2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA) | 978-1-7281-6535-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCSPA49915.2021.9385739

Facial Expression Recognition Using a Simplified Convolutional Neural Network Model

Amany Kandeel, Mina Rahmanian, Farhana Zulkernine, Hazem M. Abbas, Hossam Hassanein
School of Computing, Queen's University
{19aak6, 18mrs2, farhana.zulkernine, hazem.abbas}@queensu.ca, hossam@cs.queensu.ca

***Abstract*—Facial Expression Recognition (FER) is one of the most important information channels by which Human-Computer Interaction (HCI) systems can recognize human emotions. The importance of FER is not limited to the direct interaction between the machine and humans but can be extended to security, virtual reality, education, and entertainment. In this paper, we propose two Convolutional Neural Network (CNN) models for FER. One of these models achieved 100% accuracy for the JAFFE and CK+ benchmark datasets with lower computational complexity. We applied image augmentation techniques and image enhancement techniques with the first model. The other CNN model is an extended version of the first model that has been validated for the more challenging FER2013 dataset and we obtained 69.32% for this dataset. By comparing to the recent state-of-the-art approaches to FER, we demonstrate the superior accuracy and efficiency of the proposed approaches.**

***Keywords*—Facial Expression Recognition, Deep Learning, Convolutional Neural Network, Data Augmentation**

I. Introduction

In the last decade, Facial Expression Recognition (FER) research gained a lot of attention due to the advancement achieved in related research areas such as face detection and recognition [1][2]. These advancements encourage researchers to study facial expressions and to build real-time FER approaches. FER approaches can be divided into two main categories. The first category is the traditional approaches that extract handcrafted features using methods such as Gabor wavelets, Local Binary Patterns (LBP), and Principal Component Analysis (PCA) [3]. Subsequently, the features are categorized into the respective facial expression classes based on classification methods such as Support Vector Machine (SVM) and Nearest Neighbor (NN) [4]. The second category is deep learning-based approaches that rely on reducing the dependence on manual extraction of facial patterns and enable machines to learn directly from the input images [5]. The deep learning-based approaches are composed of three basic steps: pre-processing, feature learning, and feature extraction. The pre-processing step is employed before training to enhance the input images, such as face alignment, image cropping, and face normalization [6]. Feature learning and extraction are performed using Deep Neural Networks (DNNs), which use training data and artificial intelligence algorithms to learn the relations among the extracted features [4]. Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are the most common deep learning approaches used in this era, which enable learning spatial data patterns and temporal data patterns respectively [5]. In the CNN model, the input image is convoluted through a combination of filters; each filter has a particular set of values to produce a specified feature map.

CNN is one of the most popular algorithms for deep learning for images and videos. Like traditional neural networks, CNN is composed of three types of layers: input, output, and multiple hidden layers. These three types of layers are classified into two groups, feature detection layers and classification layers. Feature detection layers are composed of convolution layers, pooling layers, and generating activation. Convolution layers use filters to convolute or transform an image. Next pooling layers reduce the extracted features to the key feature set. Finally, the Rectified Linear Unit (ReLU) generates non-linear activations to overcome the problem of vanishing gradient, allowing models to learn faster and perform better during the model training phase. Classification layers contain a set of fully connected layers that output the predicted values. The final layer uses different activation functions such as the Softmax function, to classify the predicted values into output classes [7].

Although humans can easily recognize facial expressions, it is still a challenge for the machines. The first challenge for FER is due to the resemblance of some of the facial expressions in the facial feature space. For instance, the sad and angry expressions of the same person may have very similar representations, as shown in Fig. 1. Another challenge is that the same person may have different facial expressions for the same emotion. Some of the FER challenges are contributed by the conditions of the image. For instance, images of the same expression may differ in brightness, illumination, background, occlusion, and face pose, as shown in Fig. 2. The scarcity and the small sizes of facial expression datasets are also important challenges in FER research.

In this paper, we present two CNN models. The first model addresses the challenge of having a small training dataset by using an image augmentation technique and

978-1-7281-6535-6/20/$31.00 ©2020 IEEE

(a) Angry

(b) Sad

Fig. 1: The ambiguity between angry vs sad emotion

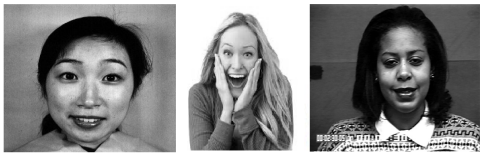

Fig. 2: Three different happy expressions

achieve an accuracy of 100% with JAFFE and CK+ datasets. The second model is designed to handle a more challenging FER2013 dataset, which includes a huge number of faces that vary in the head pose, occlusion, and illumination; our model achieves a reasonable accuracy of 69.32% for the FER2013.

The remainder of this paper is organized as follows. Section II presents the related work. Section III describes the datasets used in this research. The architecture and the performance of the two proposed models are described in Sections IV and V respectively. A critical discussion and comparison of our results with other recent FER methods are presented in Section VI. Finally, we conclude and present the future work directions in Section VII.

II. Related Work

FER attracted considerable attention from many behavioral scientists since the work of Charles Darwin in 1872 about human and animal expressions [8]. In 1971, Ekman and Friesen [9] specified six basic emotions shared by humans across different cultures: anger, disgust, fear, happiness, sadness, and surprise. Later, researchers added neutral and contempt to the list of the basic emotions [10][11]. The first attempt of automatically analyzing facial expressions in image sequences was made in 1978 by M. Suwa [12].

FER approaches are divided into two categories: conventional approaches based on manual feature extraction [3], and deep learning-based approaches, which have already proven to be more efficient [4]. In this paper, we focus on deep learning-based approaches where, given the data and the algorithm, machines automatically learn to extract features. In this section, we review some of the recent research in FER which apply deep learning approaches.

Previous research in FER focuses either on images captured in a controlled lab environment, using volunteers who follow specific instructions to express emotions [13] or on images that are captured in an uncontrolled environment such as images posted by people on the web having different poses, occlusions, and illuminations [14][15].

Lopes et al. [6] combined some of the image preprocessing steps such as intensity normalization and rotation correction with a CNN having two convolution layers. They reported accuracy of 53.44%, 96.76%, and 71.62% for JAFFE, CK+, and BU-3DFE datasets, respectively, for a seven-class classifier. Yang et al. [13] proposed a Weighted Mixture Deep Neural Network (WMDNN) to process facial grayscale images and the corresponding Local Binary Pattern (LBP) facial images. They trained the VGG16 model using the ImageNet database and obtained an average recognition accuracy of 97.02%, 92.21%, and 92.3% for CK+, JAFFE, and Oulu-CASIA datasets, respectively. Some FER approaches used static images [2], while others used videos. Zhang et al.[15] presented a hybrid deep learning model that consists of a spatial CNN, a temporal CNN, and a Deep Belief Network (DBN). Their model achieved an accuracy of 75.97% and 84.24% for FER-2013 and RAF-DB datasets, respectively.

III. Datasets

Some benchmark datasets are available publicly and are commonly used by researchers to evaluate the FER approaches. Each dataset has different characteristics and varies in image size, data type, diversity, and other image specifications such as illumination, occlusion, and poses of faces. In order to explore the different challenges of FER, we chose three datasets from the commonly used ones; namely, the Japanese Female Facial Expression (JAFFE) [10], the Extensive Cohn–Kanade (CK+) [11], and the Facial Expression Recognition Challenge (FER2013) datasets [16]. Since the JAFFE and the CK+ are small datasets (less than 1000 samples), we needed to address the data scarcity problem in FER. We applied image augmentation (as will be described later in Section IV-A), which significantly improved the model accuracy. The FER2013 is a large (more than 35,000 sample) dataset. So, we used augmentation for only one emotion to improve data distribution. FER can be applied to static images or to a sequence of video frames. While JAFFE and FER2013 contain static images, CK+ contains sequences of images extracted from video frames. Both JAFFE and CK+ are collected from subjects in a lab under controlled environments. Consequently, all the images are frontal faces and have limited variations in image conditions. On the other hand, FER2013 samples are collected from real-world Internet data. Therefore, the images vary greatly in different aspects, such as occlusion, age of the person, skin color, the position of the face, and overall image quality, which affect the recognition accuracy. We analyze and process each of the above datasets differently because of the differences in image types, sizes, distribution, and characteristics, as explained in the next subsections.

A. The JAFFE Dataset

The JAFFE database [10] contains 213 images of 7 facial expressions (six basic facial motions and the neutral expression) posed by 10 Japanese female models.

The main challenge of this dataset is its small size. Thus, we use data augmentation to overcome this problem.

B. The CK+ Dataset

We use the second version of the Cohn-Kanade database [11]. It has 593 sequences from 123 subjects, out of which only 327 sequences have labels referring to seven emotions, namely, anger, contempt, disgust, fear, happiness, sadness, and surprise.

C. The FER2013 Dataset

The FER2013 dataset was prepared for a 2013 Facial Expression Recognition challenge [16]. The images were collected from the Internet, and the faces greatly vary in age, occlusion, and pose. In addition, this dataset contains a number of invalid samples, including non-face images, incorrect face cropping, and expression labeling errors as seen in Fig. 3. These factors offer great challenges and have a significant effect on the FER models. On the other hand, the diversity in the dataset is beneficial for building a robust model [17]. The FER2013 is one of the largest FER datasets, which includes 35,887 images [18]. This dataset uses seven facial expressions (six basic expressions and the neutral expression) [19].

Fig. 3: FER2013 dataset sample

IV. First proposed model

Our goal was to build a FER model that is simple and accurate for the selected datasets. To achieve that, we used a CNN model with very few parameters. Because of the diversity of the images in the datasets, we defined two models. Our first model was developed to work for small datasets such as the JAFFE and CK+, which contain frontal faces.

One of the main challenges in developing deep learning approaches is the need for large training data to perform accurately. Therefore, in developing our first model, the goal was to overcome the small data size challenge, which we did by using data augmentation to increase the size of the dataset. In addition, we enhanced the images to improve model performance. The main stages of the data processing workflow of the first model are shown in Fig. 4, which starts with detecting the face followed by cropping, and then augmenting the number of images by rotation. Finally, these images are used to train a CNN model. The multilayered CNN model is also shown in Fig. 4. In the following section, the data processing workflow is described in more detail.

A. Data Pre-processing

Image processing tools improve the performance of FER by enhancing the quality of images and removing information that is not needed for expression recognition and thus help reduce model complexity. It is also important to perform the pre-processing image enhancements without affecting the facial expression in the image, i.e., enhancing the image while ensuring so that the emotion remains unchanged. Therefore, we first apply image cropping and normalization to enhance the images in the JAFFE and CK+ datasets. Down-sampling is exploited to reduce the model parameters, and rotation is used to create new images and increase the number of training images.

a) Image Cropping: The original images may have backgrounds containing unimportant details, which may reduce the recognition accuracy. Image cropping overcomes this problem by removing the areas which do not have any important information for FER and keeps only the face area, as shown in Fig. 4.

b) Image Normalization: Image normalization is the process of changing the range of pixel intensity values. This process makes training less sensitive to the scale of features.

c) Down-Sampling: Down-sampling is used to decrease the image size without affecting the image details. It reduces the number of parameters, which leads to alleviating the computational complexity of the model without causing image distortion. Down-sampling is applied to both JAFFE (256×256) and CK+ (640×480 and 640×490) to convert the images into 48×48 pixel size images. However, the FER2013 images already have 48×48 pixels.

d) Data Augmentation Using Rotation: Data augmentation is used to expand the size of training datasets by adding modified versions of the images to the dataset. Image augmentation may include random cropping, shifting, noise addition, flipping, and translation. We applied data augmentation to the whole JAFFE and CK+ datasets and to only images showing a specific type of disgust emotion of FER2013, where it has very few numbers of samples compared with other emotions.

One of the data augmentation methods is known as image rotation, where images are rotated by small angles without affecting the face alignment. As shown in Fig. 4, we rotate the original images by angles less than 10 degrees to reduce the effect of rotation on the expressions in the original images. Using this approach, we can obtain a large number of additional images generated from each image. Because the JAFFE dataset is very small, we increase it by 50 times using only 5 degrees of rotation at a step of 0.1 degrees. Thus we increase the size of the JAFFE dataset from 213 to more than 10,000 images. For the CK+ dataset, we generate additional images for each image, which leads to the dataset of size 9,810 images from the original 981 images. This image augmentation raises the model accuracy significantly.

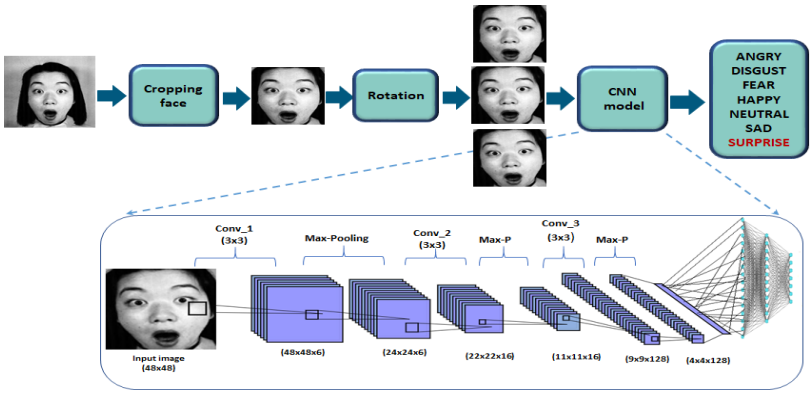

Fig. 4: The first proposed model architecture

B. CNN Model

CNN is one of the deep learning approaches that achieves impressive performances in many applications, such as image classification and recognition [7]. CNN is a common deep learning model for images due to its powerful ability to treat arrays [6]. CNN is composed of two sets of layers: 1) the feature extraction layer, which is composed of convolutional, pooling, and ReLU units; and 2) the classification layer, which is composed of fully connected layers and a final output layer. According to the architecture of the model in Fig. 4, the processed images from cropping and rotation are used to train the CNN model. The training process defines the relationships and weights in the model, after which it becomes ready to be used for scoring the unseen test images and place them into different categories based on emotion. Our CNN model is trained using three layers with different filters as described below.

1) Convolutional Layer: This layer receives the pre-processed images and extracts features according to the defined kernels in this convolution. The different kernels perform multiple operations such as edge detection and sharpening.

a) ReLU Activation Function: This is one of the most common activation functions and is used, especially in CNN and deep learning. It is used to increase the non-linearity in images to address the vanishing gradient problem in using supervised learning to train the CNN models.

b) Batch Normalization: It normalizes the input layer by adjusting and scaling the activations. It is used to speed up and stabilize training on neural networks. It also allows greater learning rates to expedite convergence towards more accurate solutions [20].

2) Pooling Layer: Pooling summarizes the convoluted area into a single value and thus helps in reducing memory consumption in deeper layers. It also converts the spatial information into features. The most common types are max-pooling and average pooling.

As shown in Fig. 4, the model uses three layers, followed by a max-pooling layer (with kernel size 2x2). The number of filters is chosen by experiments to increase the accuracy of the model performance while reducing the number of parameters. The number of filters for the first model is 6, 16, and 128, respectively.

3) Fully Connected Layers: Our model has two fully connected layers, where every node in a layer is connected to all nodes in the next layer. In the final output layer, the Softmax function is employed for classification. In the fully connected layers, the first layer has 256 nodes, and the second layer has 128 nodes. Both of them have a dropout ratio of 0.5.

C. Validation: First Model

We evaluate this model in three stages: after applying the model on original images, then on cropped images, and finally after image rotation.

1) JAFFE Results: After training the model using 85% of the JAFFE dataset and validating using 15%, the test accuracy of the model without the pre-processing or augmentation steps converged at 78.12% after 150 epochs. Using the image cropping (number of images is still 213 images), the model accuracy increased to 87.5% after the 150 epochs. Finally, after applying the data augmentation process (10,650 images), the accuracy reached 100% after 100 epochs as shown in Table I and Fig. 5. The results demonstrate the importance of the image prepossessing on model accuracy.

2) CK+ Results: To run the first model on the CK+ dataset, we extracted the last three frames from each of the 327 labeled sequences and gathered a total of 981 images. The model was trained using 85% of this data and tested using the remaining 15%. The accuracy of the model using the original dataset was 96.62% after 100 epochs. After cropping the images, the accuracy reached 97.97% after 100 epochs. Finally, using the

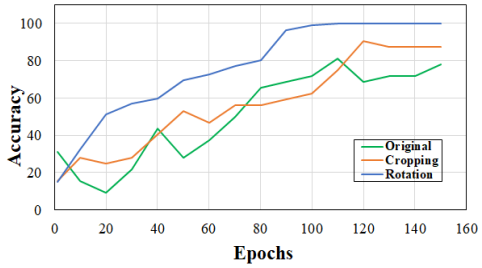

Fig. 5: The first model accuracy with JAFFE dataset

augmentation, the model achieved 100% accuracy after 10 epochs as shown in Table I and Fig. 6.

TABLE I: The first model accuracy

	Original	Cropping	Rotation
CK+	96.62%	97.97%	100%
JAFFE	78.12%	87.5%	100%

Regarding the model complexity, the total number of parameters of the model is 578,463, which is considered a small value compared to some of the other models that achieved comparable accuracy [7][13].

V. The Second Model

We evaluated the first model on a more challenging dataset, the FER2013, but obtained a low accuracy of 51.43%. Therefore, we extended this model and implemented a more complex one to address the challenges inherent in this dataset, i.e., variations in image illumination, face poses, occlusions of existing parts of the face, and images not having faces at all, as shown in Fig. 3.

A. Data Pre-processing

One of the challenges in the FER2013 dataset is imbalanced data, where the disgust emotion has only 547 samples compared to the other emotions for which the number of samples ranges between 4,000 to 9,000. To overcome this imbalanced data problem, which contributed to the poor performance, we augmented the images for disgust emotion only using just 5 degrees of image rotation at a step of 0.5 degrees to increase the number of disgust images from 547 to 5,470.

B. CNN Model

We preserved the basic structure of our first CNN model but added additional processing to each convolutional layer with more filters.

1) Feature Extraction Layers: We used three convolutional layers as in model one, each layer includes two convolution layers (with 3x3 kernel and zero padding) followed by the max-pooling layer (with kernel size 2x2) and the batch normalization function to extract the features in each layer. Additionally, we increased the number of filters in the three layers to 64, 128, and 256, respectively.

2) Classification Layers: The model contains two fully connected layers as the first model with the same number of nodes and the dropout ratio. In the final output layer, there are seven nodes that implement a Softmax function to classify emotions. Additional parameters, such as regularization, are added to improve the model performance and address the problem of overfitting.

C. Validation: Second Model

We trained and tested the model using the augmented FER2013 dataset, where 90% data was used for training and 10% for testing. The total number of parameters in this model is 2,820,039, which is five times more than that of the first model.

1) FER2013 Results: Our first model achieved 51.43% accuracy after 150 epochs with the original FER2013 dataset, while the second model achieved 64.02% accuracy with the same 150 epochs. With the partially augmented dataset having a more balanced data distribution, the second model obtained 69.32% accuracy after the same number of epochs as illustrated in Fig. 7.

VI. Comparison Of Recognition Results

We presented two different models, which were validated against three different datasets. We compare the results with some other approaches, as shown in Table II. All these approaches use CNN models. The most similar work is by Lopes et al. [6], where data preprocessing techniques were used with the CNN model. The model was evaluated using the JAFFE and CK+ datasets. But their model differs in adding noise in the data augmentation step instead of rotation. Lopes et al. used a different dataset, the BU-3DFE, instead of using FER2013. Puthanidam et al. [2] also used a similar sequence of steps with different datasets. Their model achieved 89.58%, 100% and 71.975% for KDEF, JAFFE, and combined (KDEF + JAFFE + SFEW) respectively.

As shown in the table, our first proposed model provides the highest accuracy compared to other models on the JAFFE and the CK+ datasets. Our first model succeeds in achieving two objectives: performance and simplicity. It reaches greater accuracy with a fewer number of parameters (0.56 million) than Yang et al. [13] who use the VGG16 model having 138 million parameters, and Mollahosseini et al. [7], whose model uses 9 million parameters [21]. Although our first model achieved high accuracy on small datasets using data augmentation, it did not achieve high accuracy for the FER2013 dataset (51.43%) due to the challenging nature of this dataset. However, our second CNN model obtained 69.32% for the FER2013 dataset. As shown in Table II, this result is better than that reported by most

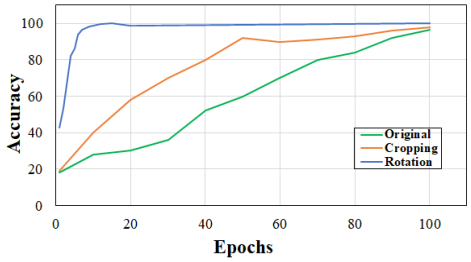

Fig. 6: The first model accuracy with CK+ dataset

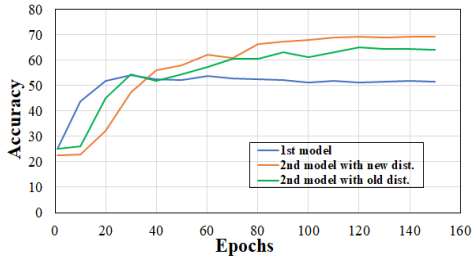

Fig. 7: The FER2013 accuracy

other researchers except Zhang et al., who implemented a complex hybrid model of two groups of CNNs (eight 5-layer CNNs and twelve 11-layer CNNs). Our second model performed considerably well with a reasonable number of parameters (2.8 million).

TABLE II: Accuracy comparison between previous approaches and the proposed models

Method	JAFFE	CK+	FER2013
Yang et al [13].	92.21%	97.02%	—
Lopes et al [6].	53.44%	96.76%	—
Yan Yan et al [22].	88.2%	98.7%	—
Mollahosseini et al [7].	—	93.2%	66.4%
Jie Shao et al [23].	—	92.86%	68%
Puthanidam et al [2].	100%	—	—
Zhang et al [15].	—	—	75.97%
Agrawal et al [24].	—	—	65.77%
Our first model	**100%**	**100%**	**51.43%**
Our second model	—	—	**69.32%**

VII. Conclusion

In this paper, we implemented two FER models. We evaluated the first model on three different FER datasets and evaluated the second on a more challenging dataset. The first model applied a combination of preprocessing steps and a CNN model. It succeeded in recognizing frontal face datasets, whether they be images or sequences of images, and achieved an accuracy of 100% for both the JAFFE and the CK+ datasets. The second model was used for a more challenging dataset, FER2013, and achieved an accuracy of 69.32% due to the challenges inherent in this dataset. However, this accuracy is reasonable compared to the other models.

In the future, we plan to validate our models against other datasets and more data augmentation methods to improve accuracy. We will try to improve the accuracy further for the FER2013 dataset. Although we used sequences of frames in the CK+ dataset, we did not consider the temporal aspect in expression recognition. So, we like to extend the model to recognize facial expressions in videos as future work.

References

[1] M. A. Berbar, H. M. Kelash, and A. A. Kandeel, "Faces and facial features detection in color images," in *Geometric Modeling and Imaging–New Trends (GMAI'06)*. IEEE, 2006, pp. 209–214.

[2] R. V. Puthanidam and T.-S. Moh, "A hybrid approach for facial expression recognition," in *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*. ACM, 2018, p. 60.

[3] J. Kumari, R. Rajesh, and K. Pooja, "Facial expression recognition: A survey," *Procedia Computer Science*, vol. 58, no. 1, pp. 486–491, 2015.

[4] Y. Miao, H. Dong, J. M. A. Jaam, and A. E. Saddik, "A deep learning system for recognizing facial expression in real-time," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 2, pp. 1–20, 2019.

[5] B. Ko, "A brief review of facial emotion recognition based on visual information," *sensors*, vol. 18, no. 2, p. 401, 2018.

[6] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[7] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–10.

[8] C. Darwin and P. Prodger, *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.

[9] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.

[10] M. Lyons, M. Kamachi, and J. Gyoba, *The Japanese Female Facial Expression (JAFFE) Database*, Apr. 1998, n.b. Unauthorized redistribution of the JAFFE database is not permitted.

[11] 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*. IEEE, 2010, pp. 94–101.

[12] M. Suwa, "A preliminary note on pattern recognition of human emotional expression," in *Proc. of The 4th International Joint Conference on Pattern Recognition*, 1978, pp. 408–410.

[13] B. Yang, J. Cao, R. Ni, and Y. Zhang, "Facial expression recognition using weighted mixture deep neural network based on double-channel facial images," *IEEE Access*, vol. 6, pp. 4630–4640, 2017.

[14] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 2015, pp. 435–442.

[15] X. Zhang and Y. Ma, "Learning of complicate facial expression categories," in *Proceedings of the 2019 International Conference on Image, Video and Signal Processing*. ACM, 2019, pp. 73–80.

[16] I. J. Goodfellow, D. Warde-Farley, P. Lamblin, V. Dumoulin, M. Mirza, R. Pascanu, J. Bergstra, F. Bastien, and Y. Bengio, "Pylearn2: a machine learning research library," 2013.

[17] H. Ma and T. Celik, "Fer-net: facial expression recognition using densely connected convolutional network," *Electronics Letters*, vol. 55, no. 4, pp. 184–186, 2019.

[18] V. V. Salunke and C. Patil, "A new approach for automatic face emotion recognition and classification based on deep networks," in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. IEEE, 2017, pp. 1–5.

[19] X. Wang, X. Wang, and Y. Ni, "Unsupervised domain adaptation for facial expression recognition using generative adversarial networks," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[20] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, "Deep learning based on batch normalization for p300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, 2018.

[21] M. Jeong and B. C. Ko, "Driver's facial expression recognition in real-time for safe driving," *Sensors*, vol. 18, no. 12, p. 4270, 2018.

[22] Y. Yan, Z. Zhang, S. Chen, and H. Wang, "Low-resolution facial expression recognition: A filter learning perspective," *Signal Processing*, p. 107370, 2019.

[23] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, pp. 82–92, 2019.

[24] 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*, pp. 1–8, 2019.