An Optimized Neural Network Model for Facial Expression Recognition over Traditional Deep Neural Networks

Pavan Nageswar Reddy Bodavarapu[1], P.V.V.S Srinivas[2]
Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, Vijayawada, India

***Abstract*—Emotions have a key role in Feedback analysis to provide a good customer service, the main seven emotions are Anger, Disgust, Fear, Happy, Neutral, Sad and Surprise. There are several advantages, an efficient Facial Emotion Recognition model can help us in self-discipline and control over the drivers, while they are driving the vehicle. Low resolution and Low-reliable images are main problems in this field. We proposed a new model which can efficiently perform on Low resolution and Low-reliable images. We created a low resolution facial expression dataset (LRFE) by collecting various images from different resources, which contains low resolution images. We also proposed a new hybrid filtering method, which is a combination of Gaussian, Bilateral, Non local means filtering techniques. Densenet-121 achieves 0.60 0.68 accuracy on fer2013 and LRFE respectively. When hybrid filtering method is combined with Densenet-121, it achieved 0.95 accuracy. Similarly Resnet-50, MobileNet, Xception models performed effectively when combined with the hybrid filtering method. The proposed convolutional neural network(CNN) model achieved 0.65 accuracy on fer2013 dataset, while the existing models like Resnet-50, MobileNet, Densenet-121 and Xception obtained 0.60 0.57 0.60 0.52 accuracies on fer2013 respectively. The proposed model when combined with hybrid filtering method achieved 0.85 accuracy. Clearly the proposed model outperforms the traditional methods. When the hybrid filtering method is combined with the CNN models, there is significant increase in the accuracy.**

***Keywords*—Facial expression recognition; deep learning; filtering techniques; convolutional neural network; emotion**

I. Introduction

The raw data consists of noise like random variation of brightness or color information, removing noise from the images drastically improves the performance of the facial emotion recognition models. To eliminate noise from images there are many denoising techniques such as gaussian blur, bilateral filter, non-local means filtering. Gaussian Blur helps in blurring the edges and reducing the contrast, but it reduces the details [1]. Bilateral Filter decreases the noise by preserving the edges by replacing the intensity of pixels with weighted average of intensity from surrounding pixels [2]. Gaussian Filter, Bilateral Filter and other traditional filtering techniques can remove image noise, but the image structure information is not retained enough. Non Local Means Filtering averages neighbours with similar neighbourhoods, with much greater clarity and smaller extent loss of detail post-filtering. The limitation of this technique is, efficiency is slightly lower when compared to traditional techniques. The computation complexity is quadratic in number of pixels in the image, so it is expensive to apply. To speed up the execution many techniques were designed, one such technique is fast Fourier transform, it determines the similarity between two pixels by speeding up the algorithm by factor of 50 and also maintains the quality of result [3][4].

When compared grayscale images with RGB images, grayscale images achieves more accuracy in object recognition field. The other benifit of using grayscale images is, cost of computation will decrease [5][6]. Due to continuous gradient updating, overfitting is one of the basic issues in neural networks. This results in poor performance of the neural network model [7]. For a deep learning model to perform well, it needs a large amount of samples. Gathering more number of samples or large dataset might be expensive, so an possible way is, to automatically generate new samples, this process is called data augmentation. Data Augmentation is used to improve neural network model performance by decreasing data bias and improving the model generalization [8]. Batch normalization is used to increase the stability of a neural network and allows us to use higher learning rates. For as much as dropout can decrease overfitting in a model, a batch normalized neural network can remove or reduce the overfitting [9]. The traditional way for training a CNN is via stochastic gradient descent [10]. Instead of decreasing the learning rate, increase the batch size during the training. This method shows identical performance with lesser parameter updates [11][12][13].

Haar feature-based cascade classifier is a machine learning approach, where the cascade function is trained on positive and negative images. This approach is useful in object detection in images [14][15]. The best way to distinguish a neutral facial emotion from other emotions is to check whether the person's mouth is open or closed. If the mouth seems to open, then that it does not belongs to neutral Facial emotion. A lot of research is being done in mobile applications for Emotion recognition tasks. Even though the present mobile devices have enough memory and processing power, when compared to previous generation smart phones, we cannot directly use solutions from computer to smart phones in respect of facial emotion recognition [16][17]. Social robots are very much needed for the society, they can behave as consort for old people, help doctors during

operations. In facial emotion recognition mouth, eyes, eyebrows play a key role for emotion recognition, Gabor filter helps to obtain these features from an image [18][19].

Our major contributions in this research paper can be outlined as: (1) designed a novel convolutional neural network, Fig. 4 represents the proposed model architecture; (2) presented hybrid denoising method; (3) low resolution facial expression (LRFE) dataset is created for facial expression recognition; (4) compared with traditional methods. We applied various filtering techniques such as Average filtering, Median filtering, Gaussian filtering, Non local means filtering, Bilateral filtering and Hybrid denoising method to both FER2013 and LRFE dataset and compared the results. The Hybrid denoising method is presented by combining Gaussian, Bilateral and Non local means filtering techniques. The proposed model is compared with traditional methods, the batch size used in this research is 32 and trained for 100 epochs. Various techniques like dropout, L2 regularization are used to avoid overfitting. We build a novel convolutional neural network because the existing methods are not working well on the test sets and are very large in size(more number of layers when) and taking more time to train them. So our proposed convolutional neural network overcomes all these problems. In section III proposed work we explained about the dataset used in this paper and the algorithm of the proposed model. Next, in experiment and result section, we pointed out all the experimental results along with graphs of all the techniques used. Table I outlines the description of FER2013 and LRFE datasets, respectively.

TABLE I. OUTLINE OF FER2013 AND LRFE DATASET

Category	FER2013	LRFE
Downloadable	YES	NO
No.of emotions	7	7
Gender	FEMALE/MALE	FEMALE/MALE
No.of images	35887	6100

II. RELATED WORK

Zhiding Yu et al [20] proposed a method based on ensemble of three face detectors, followed by a classification module with ensemble of various convolutional neural networks. Each convolutional neural network model is pre-trained and fine-tuned on Facial Expression Recognition challenge 2013 and SFEW 2.0, respectively. This method achieved 61.29% on test set of SFEW 2.0. Zhihao Zhang at al [21] designed a convolutional neural network to extract features from video clips and a feature matrix processing method is used for identifying the apex frame from such a long video. By combining feature extraction and feature matrix processing methods, the model achieved smaller Mean Absolute Error (MAE). Samira Ebrahimi Kahou et al [22] proposed a hybrid convolutional neural network (CNN) – recurrent neural network (RNN) method for facial expression recognition. Recurrent Neural Networks produced state-of-art performance on diverse set of sequence analysis tasks. The results show that higher recognition accuracy can be achieved by combining feature-level and decision-level fusion networks.

Bing Feiwu et al [23] proposed a model, which can solve the problem of customizing the general model without the label information of the testing samples. The model resulted an improve in accuracy by 3.01% 0.49% 5.33% when tested on extended Cohn-Kanade (ck+), Radboud Faces Database (RaFD) and Amsterdam Dynamic Facial Expression set (ADFES) respectively. Shamim Hossain et al [24] designed a model for mobile application, which can detect the facial emotions with less computation, since a mobile device has limited processing power we need a model which can recognize facial emotions with computationally less expensive. The proposed model takes only 1.4 seconds to recognize one instance of emotion and obtained an 99.8% 99.7% accuracies on JAFFE database and CK database respectively. Jia Deng et al [25] proposed conditional generative adversarial network approach to reduce the intra-class variations. The proposed approach consists of a generator G and discriminators (Di, Da and Dexp). For learning the generative and discriminative representations, three loss functions were designed. But there is one limitation in this approach is that the model is trained individually for each different datasets, a model which is trained on a particular dataset may result in poor accuracy on another dataset.

Hongli Zhang et al [26] designed a method based on convolutional neural network and edge detection for facial emotion recognition. For testing this they created a simulation experiment by combining the fer-2013 database with LFW dataset. The average recognition obtained by this method is 88.56% and the train speed on the training dataset is 1.5 times faster than the traditional method. Yingying Wang at al [27] proposed a hybrid transfer learning model, which is based on Convolution Restricted Boltzmann Machine (CRBM) model and a Convolutional Neural Network (CNN) model, since there are some content differences between the datasets during traditional transfer learning, which affects the classification performance of the model. In this model CRBM replaces the full connection layer in the CNN model. The added CRBM layer learns about the unique statistical characteristics of the target set. This helps in eliminating the content differences between the datasets.

Ronak Kosti et al [28] presented “Emotions in context Database” (EMOTIC), this dataset contains images of people in context in non-controlled environments with 26 emotional categories. They trained a convolutional neural network model on EMOTIC dataset that can analyze the person and the whole scene to classify the emotion states. There model is able to make notable guesses on the emotion states, when the face of the person is not visible. Jianzhu Guo et al [29] created ICV-MEFED dataset. It includes 50 classes of compound emotions (e.g., happy-disgusted and sadly-fearful) and labels that are evaluated by psychologists, since the labels that are obtained automatically by machine learning based algorithms could lead to inaccuracies. They have organized a challenge on the ICV-MEFED dataset at FG workshop 2017. After analyzing the top three methods, the experimental results indicate that pairs of compound emotions (e.g., happily-surprised vs surprisingly-happy) are more difficult to recognize.

III. Proposed Work

A. Filter Description

The main aim of our research is to compare the Facial Emotion recognition accuracy of Gaussian, Bilateral, Non local means, Average, Median, Hybrid denoising techniques. A hybrid denoising method is proposed by combining the Gaussian, Bilateral, Non-local Means denoising techniques. Guassian filter is a 2D convolution filter, which blur the image, helping in remove the noise. The only limitation with this technique is, the loss of image details is high when compared to other techniques. Bilateral is a non-linear filtering technique used to remove noise from the image by preserving the edges. The limitation of this technique is that it introduces false edges in the image. Non local means filter, unlike taking the mean value of a group of pixels, non local means takes a mean of all pixels and unlike other techniques which blur the image, non local means can restore the texture of image.Median filter is one of the non-linear digital filtering technique, used to remove the noise from the images. It removes the noise from the images by preserving the edges. For removing salt and pepper noise, median filter is most effective.Average filtering helps in removing the noise from the images by replacing each value with average of neighbouring pixels; decreases the intensity variation among neighbouring pixels.

B. Dataset Description

1) FER2013 dataset: This dataset contains nearly 35887 images of various people facial expressions. It is a publicly available dataset, which enables to do research in the field of facial expression recognition. It contains both male and female gender images. The no. of emotions in fer2013 are seven (Happy, Sad, Neutral, Fear, Disgust, Anger, Surprise). This dataset is then divided in the ratio of 80:20 for training and testing purpose. Thus, the training and testing set contains 28709 and 7178 images respectively. Fig. 1 shows the sample images of FER2013 dataset.

2) LRFE (Low resolution facial expression) dataset: We collected images and videos from different resources belonging to seven different Facial Emotions (Anger, Disgust, Sad, Neutral, Fear, Happy, Surprise). All the videos are split into images and these images are organized into their respective directories based on the Facial Emotion. The dataset contains 35000 images belonging to seven different Facial Emotions. The raw images collected are of different formats (file extensions with .png, .gif, .tiff, .jpg). So we converted all the images with file extension other than .JPG into .JPG format. Next, we converted all the images into grayscale format from RGB format. After converting the images into grayscale, we resized all the images to 48X48 pexels. Fig. 2 shows the sample images of LRFE dataset.

3) Mixed dataset: Randomly various images from each emotion category are mixed together from FER2013 and LRFE dataset to form the mixed dataset. This dataset contains nearly 35000 images belonging to seven different facial expressions (Happy, Sad, Neutral, Fear, Disgust, Anger, Surprise). Later different types of denoising techniques like Bilateral, Non local means, Gaussian, Average, Median filtering are applied to all images. Fig. 3 shows the sample images of Mixed dataset.

Fig. 1. Sample Images in FER2013 Dataset.

Fig. 2. Sample Images in LRFE Dataset.

Fig. 3. Sample Images in Mixed Dataset.

C. Algorithm

Step 1: Input of image dataset containing seven different facial emotions.

Step 2: Firstly, convert all the images into JPG format.

Step 3: Secondly, convert all the RGB images into gray scale format.

Step 4: Thirdly, re-size all the corresponding gray scale images to 48X48 pixels.

Step 5: Now, Assign labels to all the images after re-sizing.

Step 6: Split the dataset into 80:20 ratio for training and testing the model for classification of image.

Step 7: Train the model on the training set and evaluate the model on the testing set.

Step 8: Finally, output the classification of image based on the emotion expressed in the image.

D. Model Architecture

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_8 (Conv2D)            (None, 48, 48, 32)        320
_________________________________________________________________
conv2d_9 (Conv2D)            (None, 48, 48, 64)        18496
_________________________________________________________________
batch_normalization_4 (Batch (None, 48, 48, 64)        256
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 24, 24, 64)        0
_________________________________________________________________
conv2d_10 (Conv2D)           (None, 24, 24, 128)       73856
_________________________________________________________________
conv2d_11 (Conv2D)           (None, 22, 22, 256)       295168
_________________________________________________________________
batch_normalization_5 (Batch (None, 22, 22, 256)       1024
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 11, 11, 256)       0
_________________________________________________________________
flatten_2 (Flatten)          (None, 30976)             0
_________________________________________________________________
dense_4 (Dense)              (None, 1024)              31720448
_________________________________________________________________
dropout_5 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_5 (Dense)              (None, 7)                 7175
=================================================================
Total params: 32,116,743
Trainable params: 32,116,103
Non-trainable params: 640
_________________________________________________________________
```

Fig. 4. Model Architecture.

IV. Experiment and Results

A. Performance on FER2013 Dataset

The Table II (fer2013 dataset) is divided in the ratio of 80:20 for training and testing purpose, the batch size is 32 and all the models are trained for 100 epochs. During model implementation, 80 percent of fer2013 dataset is used for training the model and remaining 20 percent is divided into validation and testing the model.

Table III shows the accuracy and loss comparison of different deep learning models and proposed FerExpNet model on Fer2013 dataset. The proposed FerExpNet achieves an accuracy of 0.79 0.65 on training and testing sets of Fer2013 respectively. The results clearly indicate that the proposed FerExpNet is performing better than the state-of-models on Fer2013 dataset. The state-of-art VGG variants VGG16, VGG19 obtained 0.60 0.53 accuracy on Fer2013 dataset. The Xception model achieved only 0.52 accuracy on Fer2013 dataset, which makes it less efficient in Facial expression recognition, when compared with ResNet-50 and Densenet-121 on Fer2013 dataset. A mobile application efficient model MobileNet achieved 0.57 accuracy on Fer2013 dataset. Among all the models implemented, Xception model is not performing better on Fer2013 for facial expression recognition. The results show that Xception model train accuracy is 0.53 and train loss is 1.29, indicating that it under-performance on fer2013 for facial expression recognition. The latest EfficientNet-B7 obtained an 0.60 accuracy on Fer2013 dataset. In Fig. 5(a), 5(b) and 5(d) we can see that train loss and test loss converges quickly, as the number of epochs increases the train loss and test loss decreases quickly. In Fig. 5(c), we can see that the Xception model is taking much more number of epochs to decrease the train loss and test loss, when compared to proposed FerExpNet model. In Fig. 5(e), we can see that the DenseNet121 model is taking much more number of epochs to decrease the train loss and test loss, when compared to proposed FerExpNet model The Fig. 5(f) shows the accuracy vs loss comparison graph of proposed FerExpNet model. After analyzing all the results, the proposed FerExpNet model is performing better than the existing state-of-art models on Fer2013 dataset for facial expression recognition in terms of accuracy and loss.

Table IV shows the results of FerExpNet on Fer2013, when different filtering techniques are applied. We designed a novel Hybrid filtering method (HDM), which is a combination of Gaussian, bilateral and non-local means filtering techniques. When the proposed FerExpNet is combined with average filtering technique, the model achieved 0.70 accuracy on Fer2013 dataset. The proposed model without any filtering technique achieved 0.65 accuracy on Fer2013 dataset, there is a significant increase in accuracy after applying the average filter. When the proposed FerExpNet is combined with Gaussian filtering technique, the model achieved 0.65 0.55 on train set and test set of Fer2013, respectively. The accuracy of this approach is only 0.55 which is less, when compared to FerExpNet without filtering techniques, because when Gaussian filter is applied a lot of details in the images will be lost. When the designed hybrid filtering method (HDM) is combined with FerExpNet, the model achieved 0.87 0.85 accuracy on train and test sets of Fer2013 respectively. There is a significant increase in both accuracy and loss, when compared to FerExpNet without filtering techniques. The results in Table IV show that the FerExpNet with hybrid filtering method is performing better than other filtering techniques on Fer2013 dataset. The Fig. 6(a), 6(b) and 6(c) represents the comparison of accuracy and loss of various denoising techniques on proposed model when applied on FER2013 dataset, respectively.

TABLE II. OUTLINE OF FER2013 DATASET

Dataset	Name & No. of images in each emotion						
	Happy	Sad	Angry	Disgust	Sad	Surprise	Neutral
Fer2013	8989	6077	4953	547	6077	4002	6198

TABLE III. OUTLINE OF ACCURACY AND LOSS OF VARIOUS MODELS ON FER2013 DATASET

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train Loss	Test Loss
1	VGG16	Fer2013	0.63	0.60	1.01	1.10
2	VGG19	Fer2013	0.54	0.53	1.22	1.20
3	Resnet-50	Fer2013	0.63	0.60	0.97	1.09
4	MobileNet	Fer2013	0.59	0.57	1.10	1.15
5	Xception	Fer2013	0.53	0.52	1.29	1.39
6	EfficientNetB7	Fer2013	0.63	0.60	1.10	1.09
7	DenseNet121	Fer2013	0.61	0.60	1.06	1.08
8	FerExpNet	Fer2013	0.79	0.65	0.69	1.07

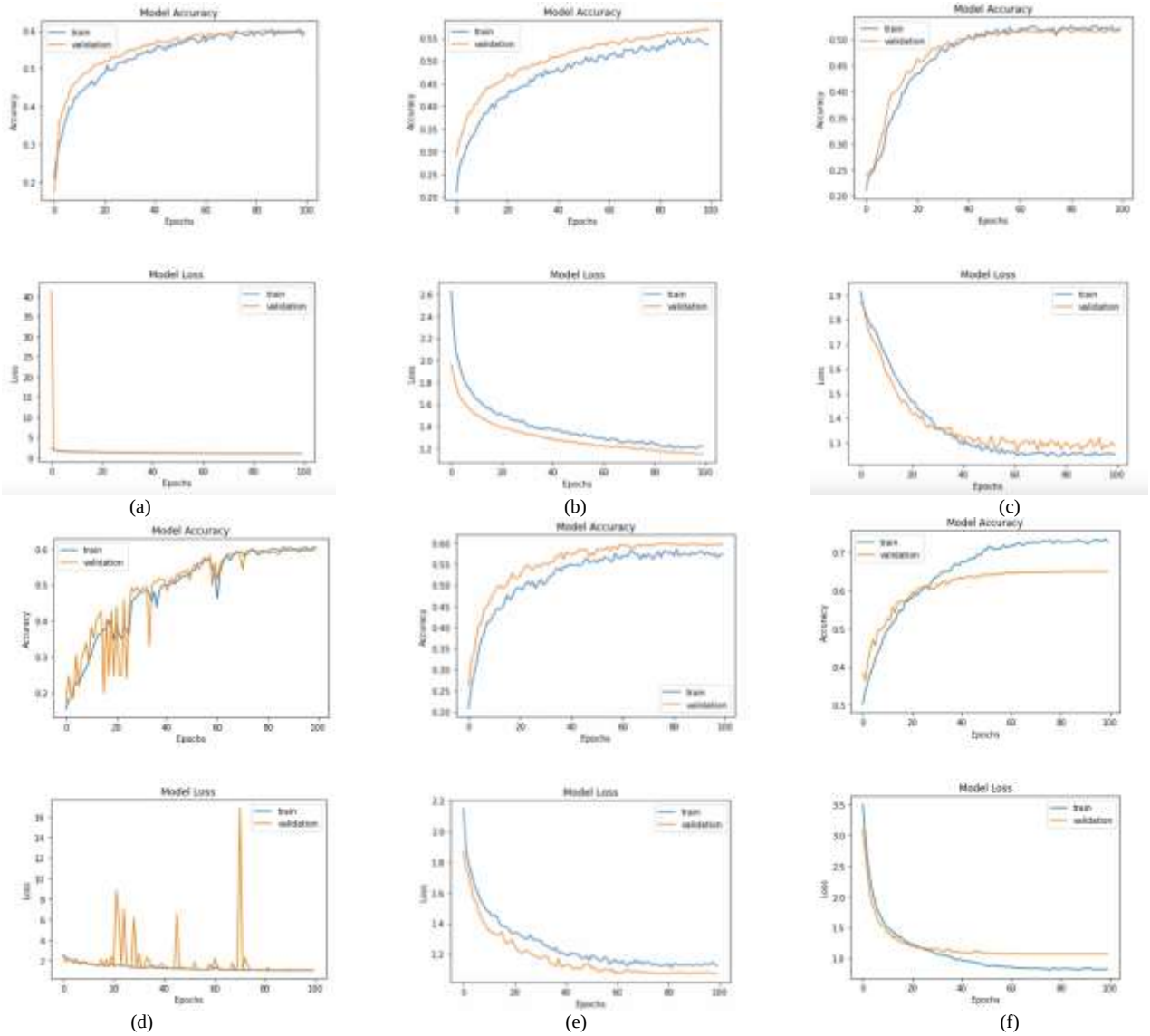
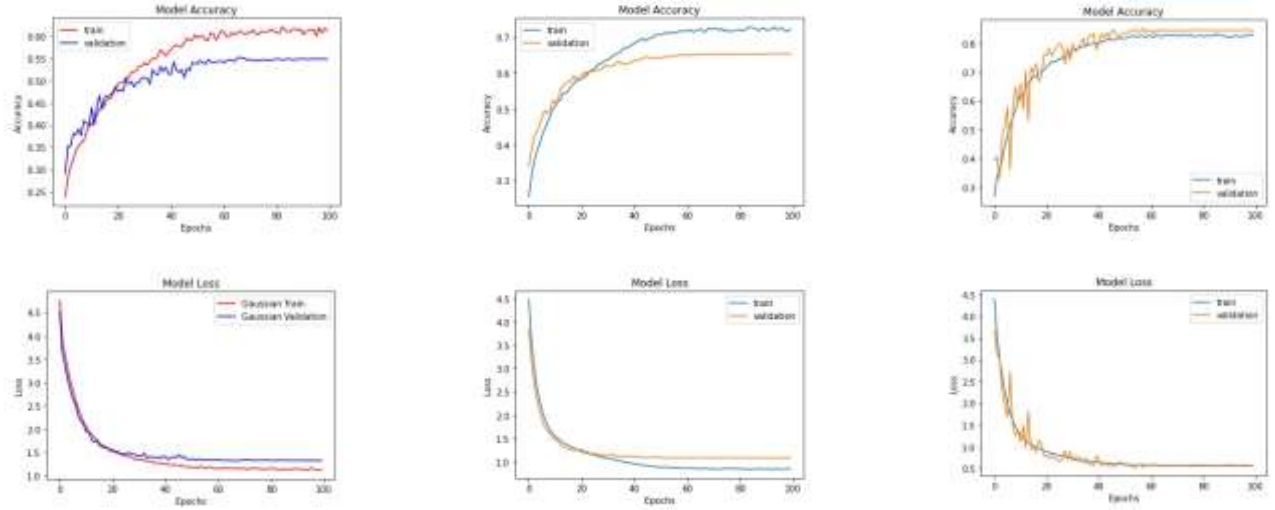


Fig. 5. (a) Accuracy and Loss of Resnet-50 on Fer2013 (b) Accuracy and Loss of MobileNet on Fer2013 (c) Accuracy and Loss of Xception on Fer2013. (d) Accuracy and Loss of EfficientNetB7 on Fer2013 (e) Accuracy and Loss of DenseNet121 on Fer2013 (f) Accuracy and Loss of FerExpNet on Fer2013.

TABLE IV. OUTLINE OF ACCURACY AND LOSS OF PROPOSED FEREXPNET ON FER2013 DATASET AFTER APPLYING VARIOUS FILTERING TECHNIQUES

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train Loss	Test Loss
1	FerExpNet_Average	Fer2013	0.91	0.70	1.45	2.39
2	FerExpNet_Median	Fer2013	0.76	0.60	0.79	1.19
3	FerExpNet_Bilateral	Fer2013	0.80	0.65	0.69	1.09
4	FerExpNet_Gaussian	Fer2013	0.65	0.55	1.05	1.33
5	FerExpNet_NonLocal Means	Fer2013	0.79	0.65	0.71	1.09
6	FerExpNet_HDM	Fer2013	0.87	0.85	0.47	0.56



(a) (b) (c)

Fig. 6. (a) Accuracy and Loss of FerExpNet_Gaussian on Fer2013 (b) Accuracy and Loss of FerExpNet_NonLocal Means on Fer2013 (c) Accuracy and Loss of FerExpNet_HDM Means.

B. Performance on LRFE Dataset

The LRFE dataset is divided in the ratio of 80:20 for training and testing purpose, the batch size is 32 and all the models are trained for 100 epochs. During model implementation, 80 percent of LRFE dataset is used for training the model and remaining 20 percent is divided into validation and testing the model.

Table V shows the accuracy and loss comparison of different deep learning models and proposed FerExpNet model on Low resolution Facial Expression (LRFE) dataset. The proposed FerExpNet achieves an accuracy of 0.95 0.71 on training and testing sets of LREF dataset, respectively. The results clearly indicate that the proposed FerExpNet is performing better than the state-of-models on LRFE dataset. The state-of-art VGG variants VGG16, VGG19 obtained 0.69 0.66 accuracy on LRFE dataset. The MobileNet model achieved only 0.65 accuracy on LRFE dataset, which makes it less efficient in Facial expression recognition, when compared with Xception and FerExpNet on LRFE dataset. The Xception model achieved 0.69 accuracy on LRFE dataset, which is second best after the FerExpNet on LRFE dataset. The latest EfficientNet-B7 obtained an 0.65 accuracy on LRFE dataset. In Fig. 7(a), 7(b) and 7(d) we can see that train loss and test loss converges quickly, as the number of epochs increases the train loss and test loss decreases quickly. In Fig. 7(c) and 7(e) we can see that the Xception and DenseNet121 models are taking much more number of epochs to decrease the train loss and test loss, when compared to proposed FerExpNet model. The Fig. 7(f) shows the accuracy vs loss comparison graph of proposed FerExpNet model. After analyzing all the results, the proposed FerExpNet model is performing better than the existing state-of-art models on LRFE dataset for facial expression recognition in terms of accuracy and loss.

TABLE V. OUTLINE OF ACCURACY AND LOSS OF VARIOUS MODELS ON LRFE DATASET

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train Loss	Test Loss
1	VGG16	LRFE	0.87	0.69	0.40	1.16
2	VGG19	LRFE	0.84	0.66	0.47	0.96
3	Resnet-50	LRFE	0.89	0.69	0.28	0.95
4	MobileNet	LRFE	0.85	0.65	0.40	0.98
5	Xception	LRFE	0.95	0.69	0.27	1.99
6	EfficientNetB7	LRFE	0.79	0.65	0.71	1.09
7	DenseNet121	LRFE	0.89	0.68	0.30	0.98
8	FerExpNet	LRFE	0.98	0.74	0.16	1.19

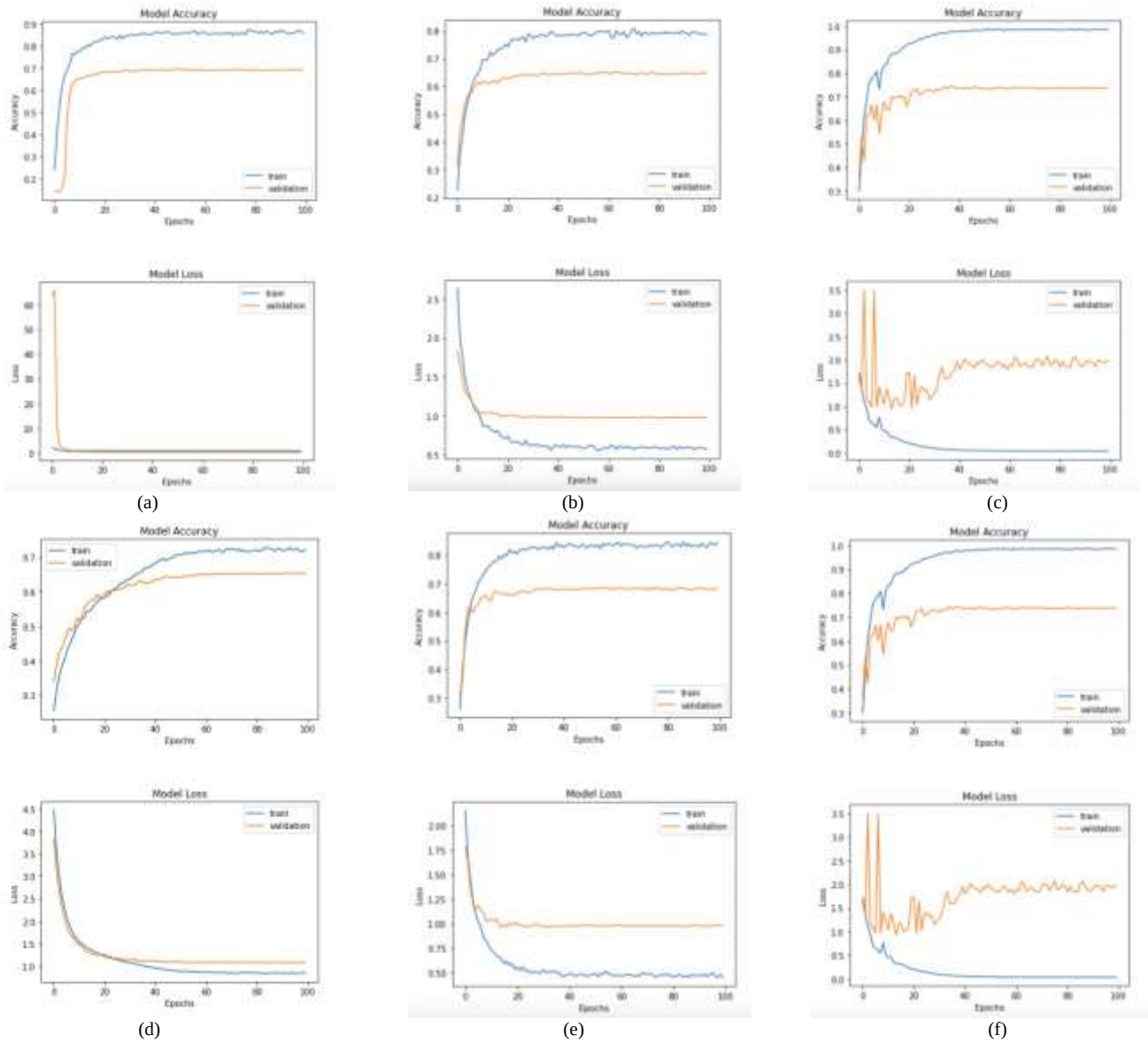


Fig. 7. (a) Accuracy and Loss of Resnet-50 on LRFE dataset (b) Accuracy and Loss of MobileNet on LRFE dataset (c) Accuracy and Loss of Xception on LRFE dataset. (d) Accuracy and Loss of EfficientNetB7 on LRFE dataset (e) Accuracy and Loss of DenseNet121 on LRFE dataset (f) Accuracy and Loss of FerExpNet on LRFE dataset.

Table VI shows the results of FerExpNet on LRFE dataset, when different filtering techniques are applied. We designed a novel Hybrid filtering method (HDM), which is a combination of gaussian, bilateral and non-local means filtering techniques. When the proposed FerExpNet is combined with average filtering technique, the model achieved 0.70 accuracy on LRFE dataset. The proposed model without any filtering technique achieved 0.65 accuracy on LRFE dataset, there is a significant increase in accuracy after applying the average filter. When the proposed FerExpNet is combined with Gaussian filtering technique, the model achieved 0.98 0.58 on train set and test set of LRFE dataset respectively. The accuracy of this approach is only 0.58 which is less, when compared to FerExpNet without filtering techniques, because when gaussian filter is applied a lot of details in the images will be lost. When the designed hybrid filtering method (HDM) is combined with FerExpNet, the model achieved 0.98 0.95 accuracy on train and test sets of Fer2013, respectively. There is a significant increase in both accuracy and loss, when compared to FerExpNet without filtering techniques. The results in Table VI show that the FerExpNet with hybrid filtering method is performing better than other filtering techniques on LRFE dataset. The Fig. 8(a), 8(b) and 8(c) represents the accuracy and loss comparison of various denoising techniques on proposed model on LRFE dataset, respectively.

C. Performance on Mixed Dataset

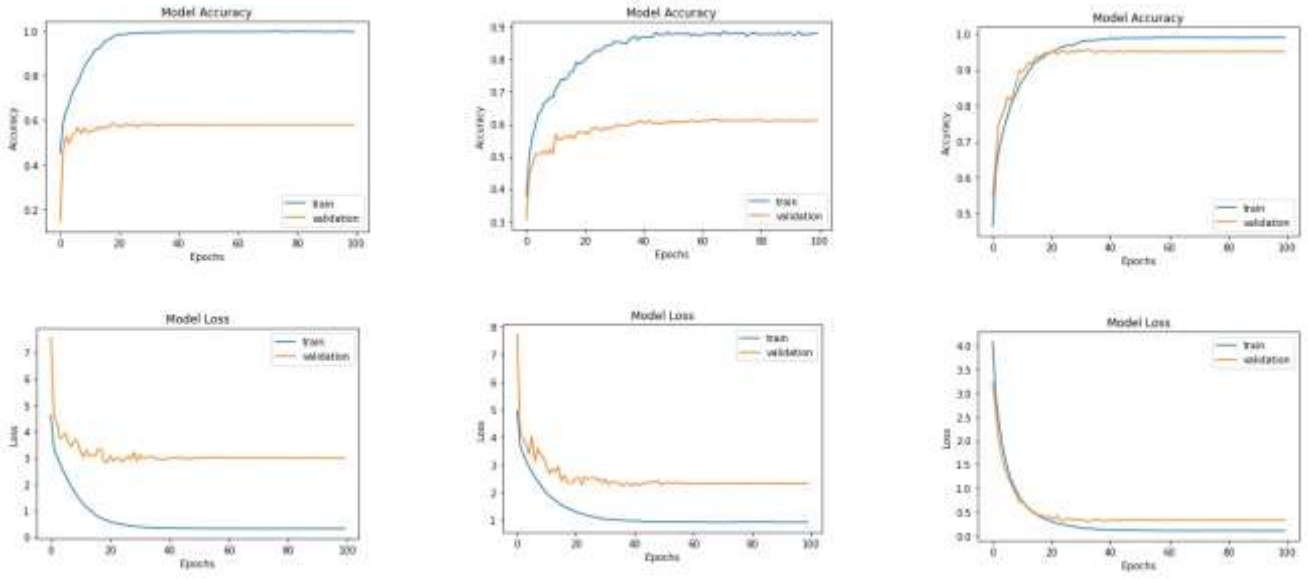
Randomly various images from each emotion category are mixed together from FER2013 and LRFE dataset to form the mixed dataset. Later, the dataset is divided in the ratio of 80:20 for training and testing purpose, the batch size is 32 and

all the models are trained for 100 epochs. During implementation, 80 percent of dataset is used for training and 20 percent of dataset is used for validation and testing purpose.

Table VII shows the comparison of various deep learning models and FerExpNet on Mixed dataset. This mixed dataset is created by mixing different emotions from Fer2013 and LRFE dataset into one directory. The dataset is then divided into training and testing sets in the ratio 80:20. The proposed FerExpNet achieved 0.96 accuracy on Mixed dataset. The state-of-art models like MobileNet, DenseNet and Xception achieved 0.91 0.95 0.92 accuracy on mixed dataset respectively. The results clearly show that the proposed FerExpNet is performing slightly better than the traditional methods on mixed dataset for Facial expression recognition.

TABLE VI. OUTLINE OF ACCURACY AND LOSS OF PROPOSED FEREXPNET ON LRFE DATASET AFTER APPLYING VARIOUS FILTERING TECHNIQUES

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train Loss	Test Loss
1	FerExpNet_Average	LRFE	0.91	0.70	1.45	2.39
2	FerExpNet_Median	LRFE	0.99	0.73	0.23	1.59
3	FerExpNet_Bilateral	LRFE	0.98	0.63	0.43	2.52
4	FerExpNet_Gaussian	LRFE	0.98	0.58	0.30	3.00
5	FerExpNet_NonLocal Means	LRFE	0.93	0.61	0.79	2.32
6	FerExpNet_HDM	LRFE	0.98	0.95	0.07	0.33



(a) (b) (c)

Fig. 8. (a) Accuracy and Loss of FerExpNet_Gaussian on LRFE (b) Accuracy and Loss of FerExpNet_NonLocal Means on LRFE (c) Accuracy and Loss of FerExpNet_HDM on LRFE.

TABLE VII. OUTLINE OF ACCURACY AND LOSS OF VARIOUS MODELS ON MIXED DATASET

S.no	Model Name	Dataset	Train Accuracy	Test Accuracy	Train Loss	Test Loss
1	FerExpNet	Mixed	0.99	0.96	0.05	0.38
2	MobileNet	Mixed	0.95	0.91	0.16	0.28
3	DenseNet	Mixed	0.99	0.95	0.05	0.19
4	Xception	Mixed	0.95	0.92	0.23	0.38

V. CONCLUSION

A Novel optimized neural network on basis of convolutional neural network is proposed in this paper. We also designed a new hybrid filtering method, which is a combination of Gaussian, bilateral and non-local means filtering techniques. This hybrid filtering method is used for removing any noise present in the images. All the datasets are divided in the ratio of 80:20 for training and testing purpose. The proposed FerExpNet achieved 0.65 accuracy on Fer2013 dataset and this model is performing better than the state-of-art models on Fer2013. When the hybrid filtering method is combined with the proposed FerExpNet, the model achieves 0.85 accuracy on Fer2013 dataset. There is a significant increase in the accuracy when hybrid filtering method is applied. The proposed FerExpNet obtained 0.74 accuracy on LRFE dataset, outperforming the existing models, similarly when the hybrid filtering method is combined with FerExpNet the accuracy increased to 0.95 on LRFE dataset. The results show that the proposed FerExpNet is performing better than

the existing models. The average time taken for each epoch on LRFE dataset is 1sec, similarly the average time taken for each epoch on Fer2013 is 5sec for FerExpNet respectively. The future work of this paper is to build a more sophisticated convolutional neural network model which can be integrated into a mobile device for wide use in real-world applications.

ACKNOWLEDGMENT

I sincerely thank P.V.V.S Srinivas for guiding me in this research work and helping through difficult times.

REFERENCES

[1] Kopparapu, Sunil Kumar, and M. Satish. “Identifying Optimal Gaussian Filter for Gaussian Noise Removal.” 2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (December 2011). doi:10.1109/ncvpripg.2011.34.

[2] Tomasi, C., & Manduchi, R. (n.d.). Bilateral filtering for gray and color images. Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271). doi:10.1109/iccv.1998.710815.

[3] Deledalle, C.-A., Duval, V., & Salmon, J. (2011). Non-local Methods with Shape-Adaptive Patches (NLM-SAP). Journal of Mathematical Imaging and Vision, 43(2), 103–120. doi:10.1007/s10851-011-0294-y.

[4] Ashok Kumar, P. M., Jeevan Babu Maddala, and K. Martin Sagayam. "Enhanced Facial Emotion Recognition by Optimal Descriptor Selection with Neural Network." *IETE Journal of Research* (2021): 1-20.

[5] Bui, H. M., Lech, M., Cheng, E., Neville, K., & Burnett, I. S. (2016). Using grayscale images for object recognition with convolutional-recursive neural network. 2016 IEEE Sixth International Conference on Communications and Electronics (ICCE). doi:10.1109/cce.2016.7562656.

[6] Videla, Lakshmi Sarvani, and PM Ashok Kumar. "Facial Expression Classification Using Vanilla Convolution Neural Network." In *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, pp. 1-5. IEEE, 2020.

[7] Salman, Shaeke, and Xiuwen Liu. "Overfitting mechanism and avoidance in deep neural networks." *arXiv preprint arXiv:1901.06566* (2019).

[8] Tran, Toan, Trung Pham, Gustavo Carneiro, Lyle Palmer, and Ian Reid. "A bayesian data augmentation approach for learning deep models." In *Advances in neural information processing systems*, pp. 2797-2806. 2017.

[9] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167* (2015).

[10] Sainath, Tara N., Brian Kingsbury, Hagen Soltau, and Bhuvana Ramabhadran. "Optimization techniques to improve training speed of deep neural networks for large speech tasks." *IEEE Transactions on Audio, Speech, and Language Processing* 21, no. 11 (2013): 2267-2276.

[11] Smith, Samuel L., Pieter-Jan Kindermans, Chris Ying, and Quoc V. Le. "Don't decay the learning rate, increase the batch size." arXiv preprint arXiv:1711.00489 (2017).

[12] Inthiyaz, Syed, M. Muzammil Parvez, M. Siva Kumar, J. Sri sai Srija, M. Tarun Sai, and V. Amruth Vardhan. "Facial Expression Recognition Using KERAS." In *Journal of Physics: Conference Series*, vol. 1804, no. 1, p. 012202. IOP Publishing, 2021.

[13] Mahesh, D. Sri Sai, T. Maneesh Reddy, A. Sai Yaswanth, C. Joshitha, and S. Sudarshan Reddy. "Facial detection and recognition system on Raspberry Pi with enhanced security." In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pp. 1-5. IEEE, 2020.

[14] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I-I. IEEE, 2001.

[15] Srinivas, P. V. V. S., and Pragnyaban Mishra. "Facial Expression Detection Model of Seven Expression Types Using Hybrid Feature Selection and Deep CNN." In *International Conference on Intelligent and Smart Computing in Data Analytics: ISCDA 2020*, pp. 89-101. Springer Singapore, 2021.

[16] Suk, Myunghoon, and Balakrishnan Prabhakaran. "Real-time mobile facial expression recognition system-a case study." In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 132-137. 2014.

[17] Abinaya, R., Lakshmana Phaneendra Maguluri, S. Narayana, and Maganti Syamala. "A Novel Biometric Approach for Facial Image Recognition Using Deep Learning Techniques." *International Journal of Advanced Research in Engineering and Technology* 11, no. 9 (2020).

[18] Ruiz-Garcia, Ariel, Mark Elshaw, Abdulrahman Altahhan, and Vasile Palade. "Emotion recognition using facial expression images for a robotic companion." In *International Conference on Engineering Applications of Neural Networks*, pp. 79-93. Springer, Cham, 2016.

[19] Jahnavi, P., Enireddy Vamsidhar, and C. Karthikeyan. "Facial expression detection of all emotions and face recognition system." *International Journal of Emerging Trends in Engineering Research* 7, no. 12 (2019): 778-783.

[20] Lopes, Nuno, André Silva, Salik Ram Khanal, Arsênio Reis, João Barroso, Vitor Filipe, and Jaime Sampaio. "Facial emotion recognition in the elderly using a SVM classifier." In *2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*, pp. 1-5. IEEE, 2018.

[21] Yu, Zhiding, and Cha Zhang. “Image Based Static Facial Expression Recognition with Multiple Deep Network Learning.” Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI ’15 (2015). doi:10.1145/2818346.2830595.

[22] Zhang, Zhihao, Tong Chen, Hongying Meng, Guangyuan Liu, and Xiaolan Fu. “SMEConvNet: A Convolutional Neural Network for Spotting Spontaneous Facial Micro-Expression From Long Videos.” IEEE Access 6 (2018): 71143–71151. doi:10.1109/access.2018.2879485

[23] Ebrahimi Kahou, S., Michalski, V., Konda, K., Memisevic, R., & Pal, C. (2015). Recurrent Neural Networks for Emotion Recognition in Video. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI ’15. doi:10.1145/2818346.2830596

[24] Wu, Bing-Fei, and Chun-Hsien Lin. “Adaptive Feature Mapping for Customizing Deep Learning Based Facial Expression Recognition Model.” IEEE Access 6 (2018): 12451–12461. doi:10.1109/access.2018.2805861.

[25] Hossain, M. Shamim, and Ghulam Muhammad. “An Emotion Recognition System for Mobile Applications.” IEEE Access 5 (2017): 2281–2287. doi:10.1109/access.2017.2672829.

[26] Deng, Jia, Gaoyang Pang, Zhiyu Zhang, Zhibo Pang, Huayong Yang, and Geng Yang. “cGAN Based Facial Expression Recognition for Human-Robot Interaction.” IEEE Access 7 (2019): 9848–9859. doi:10.1109/access.2019.2891668.

[27] Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab. “A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing.” IEEE Access 7 (2019): 159081–159089. doi:10.1109/access.2019.2949741.

[28] Wang, Yingying, Yibin Li, Yong Song, and Xuewen Rong. “The Application of a Hybrid Transfer Algorithm Based on a Convolutional Neural Network Model and an Improved Convolution Restricted Boltzmann Machine Model in Facial Expression Recognition.” IEEE Access 7 (2019): 184599–184610. doi:10.1109/access.2019.2961161.

[29] Kosti, Ronak, Jose M. Alvarez, Adria Recasens, and Agata Lapedriza. “Emotion Recognition in Context.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017). doi:10.1109/cvpr.2017.212.