Deep Learning Methods for Facial Expression Recognition

Conference Paper · October 2019

DOI: 10.1109/ICOM47790.2019.8952056

2 authors, including:

Chowdhury Mohammad Masum Refat
International Islamic University Malaysia
6 PUBLICATIONS **1** CITATION

SEE PROFILE

2019 7th International Conference on Mechatronics Engineering (ICOM)

Deep Learning Methods for Facial Expression Recognition

Chowdhury Mohammad Masum Refat
Dept. of Mechatronics Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
chowdhury.masum@live.iium.edu.my

Norsinnira Zainul Azlan
Dept. of Mechatronics Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
sinnira@iium.edu.my

***Abstract*—Deep learning is very popular methods for facial expression recognition (FER) and classification. Different types of deep learning algorithms have been used for FER such as deep belief network (DBN) and convolutional neural network (CNN). In this paper, we analyze various deep learning methods and their results. We have chosen Deep convolutional neural network as the best algorithms for facial expression detection and classification. In our study, we have tested the algorithm using Japanese Female facial expressions database (JAFFE) datasets by anaconda software. The deep convolution neural networks with JAFFE datasets accuracy rate around 97.01%.**

Keywords—Convolutional neural networks (CNN), deep belief network (DBN), Facial expression recognitions (FER), facial expression classification.

I. INTRODUCTION

One of the great psychologist Mehrabian said that " 55% message was conveyed by the facial expression in human communications while language transfer just 7%"[1]. Therefore, the expression in daily communication of humankind is important in regular human communication. It is a very important process to convey emotional information with people. Human emotion data processing becomes the computer requirement to understand this data in various application such as social robots, security monitoring, and computer interactive graphical games. For this reason, a lot of research[1] was done or processing with facial expression recognition (FER). Nowadays, FER is one of the most popular topics among research.[2]

Facial expression recognition problem is a traditional problem in computer vision area. In general, the facial expression represents the feelings of the people [2] and the facial expression recognition(FER) has a lot of applications such as human-machine interaction which had used in social robots, security monitoring and computer interactive graphical games. [3] It is used in behavioral science to provide social information (Origin, age and gender) and used in medical science for pain monitoring, treatment of mental retardation, depression and anxiety.

In this paper, we analyze the FER by using deep learning method from previous study [2]. Moreover, when we communicate with a different person for different purpose facial expression was a very important part. Nowadays for facial expression identification process the deep learning doing a great job. In general, among two deep learning methods that are popular nowadays include deep belief network (DBN) and convolutional neural network (CNN), are usually used to solve recognition problems [4]. Therefore, we review the previous study on both for facial expression recognition.

Most FER systems involve three main steps: face image pre-processing, feature extraction, and classification. Deriving an effective and robust facial representation from raw face images is extremely important for FER. Extracted facial representations can be grouped into two categories: action unit (AU)-based methods and appearance-based methods. Nowadays different methods and algorithms had applied for facial expression recognition system.[5] To resolve the problem of having an input including multiple well-known approved face datasets, Ali [5] formed a deep neural network architecture that takes recorded facial images as input and classifies all the data. To recognize the facial expressions over different facial views, Tong [6] was proposed a novel deep neural network (DNN) directed method in which called scale invariant feature transform (SIFT) which correspond to the set of position points are first obtained from each facial image. Many kinds of research have also been conveyed using engineered features (e.g. Gabor and Scale Invariant Feature Transform (SIFT)) where the classifiers hyperparameters are attuned to give best recognition efficiencies across in the single database.[6] Recently, many works on facial expression recognition successfully use Convolutional Neural Networks (CNNs) for feature classification and extraction. CNNs found in literature and the special type of multilayer perceptron (MLP) that concentrates on the local relationship between pixels by using perceptive fields.[7] They also are shown to achieve high recognition standards in different image recognition assignments these methods. This paper presents the review facial expression and technique for face detection feature extraction, classification using deep learning method. In this study we choose to implement deep convolution neural network method on JAFFE (Japanese Female Facial Expression) database. The simulation results show the train sample time, classification and accuracy.

Many facial expression recognition methods represent facial image by high-dimensional over-complete face descriptors, followed by shallow models. Mliki et al. proposed a method which is able to monitor the intensity variation of facial expression and to reduce the classification confusion between facial expressions classes. [9] The researchers used Vector field convolution to segment the facial feature contours. Lyons et al. applied the Gabor feature to calculate the expression characteristics by combining elasticity graph matching and linear analysis [9]. Cotrel et al. proposed a wavelet-based face recognition method that is robust against the facial position and light variation for real-time application.

Tong proposed a novel deep neural network (DNN) method which has a scale invariant feature transform (SIFT) feature corresponding to a set of landmark point are first extracted from each facial image [9]. Wang et al. proposed expression recognition method-based evidence theory and

978-1-7281-2971-6/19/$31.00 ©2019 IEEE

local texture where the facial image is divided into regions with important recognition features and the local binary patterns textural feature of the regions are extracted.[10]. The works in [1][2][3][4] proposed novel boosted deep belief network (DBN) to perform the three training stages iteratively in a unified loopy framework. Kahou et al. [10], in their work conducted two experiments using CNN's, the CNN with per-train standard wild facial expression dataset worked in the first experiment, and Toronto faces dataset in the second experiment. 6 universal facial expressions. Dennis H et. al. applied the facial image into two channels of CNN. The information from the two networks was combined which result in 94.4% recognition accuracy. Burkert et al. proposed a CNN architecture for facial expression recognition. Mollahosseini [11] also proposed a deep neural network architecture to address the facial expression recognition problem across multiple well-known standard face datasets.

Zhang et al.[11] proposed a novel deep neural network: DNN drive facial feature learning model which employs several layers to characterize the corresponding relationship between the (Scale-invariant feature transform) SIFT feature vectors and their corresponding high-level semantic information.

Facial expression recognition is one of the challenging topics for the researcher and is still now it was developing day by day. With the complexity of FER and the nonrigid of the face, there is my difficulty such as-

- Need to Improve the FER algorithm robustness. Nowadays, most research expression database builds up in the lab, but it is true that the facial characters in real life constantly changing and more complex.
- It was variations in the community, culture and geography.
- Expression was different for different gender
- It was quite difficult to classify in the real-time processing data.

This paper is organized as follows: Section II gives a background on existing works done by previous researches using deep learning for facial expression recognition, Section-III describes all the proposed architecture and method shortly, Section-IV reports the simulation results and discussion. Lastly, Section-V draws the conclusion.

II. FEATURE MINING METHODS

Various facial expression recognition methods proposed in the literature. However, the most popular methods include deep belief network and deep convolution neural network. In the paper, we present both methods. In this section, we are present the research outputs them venous for facial expression algorithm. In this time, there are different kinds of feature mining methods, which are divided into different sectors such as texture features, geometric shapes and the combination of the two.

A. Texture Features Based Method

Local Binary Patterns (LBP) is a local operator which is used for face recognition, image texture analysis and facial expression recognition.

Here is the equation of LBP,

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p,\ s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

Where g_c is the value of the center pixel (x_c, y_c) for local neighborhood and g_p is the gray scale value P have equally spaced pixels on the circle of radius R. Local Directional Patterns (LDP) operator is proposed for extract facial texture features. The LDP feature is subjected by LDP variance. The very important direction of LDP is calculated by Principle Component (PCA). The proposed Gradient Turning Pattern calculates the texture data in the local area by measuring the gradient path to form the facial illustration. Lyons et al. [2] using the Gabor feature for calculate the expression features, mixing with elasticity graph alike and linear analysis. Then the high-dimensional Gabor feature is a difficulty to quickly recognize facial expressions [12]. The image is divided into local subregions to extract multiscale Gabor features describes the topology configuration of human visual. The algorithm is combined Gauss Laguerre wavelet and face orientation to represent the expression feature, which is them extracted the texture feature. The latter processed geometric features.[2]

B. Geometric Features Based Method

Active Shape Model (ASM) and Facial Action Units (FAUs) 2] are the two most typical methods for geometric features.

$$S_i^k = \{(x_0, y_0)_i^k, (x_1, y_1)_i^k, \ldots, (x_N, y_N)_i^k\} \quad (2)$$

Here (x, y) is the transformed landmark coordinate position. ASM builds geometric features by separating facial feature objects. The geometric features were obtained by the improving ASM and using the triangle features in the prototypical and the method achieved very good results [13]. Besides, texture and shape, Active Appearance Mode (AAM) also used statistical model for modeling the face, to identify the face and analyze the expression [13]. The facial expression recognition algorithm is proposed, combining AAM and ASM to were used in the study improve the problem of feature objects convergence in AAM. According to FACS (facial action coding system), another important geometric feature by analyzing the constant facial points and instantaneous characteristics. In this image, different features were tracked and modeled to recognize facial expression. Combining through multiscale feature extraction and multicore absorbing algorithm, the contour of the face character was defined [14].

C. Multi feature Fusion Based Method

Multi-feature fusion is very well known methods for facial expressions recognition (FER) and the different features in many algorithms are combined to describe facial expressions. The combination of Gabor and LBP has been [4] used in, and the minimized error rate and the feature dimension is selected by genetic algorithm (GA). Li et al. Used SIFT (Scale-invariant feature transform) and PHOG (Pyramid Histogram of Oriented Gradient) to calculate local texture and shape information [14]. TC (Topographic Context)[15] is used as a featured operator to describe the static facial expression feature. The Haar-like feature was used to characterize facial

expression changes over time and was encoded to form a dynamic feature according to the binary pattern [15]. According to the main direction extracted in the video sequence and the average optical flow characteristics in the face block, MDMO (Main directional mean optical flow feature, MDMO) was proposed in[15]. Table 1 shows that previously works classification and accuracy.

TABLE 1 : PREVIOUS WORK CLASSIFICATION AND ACCURACY

Literature	***Number of facial expressions***	***Dataset***	***Rate of recognition***
[1]	6	CK+ (Cohn-Kanade) MUG (Multimedia Understanding Group), RAFD (Radboud Faces Database)	94.00%
[3]	7	JAFFE (Japanese Female Facial Expression) , CK+	95.71%
[4]	7	CK+	86.74%
[6]	7	KDEF (Karolinska Directed Emotional Faces)	93.07%
[8]	7	JAFFE, CK+	80.30%
[10]	7	CK+, JAFFE	87.01%
[12]	7	FER-2013 (Facial expression recognition)	71.69%
[14]	7	CK+	91.02%
[16]	6	Ck+, JAFFE, FACES	83.01%
[17]	6	FER-2013	93.07%

III. ARCHITECTURE AND METHODS

A. Deep learning Facial Recognition General Algorithm

Nowadays nonlinear neural network architecture has multiple hidden layers and deep learning helps us to simulates the human brain learning and analysis. For reconstructing the input data by layer, the feature design of the sample in the real space is converted into a more complex feature space. So it is very important to characterize the data set and imitate the human brain to understand the data such as sound, image and text. In this method , facial expression recognition has been done by two different methods which are deep belief network (DNN) and the convolution neural network (CNN).[16][17]

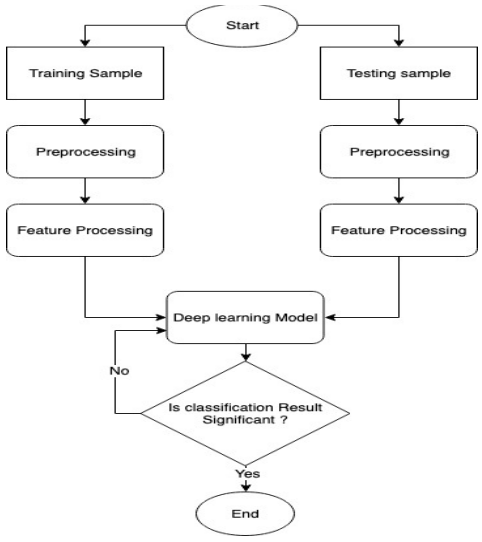

Fig. 1: Block diagram for FER based on deep learning

B. Convolution Neural Network Method

The convolution neural network was most popular for facial expression recognition [1][3][4]. Different types of facial expression recognition have been developed based on a convolution neural network that is capable to predict the local facial feature object detection and CNN can do multitask deep neural network. A novel method was developed using the combination of optical flow and DNN stacked sparse autoencoder analyze video image effectively and reduce the impact of individual character variation on FER. Some researcher combined K-nearest neighbor's (KNN) and region of interest (ROI) for making a fast and simple improved method which is also called ROI-KNN method for classifying the facial expression. Previously deep neural networks method decreases the measurement error rate and lacking data but ROI-KNN method solved this problem [20]. Another most popular CNN architecture had two hard-coded feature extraction such as convolution autoencoder (CAE) and standard CNN which was most significantly improved the accuracy and reduce the time for training data. 3D convolution neural network design consists of 3D inception-ResNet layers supported by LSTM unit commonly extracts spatial connections within facial images and it looks like temporary relation between video frames.[7] figure 2 shows the convolution neural network.

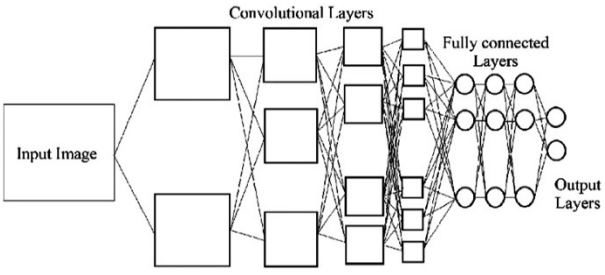

Fig. 2 : Convolution Neural Network [2]

C. Deep Belief Network Method

In machine learning technology, deep belief network (DBN) was a generative graphical design. Some researcher believes that it was alternatively deep neural network. The deep belief neural network has multiple hidden layers look

like a convolution neural network. The DBN architecture all the layers were connected to each other but not among units in each layer. When trained set the examples without administration, a deep belief network learns to probabilistically replace its inputs data. The layers then act as characteristic detectors. After this learning process the deep belief network trained with supervision to achieve classification the data. [2]

Deep Belief Network was accomplished significant results in facial expression recognition. Firstly the face was crop at various scales, and then the Histogram of Oriented Gradient (HOG) feature was extorted from this local crop, finally, deep belief network was identified within the DBN. The Gabor wavelet features were obtained from the face crops recognized by DBN, and the fused feature fusion was input in the automatic coding machine for expression identification . The strong classifier connected DBN and the AdaBoost method got a higher expression recognition accuracy standard for FER [25, 28].

DBN set of the model can increase the performance of expression classification. Modified Local Directional Patterns(MLDP-GDA), Generalized Discriminant Analysis method to extract salient features from base faces was applied to deep belief network for training various facial expressions.

Among all the algorithm, we have been chosen deep convolution neural network (CNN) because CNN train sample time, classification and accuracy are better than other algorithms. Figure 3 shows that deep belief network.

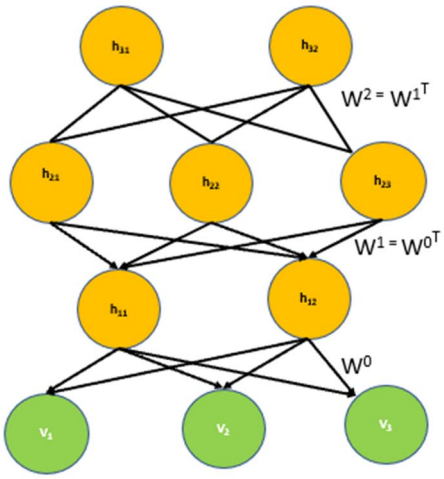

Fig. 3 : Deep Belief Network [18]

In this study, we have reviewed different paper which are related with facial Expression recognition using deep learning algorithms and we are find out which algorithm is best for facial expression recognition. We used three different parameters to determine for FER best algorithm which are train sample time, classification and accuracy. Overall, we found out that deep convolution neural network took less time to train the sample and also its classification success rate is better than other algorithm. In addition, deep convolution neural network accuracy also better than other categories. Therefore, we are using deep convolution neural network for our experiment and analysis the overall data why this algorithm is best for FER system. Table 2 show that difference between DNN and CNN train sample time, classification and accuracy.

TABLE 2 : DIFFERENCE BETWEEN DBN AND CNN

	Train sample time	*Classification success rate*	*Recognition rate %*
DBN [4]	61.736ms	6	72.78%
CNN [3]	4.3961ms	7	95.71%

IV. METHODOLOGY

Facial image classification important library OpenCV, TensorFlow, NumPy, Matplotlib, pylab, keras. Various types of dataset available now for facial expression recognition. In this paper used The JAFFE datasets. The JAFFE datasets have 213 facial expression image (grayscale) [3] and total facial expression categories are 7 such as smile, sad, anger, surprise, fear, disgusting and neutral posted by Japanese 10 females. In this training model designed by three max pooling layer and one full connected layer and this model follow the deep convolution neural network model. After training the model successful evaluating the result, if the results not acceptable retraining the model. Total data loss in the train time show in the data visualizing.

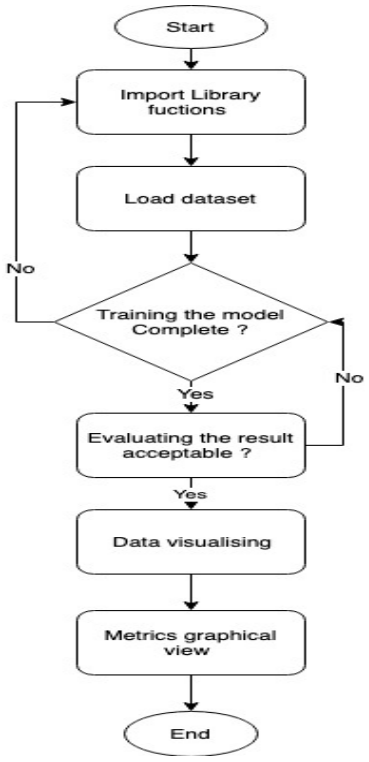

Fig. 4 : Flow chat from facial image classification by anaconda software

V. EXPERIMENTAL RESULT AND ANALYSIS

The proposed method performance was assessed by the total classification accuracy using by JAFFE datasets. The suggested system runs on Intel Core i5 processor cash 3.3GHz, RAM 8 GB and this method developed on Anaconda software. The JAFFE datasets have 213 facial expression image (grayscale) [3] and total facial expression categories are 7 such as smile, sad, anger, surprise, fear, disgusting and neutral posted by Japanese 10 females. Figure four shows that

sample of Japanese female image from the datasets. All the image size 256*256. Total image datasets were divided into 70:30 ratio, 70% for training purpose and 30 % for testing purpose.

Fig. 4 : Sample of Japanese female facial expression [19]

For trained this network for classification purpose, this dataset divided into two subsets. This two-subset ration is 70:30 among the overall dataset. Here 70 % datasets used for training purpose and the other 30% used for tasting purpose for classification facial expression detection. This CNN network was trained by the supervised backpropagation method. Figure7 shows that confusion matrix of facial recognition accuracy for the different expression used training weights information for getting the best accuracy. This system accuracy 97.1% for facial expressions recognitions.

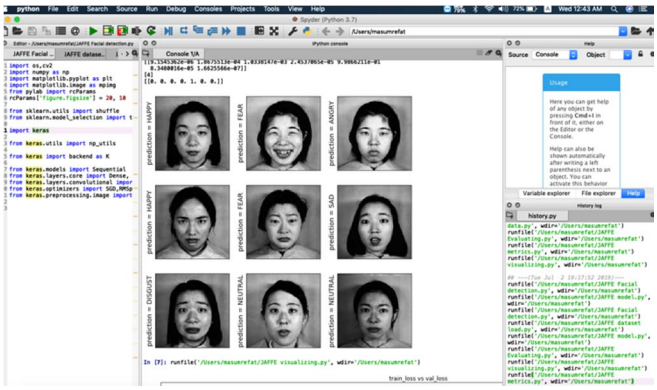

Fig. 5 : Facial expression classification by anaconda

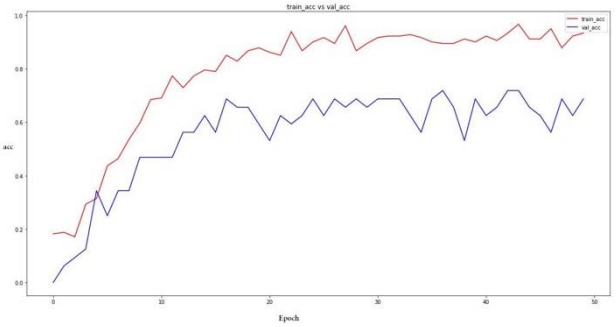

Fig. 6 : train acc vs valid acc for JAFFE datasets

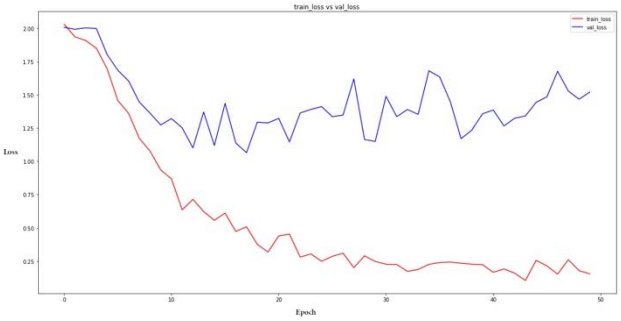

Fig. 7: Train loss Vs Valid loss for JAFFE datasets

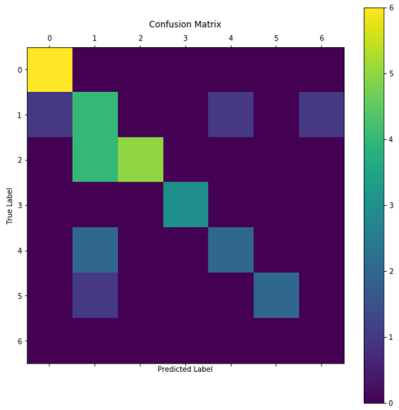

Fig. 8: Confusion matrix of facial recognition accuracy for the different expression

The algorithms average recognition is higher than other methods which methods are using in the different facial expression's classification research. Overall performance comparison for different methods for FER shows in table II.

TABLE 3 : PERFORMACE COMPARISON FOR DIFFERENT METHODS FOR FER

Methods	*Recognition Accuracy*
Two-channel [3]	94.04%
KNN (K-Nearest Neighbors) [9]	76.76%
VGG (visual geometry group) [14]	80.02%
Proposed Method	97.10%

VI. CONCULSION

In this paper, we had review different types of deep learning for facial expression recognition. We have chosen deep convolution neural network (CNN)method in our design model and compare with others method which is already done before using anaconda. This system accuracy 97.1% for total facial expression categories are 7 such as smile, sad, anger, surprise, fear, disgusting and neutral also average 22ms/step sample time need for train on 181 sample 32 validation sample. On the other, four-time higher sample time need for DNN. So, CNN sample time, classification and accuracy are better than other methods. Using anaconda software with python programing language the result is faster than others platform. From the result it can be seem that accuracy is acceptable. In this algorithm, the detected face was cropped, and input normalized in the CNN. Furthermore, we have plan to involve the application of developed system will be used for upper arm rehabilitation robot. deep learning is a useful algorithm for difficult feature object detection ability.

However, in the time all facial expression model based on deep learning took too much time to train the data. Hence, need to develop a new model which could be reduced the train time and also real-time processing.

ACKNOWLEDGMENT

The authors would like to acknowledge the Asian Office of Aerospace Research and Development (AOARD) for supporting this research work under the grant number: FA2386-18-1-4105 R&D 18IOA105.

REFERENCE

[1] A. Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning - IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) ", 2017.
[2] T. Zhang, "Facial Expression Recognition Based on Deep Learning: A Survey - International Conference on Intelligent and Interactive Systems and Applications ", 2018.
[3] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang and T. Zhang, "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts - IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech) ", 2017.
[4] H. Jung et al., "Development of deep learning-based facial expression recognition system - 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) ", 2015.
[5] S. Bazrafkan, T. Nedelcu, P. Filipczuk and P. Corcoran, "Deep learning for facial expression recognition: A step closer to a smartphone that knows your moods - IEEE International Conference on Consumer Electronics (ICCE) ", 2017.
[6] G. Kumar, R. Kumar and G. Sanyal, "Facial emotion analysis using deep convolution neural network - International Conference on Signal Processing and Communication (ICSPC)", 2017.
[7] P. Huynh, T. Tran and Y. Kim, "Convolutional Neural Network Models for Facial Expression Recognition Using 3D-BUFE Database - 10.1007/978-981-10-0557-2_44.", 2017.
[8] L. Zhang, Y. Yang, W. Li, S. Dang and M. Zhu, "Research of Facial Expression Recognition Based on Deep Learning - IEEE 9th International Conference on Software Engineering and Service Science (ICSESS) ", 2018.
[9] K. Shan, J. Guo, W. You, D. Lu and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure - IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) ", 2017.
[10] Y. Sun and Q. Li, "The Application of Deep Learning in Mathematical Education - 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII) ", 2018.
[11] M. Wang, Z. Wang, S. Zhang, J. Luan and Z. Jiao, "Face Expression Recognition Based on Deep Convolution Network - 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) ", 2018.
[12] W. Deng, J. Hu, S. Zhang and J. Guo, "DeepEmo: Real-world facial expression analysis via deep learning - Visual Communications and Image Processing (VCIP) ", 2015.
[13] T. Raksarikorn and T. Kangkachit, "Facial Expression Classification using Deep Extreme Inception Networks - 15th International Joint Conference on Computer Science and Software Engineering (JCSSE) ", 2018.
[14] I. Lee, H. Jung, C. Ahn, J. Seo, J. Kim and O. Kwon, "Real-time personalized facial expression recognition system based on deep learning - IEEE International Conference on Consumer Electronics (ICCE) ", 2016.
[15] W. Li, M. Li, Z. Su and Z. Zhu, "A deep-learning approach to facial expression recognition with candid images - 14th IAPR International Conference on Machine Vision Applications (MVA) ", 2015.
[16] A. Sajjanhar, Z. Wu and Q. Wen, "Deep Learning Models for Facial Expression Recognition - Digital Image Computing: Techniques and Applications (DICTA)", 2018.
[17] S. Sharma, K. Kumar and N. Singh, "D-FES: Deep facial expression recognition system - Conference on Information and Communication Technology (CICT) ", 2017.
[18] https://medium.com/datadriveninvestor/deep-learning-deep-belief-network-dbn-ab715b5b8afc
[19] H. Alshamsi, V. Këpuska and H. Meng, "Automated facial expression recognition app development on smart phones using cloud computing – IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON) ", 2017.
[20] Xiao, S., Ting, P., Fu-Ji, Rn: Facial expression recognition using ROI-KNN deep convolutional neural networks. Acta Automatica Sinica 42(6), 883–891 (2016)

.