Data Augmentation & Merging Dataset for Facial Emotion Recognition

Jung Hwan Kim, Dong Seog Han
Graduate School of Electronics and Electrical Engineering, Kyungpook National University

jkim267@knu.ac.kr, dshan@knu.ac.kr

Abstract

The facial emotion recognition (FER) system is desirable for many research fields such as game development, social workers, and the autonomous driving vehicles. The most popular research datasets, called FER 2013 and Extended Cohn-Kanade (CK+), were mainly tested by FER's researchers. However, testing only one dataset as FER 2013 or CK+ sometimes had limited further improvement of FER's performance even after data augmentation. Since the sampling FER datasets heavily affected to the FER's performance, we propose that merging datasets could be another method to improve the FER's performance other than using the data augmentation technique. By merging different datasets to magnify the number of training facial images, the FER performance improved 15.33% of validating accuracy.

I. Introduction

The facial emotion recognition (FER) is the future instrument for the game industry, social workers, and developing the autonomous driving vehicles. Improving FER system's performance is still in the primitive stage due to the scarcity of the FER datasets [5-6] for many FER researchers. The small number of facial images for training could lead the overfitting problem when the recent sophisticated neural networks and data augmentation were applied. Some FER researchers were decided to more facial images to improve the further FER's performance. But, others claimed that applying data augmentation on a small dataset could solve the data deficiency.

In essence, Kim *et al* [1] claimed that applying data augmentation onto the small number of facial images in FER dataset robustly solved a problem of the FER dataset's scarcity. They used the extended Cohn-Kanade (CK+) [6] dataset and applied with data augmentation to randomly manipulate facial images' transformation. Still, training with a small number of facial images could lead the bias result. We discovered that the given small number of face images to train performed well, but badly performed on newly detecting facial images even after the data augmentation.

In addition, Rosebroke [2] explained the fundamental concept of data augmentation. The data augmentation randomly added the jitteriness onto the original dataset's distribution, and magnify randomness of the dataset. The data augmentation generally solved an insufficient number of training images in a dataset, yet Sakai *et al* [3] displayed results which the large number of collected bio signals sometimes showed better performance than the small number of bio signals with data augmentation.

In this paper, we compared the performance of the Xception algorithms with and without applying the data augmentation on the FER 2013 [5] dataset. After the comparison, we collected the additional facial images and combined with different FER datasets to inspect the result of increasing the number of the dataset.

II. With and Without the Data Augmentation

FER 2013 had 48×48 pixels size of 35,813 facial images and 7 different categorical emotions: angry, disgust, fear, happy, neutral, sadness, and surprise. The first result from Fig. 1 without applying the data augmentation showed the overfitting problem during the training process. After applying the data augmentation of that data, the overfitting problem was resolved, and the performance was slightly improved from Fig. 2.

However, applying data augmentation on the small number of facial images could not improve the FER's performance further. Therefore, we were deterministic that increasing the number of facial images could potentially improve the FER's performance.

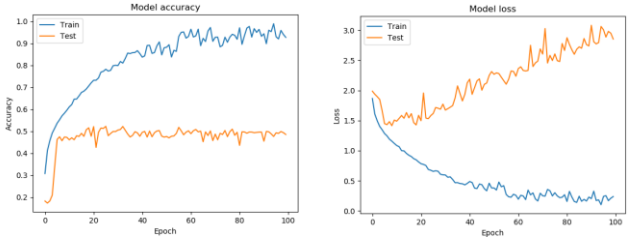

Fig. 1 Training the Xception model without data augmented FER 2013 Dataset.

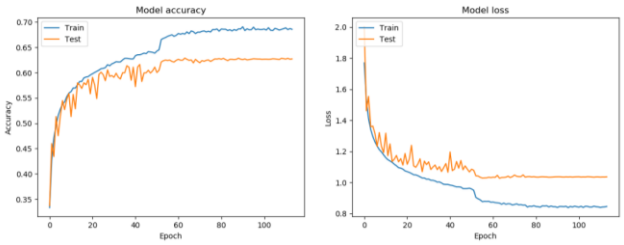

Fig. 2 Training Xception Model with data augmented FER 2013 Dataset.

III. Merging Datasets and Test Results

CK+ from Fig. 3 has 640×490 pixels size of 918 facial images within 8 categorical classifications: angry, contempt, disgust, fear, happy, neutral, sadness, and surprise. These facial images did not properly crop and a large portion of background pixels. Besides, training with the small number facial images even after applying data augmentation could lead the poor performance with the unseen dataset. The training result from CK+ dataset showed far superior than training with FER 2013 dataset, yet the model hardly detected the new emotional faces from unseen facial images. Training only CK+ dataset would not reach the robust FER performance.

We collected facial images from 60 video clips on YouTube and created intelligent signal processing lab (iSPL) dataset at Kyungpook National University from Fig. 3. The iSPL dataset contained 8,173 valid facials images and 7 categorical emotions as FER 2013. Although the performance showed better performance than the FER 2013 and CK+, the testing unseen dataset was still unable to detect the new facial emotions. Hence, we decided to merge all datasets together.

All facial images from different datasets such as FER 2013, CK+, iSPL have different size and different position of faces. To merge those different datasets without concerning of such the different sizes and position of facial images, we created the facial images threshing (FIT) machine. The FIT machine contained the multi-task cascade neural network (MTCNN) [7] and resizing program [2] that could symmetrically match to FER 2013 dataset. After all facial images became standardized to FER 2013's size and cropped faces by the FIT machine, we conducted a final experiment with the merged dataset.

The final result of the Xception algorithms and merged datasets from Fig.4 reached 76.32% of validating accuracy and increased 15.33% comparing with the 60.99% from Fig. 2. The experiment was applied with data augmentation in order to prevent from possible over-fitting problem. From Table I, applying confusion matrix's evaluation could confirm the robust improvement of the FER's performance. We used the unseen private facial images from the FER 2013 as simulating real-time testing.

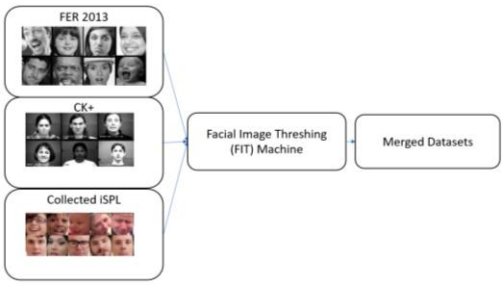

Fig. 3 Merging FER 2013, CK+, and iSPL Datasets by the FIT machine.

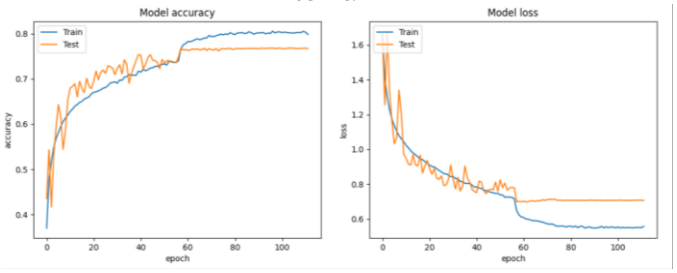

Fig.4 The final result from the merged datasets and the Xception algorithms.

Table I. The Result of Confusion Matrix from the unseen private test.

Datasets	Precision	Recall	F1 Score
FER 2013	61.6532%	58.7689%	59.4004%
Merged Dataset	**66.6236%**	**66.8845%**	**66.6779%**

IV. Conclusion

To conclude, data augmentation prevents from the over - fitting problem and also generalize the entire dataset. However, augmenting the small number of facial images by merging additional dataset proved to have further improvement of FER's performance but not with data augmentation.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding)

References

[1] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," IEEE Access, vol. 7, pp. 41273-41285, 2019.
[2] A. Rosebroke, Deep Learning for Computer Vision with Python, pp. 14-29, 2017.
[3] A. Sakai, Y. Minoda, and K. Morikawa, "Data augmentation methods for machine-learning-based classification of bio-signals," in 2017 10th Biomedical Engineering International Conference (BMEiCON), pp. 1-4.
[4] F. Chollet, "Xception: Deep learning with depthwise separable convolutions", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251-1258.
[5] P.-L. Carrier, A. Courville, I. J. Goodfellow, M. Mirza, and Y. Bengio, "FER-2013 face database," Universit de Montral, 2013.
[6] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, 2010, pp. 94-101, doi:10.1109/CVPRW.2010.5543262.
[7] J. Brownlee, Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python: Machine Learning Mastery, 2019.