Computer Systems Science & Engineering
DOI:10.32604/csse.2022.019749
Article

Tech Science Press

Facial Expression Recognition Using Enhanced Convolution Neural Network with Attention Mechanism

K. Prabhu[1,*], S. SathishKumar[2], M. Sivachitra[3], S. Dineshkumar[2] and P. Sathiyabama[4]

[1]Department of EIE, Kongu Engineering College, Perundurai, 638060, Tamilnadu, India
[2]Department of EEE, M.Kumarasamy College of Engineering, Karur, 639113, Tamilnadu, India
[3]Department of EEE, Kongu Engineering College, Perundurai, 638060, Tamilnadu, India
[4]Department of EEE, CSI Engineering College, Ketti, 643215, Tamilnadu, India
*Corresponding Author: K. Prabhu. Email: kprabhuresearch1@gmail.com
Received: 24 April 2021; Accepted: 14 June 2021

Abstract: Facial Expression Recognition (FER) has been an interesting area of research in places where there is human-computer interaction. Human psychology, emotions and behaviors can be analyzed in FER. Classifiers used in FER have been perfect on normal faces but have been found to be constrained in occluded faces. Recently, Deep Learning Techniques (DLT) have gained popularity in applications of real-world problems including recognition of human emotions. The human face reflects emotional states and human intentions. An expression is the most natural and powerful way of communicating non-verbally. Systems which form communications between the two are termed Human Machine Interaction (HMI) systems. FER can improve HMI systems as human expressions convey useful information to an observer. This paper proposes a FER scheme called EECNN (Enhanced Convolution Neural Network with Attention mechanism) to recognize seven types of human emotions with satisfying results in its experiments. Proposed EECNN achieved 89.8% accuracy in classifying the images.

Keywords: Facial expression recognition; linear discriminant analysis; animal migration optimization; regions of interest; enhanced convolution neural network with attention mechanism

1 Introduction

Human expressions are one of the most effective forms of communications between humans. An expression can provide information on the internal emotional state of a human being. Expressions can flow from a speaker to a listener in a conversation and vice versa. In case of automatic recognitions, an expression can be treated as a deformation of the human face or changes in facial pigmentations [1]. A facial expression transfers around 55% of the intent in a communication which is more than what voice and language can convey together [2]. Researchers have been analyzing emotions, behavior and psychology of humans using FER which has generated significant interest in the areas of HCI, mental health assessments [3] and intelligent transport systems [4]. Moreover, Multimedia gadgets based on FER

can recognize real time expressions and thus can be used for feedbacks from customers [5]. One major problem of using FER in videos is the extraction of discriminative spatiotemporal features from human expression in video sequence images. Thus, FER can be used not only in recognizing human emotions but also in interpersonal communications [6–7]. Automated analysis in FER is a challenge as humans can show a variety of expressions on their face while the datasets for such expressions is very small. Humans can easily recognize facial expressions but the same becomes a complex issue in automations despite recent recognition methods using high-resolution images [8]. Facial expressions change human expressions. Machine Learning (ML) has played an important role in FER based on its complexity. DLTs have also performed well in facial expression recognitions which can be classified as static images and image sequences [9]. Image sequences capture changes from a neutral point to a point of peak expression. When the expressions are on the same face, it is easier to extract expression features. Methods using static images analyses peak expression images without temporal information.

Expression extraction techniques in FER can be geometry-based or appearance-based. Appearances highlight the texture of an expression like wrinkles which can be extracted using Local Binary Patterns (LBP) or filters like Gabor wavelets filter [10], Gradient filter with histograms. Facial components like the nose can be extracted using geometry-based methods where the initial part is tracking of facial points using an Active Appearance Model (AAM) for detections. Both the methods have their own disadvantages in selecting proper features. A designer has to identify the proper filter in appearance-based methods. FER systems when applied to lab-collected datasets are found to perform poorly in their recognition of human expressions and large-scale facial expression datasets are required for researches in FER [11]. Facial occlusions pose their own set of challenges mainly because of the facial positions. Further, e occlusions are caused by other objects used by humans like glasses, scarf, mask etc which block important facial parts in FER. Variations in occlusions are difficult to cover due to limited data on these and inevitably lead to minimized accuracies in detections. A number of researchers have turned to ML approaches for FER. Deep learning techniques have also been used in FER; hence this paper proposes an Enhanced Convolution Neural Network with attention mechanism (ECNN) to address issues in occlusion. It mimics humans in recognizing facial expressions-based patches in faces. It perceives concentrates on informative and unblocked patches in the face. The rest of the paper is organized as follows: section two is Literature review, three is proposed methodology results and discussion are explained in section four and the paper concludes with section five.

2 Literature Review

Pu et al. [12] proposed a framework for FER by using Action Units (AUs) in image sequences. Facial motion was measured using AAM feature points and their displacements were estimated with Lucas–Kanade (LK) optical flow tracker. The vectors obtained in displacements between the neutral and peak expressions in motion were transformed using random forest for determining expression's Aus. Such detected AUs were again used iin classification by random forest. The experiments on Extended Cohn-Kanade(CK+) database demonstrated higher performance. Patch-Gated Convolution Neural Network (PG-CNN) was proposed in the study by Li et al. [13]. PG-CNN automatically perceived occluded regions of the face from un-occluded regions. It extracted possible regions of interest feature maps from facial landmarks. Every patch was reweighed for its importance. The technique was evaluated on largest FER datasets namely AffectNet and RAF-DB. Their results showed that proposed PG-CNN improved recognition accuracy on original and synthetically occluded faces.

Zhang et al. [14] proposed a Monte Carlo algorithm for recognizing occlusions in images. It extracts face's part images using Gabor filter and matches distances based on templates created. The resultant facial features were found to covered occlusions. The technique was evaluated using randomly placed

occlusion patches in the regions around the eyes and mouth. Training and test strategies were adopted on matched and mismatched places to analyze occlusions. Tested on the JAFFE and Cohn-Kanade databases, it demonstrated robustness and speed in processing occlusions in the mouth and eyes parts of images. A symmetric Speeded Up Robust Features (SURF) framework was used by Hu et al. [15] in FER. The symmetric SURF detected the occlusion part by locating a horizontal symmetric area. This was followed by a face in painting module based on mirror transition in an unsupervised manner. A recognition network based heterogeneous soft partitioning took the weights of each part as input for recognition in training. The weighted image was fed into a trained neural network for expression recognition. Their results on Cohn-Kanade (CK +) and fer2013 datasets showed an improvement of seven to eight percent when compared to others in FER. Histogram of Oriented Gradient (HOG), Genetic Algorithm (GA) and Support Vector Machines (SVM) wre also used by Mlakar et al. [16] in FER. Six emotions were prototyped. Features were assessed from a neutral and peak expression of a face. HOG descriptors were determined by a GA and SVM recognized the expressions. The proposed method was tested on Cohn Kanade (10 subjects) and JAFFE database (192 images) resulting in a mean recognition rate of 95% which was other video-based methods in FER.

Ren et al. [17] proposed AAM-SIFT (Active Appearance Model- Scale-Invariant Feature Transform) based on Fuzzy C-Means (FCM) clustering. AAM templates estimate feature points on the face images. The features are identified by a hybrid gradient direction histogram based on AAM and SIFT (Scale-Invariant Feature Transform) descriptors. The features are then divided into different groups and adaptive weights are obtained by FCM. The membership degree computed by FCM representing classes is input into SVM for classifications. BU-3DFE database with six facial expressions and seven poses were used for the experiments and the proposed technique was found to be effective in improving expression recognitions. Binarized Auto-encoders (BAEs) and Stacked BAEs were used by Sun et al. [18] in their work for recognition of expressions. A huge unlabeled facial dataset when subjected to Binarized Neural Networks (BNNs) can improve performances. The study combines unconstrained face normalization, a LBP descriptor variant, BAEs and BNNs to achieve achieves better performances on Static Facial Expressions. Its hardware requirements and memory consumption were found to be very low. LBP was also used by Qi et al. [19]. They used it with SVM in FER. The LBP operator extracted facial contours which was segmented the face into sub-regions using a pseudo-3-D model. Global facial expression images were mapped with LBP for feature extraction which was then classified by SVM and softmax. Cohn-Kanade (CK+) facial expression data set along with volunteer's faces were used for the tests. The results showed better recognition rates than other traditional models. Expression-specific Local Binary Pattern (es-LBP) was proposed by Chao et al. [20] where the class was regulated by class-regularized Locality Preserving Projection (cr-LPP) technique in FER, es-LBP is an improved technique in assessing particular facial points from f human faces.

The connection between facial features and expression classes was enhanced by cr-LPP which maximized class independence while preserving local features with dimensionality reduction. Their results showed effectiveness in their FER outputs. Aly et al. [21] proposed a technique called Multi-Stage Progressive Transfer Learning (MSPTL) in an automated FER for 2D images. It classified emotions in 6 basic categories namely sadness, happiness, fear, surprise, anger, and disgust in frontal and non-frontal poses. AlexNet deep convolutional neural network [22–24] tuned the outputs in a three-stage process. The initial two training stages were based on FER datasets with frontal images only while the third FER dataset include non-frontal image poses. Their results on VT-KFER and 300W databases outperformed other systems in their expression recognitions.

3 Proposed EECNN (Enhanced Convolution Neural Network with Attention Mechanism) Methodology

Automated FERs exist in a variety of applications like neuro marketing, robotics and interactive games. They generally involve three steps mainly preprocessing, feature extractions and classifications where extraction of facial expression features is the most important part efficient feature extractions result improved recognitions of facial expressions. The proposed EECNN can recognize seven types and its architecture where depicted in Fig. 1.

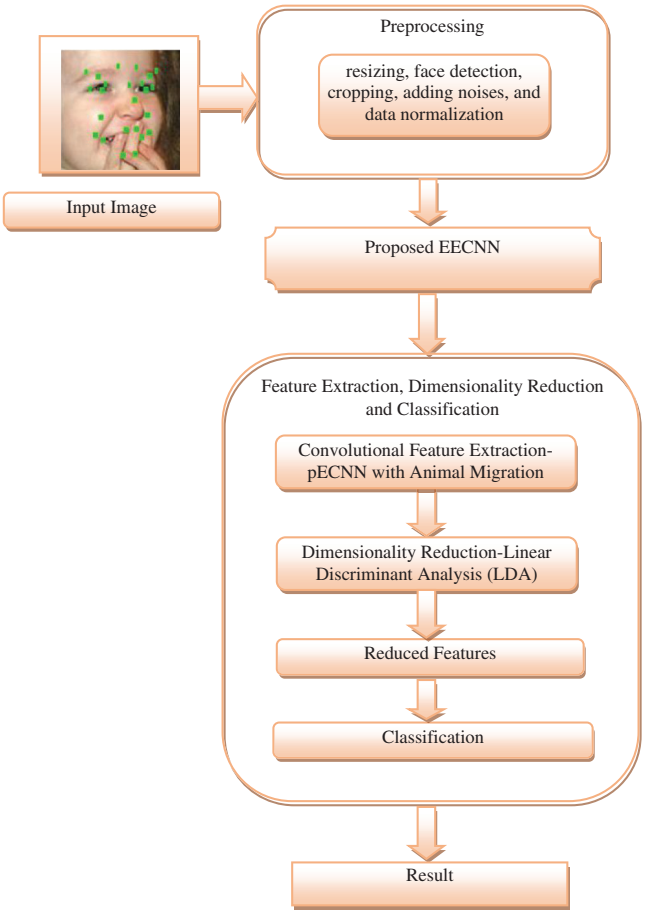

Figure 1: EECNN architecture

EECNN has been proposed to recognize facial expressions that are partially occluded. These regions are automatically conceived with a focus on informative and un-blocked regions. It consists of four steps namely Pre-processing, Feature Extraction, Dimensionality Reduction and Classification. An input image is preprocessed to reduce the effects of illuminations.

3.1 Preprocessing Stage

In this phase images are resized (128x128 pixels), cropped, noises are added and finally normalized. Preprocessing is an important step in FER as images get converted an into a normalized uniform sized image where normalization is typically on illumination. Image resizing as a pre-processing step enlarges or reduces image in pixel format for output display. Thus, even low-resolution images can be converted

into high resolution image which helps visualizing data. Image may contain artifacts or noises where noise is a random variation in color or brightness of an image. Noise can generate uncertainty in signals and thus have to be eliminated. This step is executed with available filters. Median filters are used for removing noise in this work. Normalization is required in images as they may have poor contrasts due to glare. Normalization helps visualize a picture with uniform intensity. This work uses Global Contrast Normalization (GCN) by subtracting each pixel value by the mean and then divides it by standard deviation. This eliminates problems in images with low contrast in FER. GCN is given in Eq. (1)

$$X'_{i,j,k} = s\frac{X_{i,j,k} - \bar{X}}{\max\{\epsilon,\ \sqrt{\lambda + \frac{1}{3rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{3}(X_{i,j,k} - \bar{X})^2}} \tag{1}$$

where, $X_{i,j,k}$ has row i, column j, and colour depth k, and $\bar{X}$ is the mean intensity of the entire image. EECNN further normalizes the images with a local normalization for computing local means and local standard deviations where element-wise subtraction and division are used. Finally, histogram equalization is done on the image for image enhancements based on a cumulative density function.

3.2 Feature Extraction

Feature extraction is the main phase EECNN as it is a very important stage in FER for identifying patches that are occluded. They change image pixel information into an elevated representation in terms of color or shape or motion or texture. This phase also decreases dimensionality space. The location of features is found from video sequences by dividing them into blocks D and difference between consecutive blocks P_i and P_{i+1} is determined using Eq. (2)

$$D_i = |(P_i - P_{i+1})| \tag{2}$$

Each block's Median is also calculated as Median = $Sum\ (D_i)/\ (n \times n)/2$ where $M = Fix\sqrt{Median}$. Each pixel in the video frames is divided into MSB and LSB, thus splitting the video sequence into frames and further frames into pixel size to find the location of features. The image is processes in convolutional net (VGG) which is then decomposed into sub-feature maps to obtain diverse local patches. Each local patch's is encoded as a weighed vector by a Patch-Gated Unit (PG-Unit) which uses AttentionNet. These weighed local representations are also encoded as a weighed vector by a Global-Gated Unit (GG-Unit). These two connected layers are used in classifications EECNN minimizes softmax loss by considering different interest of the local and global regions. Fig. 2 is an illustration of feature extraction. Initially each selected frame will divide into blocks (n x n) as shown in Fig and then appropriate pixels are determined by comparing consecutive blocks in the frame.

3.3 Patch Based ECNN (PECNN)-AM

pECNN-AMO focuses on local discriminative and representative patches. It is achieved in two key schemes namely region decomposition and occlusion perception. Facial expressions are activities that invoke muscle motions and localizing expression-related parts is beneficial in FER. Further, dividing the face into multiple local patches helps locate occlusions [25]. Facial landmark with 68 points are detected and 24 points are re-computed for covering the face. Patches are defined with these points as the center. Convolutional operations decrease the model size and enlarge receptive fields of neurons based on the $512 \times 28 \times 28$ resulting in 24 local regions, each with a size of $512 \times 6 \times 6$. pECNN automatically perceives blocked facial patches and concentrates on unblocked or informative patches. The structure of PG-Unit is illustrated in Fig. 3.

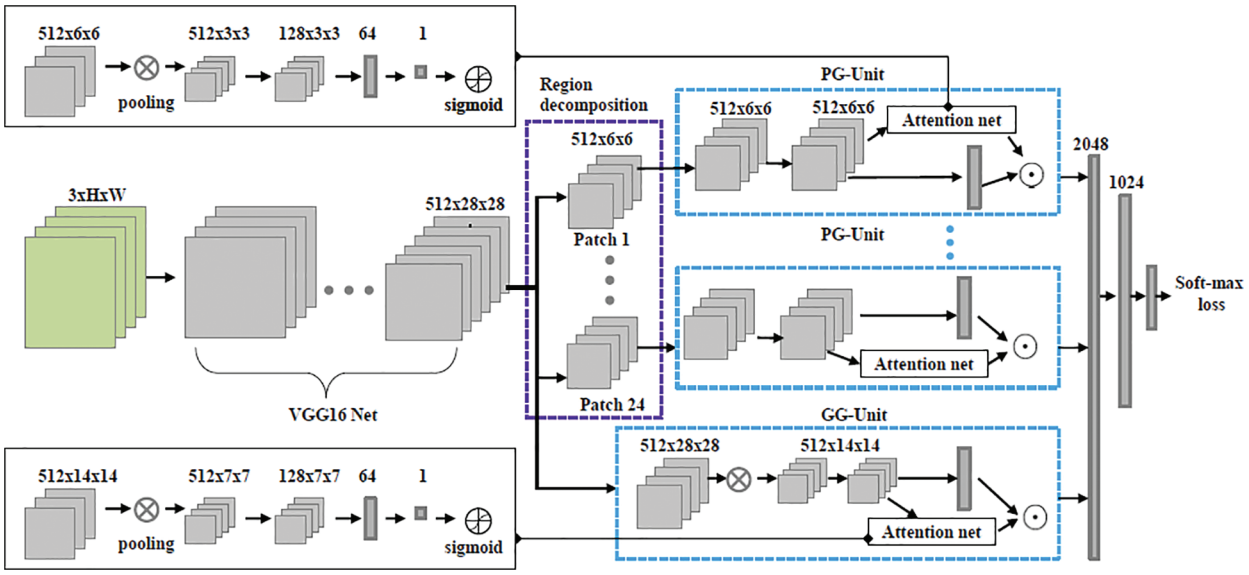

Figure 2: EECNN feature extraction roadmap

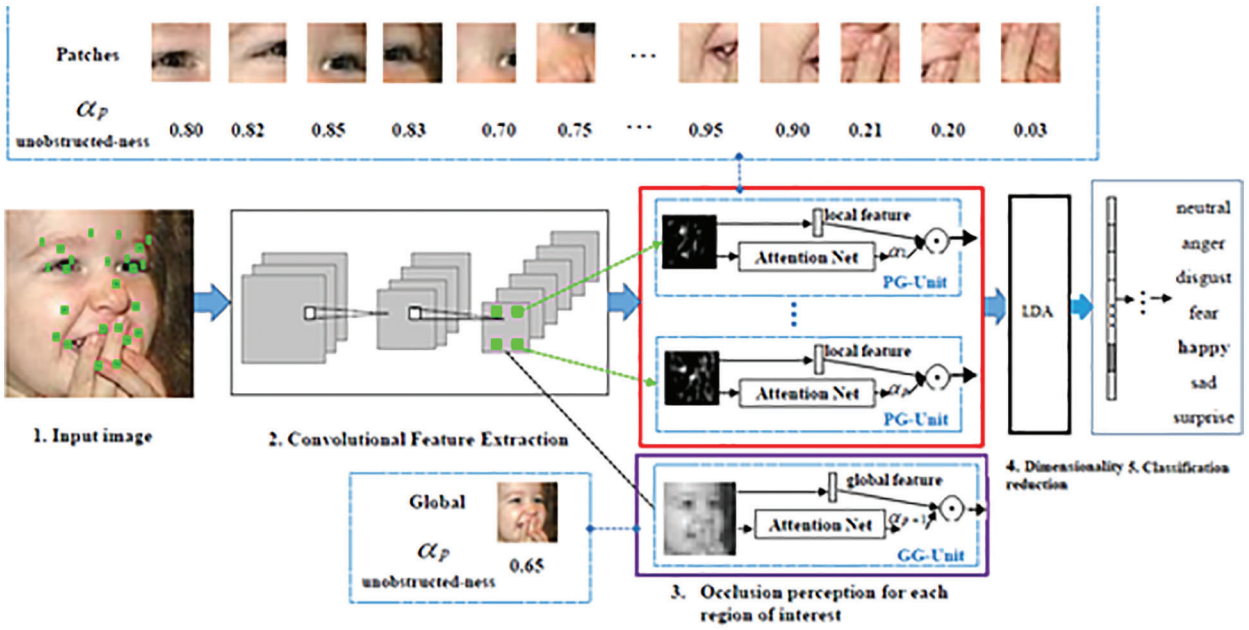

Figure 3: EECNN PG unit

Each patch's cropped local feature maps are fed into two convolution layers without decreasing spatial resolution where input feature maps are encodes as the vector-shaped local feature and then Attention Net estimates a scalar weight to denote the importance of the local patch. Mathematically, if p_i denotes the input $512 \times 6 \times 6$ feature maps of its i-th patch. $\tilde{p}_i = \tilde{\varnothing}(p_i)$ denote the last $512 \times 6 \times 6$ feature maps, the i-th PG-Unit takes the feature maps $\tilde{p}_i$ as the input, learns the local specific facial feature ψ_i: using $\psi_i = \psi(\tilde{p}_i)$ where its weight $\alpha_i = I_i(\tilde{p}_i)$. ψ_i is a vector that represents the un-weighed feature. α_i is a scalar that represent the patch i's importance or "unobstructedness". I(.) Implies the operations in the Attention Net, A sigmoid function forces the output α_i ranges in [0; 1], where 1 indicates the most salient

unobstructed patch and 0 indicates the completely blocked patch. These PG-Units can automatically learn low weights for the occluded parts and high weights for the unblocked and discriminative parts. pECNN weight α_i is optimized using AMO for improving recognition accuracy in EECNN, since, AMO is a heuristic optimization algorithm which can be divided like animal migration and updating process where values are updated by the probabilistic method [26]. Three rules are followed in the process namely avoiding collisions with neighbors, moving in the same direction as the group and remaining close to neighbors within a neighborhood topology. The function is given in Eq. (3)

$$X_{i,G+1} = X_{i,G} + \delta.(X_{neiborhood,G} - X_{i,G}) \tag{3}$$

where $X_{neiborhood,G}$ - Current facial images count of the neighborhood, δ - a random number generator controlled by a Gaussian distribution, $X_{i,G}$ - no of facial images of i^{th} individual, $X_{i,G+1}$ - no of facial images of i^{th} individual. In the update algorithm r_1, $r_2 \in [1, \ldots, NP]$ are randomly chosen integers, $r_1 \neq r_2 \neq i$. After producing the new facial images $X_{i,G+1}$, it will be evaluated and compared with the $X_{i,G}$,and choose the individual with a better objective recognition accuracy using Eq. (4). The update algorithm is shown as Fig. 4.

$$X_i = \begin{cases} X_{i,G} \text{ if } f(X_{i,G}) \text{ is better than } f(X_{i,G+1}), \\ X_{i,G+1} \text{ otherwise} \end{cases} \tag{4}$$

(1) For i=1 to NP do
(2) For j=1 to D do
(3) If rand > P_a then
(4) $X_{i,G+1} = X_{r1,G} + rand.(X_{best,G} - X_{i,G}) + rand.(X_{r2,G} - X_{i,G})$
(5) End if
(6) End For
(7) End For

Figure 4: Population update algorithm

The migration of EECNN established as the boundary of the living area using Eq. (5)

$$low = X_{best} - R(\alpha_i), \quad up = X_{best} + R(\alpha_i) \tag{5}$$

$$R = \rho.R(\alpha_i)$$

where X_{best} – Best current best patched image, R - living area radius, ρ -shrinkage coefficient, $\rho \in (0, \quad 1)$, all in $1 \times D$ row vector. R's value depends on the size of the search space. On iterations, a larger value of R improves exploration ability while a smaller value while weight parameter α_i pECNNfor improves recognition accuracy. However, pECNN-AMO might ignore complementary information in facial images and hence its integration with global representation in EECNN leads to better FER performance for occlusions. The Global-Local Attention infers local details in a global context [27]. gECNN is an ensemble learning which promotes diversity in learned features. Facial feature maps are encoded from *conv*4_2 to *conv*5_2 in VGG16 net and based on it $512 \times 28 \times 28$ feature maps are encoded in a region of $512 \times 14 \times 14$. The **Global-Gated Unit (GG-Unit)** in gECNN automatically weighs global facial representation. It initially encodes the input feature maps as vector-shaped global representations which is the learned on a scalar weight by Attention Net for a global facial representation which is again weighed by the computed weight.

3.4 EECNN Dimensionality Reduction

Feature selections result in computational complexity and effectiveness of classifications which makes it necessary to verify inputs in a reduced form. Too many features increase the complexity of training. Selecting a sub set of features improves the efficiency of a classifier while reducing its execution time. Extracting discriminating features is Dimension reduction as it increases the class variances for class separation [28]. LDA is used in EECNN as it is easy computationally with effective predictions. Also, it is less expensive when compared to others and mainly lighting changes do not affect its performance [29]. Fig. 5 depicts Dimension Reduction in a Feature Vector, while the steps of Dimensionality reduction are shown in Fig. 6.

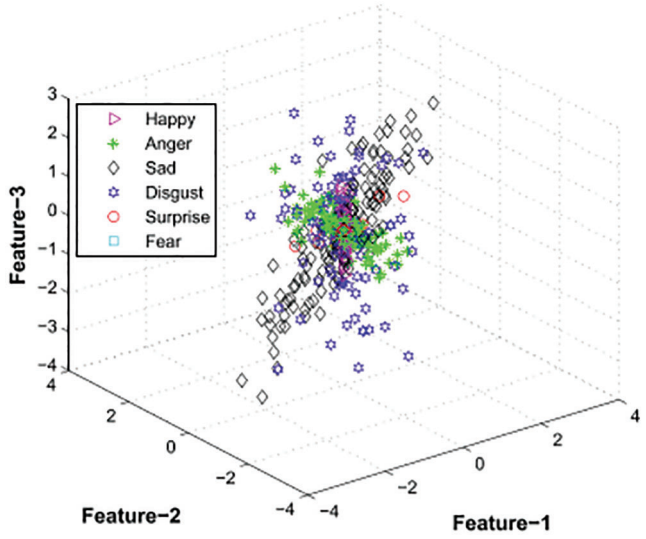

Figure 5: Dimension reduction in a feature vector

Step 1: compute the average sample values for different kinds of facial images in original space. Total number is denoted by c. x_{ij} denotes the j^{th} objects of the i^{th} class of sample

$$m_i = \frac{1}{n_i}\sum_{j=1}^{n_1} X_{ij}.X_{ij} \in R^d, \qquad \text{Where i=1, 2... c.}$$

$$m = \sum_{i=1}^{c} p^i m^i$$

Step 2: compute covariance matrix of each class: $c_i = \frac{1}{n_i}\sum_{j=1}^{n1}(X_{ij} - m_i).(X_{ij} - m_i)^T$

Step 3: Compute within-class and between-class scatter matrices:

$$c_b = \sum_{i=1}^{c} p_i\,(m_i - m).(m_i - m)^T$$

$$c_W = \sum_{i=1}^{c} c_i$$

Step 4: compute Eigen vector of matrix $c_w^{-1}c_b$ to get projection vectors. The dimensionality reduction data can be obtained by projection

Figure 6: EECNN dimensionality reduction steps

3.5 Classification

The last step of EECNN FER is Classification. It is an algorithmic approach for recognizing a given expression. The proposed e-EECNN and g-EECNN together help in classifying seven basic human emotions namely anger, neutral, disgust, happiness, fear, surprise and sadness. Classifications is done by integrating optimized the parameters of EECNN model's outputs to recognize the expressions.

4 Results and Discussion

The proposed EECNN results are detailed by way of figures and tables as necessary. Multiple datasets were used in evaluation of EECNN. RAF-DB contains which contains 30,000 facial images with annotated expressions was taken for the study. The experiments were conducted basic emotions where 12,200 images were used in training and 3,0000 images for testing. AffectNet, which contains about 400,000 images of seven discrete facial expressions and Extended Cohn-Kanade dataset (CK+) containing 593 video sequences recorded from 123 subjects was also selected. Finally, FED-RO with 400 images, a self-collected Facial Expression Dataset with Real-world Occlusions was the fourth database on which EECNN was evaluated. The synthesized occluded images were manually collected using search engine by using keywords like bread, beer, wall, hair, hand, hat, cabinet, computer, book etc. The selected items had occurrences in obstructed facial images. Tab. 1 lists category wise the no of images in the FED-RO database

Table 1: Category wise images in Fed-Ro

Dataset	Neutral	Anger	Disgust	Fear	Happy	Sad	Surprise
FED-RO	50	53	51	58	59	66	63

Tab. 2 lists comparative performance of EECNN with SVM and CNN techniques. The techniques were evaluated on both real and synthetic occlusions, including FED-RO, RAF-DB, AffectNet and Cohn-Kanade (CK+)).

Table 2: Comparative performances of techniques in FER

Dataset name	Methods/Metrics	Accuracy (%)	Error detection rate (%)
RAF-DB	SVM	83.5	16.5
	CNN	84.5	15.5
	EECNN	86.2	13.8
AffectNet	SVM	65.1	34.9
	CNN	75.7	24.3
	EECNN	78.8	21.2
CK+	SVM	84.5	15.5
	CNN	85.9	14.1
	EECNN	86.5	13.5
FED-RO	SVM	83.5	16.5
	CNN	86.1	13.9
	EECNN	89.8	10.2

The experiment result show EECNN classifier's supremacy when compared to other existing methods. Proposed EECNN classifier achieves highest accuracy in all the four mentioned datasets as seen in Fig. 7. Fig. 8 shows the error detection rates of the evaluated techniques of proposed EECNN method performs better by having lower error detection rates and in comparison with SVM and CNN. The error detection rates are shown in Fig. 8.

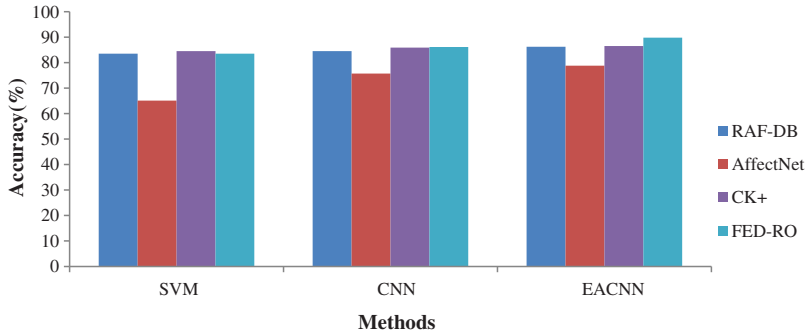

Figure 7: EECNN's comparative accuracy in FER

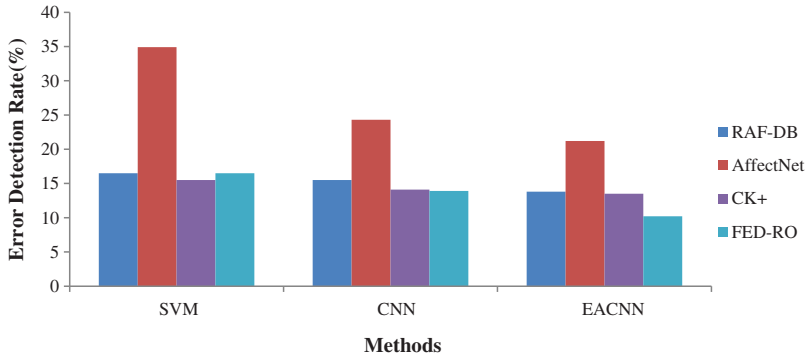

Figure 8: Co-operative error detection rates of techniques

5 Discussions

Human expressions, a form of communication provide information on the internal emotional state of a human being. However, recognizing basic emotions using computers is quite challenging as they represent delicate translations or scaling or rotation of the head in images. Recognizing emotions becomes more complex with occlusions in the face. FER in real-life scenarios with natural occlusions have been limited to the use of medical mask or sunglasses. FER evaluations on real occlusions have been lesser [30]. To problems in FER on images that are occluded this work proposes a technique using DLTs called EECNN. The proposed EECNNs rely on the detected landmarks and do not neglect any facial landmark which suffers misalignment in occlusions. It detects facial landmarks including occlusions and extracts patches irrespective of occlusions. It proves its performance on a collected and annotated a real time Facial Expression Dataset with Real Occlusions (FED-RO) which was created by mining Bing & Google search engine for occluded images with queries like “smile+face+beard” and “smile+face+glasses” etc. The preprocessing splits each image into two upper and lower regions where upper parts contain eyes and eyebrows while the lower part contains mouth and lips. Preprocessing stage involves resizing, face detection, cropping, adding noises, local normalization, global contrast normalization and histogram

equalization. After preprocessing images EECNN goes through an elaborate feature extraction, dimensionality reduction and classification steps. Feature extraction in EECNN involves patch-based extraction where EECNN combines multiple representations from a facial ROI where each representation is weighed a Gate Unit based on importance. The method's pECNN-AMO and gECNN select the feature space whose dimensions are reduced by employing LDA. Thus, in EECNN different regions of a facial image are focused and each region weighed according to its obstructions. To increase the accuracy and handle FER, this study turned towards ML and DLTs for solving issues proposed in FER.

6 Conclusions

This work has proposed an Enhanced CNN with Attention Mechanism (EECNN) for recognizing occluded faces. EECNN follows the sequential steps of Pre-processing, Feature Extraction, Dimensionality Reduction and Classification. EECNN's classifications of basic emotions have been evaluated on synthetic and real-life datasets. Its Experimental results show that EECNN improve the recognition accuracy on occluded faces. EECNN outperforms other methods as its accuracy on RAF-DB dataset is 86.2% which is higher than other methods. Thus, it can be concluded that EECNN is a viable and implementable technique for occluded faces and an improvement in FER. This work proposes to improve attention parts in faces without landmarks in its future work.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] N. Perveen, N. Ahmad, M. A. Q. B. Khan, R. Khalid and S. Qadri, "Facial expression recognition through machine learning," *International Journal of Scientific & Technology Research*, vol. 5, no. 3, pp. 91–97, 2016.

[2] K. M. Goh, C. H. Ng, L. L. Lim and U. Sheikh, "Micro-expression recognition: An updated review of current trends, challenges and solutions," *The Visual Computer*, vol. 36, no. 3, pp. 445–468, 2020.

[3] W. Zhang, Y. Zhang, L. Ma, J. Guan and S. Gong, "Multimodal learning for facial expression recognition," *Pattern Recognition*, vol. 48, no. 10, pp. 3191–3202, 2015.

[4] B. Allaert, J. Mennesson, I. M. Bilasco and C. Djeraba, "Impact of the face registration techniques on facial expressions recognition," *Signal Processing: Image Communication*, vol. 61, pp. 44–53, 2018.

[5] Y. Liu, Y. Li, X. Ma and R. Song, "Facial expression recognition with fusion features extracted from salient facial areas," *Sensors*, vol. 17, no. 4, pp. 1–18, 2017.

[6] E. Stai, S. Kafetzoglou, E. E. Tsiropoulou and S. Papavassiliou, "A holistic approach for personalization, relevance feedback & recommendation in enriched multimedia content," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 283–326, 2018.

[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, 2017.

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

[9] I. Gogić, M. Manhart, I. S. Pandžić and J. Ahlberg, "Fast facial expression recognition using local binary features and shallow neural networks," *The Visual Computer*, vol. 36, no. 1, pp. 97–112, 2020.

[10] A. R. Rivera, J. R. Castillo and O. O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1740–1752, 2012.

[11] S. Li, W. Deng and K. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proc CVPR*, Honolulu, HI, USA, pp. 2584–2593, 2017.

[12] X. Pu, K. Fan, X. Chen, L. Ji and Z. Zhou, "Facial expression recognition from image sequences using twofold random forest classifier," *Neurocomputing*, vol. 168, pp. 1173–1180, 2015.

[13] Y. Li, J. Zeng, S. Shan and X. Chen, "Patch-gated CNN for occlusion-aware facial expression recognition," in *Proc ICPR*, Beijing, China, pp. 2209–2214, 2018.

[14] L. Zhang, D. Tjondronegoro and V. Chandran, "Random gabor based templates for facial expression recognition in images with facial occlusion," *Neurocomputing*, vol. 145, pp. 451–464, 2014.

[15] K. Hu, G. Huang, Y. Yang, C. M. Pun, W. K. Lin *et al.*, "Rapid facial expression recognition under part occlusion based on symmetric SURF and heterogeneous soft partition network," *Multimedia Tools and Applications*, vol. 79, pp. 1–21, 2020.

[16] U. Mlakar and B. Potočnik, "Automated facial expression recognition based on histograms of oriented gradient feature vector differences," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 245–253, 2015.

[17] F. Ren and Z. Huang, "Facial expression recognition based on AAM–SIFT and adaptive regional weighting," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 10, no. 6, pp. 713–722, 2015.

[18] W. Sun, H. Zhao and Z. Jin, "An efficient unconstrained facial expression recognition algorithm based on stack binarized auto-encoders and binarized neural networks," *Neurocomputing*, vol. 267, pp. 385–395, 2017.

[19] C. Qi, M. Li, Q. Wang, H. Zhang, J. Xing *et al.*, "Facial expressions recognition based on cognition and mapped binary patterns," *IEEE Access*, vol. 6, pp. 18795–18803, 2018.

[20] Q. L. Chao, J. J. Ding and J. Z. Liu, "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection," *Signal Processing*, vol. 117, pp. 1–10, 2015.

[21] S. F. Aly and A. L. Abbott, "Facial emotion recognition with varying poses and/or partial occlusion using multi-stage progressive transfer learning," in *Proc SCIA*, Norrköping, Sweden, pp. 101–112, 2019.

[22] P. M. Arunkumar and S. Kannimuthu, "Mining big data streams using business analytics tools: A bird's eye view on MOA and SAMOA," *International Journal of Business Intelligence and Data Mining, Inder Science Journal*, vol. 17, no. 2, pp. 226–236, 2020.

[23] S. Kannimuthu, D. Bhanu, K. S. Bhuvaneshwari, "Performance analysis of machine learning algorithms for dengue disease prediction," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 12, pp. 5105–5110, 2020.

[24] P. M. Arunkumar and S. Kannimuthu, "Machine learning based automated driver-behavior prediction for automotive control systems," *Journal of Mechanics of Continua and Mathematical Sciences*, vol. 7, pp. 1–12, 2020.

[25] A. Dapogny, K. Bailly and S. Dubuisson, "Confidence-weighted local expression predictions for occlusion handling in expression recognition and action unit detection," *Computer Vision*, vol. 126, pp. 255–271, 2018.

[26] M. Ma, Q. Luo, Y. Zhou, X. Chen and L. Li, "An improved animal migration optimization algorithm for clustering analysis," *Discrete Dynamics in Nature and Society*, vol. 2015, pp. 1–12, 2015.

[27] A. Dhall, R. Goecke, S. Lucey and T. Gedeon, "Collecting large, richly annotated facial-expression databases from movies," *IEEE Multimedia*, vol. 19, no. 3, pp. 34–41, 2012.

[28] M. H. Siddiqi, R. Ali, A. M. Khan, Y. T. Park and S. Lee, "Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1386–1398, 2015.

[29] Y. Lu, S.Wang and W. Zhao, "Facial expression recognition based on discrete separable shearlet transform and feature selection," *Algorithms*, vol. 12, no. 1, pp. 1–13, 2019.

[30] L. Zhang, B. Verma, D. Tjondronegoro and V. Chandran, "Facial expression analysis under partial occlusion: A survey," *ACM Computing Surveys*, vol. 51, no. 2, pp. 1–49, 2018.