Mood Detection for Improving Lifestyle of Older Adults in Ambient Assisted Living Contexts

Andrea Caroppo, Alessandro Leone, Pietro Siciliano
National Research Council of Italy - Institute for Microelectronics and Microsystems
Via Monteroni, c/o Campus Università del Salento, Palazzina A3, Lecce, Italy
email: andrea.caroppo@cnr.it email: alessandro.leone@cnr.it email: pietro.siciliano@le.imm.cnr.it

Abstract— **Ambient Assisted Living (AAL) has the ambitious goal of improving the lifestyle or quality of life of elderly or vulnerable people through the use of technology. In this research, area considerable efforts have been made for the design and development of automatic solutions for the recognition of mood trough patterns of facial expressions, even using low cost and commercial vision sensors. However, there are still some open issues to be faced like the age or any situation of disability of the observed subject, the pose of the face and the environment illumination conditions. A lot of progress has been made in this topic with the emergence of deep learning methods. In the proposed work, the performance of two recent deep convolutional neural networks models are evaluated on the CIFE and FER-2013 datasets that include also facial expressions of older adults performed in uncontrolled conditions. A thorough session of experiments focused on the concept of "Transfer Learning" was carried out. The results obtained demonstrated that both deep architectures reach levels of accuracy higher than 67.8% grouping expressions of older adults into 3 categories: positive, negative and neutral. The latter classification may be sufficient for a mood detection module that could be the input of a future e-coaching platform to be integrated in the AAL context.**

Keywords- Mood Detection; Ambient Assisted Living; Older Adults; Disability; Lifestyle Improvements.

I. INTRODUCTION

In the last few years, the number of older adults is increasing. In addition, there are currently more than 2 billion disabled people in the world, that is 37.5% of the world's population. Consequently, many efforts have been made in the past and are currently being made by researchers with the aim to improve the quality of lifestyle for these categories of subjects with different frailties. Related to this issue, an Ambient Assisted Living (AAL) works to create better living conditions. AAL systems are able to continuously monitor the health status of the frailty subject through data coming from heterogeneous sensors.

In this context many potential applications, such as robotics, communications, security, medical and assistive technology, would benefit from the ability of automatically recognize facial expression [1][2], because different facial expressions can reflect the mood, the emotions and also mental activities. An automatic system capable of recognizing facial expressions could be the input of an e-coaching system, useful for example to change the living environment, implementing devices (music or video players) or changing lighting conditions based on the mood of the observed subject.

In 1971, Ekman and Friesen reported that facial expression acts as a rapid signal that varies with contraction of facial features like eyebrows, lips, eyes and cheeks. Moreover they determined that there were six basic classes in Facial Expression Recognition (FER): anger, disgust, fear, happiness, sadness and surprise [3]. A classical automatic facial expression analysis usually involves three steps: face acquisition, facial data extraction and representation (feature extraction), and classification.

FER systems can be divided into two main categories according to the feature representations: static image FER and dynamic sequence FER. In static-based methods [4], the feature representation is encoded with only spatial information from the current single image, whereas dynamic-based methods [5] consider the temporal relation among contiguous frames in the input facial expression sequence.

For the feature extraction step, the majority of the traditional methods have used handcrafted features. Generally, they are divided into the following categories: geometric-based, appearance-based and hybrid-based approaches. In particular, geometric-based features are able to depict the shape and locations of facial components such as mouth, nose, eyes and brows using the geometric relationships between facial points to extract facial features [6]; appearance-based descriptors aim to use the whole-face or specific regions in a face image to reflect the underlying information in a face image [7]; hybrid-based approaches combine the two previous types of features in order to enhance the system's performance and it might be achieved either in features extraction or classification level.

Geometric-based, appearance-based and hybrid-based approaches have been widely used for the classification of facial expressions even if it is important to emphasize how all the aforementioned methodologies require a process of feature definition and extraction very daunting. In addition, this category of methodologies easily ignores the changes in skin texture such as wrinkles and furrows that are usually accentuated by the age of the subject. Last but not least recent studies have pointed out that classical approaches used for the classification of facial expression are not performing well when used in real contexts where face pose and lighting conditions are broadly different from the ideal ones used to capture the face images within the benchmark datasets. As highlighted above, among the factors that makes FER very difficult (and consequently mood detection), one of the most discriminating is the age [8][9]. In particular, expressions of older individuals appeared harder to decode, owing to age-related structural changes in the face. Consequently, state-of-the-art approaches based on handcrafted features extraction may be inadequate for mood detection of older adults. In recent years, machine learning techniques based on convolutional neural networks (CNNs) have achieved great success in the field of computer vision. It is also a promising approach for the research of FER. Different from traditional techniques, CNNs can perform tasks in an end-to-end way, associating both feature extraction and classification steps together by training. For example, Zhang et al. [10] proposed a deep neural network (DNN) with the scale-invariant feature transform (SIFT) feature, which achieved the accuracy of 78.9% on the multi-view BU-3DFE dataset [11]. Lopes et al. [12] proposed a combination of CNN and special image pre-processing steps (C-CNN) to recognize six expressions under head pose at 0◦, whose accuracy was 90.96% on the same aforementioned dataset but its robustness was unknown under different head poses. To reduce the influence of various head poses, Jung et al. [13] proposed a jointly CNNs with facial landmarks and colour images, which achieved the accuracy of 72.5%, but the network consisted of only three convolutional layers and two hidden layers, making it be difficult to accurately learn facial features. Li and Deng [14] also presented a very interesting survey on CNN based FER techniques.

Lack of training samples is a big problem for FER in the wild using deep CNNs. In the proposed work, this problem is accentuated by having very few datasets available containing facial expressions of elderly subjects. To solve this issue, some methods used pre-trained network for classification or re-trained a network model to re-initialize the weights for new datasets. The techniques are regarded as “transfer learning” [15].

Based on the above discussion, in this paper, two pre-trained CNN, who have been successful in image recognition, are evaluated for mood detection in the wild and tested on two benchmark datasets that contain facial expressions divided by age group.

The rest of this paper is structured as follows. Section II reports the proposed pipeline emphasizing some details for pre-processing steps. The same section describes also the deep architectures and the algorithmic procedures used in this work in order to adapt the aforementioned models to the problem of mood detection in older adults. Section III presents datasets, experimental procedures and results obtained, while Section IV concludes the paper by providing an overview of the themes highlighted and the need to monitor the future.

II. Overview of the Proposed Mood Detection System

A representation of the proposed mood detection system for older adults is given in the block diagram shown in Figure 1. First, the implemented pipeline performs a pre-processing task on the input images (face detection, cropping & resizing). Once the images are pre-processed they can be classified into three categories (positive, negative or neutral expressions) using pre-trained deep networks.

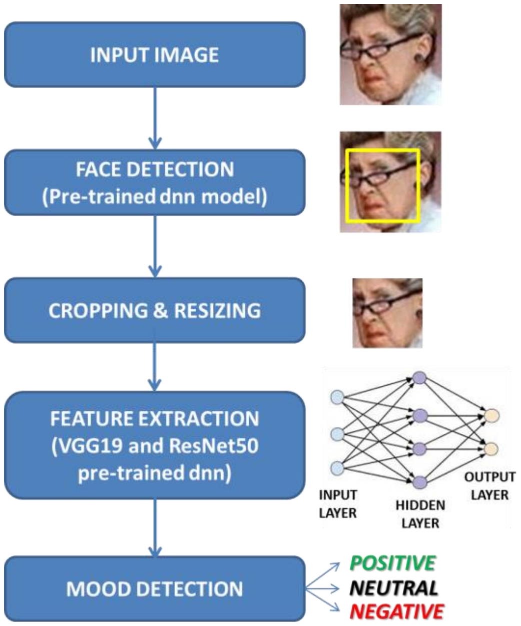


Figure 1. Overview of the proposed mood detection system for older adults in AAL context

A. Pre-Processing

The task of detecting a face in the considered scenario is not an easy problem because many difficulties arise and must be taken into account. In a living environment, for example, the face of observed subject could occupy very little area in most images if the camera is not positioned near the end user. Moreover faces can look very different depending on orientation, pose and age. Face detection went mainstream in the early 2000's when Viola and Jones [16] invented a way to detect faces that was fast enough to run on cheap cameras. Today it is very close to being the de facto standard for solving face detection tasks. Given that the original Viola-Jones face detector has limitations for multi-view face detection, in the implemented pipeline a recently methodology that has met much success is proposed to tackle this problem.

Here the library with functions that mainly aiming real-time computer vision (i.e. last version of OpenCV) is selected. In particular, starting from OpenCV 3.3 a deep neural network (DNN) architecture that performs an accurate face detector is included. The main advantage of this module is the ability to detect faces “in the wild” in real-time even if a PC without GPU is used for the processing. The aforementioned module is based on Single Shot MultiBox Detector (SSD) framework, using a reduced ResNet-10 model [17] and its output is constituted by the coordinates of the bounding box of the facial region accompanied by a confidence index, useful in case it is necessary to set a reliability threshold with respect to the detection of the facial region. Once the face has been detected a simple routine was written in order to crop the facial image. This is achieved by detecting the coordinates of the top-left corner, the height and width of the face enclosing rectangle, removing in this way all background information and image patches that are not related to the expression. Since the facial region could be of different sizes after cropping, in order to remove the variation in face size and keep the facial parts in the same pixel space, the algorithmic pipeline provides both a down-sampling step and an increasing resolution step that generate face images with the specific size required by tested deep architectures. For the down-sampling a simple linear interpolation was used, whereas a nearest-neighbor interpolation was implemented in order to increase the size of the facial images.

B. CNN Architectures For Feature Extraction

CNN is a type of deep learning model designed to automatically and adaptively learn spatial hierarchies of features, from low to high-level patterns. A typical implementation of CNN for FER encloses three learning stages in just one framework. The learning stages are: 1) feature learning, 2) feature selection and 3) classifier construction. Creating a CNN from scratch is not an easy task. So, in order to save ourselves from this over-head, in the present work the concept of "transfer learning" was adopted [15]. Transfer learning is a common and recent strategy to train a network also on a small dataset, (which is one of the main problems in the case of recognition of facial expressions of the elderly) where a network is pre-trained on an extremely large dataset, such as ImageNet [18], which contains 1.4 million images with 1000 classes, then reused and applied to the given task of interest.

In this work two deep CNNs (DCNNs), namely VGG19 and Resnet50, were evaluated as the feature extractors of the proposed method for mood detection in older adults. These DCNNs are pre-trained on a nature image dataset (ImageNet) for distinct generic image descriptors and then applied to extract discriminative features from facial images based on transfer learning theory. Below is a brief description of the two architectures.

VGG19 - The VGG networks with 16 layers (VGG16) and with 19 layers (VGG19) [19] were the basis of the Visual Geometry Group (VGG) submission in the ImageNet Challenge 2014, where the VGG team secured the first and the second places in the localization and classification tracks respectively. The VGG architecture is structured starting with five blocks of convolutional layers followed by three fully-connected layers. Convolutional layers use 3×3 kernels with a stride of 1 and padding of 1 to ensure that each activation map retains the same spatial dimensions as the previous layer. A rectified linear unit (ReLU) activation is performed right after each convolution and a max pooling operation is used at the end of each block to reduce the spatial dimension. Max pooling layers use 2×2 kernels with a stride of 2 and no padding to ensure that each spatial dimension of the activation map from the previous layer is halved. Two fully-connected layers with 4096 ReLU activated units are then used before the final 1000 fully-connected softmax layer. A downside of the VGG16 and VGG19 models is that they are more expensive to evaluate and use a lot of memory and parameters. VGG16 has approximately 138 million parameters and VGG19 has approximately 143 million parameters. Most of these parameters (about 100 million) are in the first fully-connected layer, and it was since found that these fully-connected layers could be removed with no performance downgrade, significantly reducing the number of necessary parameters.

ResNet50 - Residual Networks (ResNets) [20] are deep convolutional networks where the basic idea is to skip blocks of convolutional layers by using shortcut connections to form blocks named residual blocks. These stacked residual blocks greatly improve training efficiency and largely resolve the degradation problem present in deep networks. In ResNet50 architecture, the basic blocks follow two simple design rules: 1) for the same output feature map size, the layers have the same number of filters; and 2) if the feature map size is halved, the number of filters is doubled.

The down sampling is performed directly by convolutional layers that have a stride of 2 and batch normalization is performed right after each convolution and before ReLU activation. When the input and output are of the same dimensions, the identity shortcut is used. When the dimensions increase, the projection shortcut is used to match dimensions through 1×1 convolutions. In both cases, when the shortcuts go across feature maps of two sizes, they are performed with a stride of 2. The network ends with a 1,000 fully-connected layer with softmax activation. The total number of weighted layers is 50, with 23,534,592 trainable parameters.

C. Classification

The methodology for automated mood detection using transfer learning is shown in Figure 2. The structure illustrated consists of layers from the pre-trained model and few new layers.

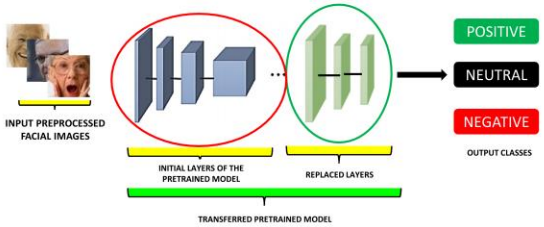

POSITIVE
NEUTRAL
NEGATIVE
INPUT PREPROCESSED FACIAL IMAGES
OUTPUT CLASSES
INITIAL LAYERS OF THE PRETRAINED MODEL
REPLACED LAYERS
TRANSFERRED PRETRAINED MODEL

Figure 2. Mechanism of transfer learning using pre-trained models for mood detection of older adults

For the present work only the last three layers of VGG19 and ResNet50 were replaced to accommodate the new image categories. Since an age-related decline in decoding facial expressions has been repeatedly reported in literature [21][22], in the proposed work the traditional facial expression classification has been modified by grouping the expressions into 3 main groups: positive (happiness), negative (fear, disgust, anger and sadness) and neutral. The expression "surprise" was not considered since it can have any valence; that is, it can be neutral/moderate, pleasant, unpleasant, positive, or negative.

Consequently, the step-by-step procedure for expression classification is outlined below:

1) resize the input images so that they are consistent with the size of the input layer of the pre-trained network model;

2) partition the data into training and test sets; 70% of images per category are taken for training and 30% as a test dataset to test the network;

3) review the network architecture (replace the final layers/modify the layers): for VGG19 and ResNet50 alter the last three layers of the pre-trained networks with a set of layers (“fully connected layer,” a “softmax layer,” and a “classification output layer”) to categorize the images into the respective classes;

4) train the network and test the new model on the target datasets.

III. RESULTS

The proposed mood detection system was tested on two benchmark datasets among the few present in the literature that include facial expressions acquired in uncontrolled conditions and containing subjects of different age groups.

The CIFE dataset [23] is composed of facial images with seven different types of expressions. The expression subsets have the following sizes: 3636, 1905, 975, 2485, 1994, 1381 and 2381 for Happiness, Anger, Disgust, Sadness, Surprise, Fear and Neutral respectively. The images are extracted from the web with gathering techniques permitted to collect in total 14757 images containing candid expression images that are randomly posed. This last detail redeems the process of expression recognition more difficult than the two previous datasets, which are made up of facial frontal expressions acquired in controlled environments.

Since CIFE dataset contains images with only the label of facial expression and without any indication about the age of the subject, it was necessary to perform an age estimation technique consolidated in the literature. The approach used in the present work is inspired from the algorithm described in [24] that permitted to group the images into four different subgroups. Table 1 presents the total number of images in the CIFE dataset, divided according to the estimated age, whereas in Figure 3 some examples of expressions performed by ageing adults are represented.

TABLE I. CIFE IMAGES BROKEN DOWN BY AGE

	Age (years)				
	<35	35-55	56-68	>68	Total
#images	5587	4828	2263	2079	14757

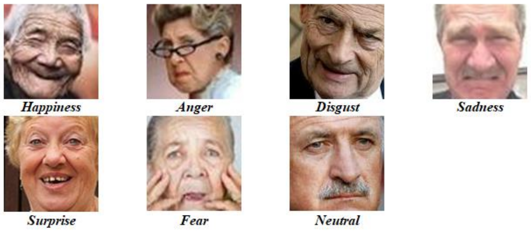


Figure.3. Some examples of expressions performed by aging adults from the CIFE dataset.

FER-2013 [25] is a large-scale FER dataset used in the ICML 2013 workshop's facial expression recognition challenge. The dataset has seven expressions including anger, disgust, fear, happy, sad, surprise, and neutral. It is comprised of 48×48 pixel grey-scale images of human faces. The training set consists of 28,709 examples, while both the test and validation sets are composed of 3,589 examples. The images of FER-2013 were collected from the Internet and the faces greatly vary in age, pose and occlusion, thus resulting in that the accuracy of human recognition is only approximately 65 ± 5%. As a powerful machine learning tools, the CNN can now surpass human beings on the FER-2013 task, and the state-of-the-art accuracy on FER-2013 is 75.42% by combining CNN extracted features and hand-crafted features for training.

Also the FER-2013 dataset does not contain the subdivision of the images into age groups, therefore the technique proposed in [24] was used again. Table 2 presents the total number of images in the FER-2013 dataset, divided according to the estimated age, whereas in Figure 4 some examples of expressions performed by ageing adults are represented.

TABLE II. FER-2013 IMAGES BROKEN DOWN BY AGE

	Age (years)				
	<35	35-55	56-68	>68	Total
#images	13560	7432	6128	5178	32298

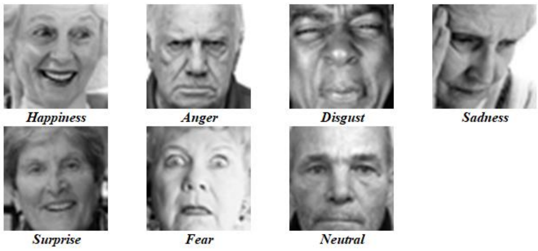


Figure 4. Some examples of expressions performed by aging adults from the FER-2013 dataset.

Various experiments were conducted to assess the FER performance of the pre-trained VGG19 and ResNet50 deep networks with transfer learning. The metric used in this work for evaluating the methodologies is the accuracy, whose value is calculated using the average of n-class classifier accuracy for each group of expressions (positive, negative and neutral):

$$Acc = \frac{\sum_{1}^{n} Acc_{\exp r}}{n} \qquad (1)$$

$$Acc_{\exp r} = \frac{Hit_{\exp r}}{Total_{\exp r}} \qquad (2)$$

where Hit_{expr} is the number of hits in the group of expression *expr*, $Total_{expr}$ represents the total number of samples of each group of expressions and n is the number of expressions to be considered (in our case *n*=3). Fig. 5 reports the average accuracy for FER on images of CIFE and FER-2013 datasets. For each dataset the accuracy obtained with VGG19 and ResNet50 is reported, depending on the classifiers used. Here, for the final classifier layer, Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR) were compared.

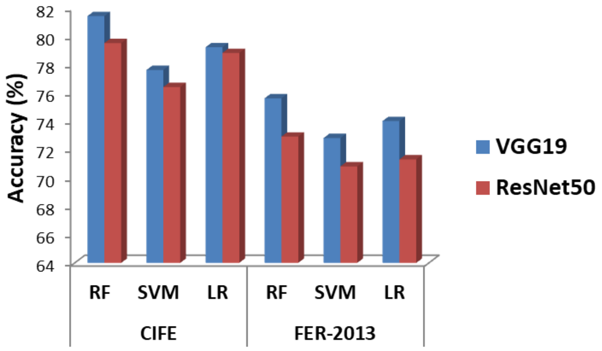


Figure 5. FER average accuracy on CIFE and FER-2013 combining VGG19 and ResNet50 with RF, SVM and LR classifier

From the results reported in the previous figure, it is evident that the recognition performances of three categories of expressions vary significantly as the dataset changes (FER-2013 is more challenging since the images are grayscale and have a resolution of 48x48 pixels) but by using for pre-training VGG 19 greater accuracy is obtained on both datasets, with the RF classifier which tends to provide an improvement in the results (about 3.8% with respect SVM and 2.2% with respect LR for CIFE and 2.8% with respect SVM and 1.6% with respect LR for FER-2013).

It is important to underline how the previous results were obtained by considering all age groups in which facial expressions were divided. Since the main objective of this work is mainly oriented towards the recognition of the facial expression category of the elderly, experiments have been carried out with the aim of measuring the performance of the methodologies by grouping the images of the datasets into different age groups. Table 3 and Table 4 report the final accuracy obtained by the proposed deep architectures with RF as classifier for each age group detected in CIFE and FER-2013 dataset respectively:

TABLE III. FER ACCURACY ON CIFE DATASET

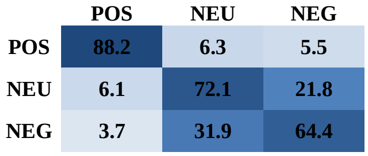
Age Group	VGG19+RF (%)	ResNet50+RF (%)
<35	86.4	85.3
35-55	84.7	81.6
56-68	79.6	77.6
>68	74.9	73.5
Average value	**81.4**	**79.5**

TABLE IV. FER ACCURACY ON FER-2013 DATASET

Age Group	VGG19+RF (%)	ResNet50+RF (%)
<35	82.5	78.8
35-55	77.1	74.7
56-68	73.2	70.3
>68	69.6	67.8
Average value	**75.6**	**72.9**

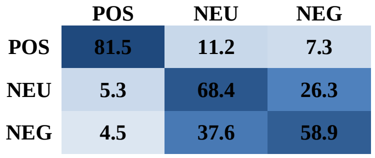
Previous results demonstrate how the age of the observed subject influences the classification of facial expressions, probably because, as demonstrated by numerous psychological studies, the elderly are less expressive and consequently even with only three classes of expressions there is a gap in the accuracy between young and older subjects.

In a multi-class recognition problem, as the FER one, the use of an average recognition rate (i.e., accuracy) among all the classes could be not exhaustive since there is no possibility to inspect what is the separation level, in terms of correct classifications, among classes (in our case positive, negative and neutral facial expressions). To overcome this limitation, for each dataset the confusion matrices obtained with VGG19 + RF model are reported in Fig. 6 and Fig. 7 (here only the facial images of older adults with more than 68 years were considered). The notation POS in tables referred to positive expression (happiness); the notation NEU referred to neutral expression whereas NEG referred to negative expressions (fear, disgust, anger and sadness).



	POS	NEU	NEG
POS	**88.2**	**6.3**	**5.5**
NEU	**6.1**	**72.1**	**21.8**
NEG	**3.7**	**31.9**	**64.4**

Figure 6. Confusion Matrix on CIFE dataset (performed by older adults with more than 68 years) using the proposed VGG19 + RF architecture.



	POS	NEU	NEG
POS	**81.5**	**11.2**	**7.3**
NEU	**5.3**	**68.4**	**26.3**
NEG	**4.5**	**37.6**	**58.9**

Figure 7. Confusion Matrix on FER-2013 dataset (performed by older adults with more than 68 years) using the proposed VGG19 + RF architecture.

The confusion matrices provide, in relation to elderly subjects, a very important information to be taken into consideration in the case of implementation of e-coaching platforms: negative expressions are confused considerably with neutral expression. This observation poses a very important problem, as negative expressions are symptomatic of the onset or aggravation of diseases.

IV. CONCLUSION

The purpose of this paper was to explore and evaluate two deep transfer learning approach for mood detection in older adults considering that the majority of the works in the literature that address this topic are based on benchmark datasets that contain facial images with a small span of lifetime (generally young and middle-aged subjects). Among the two different deep architectures tested, the pre-trained VGG19 architecture in combination with an RF classifier yielded the best performance for each considered dataset and for each age group in which the dataset has been divided considering only three main classes of facial expressions: positive, negative and neutral. It was chosen to classify the expressions in these categories because they are sufficient for the development of an integrated system capable of implementing e-coaching systems based on the mood detected.

Future work will deal with the following main aspects. A first development might be to perform the pre-training of deep architectures on datasets different from ImageNet and more specific for the topic considered. Moreover, it will be necessary to extend the number of compared deep learning approaches since a limitation of the present work is the evaluation of only two pre-trained deep architectures which have already been overcome in terms of image classification from more deeper architectures like Inception-v4 [20] and Inception-ResNet-V2 [20].

ACKNOWLEDGMENT

This work has been carried out within the project PON Si-Robotics funded by MUR-Italian Ministry for University and Research.

REFERENCES

[1] M. Pantic and L. J. M. Rothkrantz, “Automatic analysis of facial expressions: The state of the art”, IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 12, pp. 1424-1445, 2000.

[2] B. Fasel and J. Luettin, “Automatic facial expression analysis: a survey”, Pattern recognition, vol. 36, no. 1, pp. 259-275, 2003.

[3] P. Ekman and W.V. Friesen, ”Constants across cultures in the face and emotion”, Journal of personality and social psychology, vol. 17, no. 2, p. 124, 1971.

[4] A. Mollahosseini, D. Chan, and M.H. Mahoor, “Going deeper in facial expression recognition using deep neural networks”, in 2016 IEEE Winter conference on applications of computer vision (WACV), pp. 1-10, IEEE, 2016, March.

[5] X. Zhao et al., “Peak-piloted deep network for facial expression recognition”, In European conference on computer vision (pp. 425-442). Springer, Cham, 2016, October.

[6] R. Shbib and S. Zhou, “Facial expression analysis using active shape model”, International Journal of Signal Processing, Image Processing and Pattern Recognition, 2015, vol. 8, no. 1, pp. 9-22.

[7] J. Chen, Z. Chen, Z. Chi, and H. Fu, “Facial expression recognition based on facial components detection and hog features”, In International Workshops on Electrical and Computer Engineering Subfields, 2014, (pp. 884-888).

[8] G. Guo, R. Guo, and X. Li, “Facial expression recognition influenced by human aging”, IEEE Transactions on Affective Computing, 2013, vol. 4, no. 3, pp. 291-298.

[9] S. Wang, S. Wu, Z. Gao, and Q. Ji, “Facial expression recognition through modeling age-related spatial patterns”, Multimedia Tools and Applications, 2016, vol. 75, no. 7, pp. 3937-3954.

[10] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, “A deep neural network-driven feature learning method for multi-view facial expression recognition”, IEEE Transactions on Multimedia, vol. 18, no. 12, pp. 2528-2536, 2016.

[11] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, “A 3D facial expression database for facial behavior research”, In 7th international conference on automatic face and gesture recognition (FGR06) (pp. 211-216). IEEE, 2006, April.

[12] A. T. Lopes, E. de Aguiar, A. F. De Souza, 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.

[13] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, J, “Joint fine-tuning in deep neural networks for facial expression recognition”, In Proceedings of the IEEE international conference on computer vision (pp. 2983-2991), 2015.

[14] S Li and W. Deng, “Deep facial expression recognition: A survey” IEEE Transactions on Affective Computing, 2020.

[15] S. J. Pan and Q. Yang, “A survey on transfer learning”, IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[16] P. Viola and M. J. Jones, “Robust real-time face detection”, International journal of computer vision, vol. 57, no. 2, pp. 137-154 , 2004.

[17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778), 2016.

[18] O. Russakovsky, J. Deng, H. Su et al., “Imagenet large scale visual recognition challenge”, International journal of computer vision, vol. 115, no. 3, pp. 211-252, 2015.

[19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556, 2014.

[20] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning”, In Thirty-first AAAI conference on artificial intelligence, 2017, February.

[21] A.J. Calder et al., “Facial expression recognition across the adult life span”, Neuropsychologia, vol. 41, no. 2, pp. 195-202, 2003.

[22] N. C. Ebner and M. K. Johnson, “Young and older emotional faces: are there age group differences in expression identification and memory?”, Emotion, vol. 9, no. 3, p. 329, 2009.

[23] W. Li, M. Li, Z. Su, and Z. Zhu, “A deep-learning approach to facial expression recognition with candid images”, In 2015 14th IAPR International Conference on Machine Vision Applications (MVA) , 2015, (pp. 279-282). IEEE, 2015

[24] T. Wu, P. Turaga, and R. Chellappa, “Age estimation and face verification across aging using landmarks”, IEEE Transactions on Information Forensics and Security, 2012, vol. 7, no. 6, pp. 1780-1788.

[25] I. J. Goodfellow, D. Erhan, P. L. Carrier et al, “Challenges in representation learning: A report on three machine learning contests”, In International Conference on Neural Information Processing, 2013, (pp. 117-124). Springer, Berlin, Heidelberg