International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. V, No. 07, pp. 10-17, July

A Seagull Optimization Algorithm Enhanced EfficientNet-B0 Framework for Accurate Facial Expression Recognition

R. Sahila Devi[3]

[3]Assistant Professor, Department of Computer Science and Engineering,
Rohini College of Engineering and Technology,
Kanyakumari, India.

sahiladevi@gmail.com

ABSTRACT: In recent days, Facial expression recognition (FER) has a major role in machine learning to identify change of emotions in human being. Nevertheless, due to individual variances and changes in intensity of emotion, it is hard to gather precise hand-crafted elements that are closely linked to changes in expression such as illumination, occlusion, head pose and subtle variations. Hence, features that accurately details about facial expression change in human is required. To overcome the limitations, this study proposes a Seagull Optimization Algorithm (SOA) with Enhanced EfficientNet-B0 framework for precise FER. Data preprocessing is performed using Median Filter. Enhance EfficientNet-B0 approaches uses compound scaling for effective balance of network depth, weight and input resolution. Hence, to extract feature by learning discriminative and hierarchical facial representation thereby maintaining computational complexity, Enhanced EfficientNet-B0 is deployed. Optimizing model parameters is performed using SOA thereby improving discriminative capability of extracted feature. Feature learning and classification process of EfficentNet-B0 is contributed using SOA which effectively aids in searching optimal model parameters. Performance of proposed model is identified using standardized parameters such as precision, recall, accuracy and F1-score in which, values are 94%, 94%, 94% and 94% respectively.

Keywords: Facial expression recognition (FER), Seagull Optimization algorithm (SOA), Enhanced EfficientNet-B0, Median Filter, K-means clustering.

1. Introduction

Research on emotion recognition is important since it is used in clinical diagnosis, interpersonal machine learning, intelligent teaching, and mental health monitoring. FER is also playing a major role in online learning in a variety of educational domains where FER is necessary in online classes despite the fast emergence and continuous improvement of Artificial Intelligence (AI) [1,2]. In FER, extraction of accurate and robust facial expression features is required and it is effectively performed using Convolutional Neural Network (CNN) in [3]. FER is still a challenging task to perform, however, as it generally works inadequately when significant change in facial expression is not present. A variety of modalities are used in emotion recognition, including bio signal, audio signals and video streams [4]. The visual channel is especially effective because it is non-intrusive and contains a wealth of information and communication of human emotions is influenced by facial expression [5,6]. In [7], graph CNN is proposed for orienting micro-expression recognition in which expression of face is divided into regions and optical flow method is used for extracting features, but model suffers from sensitivity to image quality and limited temporal modelling. To address this challenges, accurate

FER is required in all domains such as education, clinical diagnostics, SOA which is improved with Enhanced EfficientNet-B0 model is proposed in this paper [8].

2. Recent Works

Jiayu Ye et al [2025] [8] have proposed Voluntary Facial Expression Mimicry (VFEM) by gathering facial expression data from depression, anxiety and schizophrenia. Based on VFEM, a vision transformer FER is proposed for Complex mental disorder patients (CmdVIT). This integrates facial expression through explicit and implicit mechanism. Although its performance is better in facial recognition, dataset size and diversity is limited and high computational complexity.

Bei Pan et al [2025] [9] have proposed Multiscale Sequence Information Fusion (MSSIF) model for Dynamic Facial Expression Recognition (DFER) in video sequence. Efficacy of MSSIF is evaluated using datasets namely, eNTERFACE'05, BAUM-1s and AFEW. Although using those datasets achieves greater accuracy, few limitations occur due to dependence on video quality, limited dataset availability and temporal alignment issues.

Mohammad Hameed Siddiqi et al [2025][10] have proposed healthcare monitoring systems using Neural Random Forest that make use of CNN for feature extraction process. Although recognition rate of 97.3% is being achieved, model suffers from dependence on dataset quality, less robustness to real-world situations, difficulty in mixed emotions.

Fumihiko Nakamura et al [2023][11] have proposed a FER approach with photo-Reflective Sensors which is sandwiched in Head-Mounted Display and influence of face direction affecting recognition is also performed. Although it has promising advantages of improving accuracy of classification, gathering data of vertical movements of face and gaze is difficult in improving facial expression evaluation accuracy.

Jianing Teng et al [2023][12] have proposed Typical Facial Expression Network (TFEN) that use Deep two-dimensional (2D) CNN for extracting expression and facial features from provided input videos. Residual influence of facial feature is minimized but model is sensitive to head pose and occlusion, dependence on high-quality facial sequences and datasets, difficulty in recognizing mixed emotions.

The contribution of proposed work is given in the following,

- Developing an efficient FER model by employing enhanced EfficientNet-B0 that extracts difficult spatial and facial features.
- Integrating SOA with enhanced EfficientNet-B0 to optimize network parameters and enhance classification performance.
- Evaluating proposed approach with the aid of performance metrics such as precision, recall, accuracy, F1-score and confusion matrix.

The proposed system is explained using block diagram is as follows.

3. Proposed Work Explanation

Image dataset EmotionNet-6,is provided as input which is subjected to pre-processing unit in which noise and blur control is performed using median filter, in which output of median filter is fed into segmentation unit which is performed using K-means clustering algorithm for interactive control. Extraction of feature from image dataset is performed for representation of shape and smoothens analysis. Features that are extracted is subjected for classification process, which is performed using SOA optimized using enhanced EfficientNet-B0 and accurate FER is predicted with performance metrics such accuracy, recall, precision and F1-score measure. The proposed system block diagram is represented in Figure 1.

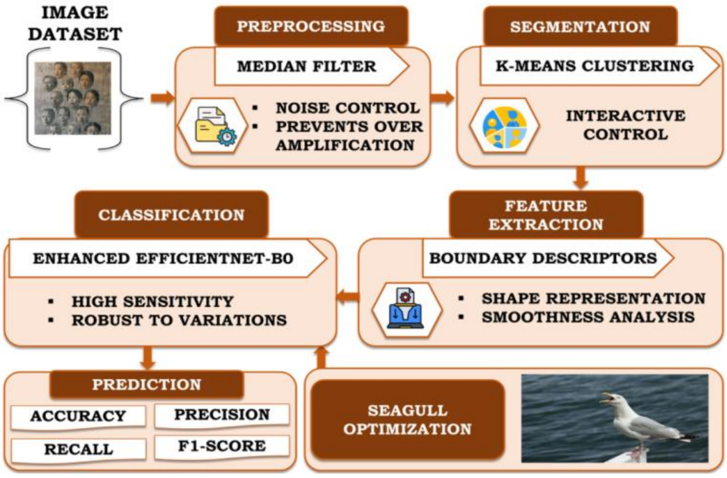

Figure 1: *Proposed system block diagram*

3.1 Median Filter

In preprocessing unit, Median filter is employed which removes noise from input image provided. This filter aids in noise suppression in images that smooth the statistical image dataset. Using this filter, pixel values is replaced with median values in a window centered at each pixel.

3.2 K-means Clustering

Unsupervised learning algorithm that aids in solving clustering problems and interactive control is performed using K-means clustering unit. Let $X = \{x_1, \dots, x_n\}$ is dataset in dimension d. K-means objective function $J(z, A)$ along with updating equation for cluster membership and centres is as follows,

$$a_k = \frac{\sum_{i=1}^{n} z_{ik} x_{ij}}{\sum_{i=1}^{n} z_{ik}} \qquad (1)$$

3.3 Enhanced EfficientNet-B0 architecture

EfficientNet-B0 is a classification approach in which network scaling, depth, resolution and breadth of images is adjusted to enhance performance. In conventional EfficientNet-B0, Convolutional Block Attention Module (CBAM) is incorporated to enhance feature extraction thereby improves classification accuracy of proposed model. This architecture employs fixed set of scaled parameters which differs image breadth, resolution and depth in equal manner. Enhanced EfficientNet-B0 provides special CNN model which number of linked layers is equal to depth of network. Width of image is determined using number of filters in convolutional layer and resolution of image is width and height of given images. Resolution, depth and breadth of images with respect to ϕ is as follows,

$$r = \Upsilon^{\phi} \qquad (2)$$

$$d = \alpha^{\phi} \qquad (3)$$

$$w = \beta^{\phi} \qquad (4)$$

Where α, $\boldsymbol{\beta}$ and Υ are constant which is applied for hyperparameter tuning mechanism. To optimize parameters of model, SOA is employed in this work and it is explained in the following.

3.4 Seagull Optimization Algorithm

Seagull birds are inhabitant of shorelines of sea and ocean, in which, length and mass of seagull is used for differentiating among themselves. Using parameter 'N', new search agent position $\overrightarrow{F_S}$ with provided information collision avoidance among agents of searching in SOA is obtained.

$$\overrightarrow{F_S} = N_x \overrightarrow{D_S}(t) \qquad (5)$$

Where current location of seagull is denoted as $\overrightarrow{D_S}$, current iteration is indicated as *t.* Modelling collision avoidance variable "N" is represented as,

$$N = E_c - \left(t * \left(\frac{E_c}{Max.Iter}\right)\right) \qquad (6)$$

Change in variable is controlled in which it is reduced from E_c to 0, collision avoidance should be selected for value of 2. Completion of collision avoidance mechanism, search agent takes step to move closer to ideal person location which is represented as,

$$\overrightarrow{M_S} = A_x(\overrightarrow{P_{bs}}(t) - \overrightarrow{D_S}(t)) \quad (7)$$

Equilibrium state is attained between tow phases namely, exploration and exploitation, randomized parameter indicated as "A" is calculated as follows,

$$A = 2 * N^2 * rand() \quad (8)$$

Then position of every search agent is changed as follows,

$$\overrightarrow{R_S} = |\overrightarrow{F_S} + \overrightarrow{M_S}| \quad (9)$$

Place to place movement of seagull is performed in three dimensions and it is indicated as follows,

$$S' = r * \cos(j) \quad (10)$$

$$T' = r * \sin(j) \quad (11)$$

$$U' = r * j \quad (12)$$

Where r denotes seagulls spiral movement radius and numbers from 0to 2 is randomly chosen. The remaining search agent's position is updated when best solution is saved.

$$\overrightarrow{D_S}(k) = (\overrightarrow{M_S} * S' * T' * U') + \overrightarrow{P_{bs}}(k) \quad (13)$$

Seagull has separate gland in their neck and bodies which is coated with white feathers. The flowchart for SOA is represented in Figure 2.

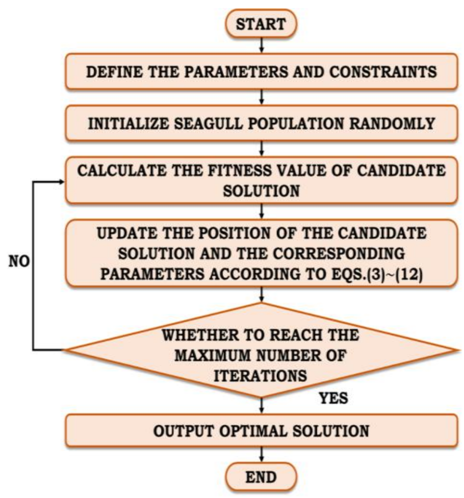

Figure 2: *Flowchart for SOA*

3.1 Mathematical Expressions and Symbols

In this section, mathematical notations and symbols have been explained that evaluates performance of classification model which is often performed using metrics like precision, accuracy, recall and F1-score measure. Ration of all correct predictions to summation of cases is stated as accuracy and it is represented as follows,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Where FN, TP, TN and FP indicate false negative, true positive, true negative and false positive respectively. F1-score provides neutral measure that combines recall and precision. Recall is stated as correctly identified actual positives. Precision is the proportion of accurate positive values in overall positive predictions. Equations (15) and (16) indicates precision and recall as follows,

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

Fi-score is evaluated using Equation (17) as follows,

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (17)$$

All these performance metrics aids to understand overall correctness of prediction and also helps in reliability of model during prediction of positive cases, especially in fluctuating datasets.

4. Results and Discussion

This section details about experimental outcomes which is executed using EmotionNet-6 dataset. The dataset is accessed using following link https://www.kaggle.com/datasets/isrcoimbra/emotionnet-6-realistic-facial-emotion-dataset and dataset name is EmotionNet-6. Training images is 2000 and testing images is 100. Figure 3 details about images based on train and test in which various emotions such as angry, fear, happy, normal and sad based on training images are shown in Figure 3(a). In Figure 3(b), test random images are shown for different emotions such as angry, fear, happy, normal and sad.

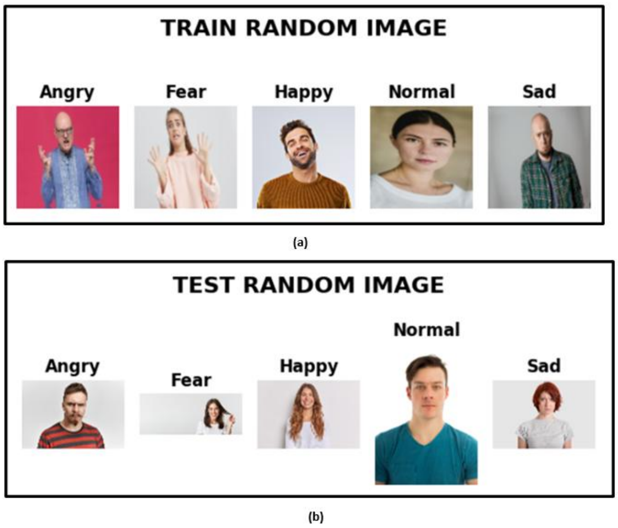

Figure 3: *Outputs (a) Train random images (b) Test Random Images*

Input image is depicted in Figure 4(a), which is resized into RGB image and Grayscale images using Median filter in preprocessing unit, which aids in noise and blur removal is shown in Figure 4(b) and 4(c).

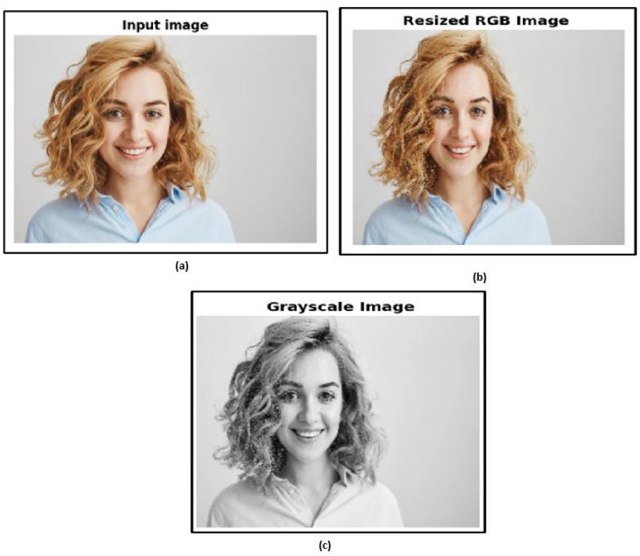

Figure 4: *(a) Input image (b) Resized RGB image (c) Grayscale image*

Output using K-means clustering which aids in interactive control of provided Emotion-6 dataset image is shown in Figure 5(a), from image obtained from K-means clustering boundary of image is extracted as shown in Figure 5(b).

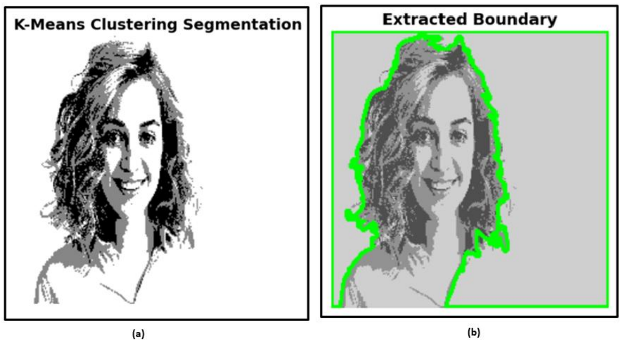

Figure 5: *Output using (a) K-means clustering (b) Extracted Boundary*

Confusion matrix is shown in Figure 6(a), which is plotted for actual class vs predicted class for various emotions like happy, sad, fear, angry and normal. ROC curve for various class is shown in Figure 6(b), such as happy, fear, sad, normal and angry which is plotted between true positive value vs false positive rate.

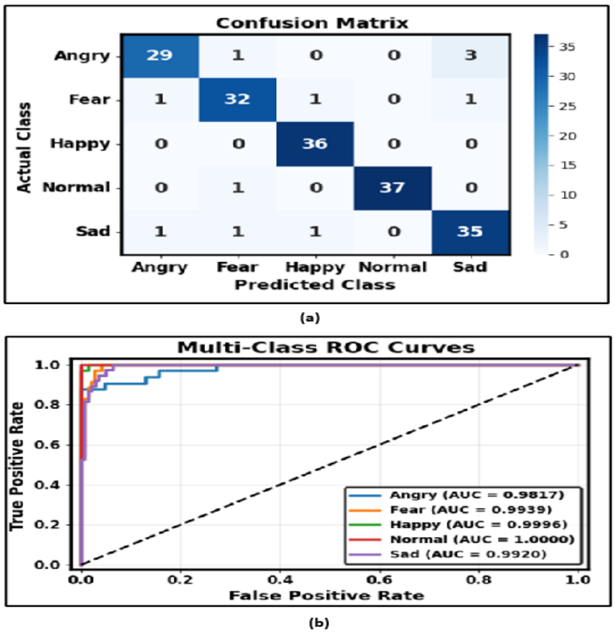

Figure 6: *(a) Confusion matrix (b) Multi-class ROC curves*

Training vs validation accuracy graph is shown in Figure 7(a), which is plotted for accuracy vs epochs and it details that accuracy of trained image increased above 400 epochs. Training vs validation loss curve is shown in Figure 7(b), in which training loss is decreased for epochs.

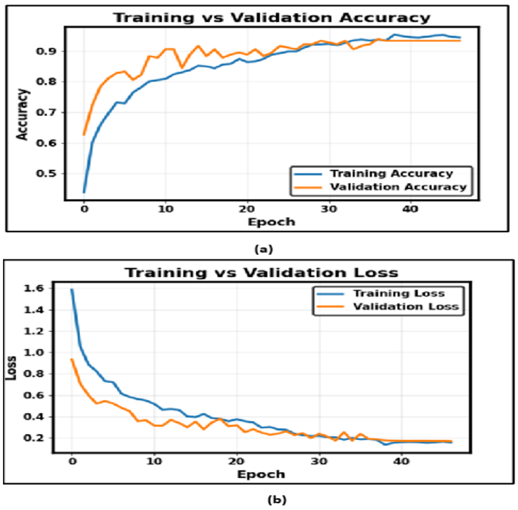

Figure 7: *Curves for (a) Training vs Validation Accuracy (b) Training vs Validation Loss*

Emotion prediction for classes fear is illustrated in Figure 8(a), which details about facial characteristics such as raised eyebrows, change in mouth and widened eyes and class angry details with narrowed eyes and tensed muscles around mouth and facial muscles are captured as shown in Figure 8(b).

Figure 8: *Emotion Prediction for (a) Fear (b) Angry*

Emotion prediction for classes happy, normal and sad is shown in Figure 9(a), 9(b) and 9(c). In Figure 9(a), features captured are broaden eyes with smiling teeth in happy class. In Figure 9(b), person with actual facial features are capture in normal class. In Figure 9(c), image with shrinked mouth and eye brows are captured in sad class.

Figure 9: *Emotion Prediction for (a) Happy (b) Normal (c) Sad*

Figure 10 details about comparison on performance metrics for various classifier model such as CNN, graph CNN and proposed model in which accuracy, precision, recall and F1-score values are 94%, 94%, 94% and 94% respectively.

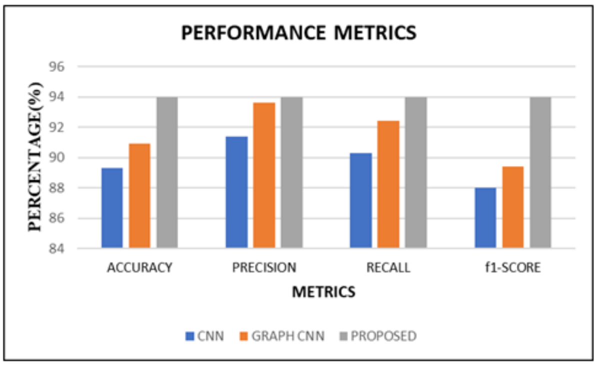

Figure 10: *Comparison graph on performance metrics*

5. Conclusion

In this paper, to recognize accurate facial expression of humans by providing input image dataset and to optimize parameters of network model, SOA which is optimized using Enhance efficentNet-B0 is proposed. The integrated framework provides balance among computational efficiency and recognition processes. Although this model is influenced by additional computational model in SOA optimization process which increases training time for datasets. Framework also depends on quality, resolution and size of different datasets limitations occurs in classification accuracy. Evaluation of classifier is predicted using performance metrics such as 94% of accuracy, 94% of recall, 94% of precision and 945 of F1-score values respectively. Accurate recognition of facial emotions enables healthcare and mental health monitoring, education in E-learning and surveillance and security purposes.

References

1. Mohammed Aly; Abdullatif Ghallab; Islam S. Fathi, Year: “Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model”, In IEEE Access, Vol: 11, pp. 121419-121433, 2023.
2. Shiwei Ren; Meng Sun; Bin Wang; Manyi Liu; Shaoyang Men, Year: 2025, “High Precision Infant Facial Expression Recognition by Improved YOLOv8”, in IEEE Access, Vol: 13, pp. 39621-39630.
3. Ying He; Yanxia Zhang; Shuxin Chen; Yu Hu, Year: 2023, “Facial Expression Recognition Using Hierarchical Features With Three-Channel Convolutional Neural Network”, in IEEE Access, Vol: 11, pp. 84785-84794.
4. Yerim Ji; Suh-Yeon Dong, Year: 2024, “Multi-Task Learning by Leveraging Non-Contact Heart Rate for Robust Facial Emotion Recognition”, in IEEE Access, Vol: 12, pp. 92175-92180.
5. Savina Jassica Colaco; Dong Seog Han, Year: 2025, “Scalable Context-Based Facial Emotion Recognition Using Facial Landmarks and Attention Mechanism”, in IEEE Access, Vol: 13, pp. 20778-20791.
6. VSGS Phaneendra Bottu; Krishnasamy Ragavan, Year: 2025, “Emotion-Based Music Recommendation System Integrating Facial Expression Recognition and Lyrics Sentiment Analysis”, in IEEE Access, Vol: 13, pp. 87740-87752.
7. Xuliang Yang; Yong Fang; C. Raga Rodolfo; Year: 2024, “Graph Convolutional Neural Networks for Micro-Expression Recognition—Fusion of Facial Action Units for Optical Flow Extraction”, in IEEE Access, Vol; 12, pp. 76319-76328.
8. Mohamed Abdel-Basset; Reda Mohamed; Laila Abdel-Fatah; Marwa Sharawi; Karam M. Sallam, Year: 2023, “Improved Metaheuristic Algorithms for Optimal Parameters Selection of Proton Exchange Membrane Fuel Cells: A Comparative Study”, in IEEE Access, Vol: 11, pp. 7369-7397.
9. Jiayu Ye; Yanhong Yu; Qingxiang Wang; Guolong Liu; Wentao Li; An Zeng; Yiqun Zhang; Yang Liu; Yunshao Zheng, Year: 2025, “CmdVIT: A Voluntary Facial Expression Recognition Model for Complex Mental Disorders”, in IEEE Transactions on Image Processing, Vol: 34, pp. 3013-3024.
10. Bei Pan; Kaoru Hirota; Yaping Dai; Zhiyang Jia; Shuai Shao; Jinhua She, Year: 2025, “Learning Sequential Variation Information for Dynamic Facial Expression Recognition”, in IEEE Transactions on Neural Networks and Learning Systems, Vol: 36, No: 6, pp. 9946-9960.
11. Muhammad Hameed Siddiqi; Irshad Ahmad; Yousef Alhwaiti; Faheem Khan, Year: 2025, “Facial Expression Recognition for Healthcare Monitoring Systems Using Neural Random Forest”, in IEEE Journal of Biomedical and Health Informatics, Vol: 29, No: 1, pp. 30-42.
12. Fumihiko Nakamura; Masaaki Murakami; Katsuhiro Suzuki; Masaaki Fukuoka; Katsutoshi Masai; Maki Sugimoto, Year: 2023, “Analyzing

the Effect of Diverse Gaze and Head Direction on Facial Expression Recognition With Photo-Reflective Sensors Embedded in a Head-Mounted Display", in IEEE Transactions on Visualization and Computer Graphics, Vol: 29, No: 10, pp. 4124-4139.

13. Jianing Teng; Dong Zhang; Wei Zou; Ming Li; Dah-Jye Lee, Year: 2023, "Typical Facial Expression Network Using a Facial Feature Decoupler and Spatial-Temporal Learning", in IEEE Transactions on Affective Computing, Vol: 14, No: 2, pp. 1125-1137.