International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Facial Expression Recognition Using Expression Generative Adversarial Network and Attention CNN

Gyanendra Tiwary[1], Shivani Chauhan[2], Krishan Kumar Goyal[3]

Submitted:21/03/2023 **Revised:**23/05/2023 **Accepted:**11/06/2023

Abstract: Facial Expressions are quite personalized and may look different for different individuals. Whereas there are certain facial muscles which shows some common features for certain human expressions across the cultures and facial shapes. Convolution Neural Network have shown tremendous success in Facial Expression Recognition task. In recent past many researchers have proposed multiple models with manageable size solution to Facial Expression Recognition task. In the current work, we have considered shape, complexion and other identity related information separate from certain specified muscle movements which are specific for emotion recognition. This is done by a novel Emotion-Generative Adversarial Network. This saves a lot of effort and simplifies the Facial Expression Recognition process. We then apply Scale Invariant Feature Transformation and vola john's face extraction method for pre-processing and face image extraction from background. This enables us to train our model accurately irrespective of scale, orientation, illumination etc and with very less training samples accurately. We feed the feature extracted facial image to an attention-based Convolutional Neural Network. This will ensure more emphasis on critical areas for expression recognition of facial image. Finally, we have used Local Binary Pattern for classification of the input image to a particular emotion class. We have tested our model on CK+, OULU- Casia and FER-2013 datasets and it is at par with performance of all major state-of-art models. Proposed model may be utilized by various automated interactive systems, such as robot to human communication, automated customer care systems etc. The proposed work may also be quite useful for observing reaction of viewers to a particular advertisement or article automatically and use this information for various purposes like user's interest, product feedback etc.

Keywords: *Facial Emotion Recognition, Generative Adversarial Network, Convolutional Neural Network*

1. Introduction

Facial Emotion Recognition (FER) is a problem of great importance for various domains from Human- Robot and Human-Computer interaction, Intelligent Marketing, AI based survey and Customer sentiment analysis for product enhancement to medical field like Automatic Depression detection, Post Traumatic Stress Disorder, Bi-Polar and other human psychological diseases detection and prevention. Specially in situations like COVID pandemic when the entire world was locked down and people including doctors are maintaining social distancing, and we spent more time with machines and computers, we felt the need of automatic diseases detection, more natural Human Computer Interaction, online counselling etc. For making any such system a reality, we need more robust and light weight models for automatic emotion recognition systems. In the current work we proposed a novel approach for automatic FER system which is robust to give good accuracy on images acquired in real-time and also it is light weight, so that it may be deployed as a mobile application or as a web-portal service. Facial image instead of being seen as a whole entity, we may look at as combination of identity and expression related information Fig-1. In the task of emotion recognition, identity related features may be avoided to make the input image light. Separating the Identity features and extract the emotional feature from the input facial image is possible using Generative Adversarial Network (GAN). Identity related features such as texture, hairstyle, beard, ornaments, any mark on face etc. are permanent and does not change frequently, whereas expression related features like eye and mouth regions are those which changes as per current mood and expression.

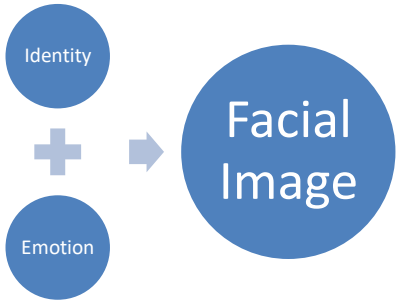

Fig-1: Facial Image: Identity + Expression

1 Department of Computer Science and Engineering, Bhagwant University
Ajmer, India, gyanendra.tiwary@gmail.com
ORCID ID : 0000-0003-1383-3717
2 Department of Computer Science and Engineering, Bhagwant University
Ajmer, India shivanichauhanbit@gmail.com
ORCID ID : 0009-0003-8674-6777
3Department of computer applications, Raja Balwant Singh Management Technical Campus ,Agra. India kgoyal@gmail.com
ORCID ID : 0000-3343-7165-777X
** Corresponding Author Email: Gyanendra.tiwary@gmail.com*

Separating identity related features will also reduce the size of input data and also save processing power on unnecessary processing of data which has nothing to do with emotion recognition. During preprocessing, multiple normalization techniques may further improve the FER accuracy. SIFT is one such transformation. Using SIFT image becomes independent of real time distortion and other issues during image acquisition. It makes the images scale and rotation invariant. SIFT also nullifies the low illumination problem during image acquisition. On this preprocessed image, we separate the facial image from background using Vola-John's method.

Recently Attention-based Convolution Neural Network (A-CNN) models have achived a lot of popularity. It achieves high accuracy through attention mechanism. Attention mechanism apply more weights to the parts of the image which are more relevant for emotion recognition, such as mouth, eyes etc, and less weights to other features. This saves a lot of unnecessary processing and put efforts only towards the relevance areas of input face. An important miles stone in the field of computer based FER task is Paul Ekman's Facial Action Coding System (FACS) and it's Action Units (AUs) [41]. In FACS a systematic approach has been developed for expressions and related facial muscles. All-important facial muscles are divided in AUs. These action units are then applied to classify a face to a particular emotion class. FACS has been used by many models since it's inception and with the emergence of CNN, it's limitation of intensive computation, which was impossible for general purpose computer has also been practically solved.

1.1. Generator

A Generative Adversarial Network (GAN) has two network components which competes with each other. The first network is Generator and other as Discriminator. Generator network generates a novel image by taking an Input image and a random number. It tries to generate a new realistic image which is of a new person. This may resemble with no existing person but an entirely new person's image.

1.2. Discriminator

Once generator network generates an image for a given input image, it goes through the discriminator network. The major task of discriminator is to check if the image looks natural or not. it calculates the errors and gives feedback to the generator back. In this way it looks like a hide and seek game between the generator and discriminator. The game between generator generating an image and discriminator detecting true or false image continues, until generator is generating such a realistic image which looks like original to the discriminator.

1.3. Linear Binary Pattern (LBP)

LBP is a classifier which can easily being implemented using CNN. It works by analyzing local features of the image. In multiple FER studies it has been applied after FACS produces AUs to take AUs as input and use it for classification of image class.

2. Related Work

FER problem can be taken in different ways. [1] Kosti et al. has considered the surrounding information along with facial expressions for detecting and rather understanding the human emotion. They used EMOTIC dataset. There are 26 emotion categories in which images were segmented. They were able to distinguish these many emotions just because they are considering the surroundings also. Two CNN modules, one for Body feature extraction and other for entire image or surrounding feature extraction has been fused finally for complete emotion recognition task.In another work [2] , Li et al. have combined the LBP and Attention mechanism for feature detection in human facial image. They have collected images of 35 persons in the age range of 20 to 25 years. They collected both RGB and depth images using MS motion sensors. They have also compared their results for CK+, JAFFE, FER 2013, OLULU datasets [3]. They have divided their work into four modules, viz: Preprocessing, Feature Extraction, Reconstruction and Classification. This work combined the CNN and LBP which further fed into attention network for classification. Wang et al., [4] proposed a system to detect expression from a single image only. For this purpose, they used SHIFT + CNN hybrid model. They considered color, texture and shape features for analyzing the image. For their experiments, they used CUHK, photo.net, EVA and Live IQ dataset and used probabilistic methods for their analysis. in his work H. Li et al. modeled a fusion network of 2D along with 3D FER. A probability-based estimation applied to fuse 2D with 3D features[5] [6].Two channel FER network has been proposed by Yang et al.. One using LBP and other using gray scale First shallow CNN used DeepID and other used VGG16 with trained on ImageNet dataset [7]. Then output of these are fused in weighted manner and softmax classification is done. In experiment they tested it on CK+, JAFFE, Oulu- CASIA which has given good results. They claimed that these two networks are complementary to each other. One is able to get local and other to get global features of FER.

Mellouk & Handouzi in a survey paper have compared and reviewed various facial expression recognition techniques using Deep CNN [8]. In another work author Connie et al. proposed a comparison of dense and simple SIFT. It has further been compared with simple CNN and combinations of all these too [9]. They have shown the supremacy of CNN + dense SIFT over all other options. An eGAN (Generative Adversarial Network) has been proposed by Lasri et al. which is being used to extract emotion components of the input facial image and map it to an identity free average face [10]. They use ResNet-101 as their base model, which was pre trained on ImageNet dataset. This identity neutral face is then fed into an attention-based CNN for Action Units detection which classify the expression using FACS.

Xia et al., proposed a GAN which takes an image I_x and an expression f_y as input and generate an output image I_y of the same identity with the input expression [11]. They modelled two stage GAN, Local and Global Perception based. [12]They emphasized on the local and global facial features which are crucial for emotion recognition. Global feature they worked on is texture. They have taken the Generator network from [13] and Discriminator network from [14]. Further they employed an attention network. The output has been split into two parallel parts:

a) Color Mask (C) and
b) Attention Mask (M)

Two local features they emphasized are eyes and mouth, as they are more important for Facial Emotion Recognition task. The local generator and discriminator works has been designed as follows:

$$G_{local} = \{G_{eye}, G_{mouth}\} \ldots\ldots\ldots\ldots$$

and

$D_{local} = \{D_{eve}, D_{mouth}\}$……..

They have used Radboud Faces Dataset (RaFD) which consists of 4824 images of size 681 x 1024 and it has been taken from 67 participants under fully controlled laboratory conditions [15]. This dataset consists of pictures from 3 angels, but they have used only frontal images. 90% images has been used for training and 10% for testing. For checking the accuracy. [16] Here authors used two evaluation parameters viz: a) Inception Score (IS) and b) FID Score. Though they have claimed their model as efficient but they have not tested it on natural images taken in uncontrolled environment. Further, their model has been developed considering only 8 discrete emotions and there is no way to represent any natural emotion or different intensities of these 8 basic emotions.

In [17], global feature space taken is the facial texture, whereas we know that there has to be correlation between different facial muscles which combinedly signifies an emotion. We also feel taking only 'eye' and 'mouth' in isolation can not accurately identify the overall facial expression. So few more facial reagons like Forehead, Chicks etc should also been considered for concluding an emotion. [18]They have used Paul Ekarman's FACS (Facial Action Coding System), but have not used that completely. They have only taken few AU's (Action Units) and concluded the emotion on that, which is not right.

[19] In this work authors Huang et al., focused on identity aware FER. They emphasized that if the difference between two person's image with same emotion id D1 and same person with different emotions is D2, then ideally D2 should be greater than D1, but in practice with various state of the art methods, D1>D2. Hence Identity can not and should not be ignored while doing FER. In another work,[20] The authors have solved two core issues in this process: 1. Synthesizing a facial image of any emotion into all 7 basic expressions, and 2. Deciding a proper matric for calculation differences between emotions in two images. First problem is solved by a Star GAN and second is addressed by a Deep CNN based network which calculates feature matric from both the images and a feature point distance based mahalanobis matric is used to calculate the difference. They have tested their method on CK+, Oulu, MMI, ISAFE, ISED datasets. [21]In this system instead of calculating the distance between feature points, they may have used discriminator network of GAN. As per our understanding, it could have given a better result.

[22] Looking at the issues in identification of human facial expression, in this work Xie et al. have proposed a Two-branch Disentangled Generative Adversarial Network (TDGAN) model. [23]Unlike other models where they try to filter and suppress other irrelevant information, in this work they tried to use these features as well to extract facial expression.[24] A two-branch model one for expression and other for remaining features is made through GAN (Generative Adversarial Network). Two independent encoders taken the input for these two branches and then fused by two decoders. In another work [25] the GANs consists of generator and discriminator. Generator takes the expression related features and Impose on a sample face. There are two Discriminators one for face and other for expression. Facial discriminator is supposed to do classify different identities whereas expression discriminator conducts expression classification. [26]They have used GANs for two tasks, expression transfer as well as recognition. [27]They have been deviated from their main objective of FER and provided an additional expression transfer feature. This feature does not fit real on all faces.

[24] In this work, Ali et al. have used facial images from four ethnic regions such as Japanese, Taiwanese, Caucasians, and Moroccans. For this purpose, they have taken three different datasets: Japanese female facial expression, Taiwanese facial expression image database, and RadBoud face database. [28]pre-processing, they applied. In this LBP and PCA is being used. Further they proposed a noble approach to combine cross cultural data.[29] In this a feed forward binary neural network is proposed which uses tan sigmoid activation function. [30] Here the authors have proposed a classifier by combining NB and Bernoulli distribution classifier to identify five universal expressions, sadness, happiness, anger, fear, and surprise as output of the proposed system. As in almost all FER systems, in [31] the model also achieve highest accuracy for Happiness, then Surprise but for other expressions their model has shown quite low performance. The major limitation of their work is the process of combining multicultural facial images and extracting feature vectors from them. [28], [32] In this work they proposed a multi-modal recurrent attention network which uses multi modal video recordings such as color, depth and thermal for facial expression recognition task. Depth and thermal features has been used as guidance priors. [33]An attention mechanism helped in focusing on emotion rich regions. The authors proposed a Multi-modal Arousal Valence Facial Expression Recognition (MAVFER) model for FER task using all three types of video data. This model takes these videos with their continuous arousal valance scores.

[34]The authors points out the limitation of categorical FER techniques (where we just take few categories of emotion like happy, surprise, sad, disgust etc) and emphasized on continuous emotion classes apart from the seven basic emotion categories. For this they have used two domains named Arousal and Valance. Arousal represents the level of engagement whereas valance represents the overall positive or negative feelings. In another work. [35] the process starts with special encoder network, which extracts frame by frame features which in turn fed into temporal decoder network which finds 'where' and 'what' of these features. [36]Authors proposed a novel Attention Guided Long Short Term Memory (AG-LSTM) to detect spatiotemporal emotionally active facial regions. Lastly it passes through a 3D-CNN for processing sequential data and emotion recognition. In this work, they have also made their multimodal dataset public and named it as MAVFER. [37]This model performed well on both RECOLA, SEWA and AFEW benchmarks

[38] In this work, Zhang et al. have done Facial emotion recognition, synthesis and face alignment in one model. The proposed model is combination of three networks: i. face alignment network, ii. Face synthesis network and iii. Facial emotion recognition. Face alignment network has two subnetworks L and Cg. These subnetworks are kept subsequent to each other. L identifies the landmarks using MSE (Mean Square Error). These landmarks are then fed into face synthesis phase as geometric code and as geometric information to FER network. [39] Another work where face synthesis network is further divided into two subnetworks, a generator formed by encoder and discriminator formed by decoder. The generator takes input features such as eyes, mouth, eyebrows, mouth etc and generates local and global identity features. These features are combined with expression code and geometry feature codes to model face. This model face is kept improving by the hide and seek game between generator and

discriminator. Finally the FER network classifies the expression on real and model image. They have tested this model on three datasets 1) Multi-PIE, 2) BU-3DFE, and 3) SFEW and it has shown satisfactory results.

[7] Here Yang et al. proposed a weighted mixture deep neural network (WMDNN) which uses VGG16 and DeepID CNNs and finally fused their outputs by a weighted network where softmax classifier is being used for final facial emotion classification. After several pre-processing like rotation, augmentation, rectification and face detection, the global expression related features are extracted by VGG16 network which takes the pre-processed grayscale facial images. The corresponding Local Binary Pattern (LBP) are obtained parallelly by DeepID network. This proposed model has been tested on ''CK+'', ''JAFFE'' and ''Oulu-CASIA'' datasets and the accuracy obtained are 0.970, 0.922, and 0.923 respectively. The authors have used existing networks and have done a little finetuning in the parameters. Finally their major contribution is the late weighted fusion network which used softmax classifier foe emotion class identification.

[40] In this work, the authors Yan et al. proposed a facial expression deep synthesis and recognition method. There are two networks first to generate facial images with different expressions using facial expression synthesis generative adversarial network (FESGAN) to increase the images in the input dataset, second CNN to detect facial expression on theses images which also uses the learning of FESGAN. Lastly the classification loss is fed back to finetune the GAN of FESGAN network using RDBP (Real Data-guided Backpropagation). They have tested their model on CK+, Oulu-CASIA and MMI datasets and satisfactory performance has been achieved. The major contribution of this work is towards generating images of same identity with different expressions and also generating new identities by changing latent face representation by random vector from prior distribution. This adds up a big number to the existing number of ladled images in the existing datasets. Using GANs they have achieved this and which in turn improved their model performance.

3. Methodology

Processing large size images through deep convolution networks requires high computing power and it usually takes a lot of time. For this reason, we have tried working on preprocessing our input image using an Expression Generative Adversarial Network (E-GAN), SIFT and Vola-John's model. Input to the proposed system will be a facial image I_{in}. During experiments authors have observed that instead of using image pixels directly, providing additional information such as facial landmarks and Histogram of Gaussian (HoG) improves the result significantly. To get this additional information, first authors have detected the face in the input image using vola-john's face detection algorithm and get facial landmarks. Then authors have calculated HoG and input it to A-CNN (Fig-2)

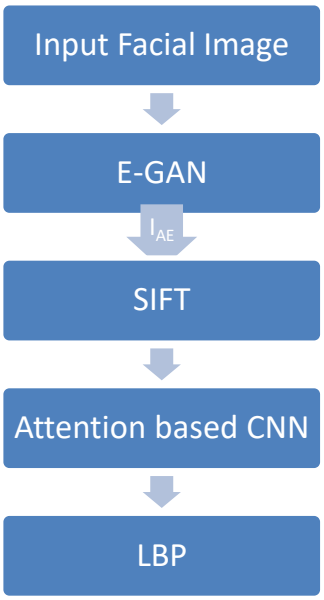

Fig-2: Steps Involved in Proposed FER Model

Facial image data can be quite noisy and there may be many unwanted things present in the background. To deal with this authors have applied vola-johns facial image detection algorithm which by far extent reduces the image size and enhances the emotion recognition accuracy. Then this facial image is fed into an EGAN (Fig-2). The purpose of this is to separate out the identity and expression related features from the input facial image. To recognize facial expression, one does not require the identity related information such as ornaments like earrings, googles etc and also features like facial texture, eye color etc. So, authors have removed these unnecessary details and kept the remaining features, which are relevant to emotion recognition task. For this purpose, the proposed novel E-GAN network takes two images I_{in} and I_{avg} as input. The generator network generates a facial image taking identity features from I_{avg} and expression information from I_{in}. Then this average expression image I_{AE} is being sent to discriminator network. Discriminator then tries to identify the expression difference between I_{in} and I_{AE} and report it back to the generator. Generator on the other hand finetunes itself to meet the expectation and bluff the discriminator. This process continues until discriminator finds almost no difference in the expression of I_{AE} and I_{in}. In this way the novel E-GAN network generates an identity free expression image for further processing.

Thereafter this identity free image is fed into Attention based Deep CNN network. Here through attention mechanism, authors put more weight on regions of greater importance such as Eyes, Mouth, Forehead etc. This significantly reduces the processing data and improves the expression detection results. The Deep CNN network basically tries to find AUs which was proposed in FACS [41]. These AUs are then fed into LBP. Finally, it classify the emotion class among 7 basic emotions: Anger, Happy, Disgust, Sad, Surprise, Neutral and Fear using LBP classifier. The final output of the proposed model is this expression information which tells about the current state of mood of the person in input image I_{in}.

4. Experimental Result and Discussion

Facial expression is a continuous data, though the major expressions may be classified into seven basic emotions. The FER

task is quite complex for a computer and achieving acceptable accuracy is quite tricky.

The flow of data in the current proposed model can be seen in Fig-3. The process starts with an input image and a average facial image.

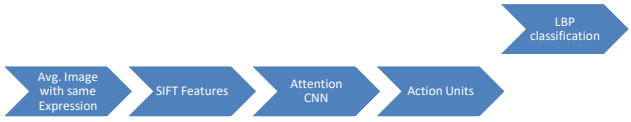

Fig-3: Proposed FER Model

The major part of the proposed model is Expression Generative Adversarial Network (E-GAN). This E-GAN model is responsible for separating out emotion and identity related features. For this E-GAN takes two images I_{in} and I_{avg} as input and produces and identity free average expression only image I_{AE} (Fig-4). The advantage of such an image is it does not have additional specific identity related features remains in the input image. This saves a lot of processing efforts and time. This type of input image has also shown better expression recognition than other state of art models on the datasets used in this study.

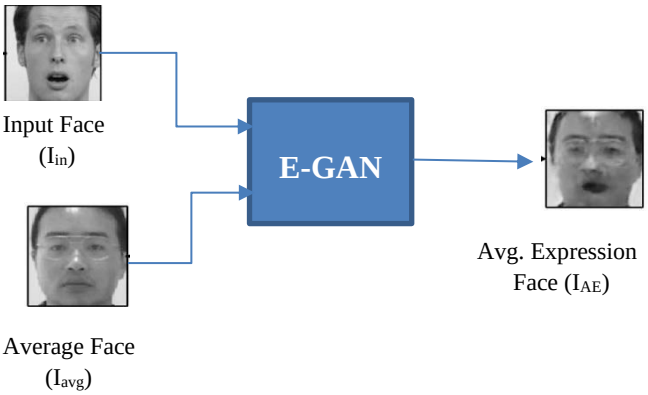

Fig-4: Expression GAN

I_{AE} is further passed through SIFT (Fig-5), which in turn normalize the rotational, scaling, illumination etc. This solves the problem of ill posed images and also issues with image acquisition. After this step images gets normalized and better ready for FER task.

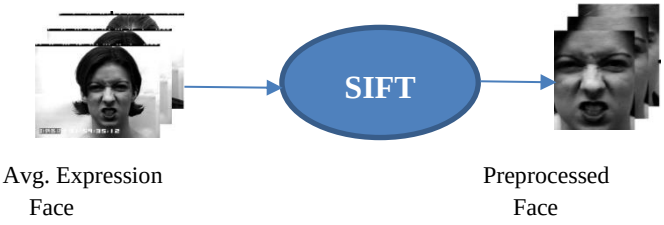

Fig-5: Scale Invariant Feature Transformation (SIFT)

Finally facial image will be separated from the background using Vola-John's face detection algorithm (Fig-6). After this step, an identity less normalized facial image is being obtained. This light weight facial image is now fit for the A-CNN.

Fig-6: Vola-John's Face Detection

Proposed A-CNN model now processes this identity free normalized facial image and detect the AUs. As shown in Fig-7, the A-CNN has multiple dropouts to avoid overfitting and the attention module to put more weight on more relevant parts of the face. By this way, FER becomes more accurate.

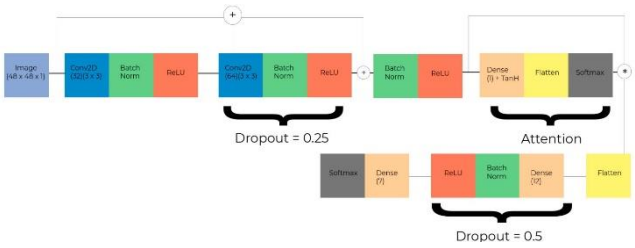

Fig-7: Attention based CNN

The concept of attention module can be better understood by a heatmap (Fig-8). The heatmap showing relevant regions for angry face (Fig-8: upper half) and happy face (Fig-8: lower half) from FER-2013 dataset.

Fig-8: Heat Map of Attention Region

One can easily observe from the heat map in Fig-8 that for happy face, mouth region has gained more attention, whereas for an angry face eye's region has gained maximum attention.

4.1. Results on CK+ dataset

The CK+ dataset has categorized 593 images of 123 subjects into six expression categories, viz: Anger, Happy, Disgust, Sad, Fear and Surprise. Authors have tested the model on this dataset and achieved quite good results. The corresponding confusion matrix has been shown in Table-1.

Table-1: The Confusion Matrix for CK+ dataset

	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	0.96	-	-	-	0.02	-	0.02
Disgust	-	0.99	-	-	-	-	0.01
Fear	-	-	0.93	0.04	-	-	0.03
Happy	-	-	-	1.00	-	-	-
Sad	0.04	-	-	-	0.83	-	0.13
Surprise	-	-	-	-	-	0.97	0.02

Neutral	-	-	-	-	-	-	0.99

4.2. Results on Oulu- CASIA dataset

This is another open access dataset with 80 subject's 2880 images. Labelled with the same six labels as in CK+, i.e. Anger, Happy, Disgust, Sad, Fear and Surprise. The proposed Model has outperformed almost all state of art models on this dataset aalso. The confusion matrix for this dataset can be found in Table-2.

Table-2: The Confusion Matrix for oulu-CASIA dataset

	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	0.95	-	-	-	0.02	-	0.02
Disgust	-	0.89	-	-	-	-	0.01
Fear	-	-	0.90	0.04	-	-	-
Happy	-	-	-	0.96	-	-	0.03
Sad	0.04	-	-	-	0.90	-	-
Surprise	-	-	-	-	-	0.93	0.02
Neutral	-	-	-	-	-	-	0.95

4.3. Results on FER-2013 dataset

This is one of the most famous dataset. This is because the images are taken in real-world conditions, for example images are taken in uneven illumination and few images are occluded too. In total 28709 training and 3589 testing images are kept in this dataset. Propose EGAN model has shown considerably effective results on this dataset too. The corresponding confusion matrix is shown in Table-3.

Table-3: The Confusion Matrix for FER-2013 dataset

	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	0.96	-	-	-	0.02	0.02	-
Disgust	-	0.88	-	-	-	0.02	-
Fear	-	-	0.94	0.04	-	0.02	-
Happy	-	-	0.07	0.95	-	-	-
Sad	-	-	0.01	-	0.90	-	0.04
Surprise	-	-	0.05	-	0.02	0.93	0.02
Neutral	0.02	-	0.03	-	-	-3	0.94

The performance of proposed EGAN model on all these public datasets are quite good. Specially expressions like Happy, Disgust and neutral are quickly identified and are quite accurate every time. The convergence of the proposed model is also tested. With increase in number of epochs, the accuracy (red) -loss (green) graph for all three datasets can be seen in Fig-9.a for the convergence on CK+ dataset, 9.b on Oulu-CASIA and 9.c on FER-2013 dataset.

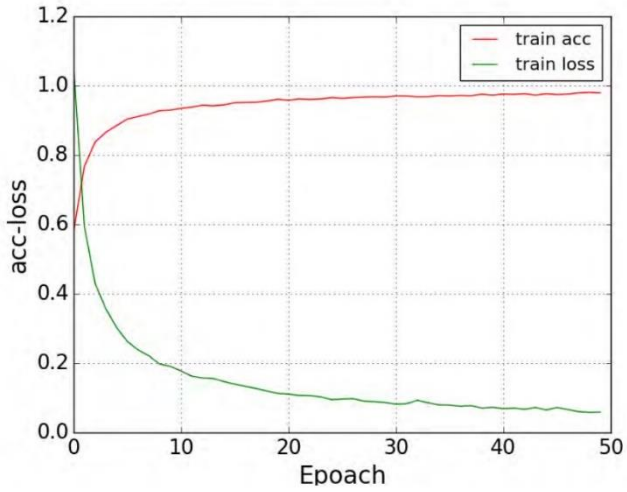

Fig-9(a): Accuracy-Loss graph for CK+

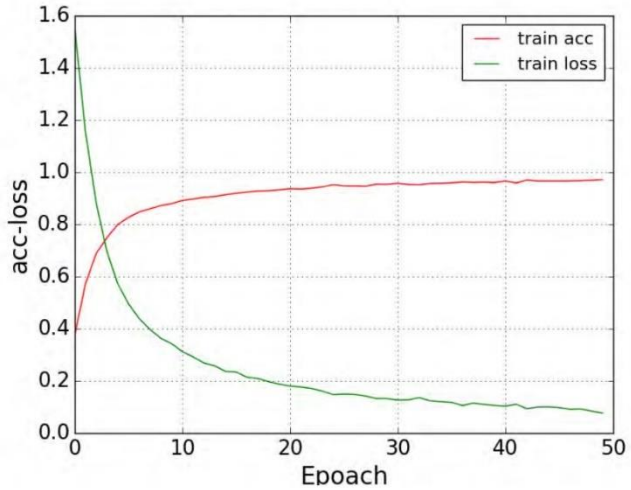

Fig-9 (b): Accuracy-Loss graph for Oulu-CASIA

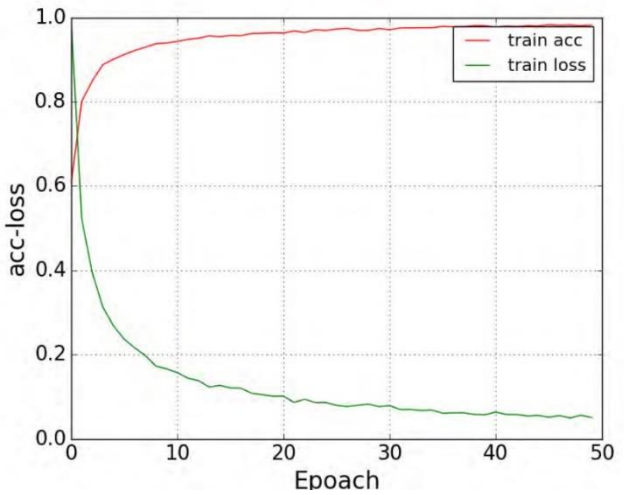

Fig-9(c) Accuracy-Loss graph for FER-2013

This shows that our model has shown accurate detection of almost all 7 basic expressions. With increase in epoach up to a certain number accuracy is increasing and loss is decreasing. After a certain limit, further increase in epoach results are not improving.

5. Conclusion & Future Work

FER using nobel E-GAN has shown quite impressive outcomes. Separating out the Identity and Expression related features has shown a promising impact on the efficiency of proposed model. Authors have tested this FER system on CK+, OULU- Casia and FER-2013 datasets. The A-CNN model has also processed the image quite effectively and classified the expression with almost accuracy using LBP. Except FER-2013, both the datasets used in this study, have laboratory controlled front facial images, hence authors are not sure how the model will perform for real time images in the wild. In future the model can be extended for facial images in the wild.

6. References

[1] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "Emotion Recognition in Context."

[2] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, "Attention mechanism-based CNN for facial expression recognition," Neurocomputing, vol. 411, pp. 340–350, Oct. 2020, doi: 10.1016/j.neucom.2020.06.014.

[3] P. P. Angelov, IEEE Computational Intelligence Society, International Neural Network Society. Morocco Regional Chapter, and Institute of Electrical and Electronics Engineers, ICDS2019 : the Third International Conference on Intelligent Computing in Data Sciences : October 28-29-30, 2019, Marrakech, Morocco.

[4] F. Wang, J. Lv, G. Ying, S. Chen, and C. Zhang, "Facial expression recognition from image based on hybrid features understanding," J Vis Commun Image Represent, vol. 59, pp. 84–88, Feb. 2019, doi: 10.1016/j.jvcir.2018.11.010.

[5] H. Li, J. Sun, Z. Xu, and L. Chen, "Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network," IEEE Trans Multimedia, vol. 19, no. 12, pp. 2816–2831, Dec. 2017, doi: 10.1109/TMM.2017.2713408.

[6] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Institute of Electrical and Electronics Engineers Inc., Nov. 2017, pp. 5967–5976. doi: 10.1109/CVPR.2017.632.

[7] 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, Dec. 2017, doi: 10.1109/ACCESS.2017.2784096.

[8] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: Review and insights," in Procedia Computer Science, Elsevier B.V., 2020, pp. 689–694. doi: 10.1016/j.procs.2020.07.101.

[9] T. Connie, M. Al-Shabi, W. P. Cheah, and M. Goh, "Facial expression recognition using a hybrid CNN-SIFT aggregator," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2017, pp. 139–149. doi: 10.1007/978-3-319-69456-6_12.

[10] Y. Xia, W. Zheng, Y. Wang, H. Yu, J. Dong, and F. Y. Wang, "Local and Global Perception Generative Adversarial Network for Facial Expression Synthesis," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 3, pp. 1443–1452, Mar. 2022, doi: 10.1109/TCSVT.2021.3074032.

[11] I. Lasri, A. R. Solh, and M. El Belkacemi, "Facial Emotion Recognition of Students using Convolutional Neural Network," in 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), IEEE, Oct. 2019, pp. 1–6. doi: 10.1109/ICDS47004.2019.8942386.

[12] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," Nov. 2017, [Online]. Available: http://arxiv.org/abs/1711.09020

[13] V. Sreenivas, V. Namdeo, and E. Vijay Kumar, "Modified deep belief network based human emotion recognition with multiscale features from video sequences," Softw Pract Exp, vol. 51, no. 6, pp. 1259–1279, Jun. 2021, doi: 10.1002/spe.2955.

[14] S. E. Kahou et al., "EmoNets: Multimodal deep learning approaches for emotion recognition in video," Journal on Multimodal User Interfaces, vol. 10, no. 2, pp. 99–111, Jun. 2016, doi: 10.1007/s12193-015-0195-2.

[15] S. Li and W. Deng, "Blended Emotion in-the-Wild: Multi-label Facial Expression Recognition Using Crowdsourced Annotations and Deep Locality Feature Learning," Int J Comput Vis, vol. 127, no. 6–7, pp. 884–906, Jun. 2019, doi: 10.1007/s11263-018-1131-1.

[16] B. C. Ko, "A brief review of facial emotion recognition based on visual information," Sensors (Switzerland), vol. 18, no. 2, Feb. 2018, doi: 10.3390/s18020401.

[17] D. Y. Liliana, "Emotion recognition from facial expression using deep convolutional neural network," in Journal of Physics: Conference Series, Institute of Physics Publishing, Apr. 2019. doi: 10.1088/1742-6596/1193/1/012004.

[18] C. Marechal et al., "Survey on AI-based multimodal methods for emotion detection," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2019, pp. 307–324. doi: 10.1007/978-3-030-16272-6_11.

[19] W. Huang, S. Zhang, P. Zhang, Y. Zha, Y. Fang, and Y. Zhang, "Identity-Aware Facial Expression Recognition Via Deep Metric Learning Based on Synthesized Images," IEEE Trans Multimedia, vol. 24, pp. 3327–3339, 2022, doi: 10.1109/TMM.2021.3096068.

[20] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," Apr. 2018, doi: 10.1109/TAFFC.2020.2981446.

[21] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," IEEE Transactions on Affective Computing, vol. 10, no. 3. Institute of Electrical and Electronics Engineers Inc., pp. 325–347, Jul. 25, 2019. doi: 10.1109/TAFFC.2017.2731763.

[22] Davis, W., Wilson, D., López, A., Gonzalez, L., & González, F. Automated Assessment and Feedback Systems in Engineering Education: A Machine Learning Approach. Kuwait Journal of Machine Learning, 1(1).

Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/102
[23] S. Xie, H. Hu, and Y. Chen, "Facial Expression Recognition with Two-Branch Disentangled Generative Adversarial Network," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 6, pp. 2359–2371, Jun. 2021, doi: 10.1109/TCSVT.2020.3024201.
[24] I. Yildirim, M. Belledonne, W. Freiwald, and J. Tenenbaum, "Efficient inverse graphics in biological face processing," 2020.
[25] G. Ali et al., "Artificial Neural Network Based Ensemble Approach for Multicultural Facial Expressions Analysis," IEEE Access, vol. 8, pp. 134950–134963, 2020, doi: 10.1109/ACCESS.2020.3009908.
[26] A. Gupta, S. Arunachalam, and R. Balakrishnan, "Deep self-attention network for facial emotion recognition," in Procedia Computer Science, Elsevier B.V., 2020, pp. 1527–1534. doi: 10.1016/j.procs.2020.04.163.
[27] D. Liu, L. Wang, Z. Wang, and L. Chen, "Novel multi-scale deep residual attention network for facial expression recognition," The Journal of Engineering, vol. 2020, no. 12, pp. 1220–1226, Dec. 2020, doi: 10.1049/joe.2020.0183.
[28] S. Rajan, P. Chenniappan, S. Devaraj, and N. Madian, "Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM," IET Image Process, vol. 14, no. 7, pp. 1227–1232, May 2020, doi: 10.1049/iet-ipr.2019.1188.
[29] T. H. Vo, G. S. Lee, H. J. Yang, and S. H. Kim, "Pyramid with Super Resolution for In-the-Wild Facial Expression Recognition," IEEE Access, vol. 8, pp. 131988–132001, 2020, doi: 10.1109/ACCESS.2020.3010018.
[30] H. Li and H. Xu, "Deep reinforcement learning for robust emotional classification in facial expression recognition," Knowl Based Syst, vol. 204, Sep. 2020, doi: 10.1016/j.knosys.2020.106172.
[31] A. John, M. C. Abhishek, A. S. Ajayan, S. Sanoop, and V. R. Kumar, "Real-time facial emotion recognition system with improved preprocessing and feature extraction," in Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 1328–1333. doi: 10.1109/ICSSIT48917.2020.9214207.
[32] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," SN Appl Sci, vol. 2, no. 3, Mar. 2020, doi: 10.1007/s42452-020-2234-1.
[33] G. P. R, E. Granger, and P. Cardinal, "Deep Domain Adaptation for Ordinal Regression of Pain Intensity Estimation Using Weakly-Labelled Videos," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.06392
[34] J. Lee, S. Kim, S. Kim, and K. Sohn, "Multi-Modal Recurrent Attention Networks for Facial Expression Recognition," IEEE Transactions on Image Processing, vol. 29, pp. 6977–6991, 2020, doi: 10.1109/TIP.2020.2996086.
[35] S. M. S. A. Abdullah, S. Y. A. Ameen, M. A. M. Sadeeq, and S. Zeebaree, "Multimodal Emotion Recognition using Deep Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 02, pp. 52–58, Apr. 2021, doi: 10.38094/jastt20291.
[36] A. Barman and P. Dutta, "Facial expression recognition using distance and shape signature features," Pattern Recognit Lett, vol. 145, pp. 254–261, May 2021, doi: 10.1016/j.patrec.2017.06.018.
[37] D. G. R. Kola and S. K. Samayamantula, "Facial expression recognition using singular values and wavelet-based LGC-HD operator," IET Biom, vol. 10, no. 2, pp. 207–218, Mar. 2021, doi: 10.1049/bme2.12012.
[38] S. Saurav, R. Saini, and S. Singh, "EmNet: a deep integrated convolutional neural network for facial emotion recognition in the wild," Applied Intelligence, vol. 51, no. 8, pp. 5543–5570, Aug. 2021, doi: 10.1007/s10489-020-02125-0.
[39] Kumar, C. ., & Muthumanickam, T. . (2023). Analysis of Unmanned Four-Wheeled Bot with AI Evaluation Feedback Linearization Method. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 138–142. https://doi.org/10.17762/ijritcc.v11i2.6138
[40] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "A Unified Deep Model for Joint Facial Expression Recognition, Face Synthesis, and Face Alignment," IEEE Transactions on Image Processing, vol. 29, pp. 6574–6589, 2020, doi: 10.1109/TIP.2020.2991549.
[41] J. Cai et al., "IDENTITY-FREE FACIAL EXPRESSION RECOGNITION USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORK," in Proceedings - International Conference on Image Processing, ICIP, IEEE Computer Society, 2021, pp. 1344–1348. doi: 10.1109/ICIP42928.2021.9506593.
[42] Y. Yan, Y. Huang, S. Chen, C. Shen, and H. Wang, "Joint Deep Learning of Facial Expression Synthesis and Recognition," IEEE Trans Multimedia, vol. 22, no. 11, pp. 2792–2807, Nov. 2020, doi: 10.1109/TMM.2019.2962317.

Author Contributions

Gyanendra Tiwary: Conceptualization, Field study, Methodology, and Implementation.
Shivani Chauhan: Data collection, Writing-Original draft preparation.
Krishan Kumar Goyal: Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.