Foreground Extraction Based Facial Emotion Recognition Using Deep Learning Xception Model

Alwin Poulose[1], Chinthala Sreya Reddy[2,3], Jung Hwan Kim[1] and Dong Seog Han[1,*]
[1]School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, 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

alwinpoulosepalatty@knu.ac.kr[1], sreyareddy2000@gmail.com[2,3], jkim267@knu.ac.kr[1], dshan@knu.ac.kr[1*]

***Abstract*—The facial emotion recognition (FER) system has a very significant role in the autonomous driving system (ADS). In ADS, the FER system identifies the driver's emotions and provides the current driver's mental status for safe driving. The driver's mental status determines the safety of the vehicle and prevents the chances of road accidents. In FER, the system identifies the driver's emotions such as happy, sad, angry, surprise, disgust, fear, and neutral. To identify these emotions, the FER system needs to train with large FER datasets and the system's performance completely depends on the type of the FER dataset used in the model training. The recent FER system uses publicly available datasets such as FER 2013, extended Cohn-Kanade (CK+), AffectNet, JAFFE, etc. for model training. However, the model trained with these datasets has some major flaws when the system tries to extract the FER features from the datasets. To address the feature extraction problem in the FER system, in this paper, we propose a foreground extraction technique to identify the user emotions. The proposed foreground extraction-based FER approach accurately extracts the FER features and the deep learning model used in the system effectively utilizes these features for model training. The model training with our FER approach shows accurate classification results than the conventional FER approach. To validate our proposed FER approach, we collected user emotions from 9 people and used the Xception architecture as the deep learning model. From the FER experiment and result analysis, the proposed foreground extraction-based approach reduces the classification error that exists in the conventional FER approach. The FER results from the proposed approach show a 3.33% model accuracy improvement than the conventional FER approach.**

***Index Terms*—Facial emotion recognition (FER), autonomous driving system (ADS), deep convolutional neural networks (DCNNs), Foreground Extraction.**

I. Introduction

In recent days, the facial emotion recognition (FER) system plays a major role in autonomous driving systems (ADS) for safe driving [1]. In ADS, the FER system provides the driver's mental conditions based on his/her emotions and the driver's emotions are useful information to reduce vehicle collision and road accidents. In FER, the system conveys the emotional state of a driver from his/her facial expressions and it determines the driver's mental health [2]. The result from the FER system is an influencing factor to determine the performance of ADS systems. In ADS, it is necessary to monitor the driver's emotions for safe driving. In recent years, many researchers proposed various techniques for emotion recognition and these techniques achieve remarkable FER performance for autonomous driving applications [3][4][5][6][7]. However, in most of the FER applications, the existing FER systems use publicly available datasets such as FER 2013 [8], extended Cohn-Kanade (CK+) [9], AffectNet [10], JAFFE [11], etc. for model training, and the model trained with these dataset has feature extraction issues for real-time implementation. The raw facial images increase the computational time of the FER system and it is necessary to perform the data preprocessing before the system uses a deep learning model for training. The lack of the features from the existing dataset creates a classification error and this error directly reflects the FER system's performance. To reduce the classification error and feature extraction issues that exist in the FER systems, we propose a foreground extraction-based FER approach, and this approach adds a high level of feature information to the FER dataset. Our experiment result and analysis show that the proposed foreground extraction-based FER approach reduces the classification error and predicts the user emotions with a high level of model accuracy and minimum loss error.

In this paper, we proposed a FER approach that predicts the current user/driver emotions. The proposed FER approach introduces a foreground extraction technique [12] for FER datasets and the dataset after foreground extraction conveys useful information for model training. The foreground extraction technique increases the feature information and the deep learning model can easily classify the user emotions with minimum error. To validate our proposed FER approach, we created a dataset with the emotions of 9 people based on the foreground extraction technique. The proposed system uses the Xception architecture [13] as the deep learning model and trained this model with our FER dataset. The FER results from our approach show that the Xception model is able to predict

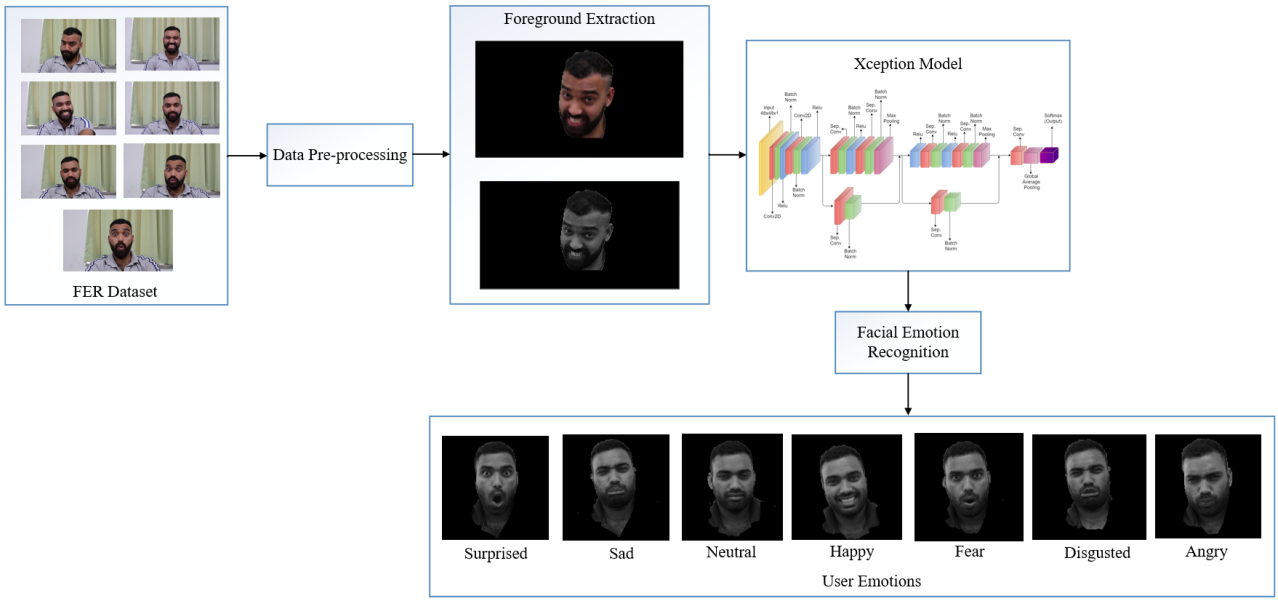


Fig. 1. Proposed foreground extraction-based FER framework.

user emotions without any computational complexity. The FER classification results indicate that our proposed FER approach has a significant role in the recent FER systems.

The rest of the paper is organized as follows; Section II presents our proposed foreground extraction-based FER system using the Xception model and Section III discusses the performance analysis of the proposed foreground extraction-based FER system with classification results. The paper is concluded in Section IV with future FER research directions.

II. Proposed Foreground Extraction-based FER System Using Xception Model

The proposed FER system consists of data collection, preprocessing, foreground extraction, which is followed by a classification model. Fig. 1 shows the proposed foreground extraction-based FER approach using the Xception model.

In the proposed FER system, we start with FER data collection from different people. When the dataset is created, we further preprocess the dataset with image processing techniques which removes the irrelevant images from our dataset. After the data preprocessing, the system performs a feature extraction technique named foreground extraction which eliminates the unnecessary information from the dataset. The output after foreground extraction are grey images that contain only useful information for the training of the Xception model. The grey images reduce the computational complexity of the FER system as compared to the raw color image-based FER datasets. The Xception model used in the proposed system uses the foreground extracted images for its training. During the training time, the Xception model interprets the foreground extracted FER images and classifies the user emotions with accurate classification results. The FER results from the proposed approach significantly reduce the classification error and improves the performance of the system.

Data Collection: In the data collection phase, we simulated the FER environment and collected emotions from 9 people. Fig. 2 shows the data collection process and the user emotions from our datasets.

During data collection, we used a camera and recorded each emotion as a 60 seconds video. The user watched some videos on his/her smartphone and simulated each emotion. We collected a FER dataset that consist of happy, sad, angry, surprise, disgust, fear, and neutral emotions from each person. Table I summarizes the data distributions of our FER dataset.

Table I: Data distributions.

Emotions	Number of Image Samples
Happy	1224
Sad	1441
Angry	1544
Surprise	1020
Disgust	1267
Fear	1327
Neutral	1273

Data Preprocessing and Foreground Extraction: Data preprocessing is a crucial step for the FER system to improve the information provided by the data. The preprocessing technique enhances the relevant features as well as diminishes the irrelevant information or noise from the datasets. During data preprocessing, we remove all the unnecessary images that are either misclassified or

Data Collection

Dataset

Fig. 2. Data collection process.

unrelated to the dataset. Next, by using the foreground extraction technique we keep any information related to the facial expressions alone by removing the unnecessary background details. The results from the foreground extraction technique are free from background noise and the model used in the FER system can easily interpret the user emotions.

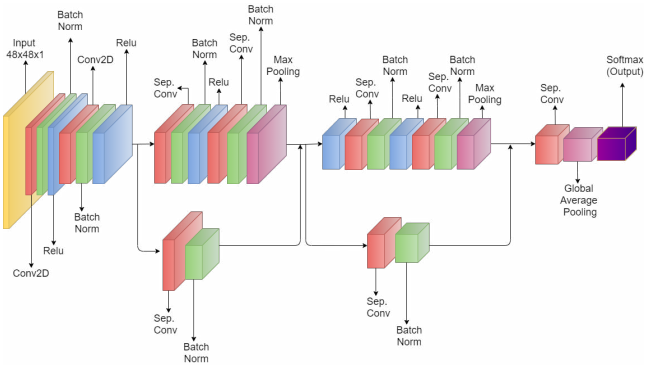


Fig. 3. Xception model used in the proposed FER system.

Deep Learning Model: Once the FER dataset is ready, the proposed system uses the Xception architecture as the deep learning model that learns patterns from the training data and uses these patterns to classify emotions. The performance of the network can be improved by hyper-parameters tuning that reduces the problem of overfitting and underfitting [14]. In this case, the model classifies the images into one of the seven emotion categories and Fig. 3 shows the architecture of the Xception model used in the proposed FER system.

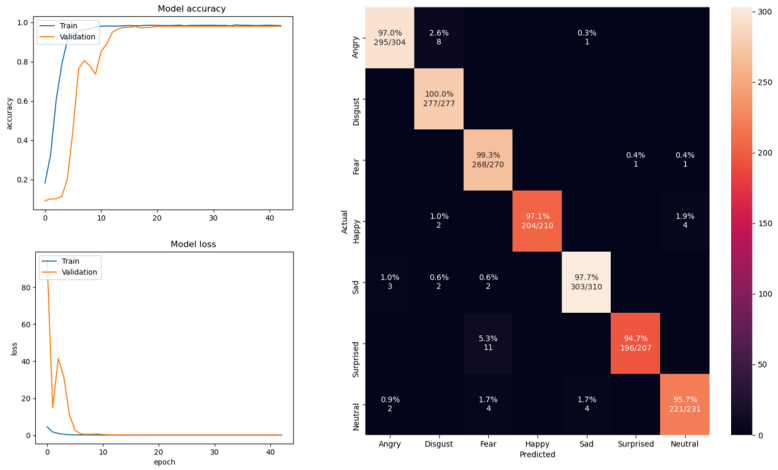
The Xception model uses a 48×48 input image size for the proposed approach and a 56×100 image size for the conventional approach. The number of the channel is set as one and the system uses Adam as the optimizer. The model uses 0.02 as the learning rate and a batch size of 128 for training and testing. The system uses 150 epochs for the model's accuracy and loss validation with an early stopping method. Table II summarizes the hyperparameter values used in the Xception model for proposed and conventional approaches.

Table II: The hyperparameter values used in the Xception model.

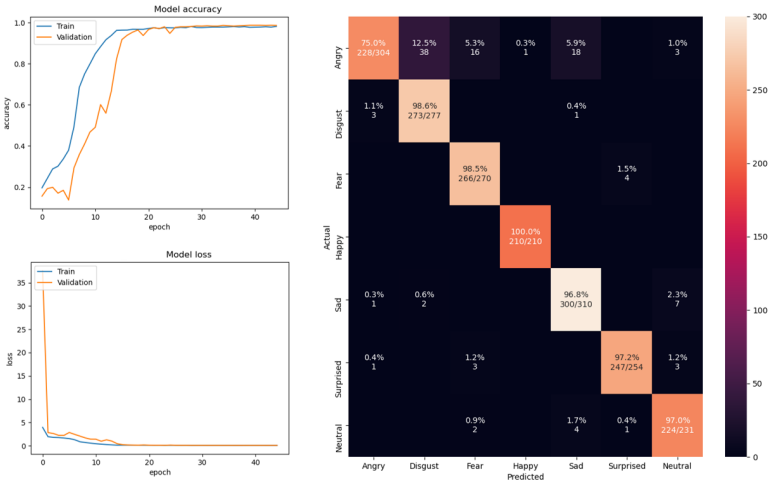
Xception Model parameter	**Value**
Input image size (Proposed approach)	48 × 48
Input image size (Conventional approach)	56 × 100
Number of channels	1
Optimizer	Adam
Learning Rate	0.02
Batch Size	128
Epochs	150

III. Experiments and Result Analysis

To validate the performance of the proposed foreground-based FER approach, we did our experiment with the



(a) Proposed foreground extraction-based FER approach.



(b) Conventional FER approach (without foreground extraction).

Fig. 4. Performance analysis of FER approaches.

Table III: Performance comparison of proposed and conventional FER approaches.

FER Approach	**Accuracy**	**Precision**	**Recall**	F_1 **Score**
Proposed foreground extraction-based FER approach	97.51	97.57	97.51	97.51
Conventional FER approach (without foreground extraction)	94.18	94.18	94.18	94

Xception model, and Fig. 4 shows the model accuracy, loss, and confusion matrix of the proposed and conventional approaches. In the case of the conventional FER approach, we used the raw image dataset without foreground extraction. The conventional approach also uses the same model configuration for training and testing.

From Fig. 4, the proposed FER approach classifies the user emotion accurately with lower model loss. The Xception model for the proposed FER approach reached 97.51% of accuracy and 0.06 of model loss. In the case of the conventional approach, the Xception model has 94.18% model accuracy and 0.08 of model loss. These results indicate that the proposed FER approach achieved a 3.33% model accuracy improvement than the conventional approach. The confusion matrices from the Fig. 4 show that the proposed approach effectively classifies the user emotion with minimum errors. The performance comparison of the proposed and conventional approaches is summarized in Table III.

From Table III, we can conclude that the proposed FER approach outperforms the conventional approach in terms of accuracy, precision, recall and F_1 score. The results indicate that the proposed foreground extraction-based technique has a high influence on the FER systems. These

results validate our proposed FER approach and show the best FER performance for user emotion classification.

IV. Conclusion

In this paper, we proposed a foreground extraction-based FER system using an Xception model. The proposed FER system reduces the system's classification error and improves the user's emotion recognition results. The Xception model used in the proposed system effectively utilizes the foreground extracted face images from the FER dataset and enhances the FER performance. The classification results from the proposed system show that it improves the classification results than the conventional FER approach. The experiment and result analysis of the proposed FER system shows that the foreground extraction-based approach is an effective FER approach for user emotion recognition. In future works, we intend to add a histogram equalization-based feature extraction technique for better FER system performance.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

References

[1] 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.

[2] D. Aouada,"Dense and Sparse 3D Deformation Signatures for 3D Dynamic Face Recognition," *IEEE Access,* vol. 9, pp. 38687-38705, 2021.

[3] D. Y. Choi and B. C. Song, "Facial Micro-Expression Recognition Using Two-Dimensional Landmark Feature Maps," *IEEE Access,* vol. 8, pp. 121549-121563, 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,* Seoul, South Korea, 13 November 2020; pp. 67–68.

[5] G. Ali, A. Ali, F. Ali, U. Draz, F. Majeed, S. Yasin, et al., "Artificial neural network based ensemble approach for multicultural facial expressions analysis," *IEEE Access,* vol. 8, pp. 134950-134963, 2020.

[6] T.-H. Vo, G.-S. Lee, H.-J. Yang, and S.-H. Kim, "Pyramid with Super Resolution for In-the-Wild Facial Expression Recognition," *IEEE Access,* vol. 8, pp. 131988-132001, 2020.

[7] 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 communications and Information Sciences (KICS) Summer Conference,* 2020, pp. 1193-1195.

[8] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, et al., "Challenges in representation learning: A report on three machine learning contests," in *International conference on neural information processing,* 2013, pp. 117-124.

[9] 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," in *2010 ieee computer society conference on computer vision and pattern recognition-workshops,* 2010, pp. 94-101.

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

[11] M. Lyons, M. Kamachi, and J. Gyoba, "Japanese female facial expression (JAFFE) database," 2017.

[12] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut interactive foreground extraction using iterated graph cuts," *ACM transactions on graphics (TOG)* vol. 23, pp. 309-314, 2004.

[13] 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.

[14] A. Hafner, P. Peer, Ž. Emeršič, and M. Vitek, "Deep Iris Feature Extraction," in *International Conference on Artificial Intelligence in Information and Communication (ICAIIC),* 2021, pp. 258-262.