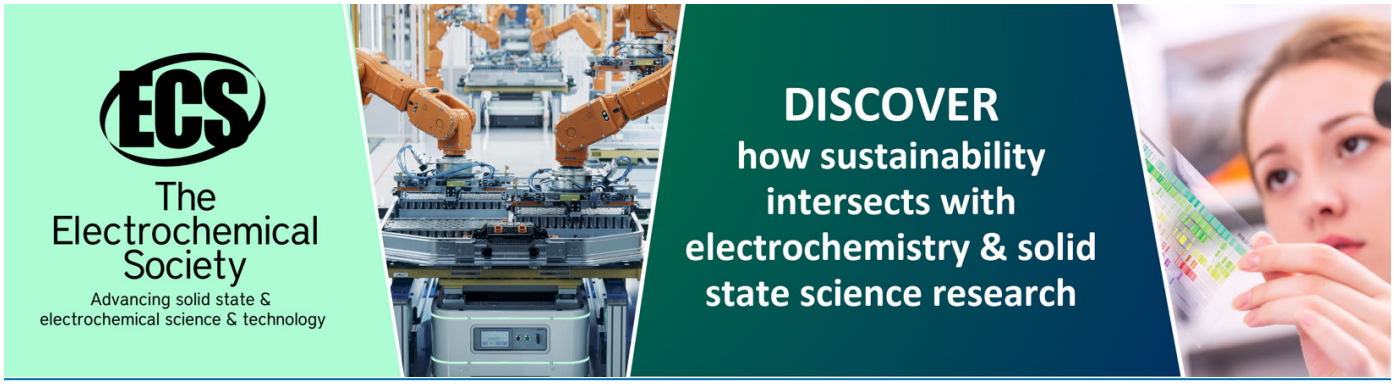
Augmentation on CNNs for Handwritten Digit Classification in a Small Training Sample Size Situation



Journal of Physics: Conference Series **1922** (2021) 012007 doi:10.1088/1742-6596/1922/1/012007

Augmentation on CNNs for Handwritten Digit Classification in a Small Training Sample Size Situation

Y Mitani[1], Y Fujita[2] and Y Hamamoto[2]

[1]National Institute of Technology, Ube College, Ube, Japan
[2]Faculty of Engineering, Yamaguchi University, Ube, Japan

mitani@ube-k.ac.jp

Abstract. In general, a deep learning needs a lot of samples. However, in a practical pattern recognition problem, the number of training samples is usually limited. We investigate the effect of an image data augmentation by a perspective transformation on a convolution neural network(CNN) for handwritten digit classification in a small training sample size situation. The experimental results show the effectiveness of the image data augmentation by the perspective transformation on the CNN for handwritten digit classification particularly in the small training sample size situation.

1. Introduction

A considerable amount of effort has been devoted to design a classifier in small training sample size situations [1],[2]. It is known that the larger training samples, the better the classification performance of a classifier. However, available samples are limited. It is difficult to correct many samples in order properly to train a classifier, because this is a time consuming task for a human. We as human must give a label or class name for each sample. Therefore, it has been desirable to develop a pattern recognition system which can work well even in small training sample size situations.

A convolution neural network(CNN) has been successfully applied in the image recognition field [3],[4]. The CNN is reported to be one of promising classifiers in a handwritten digit classification problem [5],[6]. The CNN originates from the artificial neural networks [7]. The CNN has been developed to be suitable for an image recognition problem. By using the CNN, we do not have to care what features or what classifier we should use. The CNN can automatically learn a relationship between input and output. Instead of properly training a lot of hyperparameters of the CNN, the CNN needs many training samples. It is well known that there is a trade-off between a generalization performance and an over-training. In a practical pattern recognition problem, the number of available samples is usually limited. Therefore, it is important to consider designing a CNN in small training sample size situations. R. Keshari et al.[8] have addressed the issue about small sample size training in designing a CNN. Their work focused on learning the structure and strength of filters. Recently, in designing a CNN in small training sample size situations, data augmentation [9],[10] is promising. In the data augmentation approach, a small deviation of the data is expected to lead to enhance a generalization performance of the CNN. The authors showed a positive effect of an augmentation by a perspective transformation on CNNs in classifying a cirrhosis liver on B-mode ultrasound images [11]. The liver image is a medical image which tends to be a small sample size. This is considered to be a type of texture. By the use of this augmentation on CNNs, the texture image recognition problem seems useful. We have considered to try using the data augmentation to another type of pattern

recognition problem such as a character recognition, especially in a small training sample size situation.
In this paper, we investigate the effect of an image augmentation by a perspective transformation in classifying handwritten digits on the CNN in a small training sample size situation. We try using an image data augmentation by affine and perspective transformations. The experimental results show the effectiveness of the image data augmentation by a perspective transformation.

2. Image data augmentation

2.1. MNIST dataset

In this study, we used MNIST (Modified National Institute of Standards and Technology) image dataset [12]. Figure 1 shows a part of MNIST images. There is a distortion because the letters are written by hand. The MNIST dataset is widely used for a handwritten digit classification problem.
It is also used in evaluating a generalization performance of a CNN. The MNIST dataset consists of handwritten digits with available images. Among them, we used 22,000 images. This is a 10-class problem. They are made by from 0 to 9. We used 2,200 images for each class. We assumed there were not imbalanced classes. The image size is 28x28. The image is a gray scale.

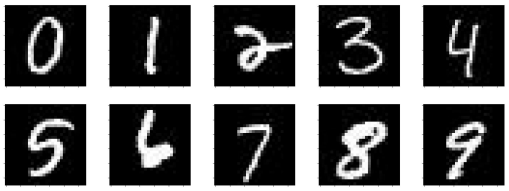


Figure 1. A part of MNIST images.

2.2. CNN architecture

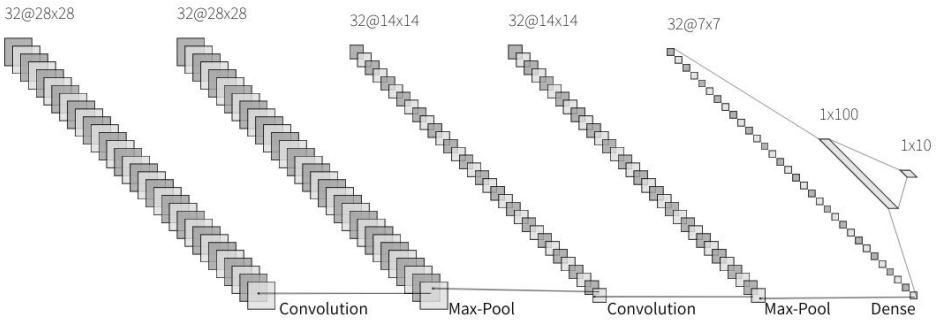


Figure 2. A network structure of the CNN we used.

We show the CNN architecture. The generalization performance of the CNN depends on its network structure and the parameters to be determined. Figure 2 shows a network structure of the CNN we used. The input of the CNN is the MNIST image of size 28x28. First, we convolve the image by using 32 filters with a 3x3 filter size. And by 2x2 maxpooling, we reduce the image size to a half-sized image, 14x14. Second, we repeatedly convolve and do max-pooling in the same manner. Then, we get 32 7x7 sized images. Third, we flatten this image into 1,568(=32x7x7) dimensional data. Finally, we make a fully connected artificial neural network. The network has one hidden layer. The number of the neurons also depends on the generalization performance of the CNN. In the experiment, we used 100 for simplicity. Then we used dropout. The rate of dropout is 0.5. The number of the outputs of the CNN is 10. Because this is a 10-class problem as we mentioned above. Therefore, the structure of the fully connected artificial neural network is 1,568-100-10. All the activation functions are ReLU except for the output. In the output, we used softmax. The learning optimizer is adam. The epochs and batch size are 100 and 100, respectively.

2.3. Image data augmentation techniques

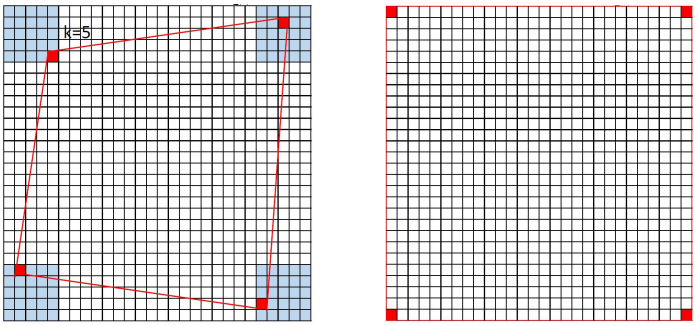


Figure 3. An example of a perspective transformation.

We show image data augmentation techniques. In this study, we have tried using affine and perspective transformations. The affine transformations include shear, rotation, width-shift, and height-shift [13]. It is expected for these techniques to improve the generalization performance of the CNN. On the other hand, we did not use image data augmentation techniques such as brightness, zoom, channel shift, horizontal flip, and vertical flip [13]. These techniques are considered to be unconcerned with the improvement of the generalization performance of the CNN.
We describe a perspective or a homography transformation. Note that the image data augmentation of [13] does not include the perspective transformation. We can see many images which differ their appearance from an image by the perspective transformation. The perspective transformation is defined by the following equation.

$$x_i' = \frac{a_1 x_i + a_2 y_i + a_3}{a_7 x_i + a_8 y_i + 1} \quad (1)$$

$$y_i' = \frac{a_4 x_i + a_5 y_i + a_6}{a_7 x_i + a_8 y_i + 1} \quad (2)$$

The coordinates (x,y) is a point of an original image. On the other hand, (x',y') is a point of an image newly generated by the perspective transformation. Figure 3 shows an example of a perspective transformation. Figs (a) and (b) are an original image when k=5 and a new image, respectively.

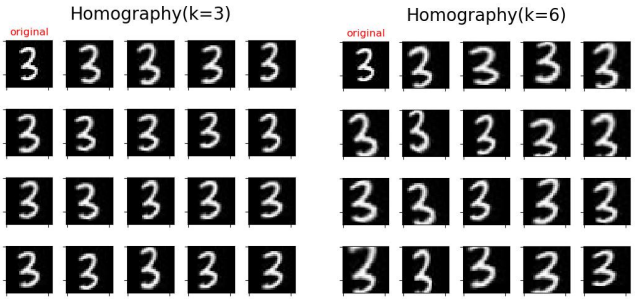


Figure 4. Examples of perspective transformations.

The perspective transformation matrix is computed by the relationship between at least 4 points between an original image and its image to be newly generated. The 4 points of the new image are set to be 4 corners, top left, top right, bottom left, and bottom right, as shown in (b). On the other hand, the 4 points of the original image are set to be within 4 rectangle areas. In an example, we randomly selected 4 points from each blue rectangle area within k x k of 4 corner points of an original image as shown in (a). Every a computation of the perspective transformation matrix, we could get an artificially newly generated image. Figure 4 shows examples of perspective transformations. Figs (a) and (b) show images when k=3 and k=6, respectively. For each of figs (a) and (b), in the top left, the image of 3 is an original image from the MNIST dataset. Another 19 images of 3 are artificially newly generated from a top left original image by perspective transformations. Each of these images differs a

little from the original image. The image distortion of (b) seems larger than that of (a). We see the parameter values of k vary image distortion. The larger the value of k is, the more the image distortion is. Therefore, in using this perspective transformation, we should take care of choosing a value of k.

In the image data augmentation, we can get an arbitrary number of images artificially generated from an image. Given 20 training images for each class, we could get 20x training images. In all the experiments of image data augmentation, we used x=100. Therefore, we used 2,000 training images per a class. Note that original images are included in these images.

3. Experiments

We used 22,000 available MNIST images [12]. In the experiments, we used 2,000 training images and 200 test images for each class. The effectiveness of the CNN is examined in terms of the error rate. We evaluate the generalization performance of the CNN by the error rate. The error rate is defined as a ratio of the number of test images misclassified to the number of all test images.

$$Error\ rate = \frac{\#\ test\ images\ misclassified}{\#\ all\ test\ images} \times 100(\%) \tag{3}$$

For error rate estimation, the holdout method has been successfully used, because it maintains the statistical independence between the training and test images [14],[15]. To evaluate the generalization performance of the CNN, the average error rate was obtained by the holdout method. By 5 repetitions, the average error rate and 95% confidence interval were obtained.

3.1. Generalization performance for the number of training images

Table 1. Average error rate(%) and 95% confidence interval of the CNN for the number of training images.

20	100	200	1000	2000
19.71±1.95	6.70±0.74	4.36±0.94	1.71±0.39	1.49±0.37

The purpose of the experiment is to investigate the generalization performance of the CNN for the number of training images in terms of the average error rate. The numbers of the training images for each class were 20, 100, 200, 1000, and 2000. The number of test images for each class was 200. Table 1 shows the average error rate and 95% confidence interval of the CNN for the numbers of training images. From Table 1, we see that the larger training images, the better the generalization performance of the CNN. When the number of training images is 2000 for each class, the generalization performance of the CNN is the best in our experiment. The average error rate was 1.49%. On the other hand, when the training image size is 20 per a class, the average error rate was the worst, 19.71%. In the following experiments, we try exploring the effect of the image data augmentation of the CNN when the training image size is 20. This assumes a practical situation of the small training sample size. We would like to improve the generalization performance of the CNN particularly in a small training sample size situation.

3.2. Generalization performance with augmentation

Table 2. Average error rate(%) and 95% confidence interval of the CNN without and with the image data augmentation.

No augmentation	Shear	Rotation	Width-Shift	Height-Shift	Perspective
19.71±1.95	20.80±5.10	16.51±3.52	18.62±3.80	18.42±2.97	*6.42±2.20*

The purpose of the experiment is to investigate the generalization performance of the CNN without and with an image data augmentation in terms of the average error rate. The numbers of the training and test images for each class were 20 and 200, respectively. The number of the image data augmentation was 2,000 per a class. As we mentioned above, 2,000 training images were artificially newly generated from 20 original images per a class. For affine transformation techniques, we used

that ranges of share and rotation were 5, 10, 20, and 30. The width- and height- shift ranges were 0.01, 0.05, 0.1, and 1.0. We used k=2, 4, and 6 for a perspective transformation technique. From the results of our experiments, we chose values which gave the best average error rates. Both the ranges of share and rotation were 20. The width- and height- shift ranges were 0.1 and 0.05, respectively. When k=4, the generalization performance of the CNN with augmented images by the perspective transformation showed the smallest average error rate. Table 2 shows the average error rate and 95% confidence interval of the CNN without and with the image data augmentation. From Table 2, we see the CNN with the image data augmentation by a perspective transformation is superior to the CNN without augmentation. The average error rate showed the smallest, 6.42%. This result seemed to be comparable with the result when the training image size is 100 as shown in Table 1. The image data augmentation by the affine transformation of rotation ranked next to that by the perspective transformation. The generalization performance of the CNN with image data augmentation by rotation, width-shift, height-shift, and perspective transformation outperforms that of the CNN without image data augmentation. Among them, we see that the effect of the perspective transformation is overwhelming. On the other hand, the generalization performance of shear seemed to be poor.

3.3. Effects of the perspective transformation for values of k

Table 3. Average error rate(%) and 95% confidence interval of the CNN with the image data augmentation by the perspective transformation for values of k.

k=2	k=4	k=6
6.86±1.56	*6.42±2.20*	7.53±2.35

The purpose of the experiment is to investigate the generalization performance of the CNN with an image data augmentation by the perspective transformation in terms of the average error rate. The numbers of training and test images for each class were 20 and 200, respectively. The number of the image data augmentation was 2,000 per a class. This situation is the same of the previous experiment. We used k=2, 4, and 6 for a perspective transformation. Table 3 shows the average error rate and 95% confidence interval of the CNN with the image data augmentation by the perspective transformation for values of k. From Table 3, we see there is an optimal value of k in terms of the average error rate. The average error rate goes worse if the value of k is too larger or too smaller. In a perspective transformation, we should take care of a selection of an optimal value of k.

4. Conclusion

In this paper, we have investigated the effect of an image data augmentation by a perspective transformation on the CNN for handwritten digit classification in a small training sample size situation. From our experimental result with the image data augmentation by a perspective transformation, the generalization performance of the CNN has been improved particularly in a small training sample size situation. We tried exploring various combinations of these image data augmentation techniques we used in the experiments, but the results lead to be poor. In this handwritten digit classification problem, the positive effect of an image data augmentation by a perspective transformation is shown like the previous work [11]. In designing a CNN, we should recommend considering the use of an image data augmentation by a perspective transformation.

In the future, we investigate applying another type of character recognition problem such as Kanji or Kuzushiji [16]. It is known that these include much distortion. This may be a more challenging task particularly in small training sample size situations. We also try exploring another type of the image recognition problem to investigate effects of the CNN with image data augmentation by the perspective transformation. We should try applying transfer learning [17] to the image recognition in small training sample size situations. Furthermore, a combination of image data augmentation and transfer learning should be investigated.

5. References

[1] Raudys S J and Jain A K 1990. Small sample size effects in statistical pattern recognition: Recommendations for practitioners and open problem. *Proc. 10th International Conference on Pattern Recognition*, pp 417-423
[2] Raudys S J and Jain A K 1991. Small sample size problems in designing artificial neural networks. *Machine Intelligence and Pattern Recognition*, **11**, pp 33-50
[3] Hinton G E, Osindero S and Teh Y W 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, **18**, no. 7, pp 1527-1554
[4] LeCun Y, Bengio Y and Hinton G 2015. Deep learning. *Nature*, **521**, no. 28, pp 436-444
[5] Kumar K and Beniwal H. 2018. Survey on handwritten digit recognition using machine learning. *International Journal of Computer Sciences and Engineering*, **6**(5), pp 96-100
[6] Sultana F, Sufian A and Dutta P. 2019. Advancements in image classification using convolutional neural network. arXiv:1905.03288
[7] Rumelhart D E, Hinton G E and Williams R J. 1986. Learning representations by back propagation errors. *Nature*, **323**, no. 9, pp 533-536
[8] Keshari R, Vatsa M, Singh R and Noore A. 2018. Learning Structure and Strength of CNN Filters for Small Sample Size Training. *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp 9349-9358
[9] Shorten C and Khoshgoftaar T M. 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, **6**.60
[10] Sato I, Nishimura H and Yokoi K. 2015. APAC: Augmented PAttern Classification with Neural Networks. arXiv:1505.03229
[11] Mitani Y, Fisher R B, Fujita Y and Hamamoto Y. 2020. Effect of an augmentation on CNNs in classifying a cirrhosis liver on B-mode ultrasound images. *Proc. 2020 IEEE 2nd Global Conference on Life Sciences and Technologies*, pp 255-254
[12] LeCun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**, no. 11, pp 2278-2324
[13] Keras Documentation. Last accessed on 2020-03-24. Image Preprocessing, Image Data Generator class. https://keras.io/preprocessing/image/
[14] Duda R O, Hart P E and Stork D G. 2001. *Pattern Classification*. Second Edition, Wiley Interscience
[15] Fukunaga K. 1990. *Introduction to statistical pattern recognition*. Second Edition, Academic Press
[16] Clanuwat T, Bober-Irizar M, Kitamoto A, Lamb A, Yamamoto K and Ha D. 2018. Deep learning for classical Japanese literature. arXiv:1812.01718
[17] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H and He Q. 2019. A comprehensive survey on transfer learning. arXiv:1911.02685