

A pre-trained model vs dedicated convolution neural networks for emotion recognition

Asmaa Yaseen Nawaf, Wesam M. Jasim

Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq

Article Info

Article history:

Received Dec 19, 2021

Revised Sep 3, 2022

Accepted Sep 24, 2022

Keywords:

Convolutional neural networks

Deep learning

Facial expression recognition

FER+ dataset

VGG16 pre-trained model

ABSTRACT

Facial expression recognition (FER) is one of the most important methods influencing human-machine interaction (HMI). In this paper, a comparison was made between two models, a model that was built from scratch and trained on FER dataset only, and a model previously trained on a data set containing various images, which is the VGG16 model, then the model was reset and trained using FER dataset. The FER+ data set was augmented to be used in training phases using the two proposed models. The models will be evaluated (extra validation) by using images from the internet in order to find the best model for identifying human emotions, where Dlib detector and OpenCV libraries are used for face detection. The results showed that the proposed emotion recognition convolutional neural networks (ERCNN) model dedicated to identifying human emotions significantly outperformed the pre-trained model in terms of accuracy, speed, and performance, which was 87.133% in the public test and 82.648% in the private test. While it was 71.685% in the public test and 67.338% in the private test using the proposed VGG16 pre-trained model.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Asmaa Yaseen Nawaf

Department of Computer Science, College of Computer Science and Information Technology

University of Anbar

Ramadi, Kirkuk Governorate, Domiz, Republic of Iraq

Email: asmaayaseen1981@gmail.com

1. INTRODUCTION

As computer technology has advanced, efforts have been made to develop smart devices capable of simulating the human mind. Human ambition was not limited to training machines to perform human tasks, but the goal became to develop devices capable of analyzing and distinguishing human emotions. Emotions can be recognized in various ways, based on heart rate variability (HRV), electro encephalography (EEG) signals, galvanic skin response signal (GSR), speech emotion recognition (SER), facial expression recognition (FER). One of the most effective methods of communication mechanisms is FER by which human machine interaction (HMI) systems can understand humans' internal emotions as facial expressions, which plays an important role in social interaction [1]. Automatic recognition of facial expressions is an important topic in computer vision research. This is due to the real value for application in many fields, such as security, interactive game, health care, patient status monitoring [2], and commercial advertisements to know the consumer's reaction [3], and in the autonomous driving system (ADS), the FER was used to identify the driver's emotions. ADS makes good use of the features of FER and improves its safety by preventing traffic accidents [4], [5]. Reaching a machine capable of distinguishing facial expressions or face detection [6], [7] is a difficult task, due to the wide variety of faces in terms of age, gender, and others [8]. Deep learning is the most powerful technology for having brought artificial intelligence (AI) devices closer to human-level intelligence. The use of deep learning in the field of facial expression recognition has given promising results.

Deep convolutional neural networks, specifically, have achieved significant results in recent public challenges [9]. Because of their large number of filters, convolutional neural network (CNN) is the best network used for image recognition tasks because they can extract special features of the input images [10], [11]. CNN has recently begun to outperform traditional methods in a variety of fields, most notably pattern recognition [12], face detection [13]. CNNs are supervised learning techniques, there are many CNN architectures. The diversity in these architectures comes from the improvements or the modifications that are applied to the original CNN architecture. Modifications are structural reformulation, regularization, and parameter optimizations. The most innovative developments in CNN architectures are focused on the use of network depth [14]. The most famous CNN structures are AlexNet, GoogLeNet, VGG, ResNet, Inception-V3 and DenseNet. VGG showed good results in the classification of images [15]. Much research relied on deep CNN to extract features from facial expressions.

A transfer learning strategy for deep CNN was used in [16]. The model was built based on the AlexNet and VGG-CNN-M-2048 and the FER2013 data set was used. Several submissions such as single fine tuning and double fine tuning were applied. The best performance was by employing dual fine-tuning tactic cascading fine-tuning approach which was 48.5% in the validation set and 55.6% in the test set. In [17] combining three different CNNs were used and tested with the FER2013 dataset. The obtained accuracy on the whole model was 65.03 %. Sang *et al.* [18] suggested a variety of CNN architectures such as BKVGG8, BKVGG10, BKVGG12 and BKVGG14 inspired by VGG's architecture with data augmentation consideration. The results show the strength of the small filter and very deep network for FER2013 dataset classification tasks. Pons and Masip [9] suggested a framework for emotion recognition relying on a supervised hierarchical committee of deep CNNs using various databases. The accuracy obtained with the proposed committee CNN was 39.3%, and 42.9% with the proposed committee VGG-16. The DenseNet-BC architecture is used in [19]. The confusion matrix and class activation map (CAM) is used to explore the effects of various facial areas on low-resolution images in the wild. They chose the FER+ dataset to test their system and obtained accuracy of 81.93%. A model of two CNNs joined as a tandem facial expression (TFE) of feature was proposed in [20]. It was tested using the FER+ dataset and the extended Cohn-Kanade (CK+) dataset with 84.3% and 99.31% accuracy respectively. The CNN was experimented using FER2013 dataset and got 60% accuracy [10]. While a simple CNN called SHCNN was trained based on FER2013 in [21] and obtained an accuracy of 69 %. Agrawal and Mittal [22] investigated the effect on CNN parameters namely kernel size and number of filters on the classification accuracy using the FER2013 dataset. Data augmentation has been applied and the tests achieved an accuracy of 65%. A neural network based on the Google Net model was proposed to construct an emotion recognition system using the Viola-Jones method for face detection. The obtained accuracy on the FER2013 dataset was 69% [23]. Kim *et al.* [24] suggested a facial image threshing (FIT) machine that improves FER system performance for autonomous vehicles by utilizing advanced characteristics of pre-trained facial recognition and training from the Xception method. In addition, the FER system problems presented when using FER 2013 and CK+ datasets were reviewed. Even powerful FER models perform poorly in real-time testing when trained without appropriate datasets. The quality of the datasets has a greater impact on FER system performance than the quality of the algorithms.

In this work, a comparison between the proposed emotion recognition convolutional neural networks (ERCNN) model, a model built from scratch, inspired by the VGG16 model, and the VGG16 pre-trained model is addressed. This paper contributes to speeding up the training by proposing the ERCNN model, minimizing the error of the training data, and improving the data set.

The paper remaining sections are organized as; section 2 illustrates the proposed network design. Comparison between the proposed ERCNN and the original VGG16 is explained in section 3. Section 4 presents the data set used in training and its augmentation. Section 5 outlines the results and discussion. Section 6 clarifies conclusions and future works.

2. THE PROPOSED NETWORK

In this paper, a comparison was conducted between two models, a modified model trained on the FER+ dataset only, and a model previously trained on a wide range of datasets, which is the VGG16 model. The pre-trained model was reset and retrained using the FER+ dataset. The results showed that the proposed ERCNN model dedicated to identifying human emotions significantly outperformed the pre-trained model in terms of accuracy, speed, and performance. The proposed networks that were used in this work will be explained in the next section.

2.1. The ERCNN model

This work is focused on increasing the depth of the CNN by increasing the number of convolution layers. Also, adding batch normalization layers and optimizing the parameters in proportion to the data used.

Table 1 shows the difference between the proposed ERCNN model and the original VGG16. The basic layers in the proposed ERCNN architecture are 24 layers as: Seventeen CONV-2D layers, each convolution layer has a kernel size of (3,3) and multiple filters in each convolution layer were used. It was started from (64 to 512) filters and the rectified linear unit (ReLU) activation function was the activation function used with Conv2D. After each Conv2D layer, the batch normalization was added to make the deep network faster and more stable by applying normalizing and standardizing operations. Five max-pooling 2D layers and pool-size (filter size) of (2,2) were used, to get the most important features from the features extracted from the convolution layer, thus reducing arithmetic operations, and preventing overfitting. Each max-pooling followed by the dropout layer with the dropout probability of (0.3), to drop the nodes in random way, this will prevent a model from overfitting. One flattening layer and one fully connected (dense) layer with its SoftMax activation function were used. Architecture for the proposed ERCNN model was shown in Figure 1.

Table 1. The difference between the architecture of the proposed ERCNN model and the original VGG16 architecture

Proposed ERCNN model					Original VGG16 model						
Layer	Feature map	Filter size	Stride	Activation function	Layer	Feature map	Filter size	Stride	Activation function		
Input	1	-----	-----	-----	Input	1	-----	-----	-----		
Block1	2Conv	64	3*3	1	ReLU	Block1	2Conv	64	-----	1	ReLU
	Batch normalization after every CONV					-----					
	Maxpooling	64	2*2	1	-----	Maxpooling	64	2*2	2	-----	
	Dropout 0.3					-----					
Block2	3Conv	128	3*3	1	ReLU	Block2	2Conv	128	3*3	1	ReLU
	Batch normalization after every CONV					-----					
	Maxpooling	128	2*2	1	-----	Maxpooling	128	2*2	2	-----	
	Dropout 0.3					-----					
Block3	4Conv	256	3*3	1	ReLU	Block3	3 Conv	256	3*3	1	ReLU
	Batch normalization after every CONV					-----					
	Maxpooling	256	2*2	1	-----	Maxpooling	256	2*2	2	-----	
	Dropout 0.3					-----					
Block4	4Conv	256	3*3		ReLU	Block4	3Conv	512	3*3	1	ReLU
	Batch normalization after every CONV					-----					
	Maxpooling	256	2*2	1	-----	Maxpooling	512	2*2	2	-----	
	Dropout 0.3					-----					
Block5	4Conv	512	3*3	1	ReLU	Block5	3Conv	512	3*3	1	ReLU
	Batch normalization after every CONV					-----					
	Maxpooling	512	2*2	1	-----	Maxpooling	512	2*2	2	-----	
	Dropout 0.3					-----					
	Flatten					Flatten					
	1 Fully (output)				SoftMax	fully					ReLU
						Dropout 0.5					ReLU
						fully					ReLU
						Dropout 0.5					
						1 Fully output					SoftMax

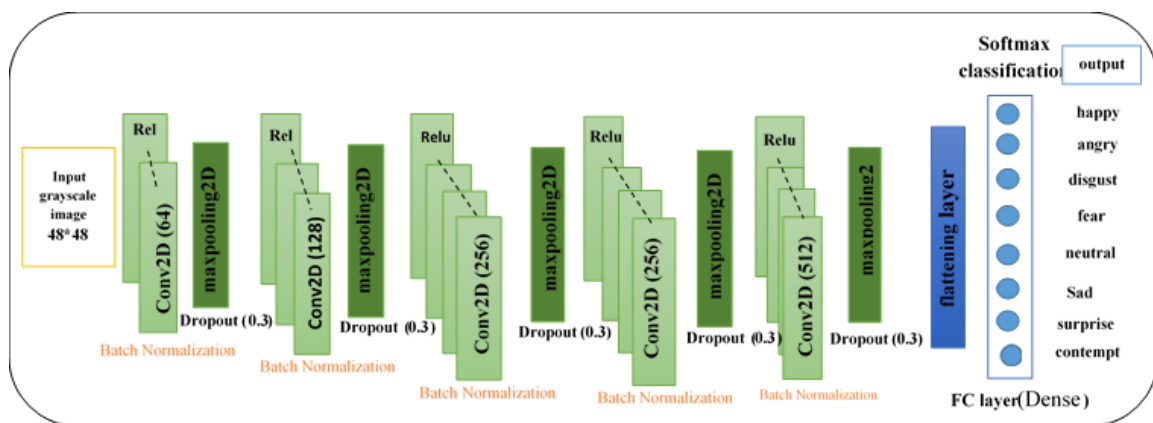


Figure 1. The proposed ERCNN architecture

The training parameter was specified as: number of the epoch is 200, the batch size 265, the width and height (48,48), the number of classes FER+ is 8 classes. Adam was chosen as an optimizer. The padding "same" option was used to avoid image size loss after applying the CNN kernel, and the stride is 1. Cross-entropy was used as a loss function to speed up the training and improve the model classifications.

2.2. VGG16-Pre-trained model

VGG16 is a convolution neural network used for images classification. It was created after training on ImageNet dataset and classified it into 1,000-class. In this paper and based on the transfer-learning approach, a pre-trained VGG16 model will be created. VGG16 is loaded from Keras, and it is fine-tuned to fit the requirements of identifying human emotions. To create an emotion recognition model, the initial layers of VGG16 are frozen. The last four layers are retrained on the extended data in order to predict the emotions. The fully connected layer is removed, and a new fully connected layer is created to meet the requirements of the emotion recognition task. At the retraining, the VGG16 pre-trained model parameters are set as: the input layer is of (224,224,3) shape. The AveragePooling2D layer and the pool-size (filter size) are (2,2). Rectified linear unit (ReLU) is the activation function used in the fully connected training layer, followed by the dropout layer with the dropout probability of (0.5). The SoftMax is the activation function in the fully connected output layer (prediction) and the number of classes is modified to match the emotions in the extended data to be 8 with FER+. The extended dataset is grayscale (single channel) images with 48×48 dimensions. Because VGG16 works with color images (3 channels), the dimensions are 224×224. Then, the grayscale images should be resized from 48×48 to 224×224 and converted into three-channel grayscale images using *ImageDataGenerator*.

3. COMPARISON BETWEEN THE ERCNN AND THE VGG16

CNN is a well-known deep learning algorithm that has been effectively employed in the identification of high-dimensional data, particularly images [25]. The entire convolution procedure involves converting a picture into another image of comparable size and convention using a weighted matrix. Furthermore, convolution is used to extract the feature map [26]. So, in this work, the focus has been on increasing the convolution layers. Table 1 shows the difference between the architecture of the proposed model and the original VGG16 architecture [27]. The number of convolution layers increased to 17, while the original was 13. One fully connected layer is used, while in the original VGG16 it was three fully connected layers and after first two fully used dropout 0.5, we added a batch normalization after each convolution layer while in the original the batch normalization does not used. The stride is 1 for max pooling with dropout 0.3 in our proposed model while it was 2 at the original VGG16. The original VGG16 used SGD as an optimizer function, but in our work, Adam was used as an optimizer function.

4. THE DATA SET

4.1. FER+ corrected dataset

Barsoum *et al.* [28] presented FER+ data set that modified with multiple labels for each face image. The wrong labels of FER2013 dataset have been corrected with crowd sourcing using 10 taggers to label each image, as shown in Figure 2, where Figure 2(a) indicates the wrong labeling in FER2013, Figure 2(b) indicates FERplus-corrected labels. The number of emotions in FER+ becomes 10 classes: neutral, happiness, surprise, sadness, anger, disgust, fear, contempt, unknown, and not face (NF). In FER+, 80% of the images were designated as training samples, 10% as validation samples (public test), and 10% as test samples (private test). In this work, the script (*FERPlus/src/generate_training_data.py*) is used to get the csv file for the corrected data from (*Microsoft/FER+*) [29]. The number of images in the FER+ dataset is 35,710 which is less than the number of images in the original FER2013 data (35,887) by 177 images. This difference is the result of deleting the NF class and the unknown class which contains blurred images.

4.2. New data

The first step in modifying the data set is to add new data to the FER+. Graduate students in machine learning at New York University created new data facial expressions [30]. The new data is made up of 13,690 grayscale images with dimensions of 48×48. We added 65% of the new data to the training data and 35% to the validation data. The number of emotions in the new data is 8 classes: anger, surprise, disgust, fear, neutral, happiness, sadness, and contempt. The emotion of the neutral contains the largest number of images (6868 images), and the emotion of contempt contains the least number of images (9 images) as shown in Table 2.



Figure 2. The wrong and corrected labels of FER dataset (a) the wrong labels of FER2013 and (b) the correct labels of FER+ [21]

Table 2. Number of images in each emotion in new data

Emotion	Angry	Disgust	Fear	Happy	Neutral	Sad	Contempt	Surprise
Images	252	208	21	5696	6868	268	9	368

4.3. Pre-processing phase

After downloading and reading the data, the data is reprocessed in the form of a set of steps, as: split data into training, validation, and test set, convert strings to lists of integers, convert data to NumPy array and normalize grayscale image with 255, and shuffle the training data. When the dataset is used with VGG16, the images is resized to 224×224. VGG16 only deals with color images (RGB images), so the grayscale image should be made as image with three-channel gray (3 channel image) using *ImageDataGenerator*.

4.4. Combined the dataset

The extended data is a combination of two sets. In order to integrate the new data to FER+, a number of processes on the new data were made; remove the user id column, make the new data labels lowercase to match FER+ labels. The new emotion labels are anger, surprise, disgust, fear, neutral, happiness, sadness, and contempt. Saving the images in the new data to FER folders and defining the percentage of the distribution of the new data, 65% to be added to the training data and 35% to the validation data. The images of a particular emotion from the new data are added to the emotion that corresponds to it in the FER+ dataset (e.g., sad to sad, angry to angry).

4.5. Apply the data augmentation

One of the most prominent problems with emotion recognition systems is the lack of data on facial expressions, especially when deep learning began to be used in the field of computer vision. Training deep learning models depends on huge data to reach satisfactory results. Data augmentation was used to avoid the over-fitting problem by increasing the emotions dataset [31]. In this work, the augmentation was added by applying a number of techniques such as rotating for 40 degrees, width–shift–range set by 0.3, height–shift–range set by 0.3, zoom range (0.3) and horizontal flip. So, different versions of the original images are created, thus increasing the diversity of the extracted features.

5. RESULTS AND DISCUSSION

5.1. Implementation of the proposed work system

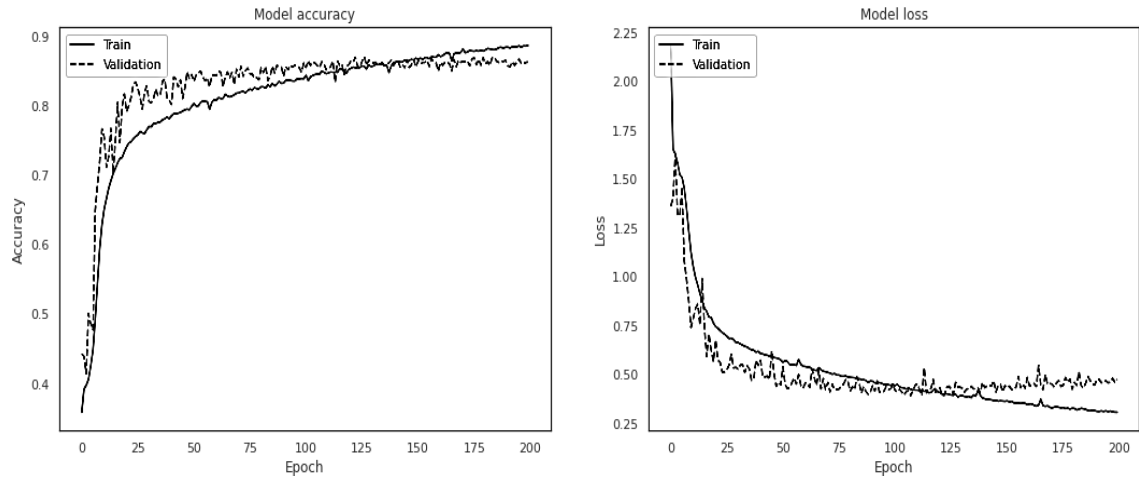
The proposed work system was implemented using the programming language Python. It was trained and tested with Keras and TensorFlow on the Kaggle platform, which allows free kernel access to Nvidia K80 GPUs. It allows the kernel to use the GPU results in a 12.5X speedup during deep learning model training. The parameters selected to be suitable for the proposed work through trial and error. After trying many parameters, the parameters used in both models are shown in Table 3.

Table 3. List of parameters used in training for the proposed work system

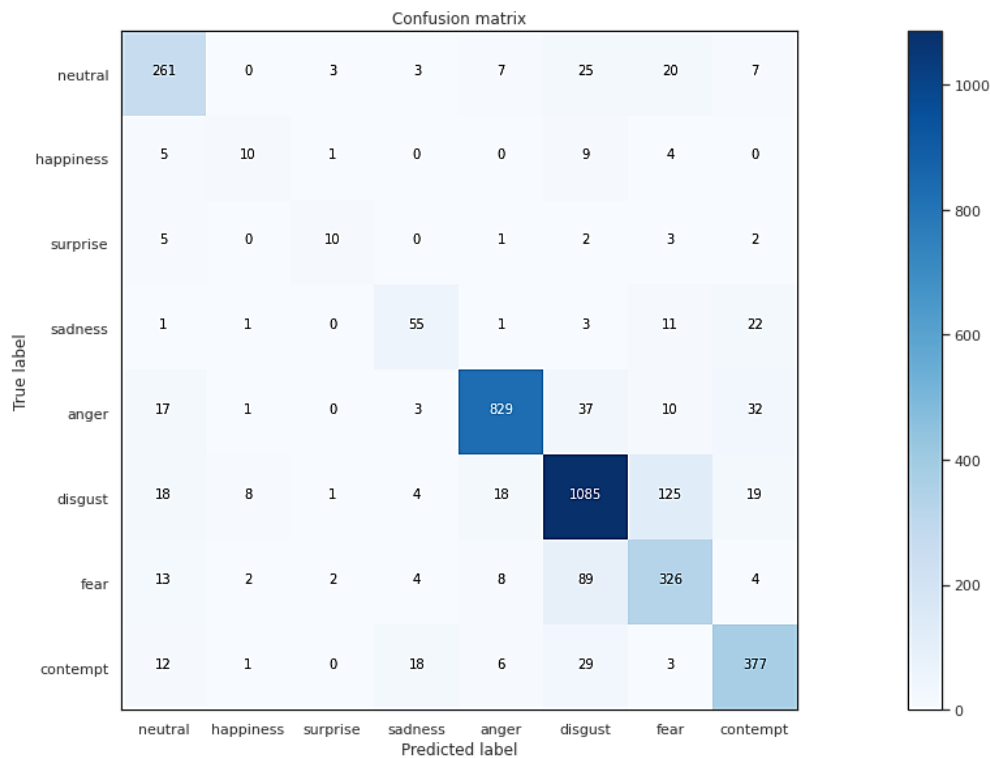
Parameters	Batch size	Number of epochs	Optimizer	Loss function	Activation function
Size, type	265	200	Adam	Cross entropy	SoftMax

5.2. Testing the proposed ERCNN with the expanded data

In this experiment, the ERCNN model is trained and tested using extended data (FER+, new data). The accuracy obtained when the percentage of adding new data was 65% to the training data and 35% to the validation data was 87.133% in the public test and 82.648% in the private test. Figure 3(a) presents the model's loss and the model's accuracy. The confusion matrix for classes prediction with the test dataset of the FER+ dataset is shown in Figure 3(b).



(a)



(b)

Figure 3. The performance of proposed ERCNN (a) model accuracy and model loss and (b) confusion matrix

Several images from the internet with various emotions were utilized to test the effectiveness of the proposed ERCNN model training using exhausted data. The images used to evaluate the model's performance are shown in Figures 4. In which, it is clear that the accuracy of the proposed ERCNN model's predictions for each emotion by looking at the images. The accuracy of a happy face is 99.92 %, a neutral face is 97.54 %,

and a sad face is 97.92 %. From the confusion matrix (the proposed ERCNN model with FER+ test dataset), the precision, recall, and F1-score were calculated for each class, the highest precision was 95.3% for anger class and the average precision of the proposed ERCNN model was 71.3%. The highest recall was 89.2% for the anger class and the average recall was 68.5%. The highest F1-score was 92.2% for anger class and the average of F1-score was 69.6%.



Figure 4. Evaluate the ERCNN using images from internet [32]

5.3. Testing VGG16 pre-trained model with expanded data

In this experiment, four layers of the VGG16 pre-trained model are trained with extended data (FER+, new data). When adding 65% of the new data to the FER+ training data, the model performance was poor as the network did not converge well as shown in Figure 5, which shows the overfitting that occurred. Also, the predictions on the internet images were poor as shown in Figure 6(a). Despite the public accuracy was 74.253%. The private accuracy was 66.498%. To improve the performance of the model and get rid of overfitting, the number of training data was increased by increasing the number of new data added to the training data by 90% and 10% to the validation data. Figure 6(b) shows some improvement in the model's performance when tested using images from the internet, in terms of emotion prediction, and the overfitting become less. Figure 7 shows the loss model and the accuracy model for the VGG16 pre-trained model. The accuracy was 71.685% in the public test and 67.338% in the private test.

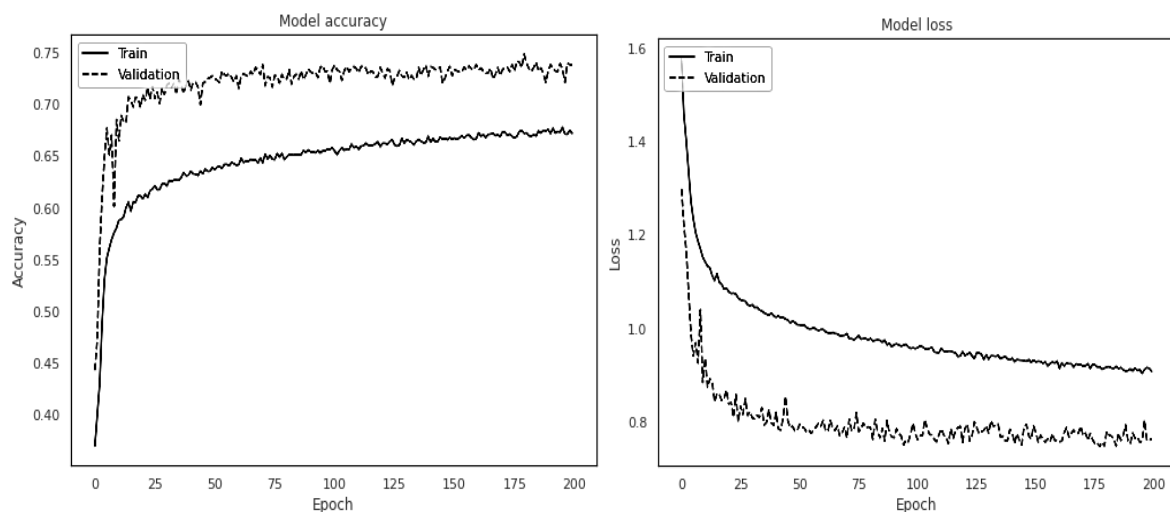


Figure 5. Model accuracy and model loss for the VGG16 pre-trained model when the percentage of adding new data was 65%



(a)



(b)

Figure 6. Evaluate the VGG16 pre-trained model using image from internet (a) when the percentage of adding new data was 65% and (b) when the percentage of adding new data was 90% [32]

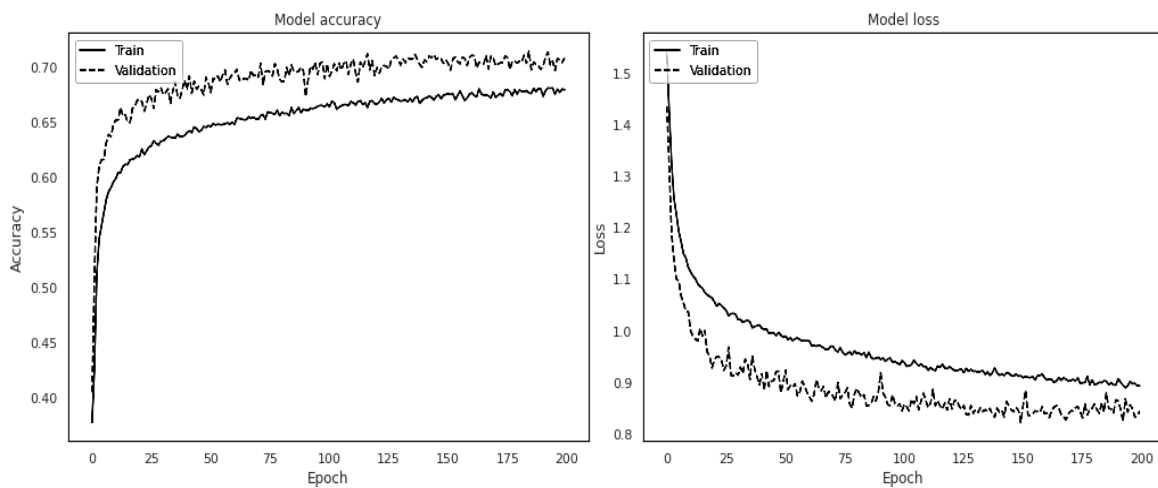


Figure 7. Model accuracy and model loss for the VGG16 pre-trained model when the percentage of adding new data was 90%

Confusion matrix for VGG16 pre-trained model with FER+ test dataset shown in Figure 8. The precision and recall for each class are calculated from the confusion matrix in Figure 8 with the VGG16 pre-trained model, the highest precision was 100% for the happiness class and the average precision of the

VGG16 pre-trained model was 60.8%. The highest recall was 84% for the anger class and the average recall was 42.5%. The highest F1-score was 75.6% for anger class and the average F1-score was 45.1%.

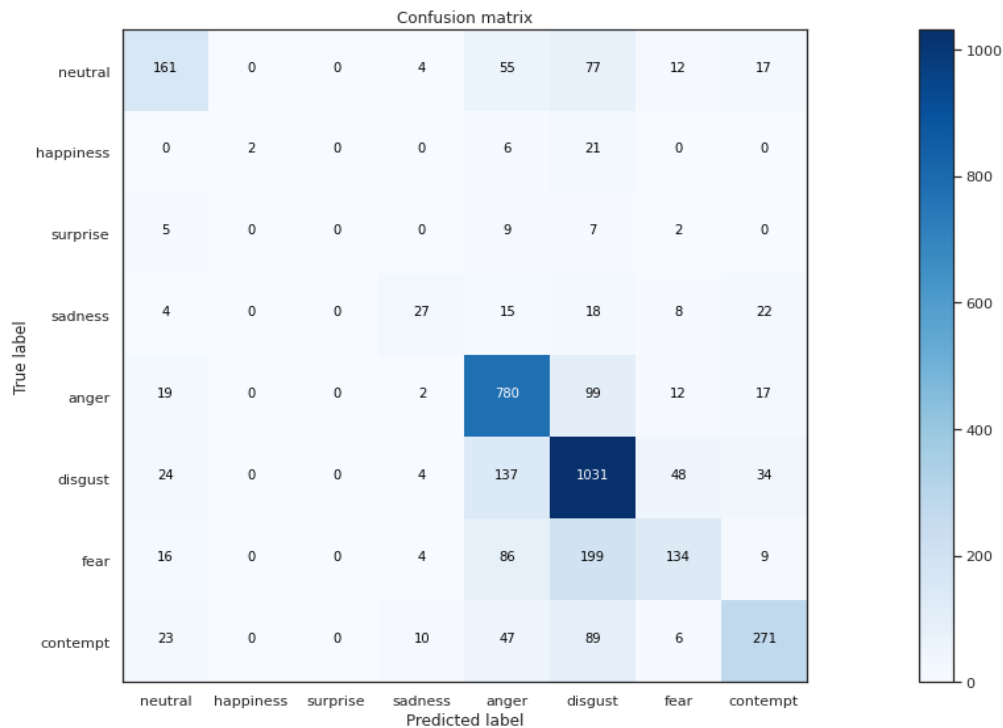


Figure 8. Confusion matrix for the VGG16 pre-trained model

5.4. Comparing the results

By comparing the results of the public test and the private test accuracy and when evaluating the performance using images from the internet, the proposed ERCNN model shows superiority over the VGG16 pre-trained model in terms of accuracy, time, and extra validation using images from the internet. The accuracy for the proposed ERCNN model was 87.133% in the public test and 82.648% in the private test, and each epoch takes 38 seconds to complete the training on the FER+ dataset (2 hours for 200 epochs). For the VGG16 pre-trained model the accuracy was 71.4685% in public test and 67.338% in private test and each epoch takes 52 seconds to complete its training (3 hours for 200 epochs). Several images from the internet with different emotions were used to test the effectiveness of training the proposed ERCNN model using the FER+ dataset. From Figure 4, it is clear that the accuracy of the predictions of the proposed ERCNN model for each emotion was correct and with a high prediction accuracy. The accuracy of the happy face is 99.92%, the neutral face is 97.54%, and the sad face is 97.92%. When evaluating the VGG16 pre-trained model using the internet images, the performance was poor even after increasing the number of training data in FER+, one face emotion is correctly predicted from three faces' emotions as shown in Figure 7. The results of the comparison are shown in Table 4.

Table 4. The accuracy, precision and recall for each model with the used dataset

Experiment name	Public test accuracy	Private test accuracy	Precision	Recall	F1-score	Time per epoch
Proposed ERCNN	87.133%	82.648%	71.3%	68.5%	69.6%	38 sec
VGG16 pre-trained model	71.685%	67.338%	60.8%	42.5%	45.1%	52 sec

5.5. Comparing with the existing work

In this section, we will compare the proposed work system with studies that used the same data set (FER+ data) in training and testing. When the ERCNN model was trained and tested using the FER+ dataset, the results were 87.133% in the public test and 82.648% in the private test. It was higher than the accuracy obtained when training and testing of the VGG16 pre-trained model with the FER+ dataset, where the accuracy in the public test was 71.685% and in the private test was 67.338%. The proposed model was also

tested (extra validation) based on images containing one or more faces in a single image with different emotions. To test the proposed ERCNN on an image containing one or more faces, the Dlib detector and OpenCV libraries were used. Excellent results were achieved, and the proposed ERCNN model has proven to be effective in predicting phase. Lian *et al.* [19] used the DenseNet-BC architecture, which has three dense blocks with 16 layers linked with global average pooling (GAP), then trained the model using the FER+ dataset, which combined the training data with the public test data, and tested the model using the private data, and got an accuracy of 81.93%. Table 5 shows the comparison of the results of the proposed model with those of other studies.

Table 5. Comparison the proposed ERCNN model with other studies

Model	Accuracy FER+
Lian <i>et al.</i> [19]	81.93%
Barsoum <i>et al.</i> [28]	83.852%
Proposed ERCNN model	87.133% in public test 82.648% in private test
Proposed VGG16 pre-trained model	71.685% in public test 67.338% in private test

4. CONCLUSION

In this paper, two models for the recognition of human emotions were compared: a pre-trained model which was trained on a large set of images (ImageNet dataset and classified it into 1,000-class) then retrained by FER dataset and a ERCNN model based on VGG16 network that was built from scratch and trained on data specific to facial expressions only. The obtained results proved the effectiveness of the proposed model for the recognition of human emotions. The proposed ERCNN model outperformed VGG16 in terms of accuracy and time, as well as when evaluating models using images from outside the data used in training and testing. The main goal of this thesis is to enhance the accuracy of identifying human emotions through facial expressions by using CNN's ability to extract features from images (FER+) and classify images. Our next step in this research will be to train and test the proposed model using other data. We also aspire to create a hybrid network that combines the proposed ERCNN model with the VGG16 pre-trained model to obtain a diversity of the extracted features.




REFERENCES

- [1] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar. 2016, pp. 1–10, doi: 10.1109/WACV.2016.7477450.
- [2] S. Shaees, H. Naeem, M. Arslan, M. R. Naeem, S. H. Ali, and H. Aldabbas, "Facial emotion recognition using transfer learning," in *International Conference on Computing and Information Technology (ICCIT-1441)*, Sep. 2020, pp. 1–5, doi: 10.1109/ICCIT-144147971.2020.9213757.
- [3] D. Indira, L. Sumalatha, and B. R. Markapudi, "Multi expression recognition (MFER) for identifying customer satisfaction on products using deep CNN and haar cascade classifier," *IOP Conference Series: Materials Science and Engineering*, vol. 1074, no. 1, Feb. 2021, doi: 10.1088/1757-899X/1074/1/012033.
- [4] A. Poullose, J. H. Kim, and D. S. Han, "Feature vector extraction technique for facial emotion recognition using facial landmarks," in *International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2021, pp. 1072–1076, doi: 10.1109/ICTC52510.2021.9620798.
- [5] A. Poullose, C. S. Reddy, J. H. Kim, and D. S. Han, "Foreground extraction based facial emotion recognition using deep learning exception model," *International Conference on Ubiquitous and Future Networks*, pp. 356–360, Aug. 2021, doi: 10.1109/ICUFN49451.2021.9528706.
- [6] T. S. Arulananth, M. Baskar, and R. Sateesh, "Human face detection and recognition using contour generation and matching algorithm," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 16, no. 2, pp. 709–714, 2019, doi: 10.11591/ijeecs.v16.i2.pp709-714.
- [7] A. H. Ahmad *et al.*, "Real time face recognition of video surveillance system using haar cascade classifier," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 21, no. 3, pp. 1389–1399, Mar. 2021, doi: 10.11591/ijeecs.v21.i3.pp1389-1399.
- [8] K. Shirisha and M. Buddha, "Facial emotion detection using convolutional neural network," *International Journal of Scientific & Engineering Research*, vol. 11, no. 3, 2020.
- [9] G. Pons and D. Masip, "Supervised committee of convolutional neural networks in automated facial expression analysis," *IEEE Transactions on Affective Computing*, vol. 9, no. 3, pp. 343–350, Jul. 2018, doi: 10.1109/TAFFC.2017.2753235.
- [10] A. Saravanan, G. Perichetla, and D. K. S. Gayathri, "Facial emotion recognition using convolutional neural networks," *arxiv.org/abs/1910.05602*, Oct. 2019, [Online]. Available: <http://arxiv.org/abs/1910.05602>
- [11] M. Berrahal and M. Azizi, "Augmented binary multi-labeled CNN for practical facial attribute classification," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 23, no. 2, pp. 973–979, Aug. 2021, doi: 10.11591/ijeecs.v23.i2.pp973-979.
- [12] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.




- [13] Y. Aufar and I. S. Sitanggang, "Face recognition based on Siamese convolutional neural network using Kivy framework," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 26, no. 2, pp. 764–772, May 2022, doi: 10.11591/ijeecs.v26.i2.pp764-772.
- [14] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
- [15] W. Wang and Y. Yang, "Development of convolutional neural network and its application in image classification: a survey," *Optical Engineering*, vol. 58, no. 4, Apr. 2019, doi: 10.1117/1.OE.58.4.040901.
- [16] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, Nov. 2015, pp. 443–449, doi: 10.1145/2818346.2830593.
- [17] K. Liu, M. Zhang, and Z. Pan, "Facial expression recognition with CNN ensemble," in *International Conference on Cyberworlds (CW)*, Sep. 2016, pp. 163–166, doi: 10.1109/CW.2016.34.
- [18] D. V. Sang, N. Van Dat, and D. P. Thuan, "Facial expression recognition using deep convolutional neural networks," in *9th International Conference on Knowledge and Systems Engineering (KSE)*, 2017, pp. 130–135, doi: 10.1109/KSE.2017.8119447.
- [19] Z. Lian, Y. Li, J. Tao, J. Huang, and M. Niu, "Region based robust facial expression analysis," in *First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, May 2018, pp. 1–5, doi: 10.1109/ACIIAsia.2018.8470391.
- [20] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial expression recognition with identity and emotion joint learning," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 544–550, 2021, doi: 10.1109/TAFFC.2018.2880201.
- [21] S. Miao, H. Xu, Z. Han, and Y. Zhu, "Recognizing facial expressions using a shallow convolutional neural network," *IEEE Access*, vol. 7, pp. 78000–78011, 2019, doi: 10.1109/ACCESS.2019.2921220.
- [22] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *The Visual Computer*, vol. 36, no. 2, pp. 405–412, Feb. 2020, doi: 10.1007/s00371-019-01630-9.
- [23] E. Ivanova and G. Borzunov, "Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions," *Procedia Computer Science*, vol. 169, no. 2019, pp. 244–248, 2020, doi: 10.1016/j.procs.2020.02.143.
- [24] J. H. Kim, A. Poullose, and D. S. Han, "The extensive usage of the facial image thresholding machine for facial emotion recognition performance," *Sensors*, vol. 21, no. 6, Mar. 2021, doi: 10.3390/s21062026.
- [25] T. Jing Wei, A. R. Bin Abdullah, N. B. Mohd Saad, N. B. Mohd Ali, and T. N. S. B. Tengku Zawawi, "Featureless EMG pattern recognition based on convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 14, no. 3, pp. 1291–1297, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1291-1297.
- [26] M. A. Obaid and W. M. Jasim, "Pre-convoluted neural networks for fashion classification," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 10, no. 2, pp. 750–758, 2021, doi: 10.11591/eei.v10i2.2750.
- [27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arxiv.org/abs/1409.1556*, Sep. 2014, [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [28] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, Oct. 2016, pp. 279–283, doi: 10.1145/2993148.2993165.
- [29] E. Barsoum, "Microsoft/ FERPlus: this is the FER+ new label annotations for the emotion FER dataset," *GitHub*. 2017, Accessed: Sep. 29, 2021. [Online]. Available: <https://github.com/microsoft/ferplus>.
- [30] B. L. Y. Rowe, "GitHub - muxspace/facial_expressions: a set of images for classifying facial expressions," *GitHub*. 2016, Accessed: Sep. 29, 2021. [Online]. Available: https://github.com/muxspace/facial_expressions.
- [31] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
- [32] "Communique production: image," *WordPress*. <https://communiqueproduction.files.wordpress.com/2014/09/cropped-screen-shot-2014-09-18-at-2-42-53-am.png> (accessed Nov. 11, 2021).

BIOGRAPHIES OF AUTHORS



Asmaa Yaseen Nawaf    holds a Bachelor of Computer Science and is now a master's student. She is currently a teacher at Al-Baida High School in Kirkuk Governorate. Her current research interests include artificial intelligence and deep learning. She can be contacted at asmaayaseen1981@gmail.com.



Wesam M. Jasim    received the B.Sc. and M.Sc. degrees in control automation engineering from University of Technology, Baghdad, Iraq, and Ph.D. degree in Computing and Electronics from University of Essex, Essex, UK. Currently, he is an assistant professor with the College of Computer Science and Information Technology, University of Anbar. His current research interests include robotics, multiagent systems, cooperative control, robust control, linear and nonlinear control, deep learning. He has published research papers at national and international journals, conference proceedings as well as chapters of books. He can be contacted at co.wesam.jasim@uoanbar.edu.iq.