of images for AI training. However, general-purpose computational hardware performance for AI model computation has reached its upper limit, and innovation of GPU performance or reduction the amount of calculation is required. Therefore, in future work, the generation of AI for hematological morphology diagnosis will also require approaches both diversification of the images and optimization of the number of images in the training dataset. Our approach shows that a suitable data augmentation method should be selected to obtain an optimized image dataset for AI training from a small number of images in mature leukocyte recognition. In our future work, we will study effective data augmentation methods for the recognition of abnormal and reactive leukocytes.

Conclusion

The effect of data augmentation with rotation was the only significant effective technique in AI model generation for leukocyte recognition. On contrast, the effect of data augmentation with image distortion or image scaling was poor. Both image distortion and image scaling did not show significant difference. Furthermore, accuracy improvement was limited to specific leukocyte categories. Although data augmentation is one effective method for high accuracy in AI training, we consider that a highly effective method should be selected.

Acknowledgments

This work was supported by SCOPE of the Japan Ministry of Internal Affairs and Communications and JSPS KAKENHI 19K21737.

References

- [1] Miotto R, Wang F, Wang S, Jiang X, and Dudley JT, Deep learning for healthcare, *Brief Bioinform* 19(6) (2018), 1236-1246. doi: 10.1093/bib/bbx044.
- [2] Chan HP, Samala RK, Hadjiiski LM, and Zhou C, Deep Learning in Medical Image Analysis, *Adv Exp Med Biol* 1213 (2020), 3-21. doi: 10.1007/978-3-030-33128-3_1.
- [3] J Wang, and L Perez, The effectiveness of data augmentation in image classification using deep learning, (2017). <https://arxiv.org/abs/1712.04621> (accessed May 14, 2021).
- [4] A Mikołajczyk, and M Grochowski, Data augmentation for improving deep learning in image classification problem, 2018 International Interdisciplinary PhD Workshop, (2018), 117-122. doi: 10.1109/IIPHDW.2018.8388338.
- [5] R Takahashi, T Matsubara, and K Uehara, Data Augmentation Using Random Image Cropping and Patching for Deep CNNs, *IEEE Transactions on Circuits and Systems for Video Technology* 30(9) (2020), 2917-2931. doi: 10.1109/TCSVT.2019.2935128.
- [6] A Hayakawa, M Ishii, Y Kobayashi, A Nakamura, T Narihira, Y Obuchi, A Shin, T Yashima, and K Yoshiyama, Neural Network Libraries: A Deep Learning Framework Designed from Engineers' Perspectives, (2021). <https://arxiv.org/abs/2102.06725> (accessed May 14, 2021).
- [7] K He, X Zhang, S Ren, and J Sun, Deep Residual Learning for Image Recognition, (2015). <https://arxiv.org/abs/1512.03385> (accessed May 14, 2021).
- [8] H Shin, N A. Tenenholtz, J K. Rogers, C G. Schwarz, M L. Senjem, J L. Gunter, K P. Andriole, and M Michalski, Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks, *Simulation and Synthesis in Medical Imaging* (2018), 1-11.
- [9] D Mahajan, R Girshick, V Ramanathan, K He, M Paluri, Y Li, A Bharambe, and L Maaten, Exploring the Limits of Weakly Supervised Pretraining, (2018). <https://arxiv.org/abs/1805.00932> (accessed May 14, 2021).
- [10] Reena MR, and Ameer PM, Localization and recognition of leukocytes in peripheral blood: A deep learning approach, *Comput Biol Med* 126 (2020), 104034. doi: 10.1016/j.combiomed.2020.104034
- [11] Wang Q, Bi S, Sun M, Wang Y, Wang D, and Yang S, Deep learning approach to peripheral leukocyte recognition, *PLoS One* 14(6) (2019), e0218808. doi: 10.1371/journal.pone.0218808.
- [12] K Kimura, Y Tabe, T Ai, I Takehara, H Fukuda, H Takahashi, T Naito, N Komatsu, K Uchihashi, and A Ohsaka, A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA, *Sci Rep* 9(1) (2019), 13385. doi: 10.1038/s41598-019-49942-z.

Address for correspondence

Hiroyuki Nozaka
Hirotsuki University Graduate School of Health Sciences
66-1, Honcho, Hirotsuki-shi, Aomori 036-8564, Japan
E-mail: hnozaka@hirosaki-u.ac.jp
Phone: +81-172-39-5918