An Application of Removing Biased Result for Deep Learning

Jung Hwan Kim[1], Savina Jassica Colaco[1], Alwin Poulose[2], Suresh Neethirajan[3], and Dong Seog Han[1*]
[1]Graduate School of Electronic and Electrical Engineering, Kyungpook National University, Republic of Korea
[2]Center for ICT and Automotive Convergence, Kyungpook National University, Republic of Korea
[3]Department of Animal Sciences, Wageningen University and Research, Netherland
jkim267@knu.ac.kr, savinacolaco@gmail.com, alwinpoulosepalatty@knu.ac.kr, sneethir@gmail.com, dshan@knu.ac.kr*

Abstract

Emotion Recognition reads personal emotional feelings by capturing specific gestures or facial expressions. The goal of emotion recognition is to minimize people's stress and improve the welfare conditions for both humans and animals. However, obtaining the specific group dataset is rarely accomplishable without teaming up with different farming research groups. Unlike a human being, locating the vital changes may be difficult for many neural network machines without psychological expertise. Especially, some research institutions usually bring a poor-quality dataset, and the dataset produces biased experimental results. To improve the quality of the emotion dataset and prevent unfair conclusions, we proposed the semi-shuffled dataset after applying the pretrained human and animal detector's model as the megadetector v.3. Thus, our proposed dataset may have an improved representation of the real-time performance than fully-shuffled without applying the pretrained detector's model.

I. Introduction

The demand for emotion recognition has been increasing these days. Also, the demand for self-driving vehicles, social workers, and even husbandry affairs have significantly increased for improving welfare and minimizing stress. Thus, obtaining the emotion dataset for a specific group may be challenging, and have mostly limited sources for further improvement.

Frequently, some emotion datasets are obtainable by collaborating with different research institutions. Our research partnering team from Wageningen University and Research provides us with video clips. Yet, videos may contain a massive amount of irrelevant pixelized features or simply no subjects in the sample images for training and testing. Although the performing result exceeds more than 95%, the pretrained Xception model could still be useless for real-time affairs. In this paper, we propose to exploit a pretrained detector as Megadetector v. 3 models [1] as we experiment on the semi-shuffled dataset. Although our proposed experiment degrades the classifying performance, our proposed approach can even more precisely inspect experimental results. In our previous research [2-7], we successfully improved the performance by removing the irrelevant features by applying the pre-trained model and focusing on vital parts for training and testing.

II. Proposed Method

The irrelevant features are turmoil to classifying neural network since it renders to confuse our classifying neural network and increases the unnecessary computation. Megadetector v3 has similar structures of Fast-RCNN, Inception, and Resnet architectures. The pretrained Megadetector v. 3 detector model locates the region of the subjects in the image and enables us to extract the vital pixels. After finding the vital features in the pictures, we can extract those subtle regions and remove all irrelevant features. Our proposed emotion recognitional neural network is the Xception [9] architecture to have the same results from Fig. 3 and 4.

III. Experimental Results

As shown in Figs. 3, the confusion matrix display robust performance since the time-sequential raw images are completely shuffled among the training, validating, and testing dataset gourps. The Xception model perceived familiar images from testing dataset as the raw images from training and validating testing datasets is almost identical.

However, from Fig. 4 and Table I, the performance displays defectively, yet the model's performance is more precisely evaluated as representing the real-time performance, unlike the evaluation from Fig 3. The Xception model from Fig. 3 is suspected to be trained with the irrelevant segments so that the experimental results may display robust performance from the time-sequential images. Thus, the emotion recognition researchers should cautiously handle the time sequential images by not mixing among all training, validating, and testing datasets. They may mix with the training and validating testing dataset to perform better. Still, to have unbiased results from Fig. 4, the testing dataset must not be randomly blended among training and validating datasets. Then, the Megadetector model is applied before starting the Xception neural network's training and validating procedures.

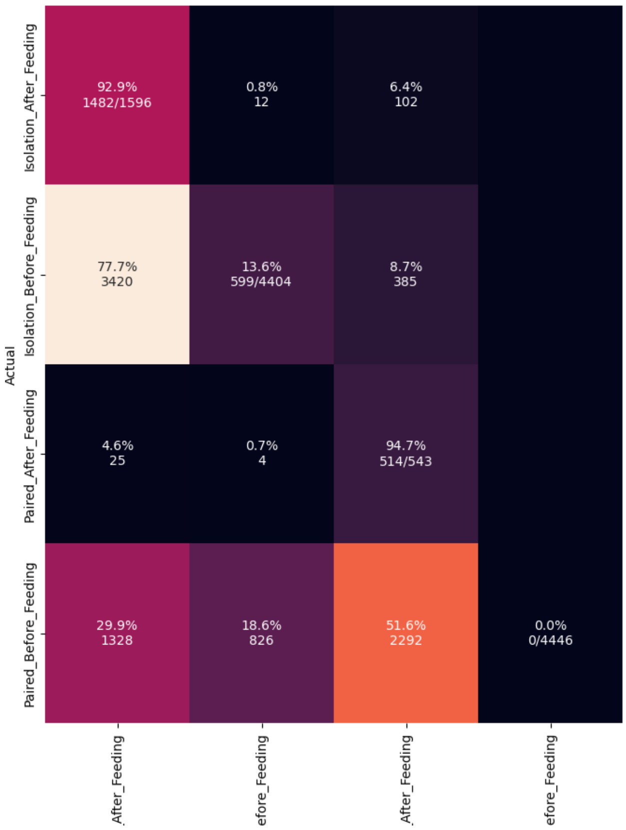

Fig. 3 the classification performance before applying with the megadetector's model and fully-shuffling.

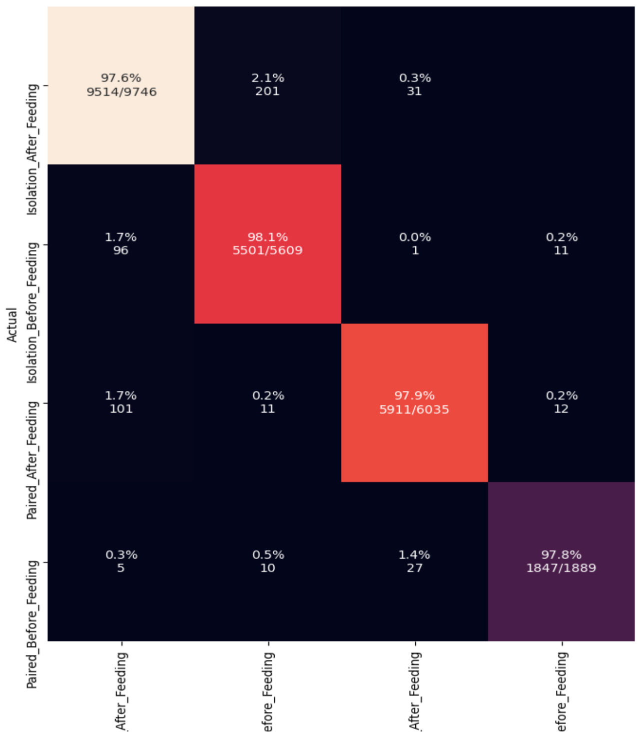

Fig. 4 the classification performance after applying with the megadetector's model and semi- shuffling.

Table I. THE CONFUSION MATRIX SUMMARIZES THE PERFORMANCE GRAPH FROM FIG. 3 AND 4.

Approaches	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Xception Model from Fig. 3	97.8264	97.8389	97.8264	97.8292
Xception Model from Fig. 4	23.6145	20.8714	23.6145	15.0215

IV. Conclusion

In this paper, we proposed the semi-shuffled sequential dataset and refined the raw images by applying the megadetector's model to prevent inaccurate results. Although our approaches degrade the classifying performance, the proper evaluation is significant as the Xception model can be applied in real-time husbandry affairs.

ACKNOWLEDGMENT

This research was performed with the support of the 5G-based autonomous driving convergence technology demonstration platform project (P0013840) supported by the Ministry of Trade, Industry and Energy and the Korea Institute of Industrial Technology Promotion.

References

[1] S. Beery, D. Morris, and S. Yang, "Efficient pipeline for camera trap image review," arXiv preprint arXiv:1907.06772, 2019.

[2] J. H. Kim, A. Poulose, and D.S. Han, "The Extensive Usage of the Facial Image Threshing Machine for Facial Emotion Recognition Performance, " Sensors, vol 21, p. 2026, 2021.

[3] J. H. Kim, R. Mutegeki, A. Poulose, and D. S. Han, "A Study of a Data Standardization and Cleaning Technique for a Facial Emotion Recognition System," In proceedings of the Symposium of the Korean Institute of communication and Information Science (KICS) Summer Conference, pp. 1193-1195, 2020.

[4] J. H. Kim, A. Poulose, and D. S. Han, "Facial Image Threshing Machine for Collecting Facial Emotion Recognition Dataset," in Proceedings of the Symposium of the Korean Institute of communications and Information Sciences (KICS) Fall Conference, 2020, pp. 67-68.

[5] J. H. Kim and D. S. Han, "Data Augmentation & Merging Dataset for Facial Emotion Recognition," in Proceedings of the Symposium of the 1st Korea Artificial Intelligence Conference, Jeju, Korea, 2020, pp. 12-16.

[6] J. H. Kim, A. Poulose, C. S. Reddy, and D. S. Han, "Segregation of AffectNet Dataset for Facial Emotion Recognition," in Proceedings of the Symposium of the Korean Institute of communications and Information Sciences (KICS), pp. 1210-1211, 2021.

[7] 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.