Bulletin of Electrical Engineering and Informatics
Vol. 11, No. 3, June 2022, pp. 1350~1357
ISSN: 2302-9285, DOI: 10.11591/eei.v11i3.3722

Facial expression recognition using HOG and LBP features with convolutional neural network

Nadia Shamsulddin Abdulsattar, Mohammed Nasser Hussain
Department of Computer Engineering, Faculty of Engineering, Al-Iraqia University, Baghdad, Iraq

Article Info

Article history:

Received Feb 23, 2022
Revised Apr 29, 2022
Accepted May 18, 2022

Keywords:

CNN
FER
HOG
LBP

ABSTRACT

In computer vision, automatic facial expression recognition (FER) continued a difficult and interesting topic. The majority of extant techniques are based on traditional features descriptors such as local binary pattern (LBP) and histogram of oriented gradient (HOG), in which the classifier's hyperparameters are tailored to produce the best recognition accuracies across a single database or a small set of similar databases. This paper integrates the power of deep learning techniques with the LBP and HOG. The LBP and HOG are estimated from each image in the dataset. The resulting dataset is applied to a convolutional neural network (CNN). The architecture of this CNN constitutes three convolutional layers and three max-pooling layers. The output layers involve BatchNormalization, three dense layers, and two dropout layers. The proposed architecture is validated on the extended cohn-kanade dataset (CK+). We obtain improvement in the accuracy of the CNN model from 0.9593 to 0.967 and 0.975 after using the LBP and HOG respectively.

Corresponding Author:

Nadia Shamsulddin Abdulsattar
Department of Computer Engineering, Faculty of Engineering, Al-Iraqia University
Baghdad, Iraq
Email: nadia.sh.a.sattar@students.aliraqia.edu.iq

1. INTRODUCTION

Automatic facial expression detection has vital applications in a wide range of fields, including human-computer interaction (HCI), although it is a difficult but fascinating topic. Facial expression recognition (FER) is gaining more and more attention because it is extensively used in many domains such as security [1], health care [2], human-robot interaction [3], smart living [4], the safety of the driver, animation [5], and e-learning. The primary goal of FER is to classify facial expressions into a variety of emotions, including surprise, contempt, sad, disgust, anger, fear, and happy. The local binary pattern (LBP) is a strong feature to represent texture. The activities of the corresponding muscles on the face produce varied textures when a facial expression appears. The histogram of oriented gradient (HOG) is employed to extract features from the image. In FER, the HOG and LBP have proven to be effective descriptors and give good results [6].

Deep learning techniques have been extensively employed and confirmed to be effective in a variety of application domains. Convolutional neural network (CNN) is a type of well-known deep learning algorithm that learns directly from the input without the need to extract human features; this property gives it a competitive advantage over traditional networks. Recently, CNN has been efficient in the FER [7], [8]. A conventional CNN model consists of several layers, each of which performs a specific function. The convolution layer is responsible for estimating features, the pooling layer for reducing the size of the preceding layer's features, and the fully-connected layer (dense layer) for eliciting high-level features and predicting the model's output. As well, the core structure of CNN uses several activation functions including ReLU, Sigmoid, and Tanh.

In this paper, we implemented a new CNN model to solve the problem of FER. The HOG/LBP are employed with the proposed model to improve the accuracy using the cohn-kanade dataset (CK+) dataset. Oliver *et al*. [9] suggested a way to decrease false positives for the recognition of mammography images employing LBP. The results show that LBP features are efficient at reducing false positives across a wide range of mass sizes and that the LBP outperformed the current methods. Khandait *et al*. [10] discovered that the height and width of the face sections have proven to be obvious features in facial expression recognition. Depend on the elements of the face and the movements of the muscles. The results gave a good performance and the accuracy was 95.26% on the JAFFE dataset. Liu *et al.* [11] suggested a deep architecture called an “AU-Aware” comprising three sequential modules. The experiments employed three datasets MMI, CK+, and SFEW. The results demonstrate that the features produced by “AU-Aware” are good and competitive with features using HOG, SIFT, Gabor, and LBP. Liu *et al.* [12] utilized the principal component analysis (PCA) to minimize the dimensionality of a huge number of features combined with HOG and LBP. The results gave a good performance on JAFFE and CK+ datasets.

Kumar *et al*. [13] have published a comparative study between LBP, deep features, and bag-of-visual-words (BoVW) for the classification of histopathological images. The obtained accuracy was 90.62% for using LBP and 94.72% for using deep features, While BoVW gave the best accuracy of 96.50%. Alhindi *et al.* [14] compared three classification models and employed one of the following feature extractors: HOG, LBP, and deep features from a pre-trained model (VGG19). The experiments are performed on the KIMIA Path960 dataset. The results gave a good accuracy of 90.52% using LBP. Nigam *et al.* [15] implemented feature extraction by recovering the HOG feature in the DWT domain and an SVM employed for recognition of the expression. The suggested method employed three datasets JAFFE, CK+, and yale face. The results of the suggested approach are efficient for FER and are better than the existing approaches.

Xie *et al.* [16] incorporated scattered features into deep feature learning to improve the ability of generalization of a CNN to recognize facial emotions. The results revealed that the suggested method produced good performances on four datasets CK+, Oulu-CASIA, FER2013, and MMI. Zhang *et al*. [17] utilized the shape geometry of the face image by suggesting an end-to-end deep learning model. The suggested model depends on a generative adversarial network (GAN) and employs three datasets Multi-PIE, BU-3DFE, and SFEW. The results of three datasets display the effectiveness of the model. Sharifnejad *et al.* [18] compared the performance of histograms of oriented gradients, local binary pattern methods, and their combination on several regions of a face image. The results produced an accuracy of 95.33% using three regions of the face: the eyes, mouth, and nose employing the CK dataset.

2. METHOD

In this work, we suggested a new CNN model and employed the HOG and LBP to improve the accuracy of this model. The suggested model contains three convolutional layers and three max-pooling layers. The output layers involve BatchNormalization, three dense layers, and two dropout layers. The input of the model is 224×224 pixels. The convolutional layers employ 32, 64, and 128 filters of size 3x3. The type of padding is ‘same’, and the activation function is ReLU. The max-pooling layers employ kernels of size 2x2. The new output layers comprise three dense layers and two dropout layers. The activation of the first two dense layers is implemented using the ReLU function while the activation for the last dense layer is the softmax function. The number of neurons for the three dense layers is 128, 64, and 7, respectively. On the other hand, the percentage of dropout is 0.5. Figure 1 clarifies the stages of the proposed system for recognizing facial expressions.

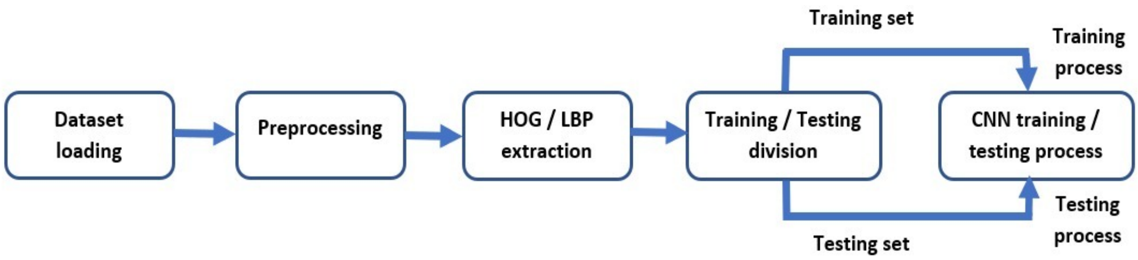


Figure 1. The proposed FER system

The first stage involves the preprocessing process in which each image in our selected FER dataset (CK+) is resized to 224×224 pixels. Pre-processing can be performed before the feature extraction procedure. The second stage explains the use of (HOG/LBP) that are employed in the proposed model. In the third stage, we train and test three models separately: the proposed CNN, LBP with CNN, and HOG with CNN. The fourth stage involves computing the classification results, which involves assigning each image in the testing set to one of seven expressions (surprise, fear, contempt, disgust, angry, happy, and sad).

HOG: It describes features by computing the occurrence of gradient orientation appearing in a certain area of an image. HOG divides the image into various cells and computes the gradients over them. It was suggested by Dalal and Triggs [19]. Suppose the intensity (grayscale) function represent (I), which describes the image to be analyzed. The image is partitioned into cells of size K×K pixels (as depicted in Figure 2(a)), for the x-axis: $G_x = I(x+1, y) - I(x-1, y)$ and for the y-axis: $G_y = I(x, y+1) - I(x, y-1)$, the magnitude of the gradient is calculated as:

$$M(x,y) = \sqrt{{G_x}^2 + {G_y}^2} \tag{1}$$

And the gradient's Q (x, y) orientation is calculated in each pixel (see Figures 2(b) and (c)) as:

$$Q(x,y) = tan^{-1}\frac{G_y}{G_x} \tag{2}$$

An M-bins histogram of orientations is created by computing the orientation of all pixels and accumulating it (see Figures 2(d) and (e)). To generate the final features vector, all cell histograms are concatenated (see Figure 2(f)) [20]. HOG is utilized in image processing and computer vision. It performs good results on feature detection of FER [21].

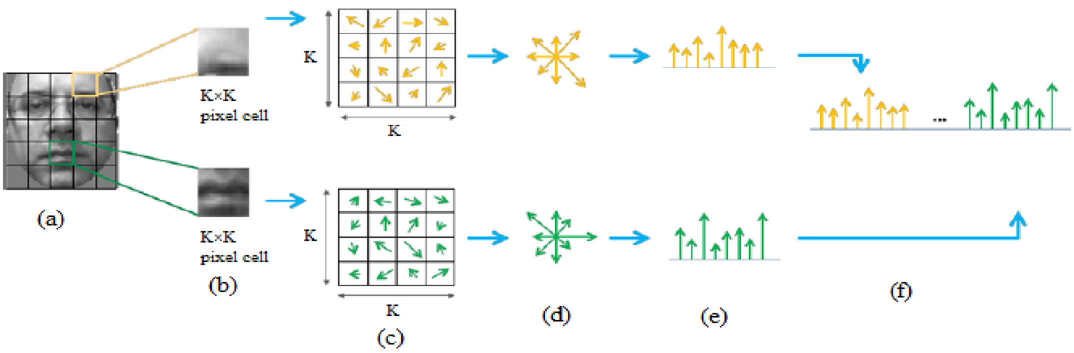


Figure 2. The process of extracting HOG features [20], (a) original image, (b) image cell, (c) gradient orientation, (d) orientation accumulation, (e) cell histogram, and (f) histogram concatenation

LBP: It is a texture descriptor, so it is a useful feature for image texture classification [22]. It was presented by Ojala *et al.* [23]. LBP names the pixels in an image by specifying the vicinity of each pixel with the central value and treating the result as a binary number. Concatenating all the binary codes in a clockwise direction, starting with the top-left one, yields a binary number, and the associated decimal value is employed for labeling [24]. In decimal form, the resulting (LBP code) is as:

$$LBP_{m,b} = \sum_{m=0}^{m-1} y(j_m - j_i)2^m \tag{3}$$

Where j_m represents the intensity value of the neighboring pixel and j_i represents the intensity value of the central pixel. The m represents the number of pixels in a circular neighborhood and b represents the radius to the circular neighborhood. The threshold function is defined as:

$$y(v) = \{ \begin{smallmatrix} 1 & if\ v>=0 \\ 0 & if\ v<0 \end{smallmatrix} \tag{4}$$

For further analysis, these codes' histogram is then employed. Figure 3 shows the LBP encoding method. Firstly, the image is encoded as an LBP image, then separated into patches, with one LBP histogram derived from each patch. The final feature vector is created by concatenating the LBP histograms of all patches [25]. LBP could be employed to describe the expressions of the facial, which gives good results in face recognition [26].

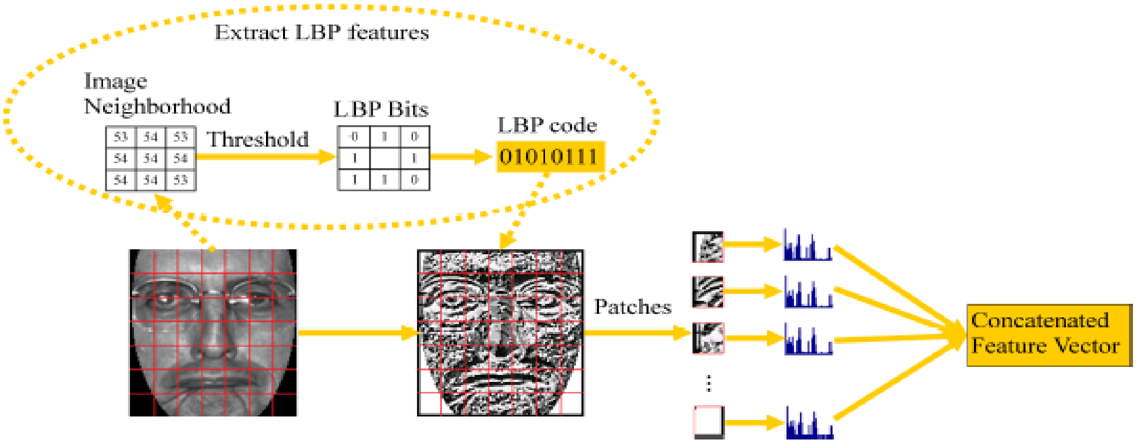


Figure 3. The LBP encoding method [25]

3. RESULTS AND DISCUSSION

3.1. Database

CK+: the extended cohn-kanade, this database involves 593 sequences of images from 123 subjects, 327 images are labeled with one of the seven main facial expressions (surprise, fear, contempt, disgust, angry, happy, and sad). This dataset is recorded under controlled conditions in the laboratory. The most typical data selection strategy for static-based methods extracts the final one to three frames of each sequence with peak creation and the beginning frame (neutral face) [27]. Figure 4 depicts a set of training samples chosen from the CK+ dataset.

Figure 4. A set of training samples was chosen from the CK+ dataset

3.2. Experimental results

This section devotes to study the effect of using feature descriptors (LBP and HOG) on the accuracy of FER when used with a suggested model. To achieve this, we explore the efficiency of three models: CNN alone, HOG+CNN, and LBP+CNN. The test accuracy of the proposed CNN model alone, reaching 0.9593. After that, we trained the LBP technique with CNN, and we get an accuracy of 0.9675. In the third experiment, we combined the HOG technique with CNN. The accuracy further increased to 0.9756. The validation accuracy and loss are outlined in Table 1.

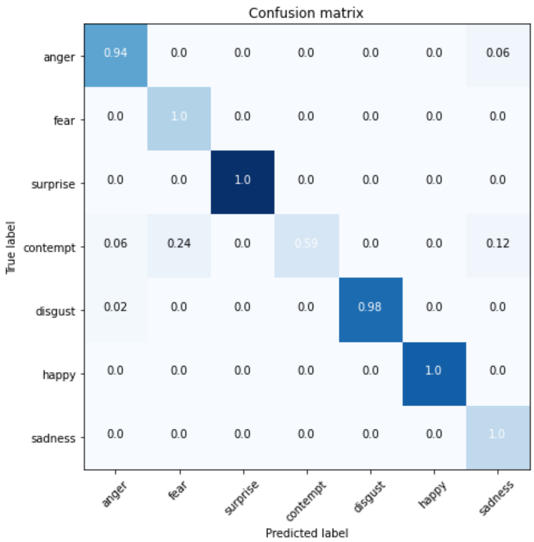
Table 1. The validation accuracy and loss for three models

Model	Val_Accuracy	Val_Loss
CNN model	0.9593	0.8281
LBP+CNN model	0.9675	0.1154
HOG+CNN model	0.9756	0.7213

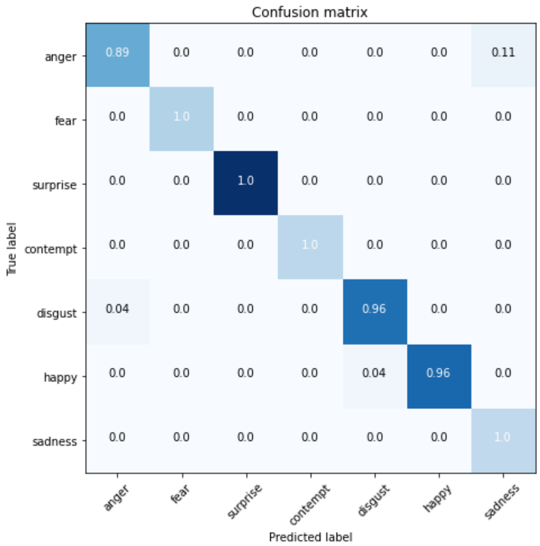
The loss of using the LBP technique is less than the HOG technique, as shown in Table 1. We trained the proposed model employing Adam optimization algorithm by splitting the CK+ dataset (75% used for training and 25% for testing). The other parameters employed in the training process are depicted in Table 2. The models' behavior with accuracy and loss of the CNN, LBP + CNN, HOG + CNN are demonstrated in Figures 5(a), (b), and (c) (in appendix) respectively. Furthermore, the confusion matrices of the CNN, LBP + CNN, HOG + CNN models are depicted in Figures 6(a), (b), and (c) respectively.

Table 2. The parameters that the model employed for training

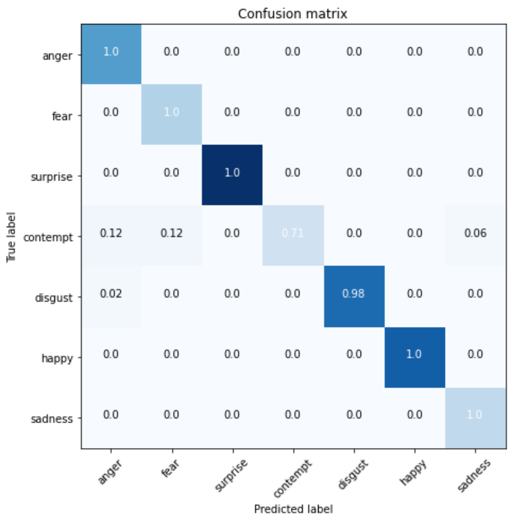
Parameters	The value
Epochs	25
Optimizer	Adam
Batch size	64
Learning rate	0.0001



(a)



(b)



(c)

Figure 6. The confusion matrix of three models (a) CNN model, (b) LBP+CNN model, and (c) HOG+CNN model

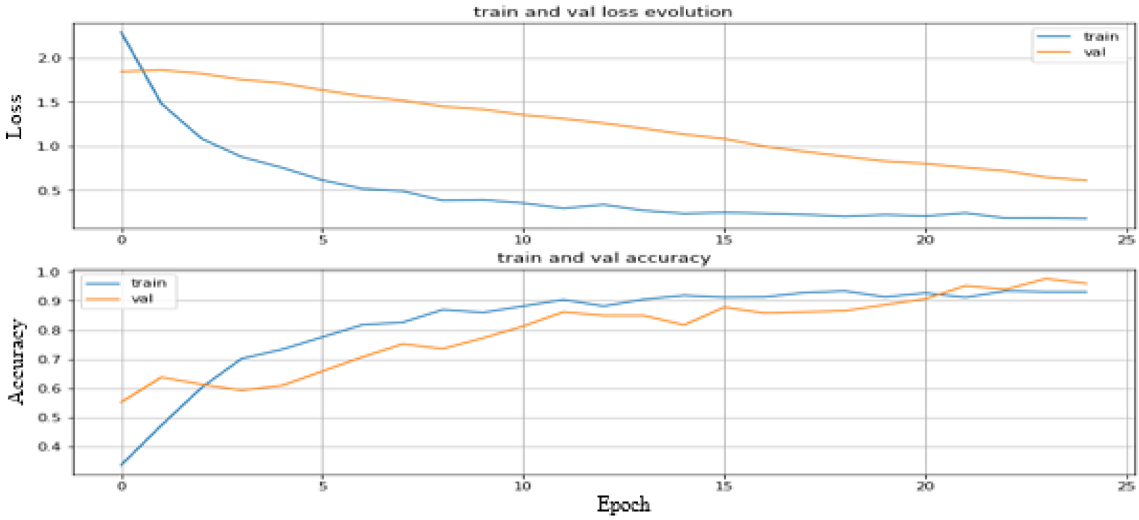
4. CONCLUSION

In this work, we presented a new neural network architecture to recognize facial expressions. In addition, we used HOG, LBP techniques with the CNN to increase the accuracy of the proposed model. The initial phase in our model is feature extraction. The LBP is known as an efficient feature for analyzing facial images (texture descriptor). In a grayscale range, HOG depicts the necessary features from an image. We trained three models: CNN, LBP+CNN, and HOG+CNN. All the assessments were implemented on the CK+ dataset. The HOG+CNN achieved a high accuracy of 0.9756, while the other models also achieved a good accuracy of 0.9675 for using LBP and 0.9593 for using a new CNN model. The results are elementary and may be improved in the future as our efforts to acquire better results continue to look into additional factors that may influence accuracy.

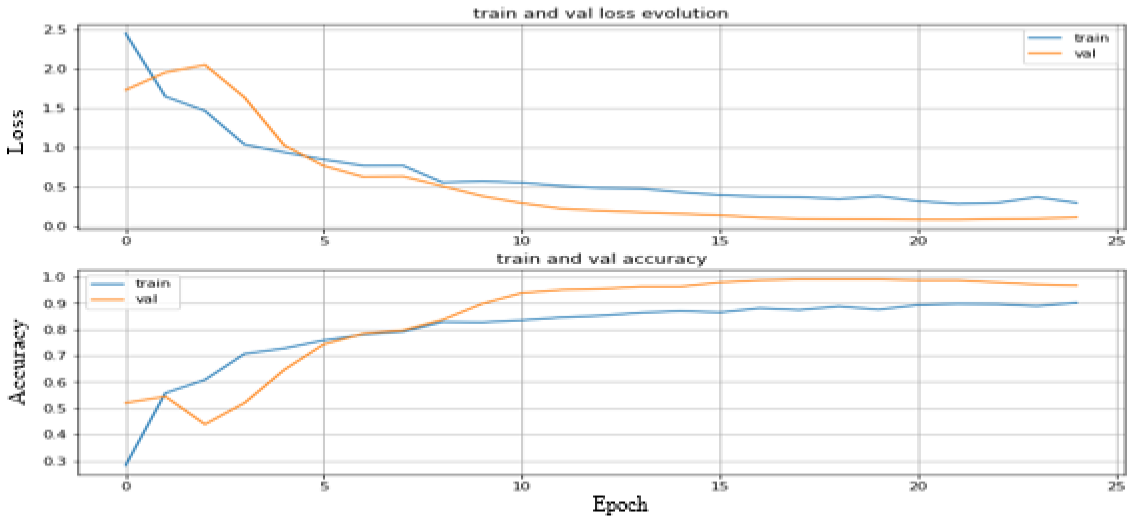
ACKNOWLEDGMENT

We would like to thank the efforts of Al-Iraqia and Mustansiriyah universities for their efforts in supporting this work.

APPENDIX

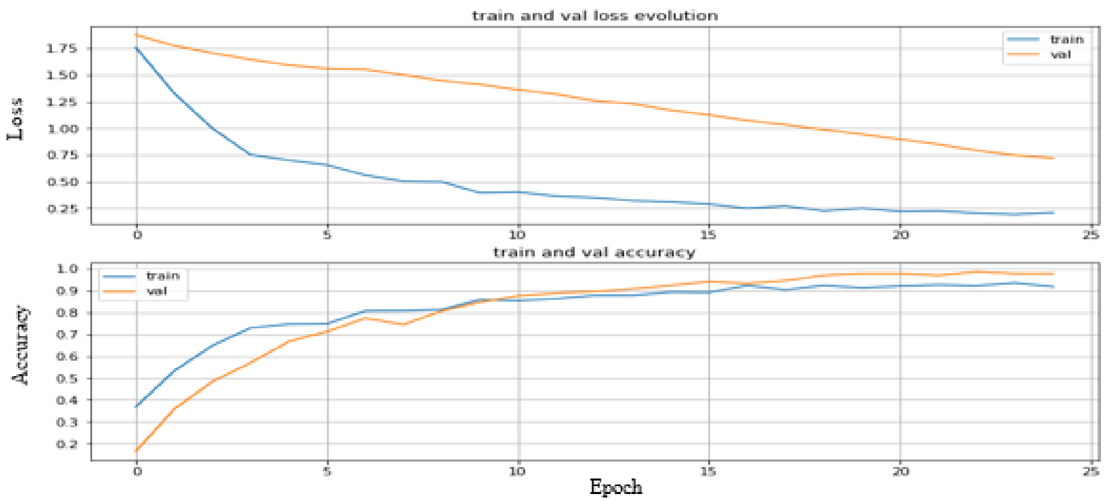


(a)



(b)

Figure 5. The development of the model performance as the epoch progresses in three models (a) CNN model, (b) LBP+CNN model



(c)

Figure 5. The development of the model performance as the epoch progresses in three models (c) HOG+CNN model (continue)

REFERENCES

[1] S. T. Saste and S. M. Jagdale, "Emotion recognition from speech using MFCC and DWT for security system," *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA),* 2017, pp. 701-704, doi: 10.1109/ICECA.2017.8203631.

[2] A. F. Caballero *et al*., "Smart environment architecture for emotion detection and regulation," *Journal of Biomedical Informatics*, vol. 64, pp. 55–73, Dec. 2016, doi: 10.1016/j.jbi.2016.09.015.

[3] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Image and Vision Computing,* vol. 28, no. 5, pp. 807–813, May 2010, doi: 10.1016/j.imavis.2009.08.002.

[4] Y. Yaddaden, A. Bouzouane, M. Adda, and B. Bouchard, "A new approach of facial expression recognition for ambient assisted living," in *ACM International Conference Proceeding Series*, no. 14, pp. 1-8, Jun. 2016, doi: 10.1145/2910674.2910703.

[5] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, "Modeling stylized character expressions via deep learning," In *Asian Conference on Computer Vision*, 2016, pp. 136–153, doi: 10.1007/978-3-319-54184-6_9.

[6] M. Ghorbani, A. T. Targhi and M. M. Dehshibi, "HOG and LBP: Towards a robust face recognition system," *2015 Tenth International Conference on Digital Information Management (ICDIM)*, 2015, pp. 138-141, doi: 10.1109/ICDIM.2015.7381860.

[7] A. R. Garcia, M. Elshaw, A. Altahhan, and V. Palade, "Deep Learning for Emotion Recognition in Faces," In *Artificial Neural Networks and Machine Learning – ICANN 2016*, vol. 9887, pp. 38-46, Aug. 2016, doi: 10.1007/978-3-319-44781-0_5.

[8] A. Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning," *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 2017, pp. 745-750, doi: 10.1109/AICCSA.2017.124.

[9] A. Oliver, X. Llad, J. Freixenet, and J. Mart, "False Positive Reduction in Mammographic Mass Detection Using Local Binary Patterns," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2007*, vol. 4791, pp. 286–293, 2007, doi: 10.1007/978-3-540-75757-3_35.

[10] S. P. Khandait, R. C. Thool, and P. D. Khandait, "Automatic facial feature extraction and expression recognition based on neural network," in *International Journal of Advanced Computer Science and Applications,* vol. 2, no. 1, pp. 113-118, Apr. 2012, doi: 10.48550/arXiv.1204.2073.

[11] M. Liu, S. Li, S. Shan and X. Chen, "AU-aware Deep Networks for facial expression recognition," *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG),* 2013, pp. 1-6, doi: 10.1109/FG.2013.6553734.

[12] Y. Liu, Y. Li, X. Ma, and R. Song, "Facial expression recognition with fusion features extracted from salient facial areas," *Sensors (Switzerland)*, vol. 17, no. 4, pp. 1-18, Mar. 2017, doi: 10.3390/s17040712.

[13] M. D. Kumar, M. Babaie, S. Zhu, S. Kalra and H. R. Tizhoosh, "A comparative study of CNN, BoVW and LBP for classification of histopathological images," *2017 IEEE Symposium Series on Computational Intelligence (SSCI),* 2017, pp. 1-7, doi: 10.1109/SSCI.2017.8285162.

[14] T. J. Alhindi, S. Kalra, K. H. Ng, A. Afrin and H. R. Tizhoosh, "Comparing LBP, HOG and Deep Features for Classification of Histopathology Images," *2018 International Joint Conference on Neural Networks (IJCNN),* 2018, pp. 1-7, doi: 10.1109/IJCNN.2018.8489329.

[15] S. Nigam, R. Singh, and A. K. Misra, "Efficient facial expression recognition using histogram of oriented gradients in wavelet domain," in *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28725–28747, May 2018, doi: 10.1007/s11042-018-6040-3.

[16] W. Xie, X. Jia, L. Shen, and M. Yang, "Sparse deep feature learning for facial expression recognition," *Pattern Recognition*, vol. 96, no. 1, Jul. 2019, doi: 10.1016/j.patcog.2019.106966.

[17] F. Zhang, T. Zhang, Q. Mao and C. Xu, "Geometry Guided Pose-Invariant Facial Expression Recognition," in *IEEE Transactions on Image Processing*, vol. 29, pp. 4445-4460, 2020, doi: 10.1109/TIP.2020.2972114.

[18] M. Sharifnejad, A. Shahbahrami, A. Akoushideh, and R. Z. Hassanpour, "Facial expression recognition using a combination of enhanced local binary pattern and pyramid histogram of oriented gradients features extraction," in *IET Image Processing*, vol. 15, no. 2, pp. 468-478, 2021, doi: 10.1049/ipr2.12037.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[20] P. Carcagnì, M. del Coco, M. Leo, and C. Distante, "Facial expression recognition and histograms of oriented gradients: a comprehensive study," *SpringerPlus*, vol. 4, no. 1, pp. 1-25, 2015, doi: 10.1186/s40064-015-1427-3.

[21] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial expression recognition based on facial components detection and hog features," In *Proceedings of the International Workshops on Electrical and Computer Engineering Subfields*, 2014, pp. 884–888.

[22] I. Jung and I. Oh, "Local Binary Pattern-Based Features for Text Identification of Web Images," *2010 20th International Conference on Pattern Recognition*, 2010, pp. 4320-4323, doi: 10.1109/ICPR.2010.1050.

[23] T. Ojala, M. Pietikainen and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," *Proceedings of 12th International Conference on Pattern Recognition*, 1994, pp. 582-585 vol.1, doi: 10.1109/ICPR.1994.576366.

[24] D. Huang, C. Shan, M. Ardabilian, Y. Wang and L. Chen, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 765-781, Nov. 2011, doi: 10.1109/TSMCC.2011.2118750.

[25] J. Ren, X. Jiang, and J. Yuan, "Face and facial expressions recognition and analysis," In *Context Aware Human-Robot and Human-Agent Interaction*, pp. 3-29,2015, doi: 10.1007/978-3-319-19947-4_1.

[26] C. Shan, S. Gong, P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," In *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, May 2009, doi: 10.1016/j.imavis.2008.08.005.

[27] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

BIOGRAPHIES OF AUTHORS

Nadia Shamsulddin Abdulsattar has a B.Sc. degree in computer engineering from the Technology University in Baghdad. She lectures in the image processing laboratory at Mustansiriyah University. Currently working on obtaining a master's degree in computer engineering from the Iraqia University. My current research interests are image processing, computer vision, deep learning. She can be contacted at email: nadia.sh.a.sattar@students.aliraqia.edu.iq.

Mohammed Nasser Hussain received the B.Sc. degree in Electrical Engineering – College of Engineering – Al_Mustansirya University. (July, 1997), the MSc. Degree in Computer and Control – Electrical Engineering Department College of Engineering – University of Baghdad, Speaker Recognition (Identification & verification) based upon Phonemes Using Wavelet Packet Transform. (October, 2000), and the PhD. Degree in Computer communication & Control – Electrical Engineering Department – College of Engineering – University of Baghdad, in the implementation of one- and two-dimensional Wavelet and Wavelet Packet Transforms on Using Field Programmable Gate Array (FPGA) electronic Platforms (June 2004). He can be contacted at email: mohammed_alturfi@yahoo.com.