Segregation of AffectNet Dataset for Facial Emotion Recognition

Jung Hwan Kim[1], Alwin Poulose[1], Chinthala Sreya Reddy[2,3], and Dong Seog Han[1,*]
[1]School of Electronics and Electrical Engineering, Kyungpook National University
Republic of Korea
[2]School of Computer Science and Engineering, Kyungpook National University, Daegu,
Republic of Korea
[3]Department of Computer Science, CHRIST University, Bangalore, India
jkim267@knu.ac.kr, alwinpoulosepalatty@knu.ac.kr, sreyareddy2000@gmail.com, dshan@knu.ac.kr*

Abstract

The facial emotion recognition (FER) system has a significant role in preventing car accidents related to the road-rage. However, improving the FER's performance requires obtaining a large amount of the dataset. Many FER novice researchers obtained the largest dataset called AffectNet. It was not directly separated and was hard to find any information on the preprocessing method for many beginners in the FER research. In this paper, we propose a readable AffectNet (RAN) unit that preprocesses the given dataset specially AffectNet. It uses the facial image threshing (FIT) machine which extracts the facial segment. The proposed RAN unit achieves a validating accuracy of 75% and an f1-score of 59%.

I. Introduction

Facial emotion recognition (FER) could prevent road rage-related accidents. However, reliable performance on the FER system requires a sufficient number of facial images. Therefore, many FER researchers searched and obtained the AffectNet dataset [1] to test their FER system's performance. Nonetheless, the FER researchers had to face the difficulty of preprocessing the AffectNet dataset. The major difficulty of preprocessing the AffectNet dataset is that it has mixed facial images. The AffectNet dataset's developers did not directly segregate the emotions but provided only the annotated formatted text files to segregate for classification.

The written description did not directly explain how to properly preprocess the facial images before testing the FER system's performance. In addition, searching for the preprocessing techniques for the AffectNet dataset was not vastly available on the web. Secondly, even though AffectNet's team aligned and preprocessed the facial images, the unnecessary background segmentation still existed and wasted a lot of hardware memory. Besides, the 224 × 224 pixels of each facial image were considered large in size for many models to train and slowed down the detection of the driver's emotion within a split second. Finally, the split ratio of the training and testing from the AffectNet was dramatically disproportioned. The initial split ratio was 99 to 1 for training and testing, while the ratio was supposed to be 70 to 30. As a solution to the problems stated above, we propose an approach that gives the guidelines to many beginners on the FER system with the AffectNet dataset. We introduce a readable AffectNet (RAN) unit which exploits the facial image threshing (FIT) machine [2] [3] and the Xception algorithm [4].

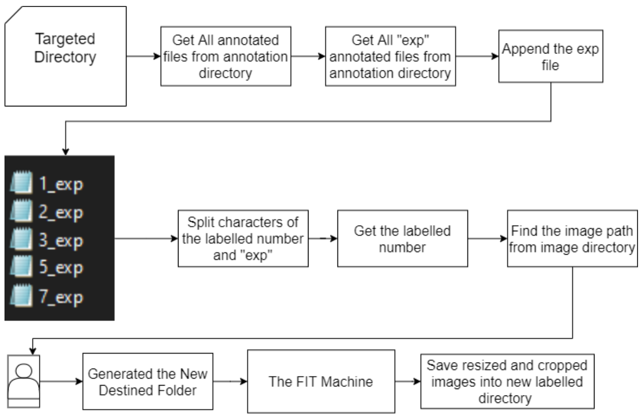

Fig. 1 The proposed RAN unit's diagram.

II. RAN Unit

Fig 1. shows the diagram of the proposed RAN unit. The RAN unit first retrieves the targeted directory, which contains the annotated text files and facial images. Then, it scans all the annotated files within the targeted directory and finds the expression-related annotated text files. Once the RAN unit finds the expressions-related text files, it takes the numerical part of its name (ignoring the "_exp" as shown in Fig. 1) to find the facial image counterparts.

While the RAN unit accesses the expression-related text files, it converts the Numpy values into an integer. The annotated expression files contain unreadable machine codes. The machine codes contain the numerical characters which are the emotion's labels. Since the machine code is comprised of Numpy values, we apply the Numpy.load() function. The returned numerical character from the function is converted into an integer. Thus, the

RAN unit could convert the Numpy values into a classified integer from lists of text files.

After the RAN unit successfully locates the facial image file, it generates new facial images into the new classified directory. Finally, the FIT machine extracts the facial features and removes the unnecessary background segmentation. Then, the FIT machine resizes the colored facial images into grayscale. The images are resized to 48 × 48 pixels by the end of the RAN unit's process.

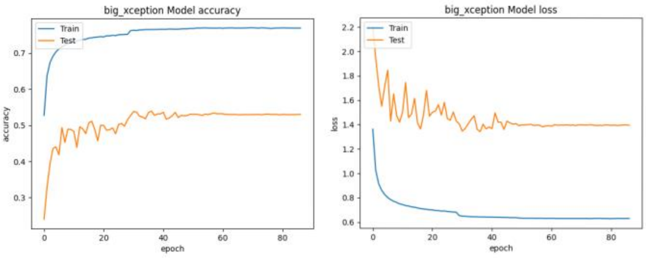

Fig. 2 Training the Xception model with a 99 to 1 ratio distribution of the training and testing datasets.

III. Test Results

Our hardware computer consists of Intel ® Core ™ i5-10600K CPU @ 4.10GHz, 32 GB RAM, and GeForce RTX 2070. The Python V 3.8 application programming interface (API) has Tensorflow-GPU Versioned 2.2 with the Keras's API. We had the unseen private testing dataset from Kim *et al.* [2][5] which was initially preprocessed by the FIT machine from the private testing of the FER 2013 dataset. The unseen private testing dataset was not used during the algorithm's training.

Originally, the number of facial image samples from the AffectNet is supposed to be 420,299. Anyhow, we only obtained 285,650 due to their intellectual property protection. The ratio of the given training and testing facial image samples is 99 to 1 which is 287,651 for training and 3999 for validation.

As shown in Fig. 2, we trained the Xception algorithm with the imbalanced data distribution of the training and testing datasets. The ratio of data distribution was 99 to 1. As a result, the validation accuracy was 54%, and the validating loss was 1.3423. The unseen private testing dataset showed an f1-score as 58% (Table I).

The ratio distribution of the training and testing facial samples is 70 to 30 as shown in Fig 3. The validating accuracy was significantly improved to 75%, and the validating loss was less than 0.67. The unseen private testing dataset marginally improved the precision, recall, and f1-score by approximately 1% (Table I).

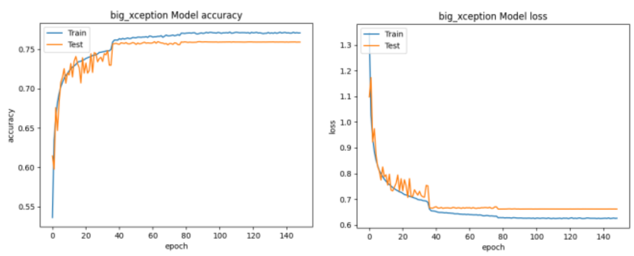

Fig. 3 Training the Xception model with a 70 to 30 ratio distribution of the training and testing datasets.

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

Split Ratio	Precision	Recall	F1 Score
99: 1	58.9924 %	60.6602 %	58.3031 %
70: 30	**59.7503 %**	**61.3823 %**	**59.1204 %**

IV. Conclusion

In this paper, we proposed a RAN unit for preprocessing the AffectNet dataset. The RAN unit improved the performance of FER's system and also the quality of the AffectNet dataset. Many beginners of the FER system can use the RAN unit to preprocess the AffectNet more easily. From these experimental results, we should reconsider the ratio distribution of the training and testing samples, so that the FER system's performance can improve further and has better stable training performance.

ACKNOWLEDGMENT

This study was carried out to support the 5G-based autonomous driving convergence technology demonstration platform project (task number: 1415169669) supported by the Ministry of Trade, Industry and Energy, and the Korea Institute for Advancement of Technology.

References

[1] A. Mollahosseini, B. Hasani, and M.H. Mahoor, "Affectnet: A database for facial expression, valance, and arousal computation in the wild," IEEE Transactions on Affective Computation, vol. 10pp. 18-31, 2017

[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, 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 communication and Information Science (KICS) Fall Conference, pp. 67-68, 2020.

[4] F. Chollet, "Xception: Deep learning with depthwise separable convolution," in Proceeding of the IEEE conference on computer vision and pattern recognition, 2017, pp.1251-1258.

[5] 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.