Human Emotion Detection Based On Facial Expression Using Convolution Neural Network

Penke Satyanarayana, Pathan Madhar Khan, Shaik Junez Riyaz

Abstract*: *Deep learning is an achievement inside the field of computer vision. This paper deals with deep learning frameworks to see outward appearances that address human feelings. Face feelings are the impressions of the inner emotions of a human. The human expressions play an essential role in nonverbal communication. Our article deals with eight standard feelings happiness, angry, sadness, fearing, surprising, disgusting, contempt and neutral. various researches have been performed, in appearing sagacious computer vision which can see the human's tendency. The proposed work achieves improved performance model with fewer epochs. To implement this, efficient algorithms and techniques are used while generating the model. In the preprocessing methodology, Histogram equalization has been applied to the raw input images. Batch Normalization technique is used in the proposed model for better learning rate. CK+ dataset is used for training and testing the model. To test the model in real time harr feature-based cascade classifier is used for detecting the face. the model was trained on Google Colab with a GPU.

Index Terms*: *Batch Normalization, Convolution Neural Network, Emotion Detection, Histogram equalization.

I. INTRODUCTION

The human facial emotion detection is a trending topic in the present world. Moreover, it is one of the outstanding and dynamic research focuses in the field of image processing. different strategies, structures, and estimations have been made to recognize and see facial emotion detection [9]. To recognize any facial emotion using convolution neural network it must undergo three steps pre-processing, training and testing. For any kind of application developed using deep learning, data is important. Data plays a major role in the efficient working of the application. The dataset used in this work is CK+. Which has 8 emotion labels, they are Neutral, Angry, Disgust, Fear, Happy, Sad, contempt and Surprise [2]. The images in the dataset are a series of photographs changing emotion state from neutral to target emotion. According to labels given in the dataset, photographs are separated with respect to the emotion. A model was created through training. In the proposed paper a CNN [3] in the form of a sequential model was developed. A model is a sequence of layers arranged to form the network. The network involves 10 layers two 2D convolution layers with ReLU as an activation function followed by max pooling layer and two 2D convolution layers again max pool followed by Flatten, Dense and Drop out finally a Dense layer with 8 output nodes with an activation function SoftMax is used. These 8 output nodes represent 8 different emotions of humans. Our technique draws its quality from making normalization a piece of the model designing and playing out the standardization for each mini-batch. Batch Normalization enables us to use significantly higher learning rates and be less cautious about initialization. Max pooling is used to reduce the pixel values in the image by selecting a maximum pixel value in the sliding window. Training the model is very hard on normal desktops. For faster and efficient training, it requires GPUs or TPUs. So, the model was trained on Google Colab with a GPU.

II. RELATED WORKS

Shin et.al [5] in their work tested different combinations of dataset with different structures of standard CNN classifier with different preprocessing algorithms. They tried on different data inputs like raw images, isotropic smoothing, applying histogram equalization, diffusion-based normalization, difference of Gaussian and analysis were made. they use 4 different networks and 5 data inputs form 20 different combinations to know the best performer. Cropping and flipping were applied on the images. Tests were executed on five check units SFEW2.0, FER-2013, CK+, jaffe, KDEF and the network structure of four distinct community structures kahou, tang, Yu, and ImageNet Hist-eq method confirmed the very notable popular overall performance for all four exquisite community candidates. Histogram equalization is evaluation enhancement approach that normally will grow the international evaluation of pix. This method is good even as the historical past's and foreground's brightness are identical. they use deep analyzing NVIDIA with GPU. Rajesh Kumar et.al [7] proposed a work based on CNN for discriminating between a genuine and fake smile. The proposed system is trained on FER 2013 dataset which has eight standard emotions. The training data set contains a total of 28,709 samples in which 3,589 are public test cases. Proposed CNN comprises of 8 layers which contains 3 pairs of convolutional and max pooling layers and 2 SoftMax layers. ReLU is used as activation function in pooling layers. At present there are no specific data sets for this purpose but upon testing the generated model result obtained are comparatively good.

Revised Manuscript Received on March 20, 2019.

Penke Satyanarayana, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India.

Pathan Madhar Khan, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India.

Shaik Junez Riyaz, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

This can further be used in human understandability and analyzing customer satisfaction.

III. PROPOSED METHOD

In the proposed work a deep learning model has been developed where data plays a major role for efficient working of application. The dataset used in this work is CK+. Which has 8 emotion labels, they are Neutral, Angry, Disgust, Fear, Happy, Sad, contempt and Surprise. CK+ contains a total of 123 participants who are between 18 and 50 years old out of which 69% are female, 81% are Euro-American, 13% are Afro-American, and 6% are from other groups [2]. The CK+ dataset consists of a series of photographs changing emotion state from neutral to target emotion. According to labels given in the dataset, photographs are separated with respect to the emotion. From the data set 80% of images are treated as training data set and the remaining for testing. The images separated are raw images which also has unwanted area. Removal of unwanted area and optimizing the image was performed in preprocessing.

A. *Pre-processing*

In the Preprocessing of the model, 4 major steps [6] were involved. They are, face detection, adjusting Max area of face [1], cropping image and Histogram Equalization. the face is detected using open cv library consists of Haar Cascade Classifiers. The boundary box is adjusted to get maximum optimal face so that every feature in the face is considered and cropping is performed on the image to avoid unnecessary information in the image. Researchers has used many techniques to improve performance of the model in preprocessing the dataset one of the efficient techniques is applying histogram equalization on cropped image as in Fig.1.

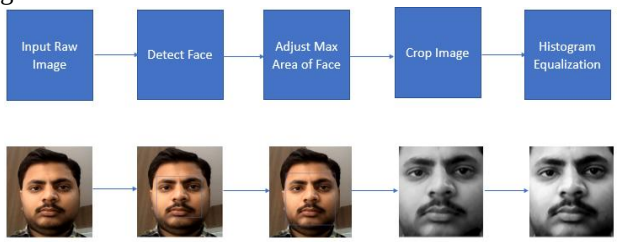


Fig. 1. Stages of Image Preprocessing

In the preprocessing technique instead of instead of applying histogram equilization on the croped image sharpning and unsharpning is applied and generated a new data set as in Fig. 2.

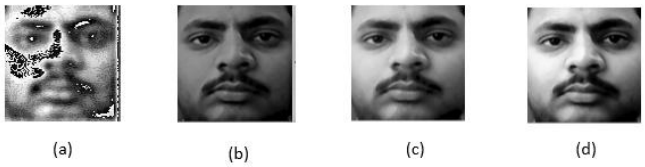


Fig. 2. Image processing technique (a)Sharped Image (b) Unshaped Image (c)Raw Image (d) Histogram equalized Image

Histogram Equalization is the process taking a low contrast picture and developing the contrast nature between the picture's relative highs and lows to bring out straight forward separations in shade and have a higher impact picture. The outcomes are striking, particularly for grayscale pictures.

B. *Training the model*

In this article a Convolutional Neural Network in the form of sequential model has been developed which consists of sequence of layers [12] arranged to form a network. The proposed model aims to reduce the complexity of the structure and training period while improving the efficiency of the model. As a base VGG16 neural structure is considered and several layers was bringing down and batch normalization. Four different models was generated using Raw Images, Sharped Images, Unshaped Images and with Histogram equalized images. Training the model is very hard on normal desktops. For faster and efficient training, it requires GPUs or TPUs. So, the model was trained on Google Colab with a GPU (graphical processing unit).

IV. CNN MODEL LAYERS

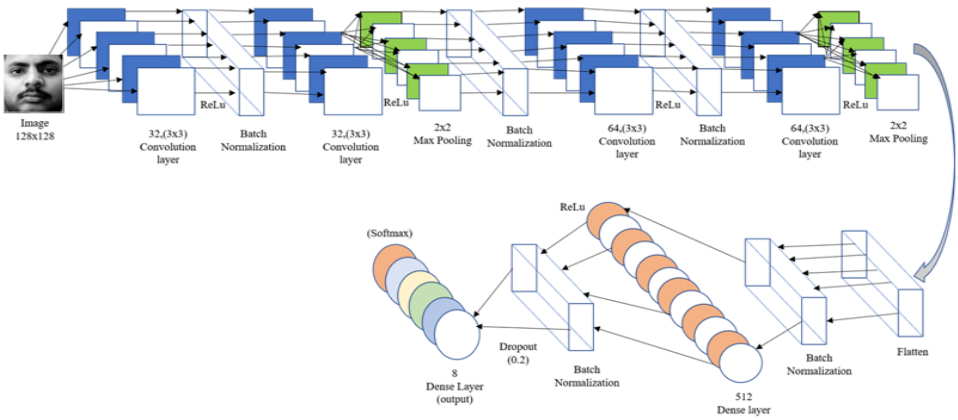


Fig. 2. Proposed network Architecture

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The proposed network involves several layers initiated by feeding an image of size 128X128 and a kernel size of 5X5 to a 2D convolutional layer with 32 filters of size 3x3 with ReLU as an activation function. This layer generates an output shape of 126x126x32 is passed through Batch Normalization. The results of this steps are convolved with second convolutional layer with 32 filters of with ReLU as an activation function. The second layer generates an output shape of 124x124x32 was send to Max Pooling layer having kernel of 2x2 size. This layer generates an output shape of 62x62x32 is passed through Batch Normalization. The results of this steps are convolved with third convolutional layer with 64 filters of size 3x3 with ReLU as an activation function. This layer generates an output shape of 60x60x64 is passed through Batch Normalization. The results of this steps are convolved with fourth convolutional layer with 64 filters of size 3x3 with ReLU as an activation function. This layer generates an output shape of 58x58x64 was send to Max Pooling layer having kernel of 2x2 size. This layer generates an output shape of 29x29x64. Which is passed through Flattener as a result the image is converted to single dimension is passed through Batch Normalization followed by Dense layer with 512 nodes and ReLU as an activation function is again passed through Batch Normalization, to avoid overfitting a Dropout of 0.2 is considered. the last layer has 8 nodes consider as output layer with SoftMax as an activation function represents 8 basic emotions as exhibit in Fig.2.

Deep Neural Network [9] is a combination of different layers. The inputs which are provided to the input layer are passed to the output layer through hidden layers adjusting the weights and bias. the sequential process of execution will decrease the speed of learning rate and it is very difficult to train the model and this can be repaid by normalization. our methodology attracts strength by applying normalization for every mini-batch. The Batch Normalization is used to increase in learning rate. and the Max pooling is used to reduce the size of the image by selecting a maximum of the sliding window as a result parameter will be reduced for the next layers. Max pooling is used to reduce the complexity in handling large number of parameters which has become a common part in any convolution neural network. the working of max pooling is exhibit in Fig.3

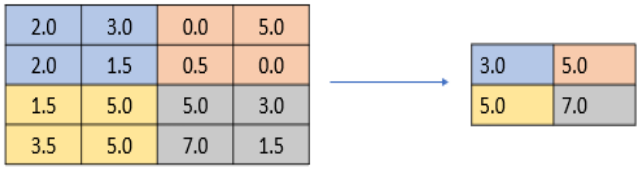


Fig. 3. Working of max pooling

V.EXPERIMENTS RESULTS

As an Experiment results, The Accuracy of the model increases with increase in epochs the same is revealed in Fig.4. The proposed model was trained in google colab and tested using jupyter notebook. The proposed algorithm shows an average accuracy of 99.54%, validation accuracy of 98.22% and the test accuracy rate of generated model when it is tested for facial expression to identify human emotion as 97.77%.

```
Epoch 1/10
500/500 [==============================] - 110s 219ms/step - loss: 0.2408 - acc: 0.9202 - val_loss: 0.1044 - val_acc: 0.9755

Epoch 00001: loss improved from inf to 0.23640, saving model to modelck2m_acc1.h5
Epoch 2/10
500/500 [==============================] - 105s 211ms/step - loss: 0.0317 - acc: 0.9895 - val_loss: 0.1069 - val_acc: 0.9733

Epoch 00002: loss improved from 0.23640 to 0.03200, saving model to modelck2m_acc1.h5
Epoch 3/10
500/500 [==============================] - 105s 211ms/step - loss: 0.0252 - acc: 0.9917 - val_loss: 0.2120 - val_acc: 0.9354

Epoch 00003: loss improved from 0.03200 to 0.02494, saving model to modelck2m_acc1.h5
Epoch 4/10
500/500 [==============================] - 105s 210ms/step - loss: 0.0200 - acc: 0.9933 - val_loss: 0.0963 - val_acc: 0.9710

Epoch 00004: loss improved from 0.02494 to 0.02072, saving model to modelck2m_acc1.h5
Epoch 5/10
500/500 [==============================] - 105s 210ms/step - loss: 0.0139 - acc: 0.9942 - val_loss: 0.0947 - val_acc: 0.9800

Epoch 00005: loss improved from 0.02072 to 0.01421, saving model to modelck2m_acc1.h5
Epoch 6/10
500/500 [==============================] - 105s 211ms/step - loss: 0.0190 - acc: 0.9932 - val_loss: 0.0942 - val_acc: 0.9710

Epoch 00006: loss did not improve from 0.01421
Epoch 7/10
500/500 [==============================] - 105s 210ms/step - loss: 0.0118 - acc: 0.9950 - val_loss: 0.0924 - val_acc: 0.9777

Epoch 00007: loss improved from 0.01421 to 0.01212, saving model to modelck2m_acc1.h5
Epoch 8/10
500/500 [==============================] - 105s 210ms/step - loss: 0.0218 - acc: 0.9920 - val_loss: 0.0988 - val_acc: 0.9800

Epoch 00008: loss did not improve from 0.01212
Epoch 9/10
500/500 [==============================] - 105s 210ms/step - loss: 0.0098 - acc: 0.9952 - val_loss: 0.0957 - val_acc: 0.9733

Epoch 00009: loss improved from 0.01212 to 0.01020, saving model to modelck2m_acc1.h5
Epoch 10/10
500/500 [==============================] - 105s 210ms/step - loss: 0.0094 - acc: 0.9954 - val_loss: 0.0993 - val_acc: 0.9822

Epoch 00010: loss improved from 0.01020 to 0.00954, saving model to modelck2m_acc1.h5
```

Fig. 4. Result obtained at successive Epochs (Jupyter Notebook)

```
Test loss: 0.3589778907628336
Test accuracy: 97.7728285077951
```

Fig. 5. Result obtained after testing (Jupyter Notebook)

confusion matrix [8] is used to know the standard of the output of a classifier. The diagonal components represent the percentage of the predicted label is up to actuality label, while off-diagonal components represent the misbranded by the classifier. the higher the diagonal values of the confusion matrix [4] indicating more correct predictions. The results in the confusion matrix show peak accuracy towards original expression.

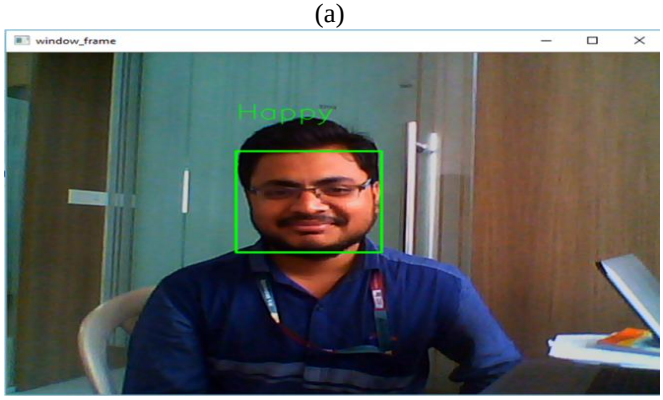
Table. 1. Confusion matrix of emotions.

Happy	92.84%	1.05%	0.05%	1.13%	0.26%	1.35%	2.30%	1.02%
Anger	0.90%	91.60%	0.83%	0.70%	1.03%	2.09%	1.05%	1.80%
Fear	1.44%	2.05%	89.47%	1.83%	1.04%	2.20%	1.07%	0.90%
Sad	0.64%	1.27%	1.97%	91.67%	2.82%	0.45%	0.16%	1.02%
Disgust	3.03%	0.88%	2.01%	1.71%	89.08%	0.96%	1.03%	1.30%
Surprise	0.41%	1.01%	2.92%	0.70%	1.06%	91.40%	1.70%	0.80%
Neutral	0.41%	0.21%	0.95%	0.16%	3.01%	1.05%	90.73%	1.48%
Contempt	0.33%	1.93%	1.80%	2.10%	1.70%	0.50%	1.96%	91.68%
	Happy	Anger	Fear	Sad	Disgust	Surprise	Neutral	Contempt

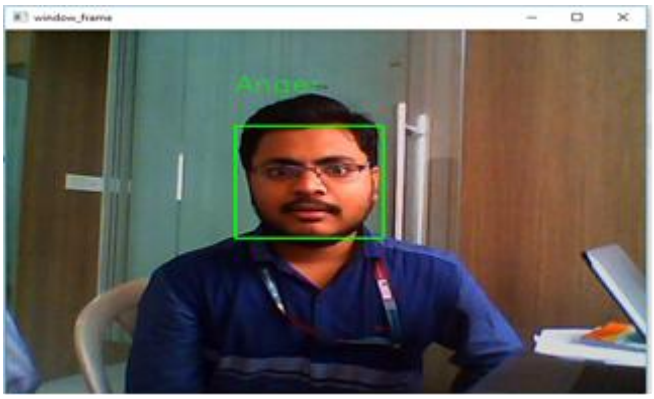
A.*Testing*

The proposed model was tested in real time using a webcam. The faces were extracted from the webcam using Haar front face classifier [8] which is available in openCV [6]. The detected area is drawn with bounding box. The extracted face was feed to the trained model to predict the emotion in the face. The predicted emotion is written as text on bounding box. Some of the samples are in Fig.6.

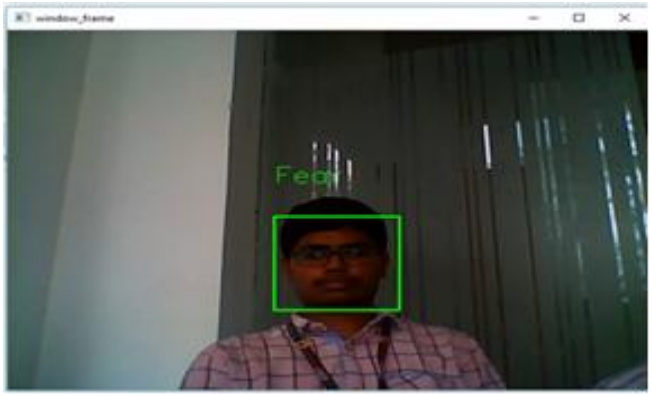
(a)



(b)

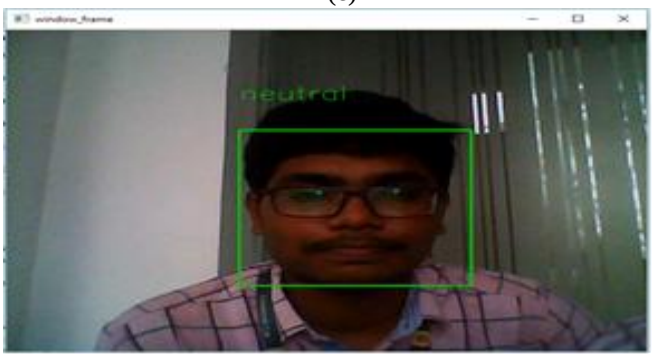


(c)



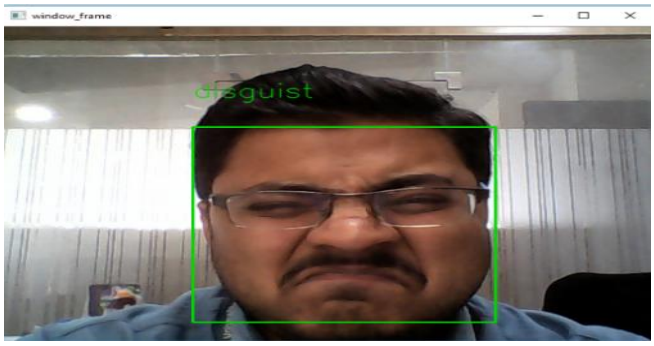
(d)

(e)



(f)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



(g)

Fig. 6. Facial expression of designed CNN based system (a)Sad (b) Happy (c)Anger (d) Fear (e) Surprise (f) Neutral (g)disgust

B. *Accuracy*

While training the proposed model the accuracy of the model increases with increase in epochs. there is a sudden increase in the accuracy which cross 90% at the second epoch and saturate about 99% in Unsharp followed by Histogram equalization and Sharp dataset.

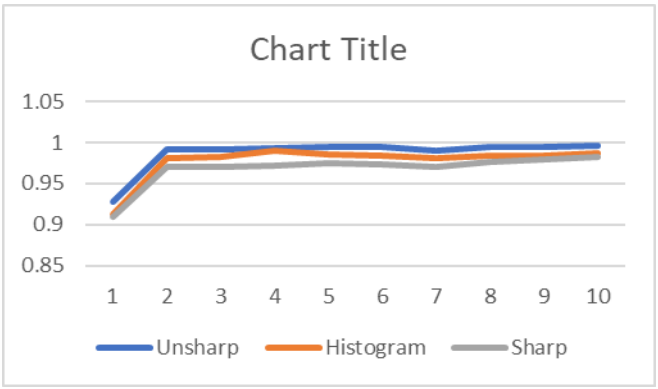


Fig. 7. Curve of Epochs vs Accuracy

C. *Loss*

While training the proposed model, the loss factor of model decreases with the increase in epochs and reaches near to zero indicating the efficiency of model, the same is shown in Fig.8. The factor loss in deep learning defines how bad the model is predicting the output. In the proposed model the loss factor is leading to zero, indicating the perfectness of the model. the loss factor is less in Unsharp followed by Histogram Equilization and Sharp dataset.

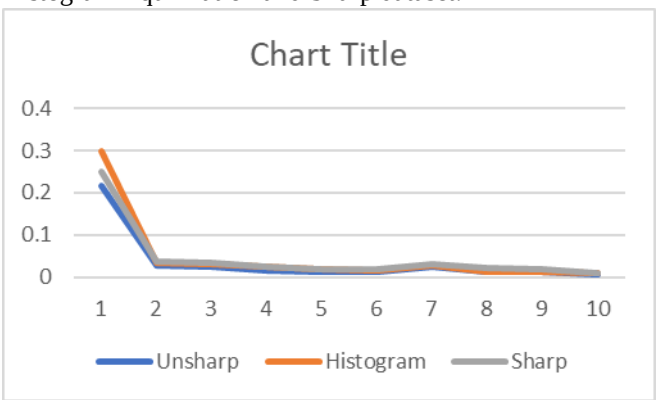


Fig. 8. Curve of Epochs vs Loss

D. *Batch normalization*

The normalization will be performed on the initial values where mean and variance are 0 and 1 respectively. The training parameters are updated as a result there will be a loss in normalization this will have an impact on training time as the network becomes deeper. Batch normalization maintain these normalizations for every mini batch and backpropagate

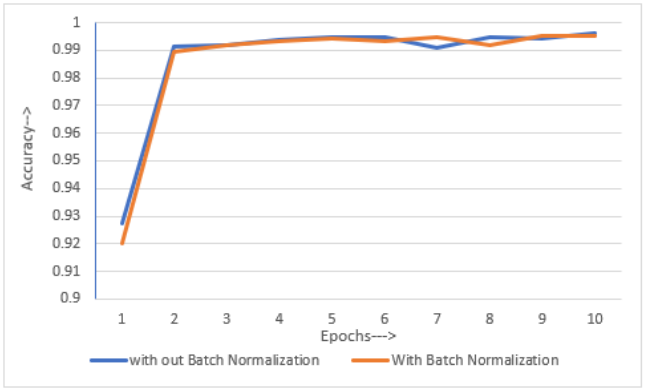


Fig. 9. Curve of Batch Normalization Epochs vs Accuracy

The proposed CNN model has been trained with and without Batch Normalization and the proposed model shows good results for Batch Normalization reducing internal covariant shift and the learning rate is more stabilized in Fig. 9 and Fig. 10.

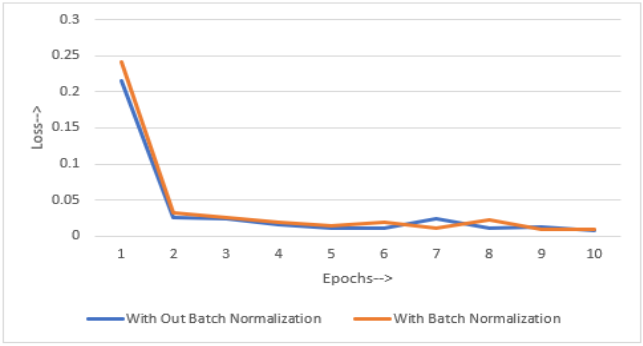
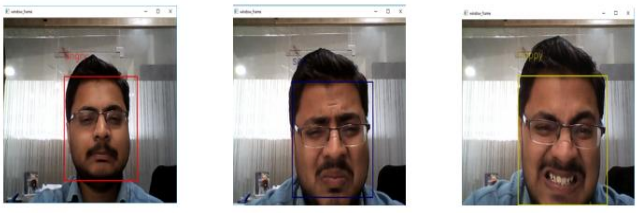


Fig. 10. Curve of Batch Normalization Epochs vs Loss

VI. FAILURE CASES

The proposed model was tested with an expression in a confused state. Where the expressions are fluctuating leads to failure test cases [1]. The proposed model was not able to predict the facial expressions which are in transforming phase from one expression to other



(a) (b) (c)

Fig. 7. Facial expression results in failure cases of our devised CNN based system (a)Sad as Anger (b) Fear as Sad (c)Anger as Happy

VII. CONCLUSION

This paper proposed an optimised convolution neural network model to detect and analyze different types of human emotion based on facial expression that includes efficient techniques and algorithms which are implemented at various levels, histogram equalization, Sharping and Unsharring the image in pre-processing and batch normalization in training the model helps to increase the accuracy. Furthermore, a webcam was used to detect human facial emotion in real time using harr cascade. The relative outcomes demonstrate that performing Unsharp images shows a better performance of the model when compared to training the model with sharped and Histogram Equalized and raw images. This paper finds the attention scores of different emotions using Cohn Kanade and Cohn Kanade + datasets.

REFERENCES

1. Rajesh Kumar G, Ravi Kant Kumar and Goutam Sanyal (2017),” Facial Emotion Analysis using Deep Convolution Neural Network”, International Conference on Signal Processing and Communication (ICSPC’17).
2. Diah Anggraeni Pitalokaa, Ajeng Wulandaria, T. Basaruddina and Dewi Yanti Lilianaa (2017), “Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition”, 2nd International Conferences on Computer Science and Computational Intelligence.
3. Xiao Liu and Kiju Lee (2018), “Optimized Facial Emotion Recognition Technique for Assessing User Experience”, 2018, IEEE Games, Entertainment, Media Conference.
4. Julio Cesar Batista and Vitor Albiero and Olga R. P. Bellon and Luciano Silva (2017),” AUMPNet: simultaneous Action Units detection and intensity estimation on multipose facial images using a single convolutional neural network”, IEEE 12th International Conference on Automatic Face & Gesture Recognition.
5. Minchul Shin, Munsang Kim and Dong-Soo Kwon (2016),” Baseline CNN structure analysis for facial expression recognition”, 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).
6. Heechul Jung, Sihaeng Lee, Sunjeong Park, Byungju Kim, Junmo Kim, Injae Lee and Chunghyum (2017),” Development of Deep Learning-based Facial Expression Recognition System”.
7. Rajesh Kumar G A, Ravi Kant Kumar and Goutam Sanyal (2017), “Discriminating Real from Fake Smile Using Convolution Neural Network”, International Conference on Computational Intelligence in Data Science (ICCIDS).
8. Justus Schwan, Esam Ghaleb, Enrique Hortal and Stylianos Asteriadis (2017),” High-Performance and Lightweight Real-Time Deep Face Emotion Recognition”.
9. Saqib Nizam Shamsi, Bhanu Pratap Singh and Manya Wadhwa (2018).” Group Affect Prediction Using Multimodal Distributions”, 2018 IEEE Winter Conference on Applications of Computer Vision Workshops.
10. Biao Yang (2015), “Facial Expression Recognition using Weighted Mixture Deep Neural Network Based on Double-channel Facial Images”, Journal of LaTeX class files, vol 14 and No. 8.
11. Paul Viola and Michael Jones (2001), "Rapid Object Detection using a Boosted Cascade of Simple Features", Conference on computer vision and pattern recognition.
12. Dolly Reney, Dr.Neeta Tripaathi,” An Efficient Method to Face and Emotion Detection”, 2015, Fifth International Conference , Communication Systems and Network Technologies

AUTHORS PROFILE

Penke Satyanarayana, received a B. Tech in Electronics and Communication Engineering (ECE) from Koneru Lakshmaiah Education Foundation, Vaddeswaram, Vijayawada, India; M. Tech in Computers and Communications Systems from JNTU, Hyderabad, India; and Ph.D. in ECE from JNTU, Kakinada, India. He is working as a Professor in the Department of ECE, K L E F, Guntur, India. He has published over 20 research papers in reputed international and national journals and conferences. He is a member in Indian Society of Technical Education (ISTE). His research interests include Wireless Communication and Signal processing, Embedded Systems and VLSI.

Pathan Madhar Khan is currently pursuing B. Tech in Electronics and Communication Engineering from, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Vijayawada, India. His research interest includes Embedded Systems and Artificial Intelligence.

Shaik Junez Riyaz, is currently pursuing B. Tech in Electronics and Communication Engineering from, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Vijayawada, India. His research interest includes Embedded Systems and Artificial Intelligence.