A Study of a Data Standardization and Cleaning Technique for a Facial Emotion Recognition System

Jung Hwan Kim, Ronald Mutegeki, Alwin Poulose, and Dong Seog Han*

School of Electronics Engineering, Kyungpook National University

jkim267@knu.ac.kr, rmutegeki@knu.ac.kr, alwinpoulosepalatty@knu.ac.kr, dshan@knu.ac.kr*

Abstract

In this paper, we introduce a data standardization and cleaning technique with different facial emotion recognition (FER) datasets in order to improve the FER model system. We discovered some irrelevant facial images in FER 2013 dataset after we successfully converted the comma-separated values (CSV) file into actual images. For FER, we employ the mini-Xception model and in order to improve its performance, all irrelevant facial images must be discarded. After discarding these irrelevant images, we saw the gap between training and testing accuracy and error reducing. Moreover, we obtained a variety of FER datasets such as Cohn-Kanade, Japanese Female Facial Expression (JAFFE), and EmotionNet. We cropped the images using multi-task cascade neural network (MTCNN), and resized images using Pyimagesearch, a tool for resizing images, because each dataset that we obtained had images of different sizes and was not cropped. This technique for data standardization and cleaning enables us to use the Cohn-Kanade dataset for testing and the FER 2013 for training even though initially, they have images of different sizes. As a result, the validation accuracy is increased by 5%, and the validation loss decreases by 2%. We recommend data standardization and cleaning to obtain and use more raw data images in order to improve the performance of the FER model produced by mini-Xception Algorithm.

I. Introduction

A facial emotion recognition (FER) system can be useful in detecting pwesonal emotion, helping psychotherapists, and tracking child's emotion development. In our research, we aim to detect human emotion and improve the performance of a popular FER system. Since AlexNet [1], the convolution neural network (CNN) has become prevalent in object recognition, detection, and segmentation tasks, and has made these tasks trivial. We used mini-Xception [2], an algorithm from 2016 whose performance we try to improve on the performance by obtaining more datasets.

We use the facial emotion recognition 2013 (FER 2013) dataset in a comma separated values (CSV) format for the FER system. CNNs require very high volumes of training data and obtaining more datasets besides FER 2013 was paramount. To reach peak performance, we are compelled to obtain a significant number of good quality image datasets that should ideally outperform a larger number of bad quality datasets [3]. However, there were several challenges in improving the FER system. Most datasets have different image sizes and faces themselves may not be cropped properly. Add to that the fact that CNNs do not have flexible input sizes. Additionally, the size of our dataset is large and so to resize and crop individual images manually will not suffice.

In this paper, we propose a data standardization and cleaning technique and show the results after applying it. The rest of the paper is organized as follows. Section II presents the data cleaning strategy. In Section III, we discuss the data standardization technique and analyze the results and conclude the paper in Section IV.

II. Data Cleaning

The FER2013 CSV formatted dataset contains seven classes; angry, disgust, fear, happiness, neutral, sadness, and surprised. We choose and tested mini-Xception since the Xception algorithm showed better performance than inception algorithm [2], and it is faster training than the big-Xception algorithm. We discovered that the mini-Xception algorithm has a limitation to reach to the ideal performance of FER. We converted the CSV file into image files in Python and then manually inspected these image files. We expected all these images to be similar to Fig. 1 which are relevant to facial emotions but some images as shown in Fig. 2, are not relevant to human facial emotions. Out of the 30,000 images in the entire dataset, we discovered 83 irrelevant images that we discarded.

Fig. 1 Example of facial images from FER2013.csv.

Fig. 2 Example of irrelevant images from FER 2013.csv.

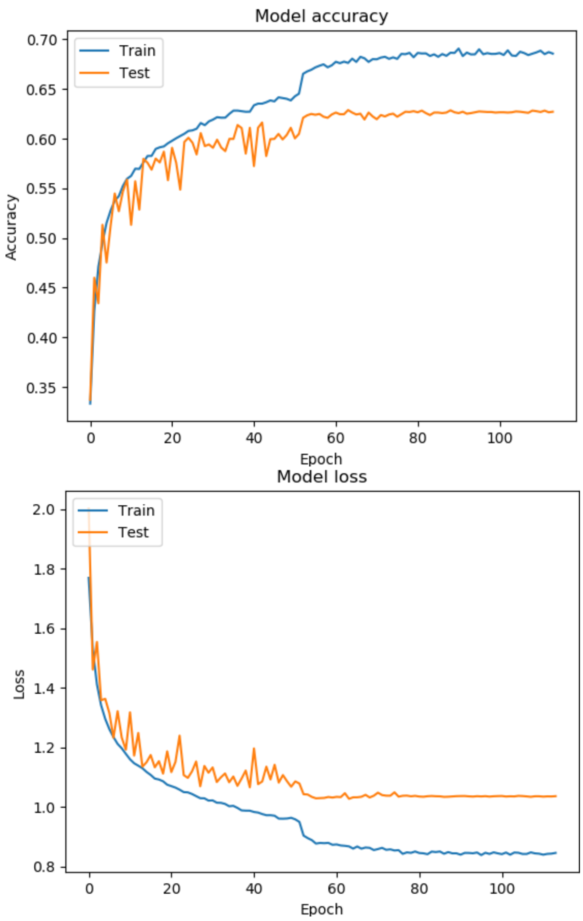

Fig. 3 FER Performance before discarding the irrelevant images.

After we discarded the irrelevant images manually, we discovered that the gaps between the training and testing losses and accuracies are reduced as can be seen in Figs. 3 and 4. The training accuracy and loss slightly decrease while the validation accuracy and loss slightly improve.

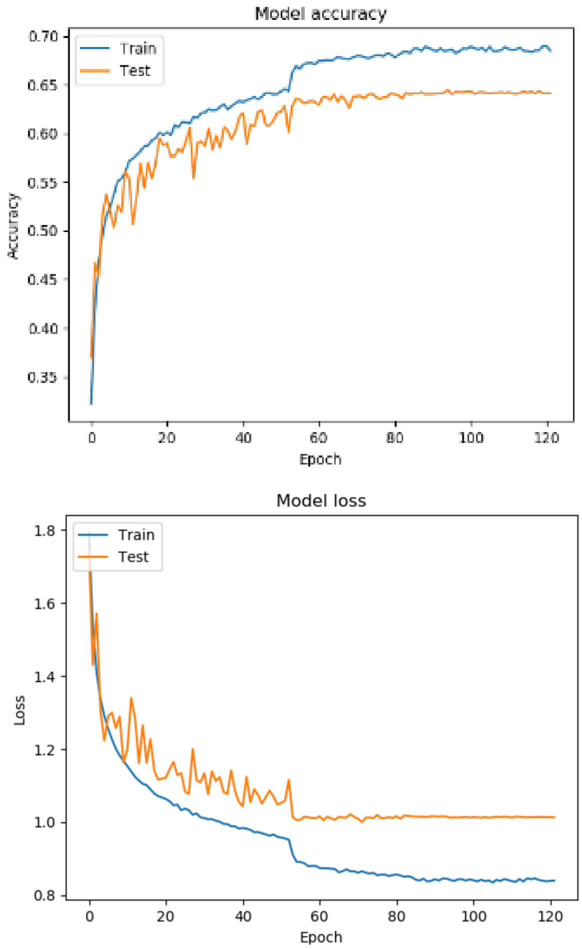

Fig. 4 FER performance after discarding the irrelevant images.

III. Data Standardization

We are compelled to add additional number of good quality image datasets into the existing FER 2013 dataset. Then, the new performance of the mini-Xception model system should ideally outperform than the previous system which is trained with the existing FER 2013 dataset only. So, we obtained the Cohn-Kanade Plus (CK+), and JAFFE, and EmotionNet from the datasets referenced by Li and Deng [3]. However, adding any datasets into FER 2013 dataset required to resize and crop the raw images because the input's width and height of all images has to match the size of FER which is 48 by 48 pixels.

The size of CK+, on the other hand, is 640 by 490 pixels in width and height respectively and does not match FER 2013's image size. Adding more data without matching the image size of FER 2013, the reference dataset, and the presence of a large portion of the background as part of the facial images from CK+ leads the model to perform poorly.

Fig. 5 CK+ dataset without data standardization technique.

Fig. 6 CK+ dataset with data standardization.

Fig. 5 shows the images from CK+ with a large portion of the background which is not relevant to the face. FER 2013 images do not have much of a background unlike the default CK+ images. With the OS walk function, the MTCNN [5], and resizing technique [4], we easily cropped and resized this large number of images and matched these images to the size of FER 2013 dataset as shown in Fig. 6.

After we added the CK+ dataset as a testing dataset for FER 2013, the validation accuracy increased by 5%, and loss decreased by 2% as shown in Fig. 7. Therefore, comparing with the results in Fig. 3, we realize a better performance after the added testing data with the data standardization and cleaning technique although we applied with the data augmentation technique.

IV. Conclusion

We have learned why a larger number of good datasets is important to reach peak performance of any model system with FER as our benchmark. We also learned that a massive number of datasets can easily be standardized using the data standardization and cleaning technique proposed in this paper without wasting a precious time of manual cropping and resizing a large number of images. In our future research, we intend to obtain additional datasets to see how we can improve the different performance from FER model systems.

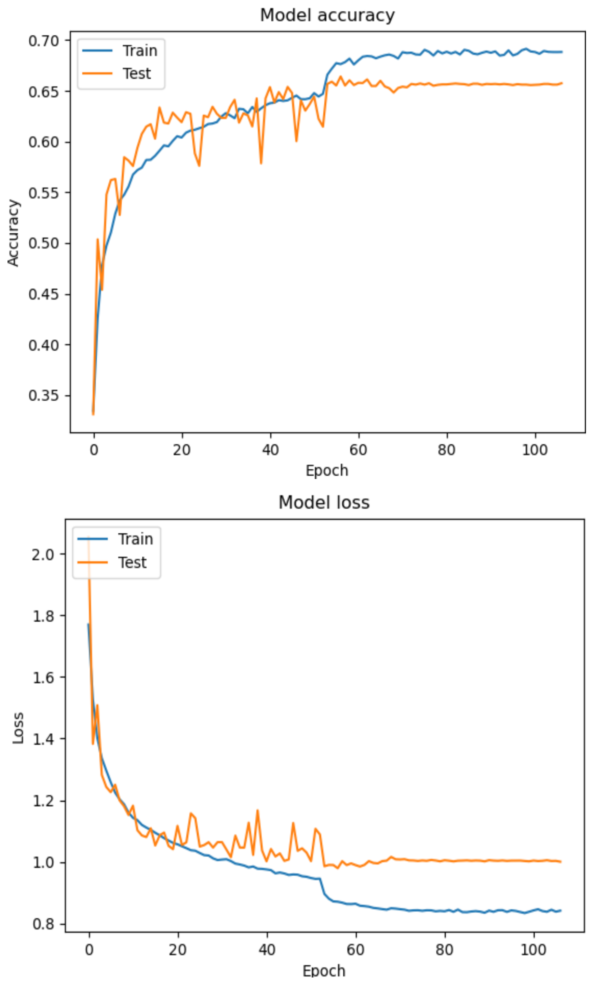

Fig. 7 FER performance after adding CK+ into FER2013 testing dataset.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (R7124-16-0004, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

References

[1] Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[2] 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.

[3] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Transactions on Affective Computing, 2020.

[4] A. Rosebroke, *Deep Learning for Computer Vision with Python*, pp. 18-23, 2017.

[5] J. Brownlee, *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python: Machine Learning Mastery*, 2019.