2019 IEEE 4th International Conference on Signal and Image Processing

Detecting Negative Emotional Stress Based on Facial Expression in Real Time

Jin Zhang
College of Electrical Engineering and Control Science
Nanjing Tech University
Nanjing, China
e-mail: 2291674590@njtech.edu.cn

Xue Mei*, Huan Liu, Shenqiang Yuan, Tiancheng Qian
College of Electrical Engineering and Control Science
Nanjing Tech University
Nanjing, China
e-mail: seraph_mx@163.com

***Abstract*—Negative emotional stress can be seen as a physiological response to mental and physical challenges. Exposure to stressful situations with a long time can have adverse effects on people, such as depression, which finally results in suicide in severe case, so it is important to monitor stress in real time and treat it properly. In this paper, we propose a new framework for stress detection in real time. The framework detects stress by recognizing three stress related facial expressions, anger, fear and sadness. We also propose a connected convolutional network, which combines low-level features with high-level features to train the deep network to recognize facial expressions. If the number of stress related frames exceeds a threshold value, the framework will remind people to take a break to relax. The experiment results demonstrate that our proposed method has better performance on facial expression recognition and realizes high-performance stress detection.**

Keywords-stress detection; facial expression; real-time recognition; convolutional neural network

I. INTRODUCTION

Negative emotional stress is a mental health problem affecting the life of one in four people [1]. Long-term negative emotional stress may potentially lead to poorer health outcomes, a decreased quality of life, and increased health care usage [2]. It also increases the risk of cardiovascular diseases and somatic complaints [3]. Due to the adverse effects of stress in our daily life, it is important to monitor such an unhealthy state in a timely manner and treat it properly [4].

There are many methods of studying stress like biomedical means [5], self-reporting questionnaire [6] and Biomarkers method [7]. However, these methods are not very practical in the case for real-time applications. Biomedical methods are more intrusive and require specific instruments. Self-reporting questionnaire methods require subjects to spend time and effort reporting symptoms unbiasedly. Biomarkers method take lot of time to collect and analysis samples, which may not be applicable.

Facial expression(FE) holds a lot of information and can reflect instantaneous stress [10]. So, FE can be instrumental to understanding stress without intrusion, clutter or individual time. Negative emotional stress is sensitive to three facial expressions, anger, sadness, and fear. Collecting these expressions at a FE sequence can predict stress more accurately [8]. Therefore, we propose a framework to detect stress based on facial expressions, which achieves a high performance with a fast speed.

II. RELATED WORK

Delmastro et al. [9] applied vibro-acoustic therapy (VAT) to detect the relations between physiological signals and stress condition. Liu et al. [11] presented a method for exploring the electro dermal activity (EDA) signal which aims at discriminating emotional stress. In [12], the researchers developed a method by monitoring skin temperature to monitor stress state. Shan et al. [13] proposed a framework for detecting and classifying human stress based on respiratory signals measured remotely by using a Kinect sensor. Sakri et al. [14] utilized the mean of the heart rate and the sum of all skin response's width-prominence products to monitor emotional stress. In [15], the researchers combined visual evidence and skin temperature as stress detecting signals. However, the collection of these data sequences in daily life conditions is very difficult for developing practical applications and intrusive devices are not friendly to people.

In recent years, different methods have been used in FER to extract appearance features of image sequences, including local binary patterns (LBP-TOP) [16], 3D histograms of oriented gradients (3DHOG) [17], weighted random forest(WRF) [18], and 3D scale-invariant feature transform (3DSIFT) [19]. Deep learning network has become the most widely used method in FER research because of its powerful feature extraction ability. 3DCNN [21] performed 3D convolution on image sequences with constraints based on deformable action parts. Jung et al. [20] combined DTAN with DTGN as DTAGN to achieve better results in FER. Ouyang et al. [26] employed a novel network called ResNet-LSTM to capture spatio-temporal information, which combine lower features to LSTMs directly. Hasani et al. [27] presented 3D Convolutional layers followed by an LSTM unit, extracting the spatial-temporal relations between different images in the facial sequences.

Most of works on FE focus on the automatic recognition, but reports on stress detection are fewer. Gao et al. [22] applied designed descriptors to analysis FE, which used to detect stress, and developed a real-time non-intrusive monitoring system. The system can alert drivers when detected stress, preventing traffic accidents due to narrow

978-1-7281-3660-8/19/$31.00 ©2019 IEEE

attention. Inspired by [22] and [28], we propose a shallow connected network which can reduce overfitting caused by limited training databases, to detect stress.

III. The Proposed Method

A. Problem Formulation

The stress detection task can be formulated as a FE classification problem according to the relationship between stress and facial expressions [8]. In this task, the objective is assign a single FE label to each input frame. If two-thirds of the labeled frames are stress related frames, which simulates people under stress for a long time, the framework will remind people to take a break to relax.

B. Overview of the Framework

Face detection & alignment is performed with Multi-task Cascaded Convolutional Networks (MTCNN) [23]. After alignment face, facial part is cropped and its dimension is resized to 224px×224px which is input dimension of our CNN model. Features from the input face sequences are extracted and classified by our CNN model. Then, finally we get FE recognition result (see Fig. 1). If stress related facial expressions lasting time exceeds a specific time, the framework will output a warning message to remind people.

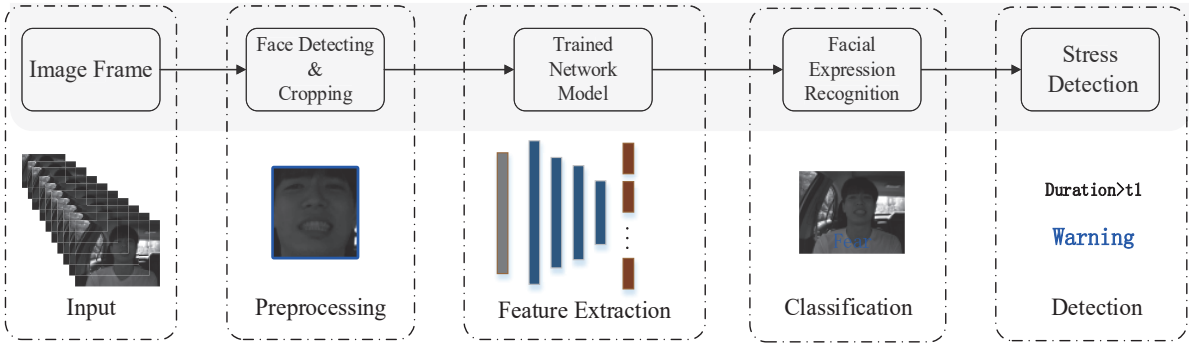

Figure 1. Overview of the proposed method for facial expression recognition.

C. Face Detection and Alignment

When the model initialized, face detection is used to find a face for every frame. Then once it succeeds to detect a face, the key points of the eyes, nose and mouth are used for face alignment module.

MTCNN adopts a cascaded structure with three stages of convolutional networks that predict face and landmark location in a coarse-to fine manner. We used MTCNN for face detection and key landmark location, then used affine transformation for face alignment.

D. Facial Expression Recognition

1) The Proposed Network Architecture

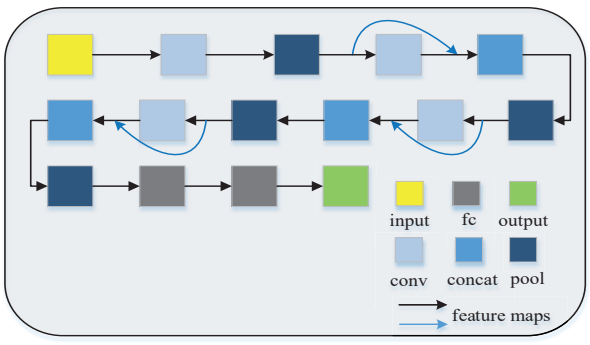

Figure 2. Our CNN model architecture.

Fig. 2 shows the CNN model structure, with one input layer, four convolutional layers, two fully-connected layers and an output layer. Inspired by [28], we propose connect low-level features with high-level features to enhanced the capacity of network. Detail specification of parameters is listed on Table I. We set the initial convolutional filters size to be 7*7, and the filter numbers for each layer to be 32,32,64,64, respectively. Implementation of CNN was done with public deep learning library TensorFlow.

TABLE I. Parameters of CNN Model Architecture

NAME	KERNEL	OUTPUT_SIZE	STRIDE
INPUT	/	224*224*1	/
CONV1	7*7	224*224*32	1
POOL1	2*2	112*112*32	2
CONV2	7*7	112*112*32	1
CONCAT1	/	112*112*64	/
POOL2	2*2	56*56*64	2
CONV3	5*5	56*56*64	1
CONCAT2	/	56*56*128	/
POOL3	2*2	28*28*128	2
CONV4	5*5	28*28*64	1
CONCAT3	/	28*28*192	/
POOL4	2*2	14*14*192	2
FC1	/	1*1*4096	/
FC2	/	1*1*2048	/
OUTPUT	/	1*1*6	/

2) Training Phase

To avoid overfitting caused by lack of labeled training data in small FE dataset, we apply data augmentation to

expand datasets, which using various transformations to generate small changes in appearances and postures. We applied five filters and five affine transform matrices. The five filters are blur, medium, □aussian, unsharp and smooth filters, and the five affine transforms are formalized by adding slight geometric transformations to identity matrix □Fig.3□. Order of images se□uences in the training set is randomly shuffled before training started.

Figure 3. Image augmentation□from 1 to 30.

3) Stress Detection

On testing phase, the CNN model received an image se□uence from testing dataset, and outputted the predicted expression of every frame by using the final trained network weights. If the predicted images are stress related facial expressions, and the number is two-thirds more than total frame number, we consider the people is under stress and warning him/her to take a break.

IV. E□PERIMENTS

A. Datasets

To verify the effectiveness of the proposed CNN architecture, we experiment with three databases, CK□, Oulu-CASIA and KMU-FED □Fig. 4.□.

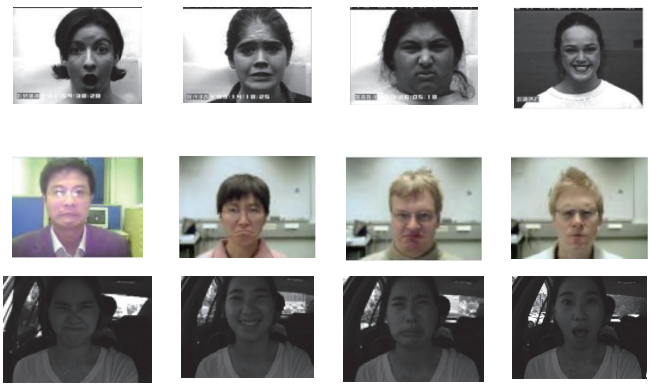

Figure 4. Examples of the datasets used in our experiments. Top□CK□□ Middle□Oulu-CASIA□Bottom□KMU-FED.

1) CK+ Database

This dataset has six basic emotion classes, including anger□An□, disgust□Di□, fear□Fe□, happiness□Ha□, sadness□Sa□ and surprise□Su□. In addition, there is another special expression called □contempt□. As we can see in Table II, the dataset contains 309 annotated se□uences with the six basic expression labels from 123 sub□ects [24]. We divide these se□uences into 10 groups, of which 9 groups are used for training and the rest for testing. In this way, we can run various FER methods multiple times and obtain an averaged accuracy for evaluation.

TABLE II. THE NUMBER OF IMA□E SE□UENCES IN EACH OF THE SI□ BASIC EMOTION CLASSES□AN□ER□AN□, DIS□UST□DI□, FEAR□FE□, HAPPINESS□HA□, SADNESS□SA□AND SURPRISE□SU□

	An	Di	Fe	Ha	Sa	Su	All
CK□	45	59	25	69	28	83	309
Oulu CASIA	80	80	80	80	80	80	480
KMU-FED	18	11	19	20	18	20	106

2) Oulu-CASIA Database

Oulu-CASIA database[16] is a little bit more challenging than CK□. It includes 80 sub□ects between 23 and 58 years old, with six basic expressions □i.e. An, Di, Fe, Ha, Sa, and Su□of each person under normal illumination conditions, as we can see in Table II. Each se□uence starts at a neutral face and ends at the apex of expression as the same settings in CK□. The same as in CK□, a 10-fold cross validation is performed to evaluate various FER methods.

3) KMU-FED Database

To verify the effectiveness of our framework in a real daily-life environment, we use KMU-FED [25] database which is captured in an actual driving environment. It contains 106 image se□uences from 12 sub□ects which include various changes in illumination and partial occlusions caused by hair or sunglasses, labeled with 6 basic emotions □i.e. An, Di, Fe, Ha, Sa, and Su□.

B. Processing Time

When the face is not detected, only detection model is activated. Even though CNN model is not executed, it also takes 10ms to process a single frame. It means that in the worst case, our model can take 10ms in response to unexpected sub□ect appearance. After detecting a face, processing time drops to 4.3ms average to process a single frame. It is because emotion classification takes much shorter processing time than face detection. Processing times were measured on an Nvidia □eForce □T□ 1080 Ti □PU.

C. Results and Discussion

A CNN that has the same basic structure as the proposed model, but no connection between layers, is trained and tested on the same FE databases for comparison with the proposed convolutional architecture.

Fig.5 compares our proposed model with some related works on CK□ dataset and Oulu-CASIA dataset. Due to the CK□ is easy to recognize, our method performs well in anger, disgust and surprise. Because of the imbalance of training data, only 28 sadness and 25 fear among dataset, result in the proposed model did not perform well enough for sadness and fear. We achieve an average recognition accuracy of 93.6□ on CK□.

Oulu-CASIA is more difficult to classify than CK□. The performance for anger and disgust is slightly poor. That

means the proposed model is easy to be confused with disgust and anger in this dataset. These two expressions could have very similar appearance features such as a frown or a twisted nose.

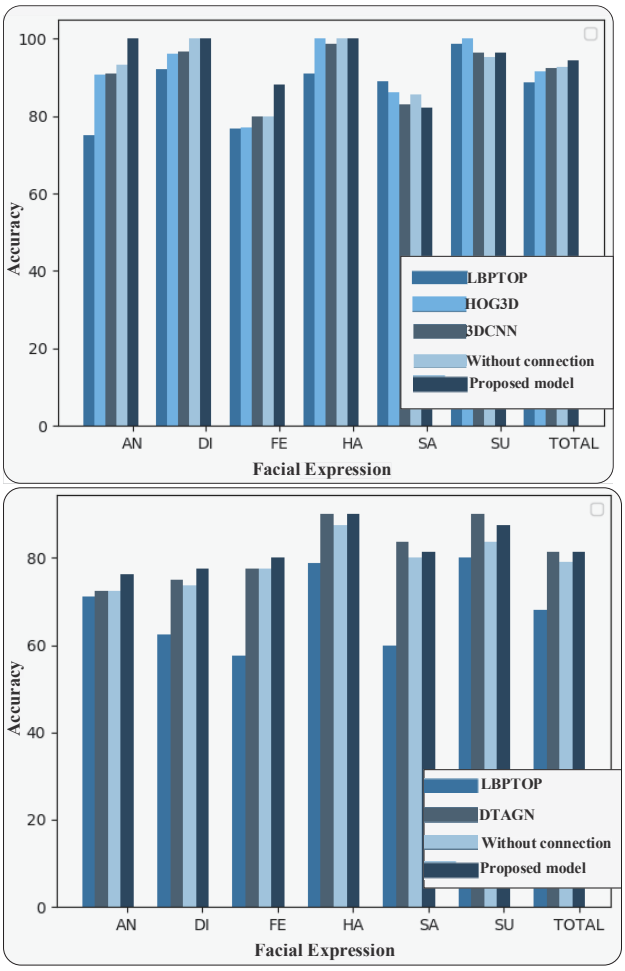

Figure 5. Confusion matrix of the proposed model on dataset, the column represents true labels and the row represent predicted labels. Left□ CK□□Right□Oulu-CASIA.

TABLE III. PERFORMANCE COMPARISON ON THE CK□DATASET

Method	Accuracy
LBPTOP[16]	89.0
HO□3D[17]	91.4
3DCNN[21]	92.3
Proposed model without connection	92.6
Proposed model	**93.6**

TABLE IV. PERFORMANCE COMPARISON ON THE OULU-CASIA DATASET

Method	Accuracy
LBPTOP[16]	68.1
DTA□N [20]	81.4
Proposed model without connection	79.1
Proposed model	**81.3**

As we can see in Table III and Table IV, the proposed method achieves the best accuracy in both of the CK□ and Oulu-CASIA databases. The proposed model without connection performs worse than the proposed model, proves the connection of the proposed model is effective.

To verify the effectiveness of our proposed CNN model in a real daily-life environment, we use KMU-FED database for specify emotion recognition in an actual driving environment, including problems that may occur on a real-life road. As we can see in Table V, the proposed model performs well on real driving environment.

TABLE V. PERFORMANCE COMPARISON ON THE KMU-FED DATASET

Method	Accuracy
CRF [18]	94.0
Proposed model without connection	96.2
Proposed model	**97.3**

TABLE VI. STRESS DETECTION ON FER SE□UENCES B□ PROPOSED METHOD

Datasets	Stress Accuracy
CK□	91.2
Oulu-CASIA	80.4
KMU-FED	99.3

We also conduct simulation experiments real-time stress detection on these datasets. When all frames in a se□uence are recognized, we count the number of stress related expression frames. When the number of stress frames exceeds two-thirds of the number of se□uence frames, we assume that the stress state lasts for a long time and needs to be warned. Table VI shows the accuracy of the stress detection and warning. Experiments show that the proposed framework can recognize facial expressions fast and detect negative emotional stress accurately in both laboratory environment and daily-life driving environment.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a real-time stress detection framework composed by face detection module, facial expression recognition module based on the connect convolution neural network and negative emotional stress detection module. To avoid overfitting due to the lack of training data, we apply data augmentation to extend datasets. Extensive experiments show that the proposed CNN architecture achieves competitive facial expression recognition performance on the CK□ and Oulu-CASIA datasets. We also conduct a simulation experiment on stress detection, counting FE frames which classified to stress expressions. The proposed framework will send a warning message if the number of stress frames exceeds a threshold value. And in daily-life driving environment like KMU-FED dataset, the framework also achieves high stress detection accuracy and low processing time.

In the future, we will explore transfer representation learned from face recognition dataset to facial expression

recognition, which can improve the performance due to the similarity of datasets. We believe it is also important to study the definition and characteristics of emotional stress in the future. More cues, such as acoustic signals, could be integrated to train a personalized stress detection model, achieving better performance.

ACKNOWLEDGMENT

The authors would like to thank the Beijing Advanced Innovation Center for Intelligent Robots and Systems for the support of project NO.2018IRS20.

REFERENCES

[1] Communications, N. World health report. 2001. URL: www.who.int/whr/2001/media_centre/press_release/en/.

[2] J. Thomas. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. Current hypertension reports, 3(3):249–254, 2001

[3] Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. Galvanic skin response (gsr) as an index of cognitive load. In CHI'07 extended abstracts on Human factors in computing systems, pages 2651–2656. ACM, 2007.

[4] W. Liao, W. Zhang, Z. Zhu and Q. Ji, A Real-Time Human Stress Monitoring System Using Dynamic Bayesian Network. CVPR, 2005, pp. 70-70.

[5] J. Bradbury, Modelling Stress Constructs with Biomarkers: The Importance of the Measurement Model. Clinical and Experimental Medical Sciences, Vol.1, No.3, pp. 197-216, 2013.

[6] M. Nübling et al., Measuring psychological stress and strain at work – evaluation of the COPSOQ I Questionnaire in Germany. GMS Psycho-Social-Medicine, Vol.3, pp. 1-14. 2006.

[7] R. Subhani, L. Xia, and A. S. Malik. EEG signals to measure mental stress. 2nd International Conference on Behavioral, Cognitive and Psychological Sciences, 2011.

[8] D. Suvashis and K. Gamada. Evaluating Instantaneous Psychological Stress from Emotional Composition of a Facial Expression. ACIII 17 (2013): 480-492.

[9] F. Delmastro, F. D. Martino and C. Dolciotti. Physiological Impact of Vibro-Acoustic Therapy on Stress and Emotions through Wearable Sensors. 2018 IEEE International Conference on Pervasive Computing and Communications Workshops, 2018, pp. 621-626.

[10] D.F. Dinges, et al., Optical computer recognition of facial expressions associated with stress induced by performance demands, Aviation, Space, and Environmental Medicine 76 ,2005, pp. B172–B182.

[11] Yun Liu, Sing Du. Psychological stress level detection based on electro dermal activity. Behavioral Brain Research, Volume 341,2018, pages 50-53.

[12] H. Kataoka, H. Yoshida, A. Saito, M. Yasuda, andM. Osumi. Development of a skin temperature measuring system for non-contact stress evaluation. The 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1998, vol. 2, pp. 940–943.

[13] Q. Shan, T. Chen, L. Yao, Z. Wu, W. Wen and Q. Liu. Remote Detection and Classification of Human Stress Using a Depth Sensing Technique. 2018 First Asian Conference on Affective Computing and Intelligent

[14] O. Sakri, C. Godin, Q. Vila, E. Labyt, S. Charbonnier and A. Campagne. A Multi-User Multi-Task Model for Stress Monitoring from Wearable Sensors, 2018 21st International Conference on Information Fusion, Cambridge, 2018, pp. 761-766.

[15] A. Kolli, A. Fasih, F. Al Machot, and K. Kyamakya. Non-intrusive car driver's emotion recognition usingthermal camera. Joint 3rd Int'l Workshop on Nonlinear Dynam ics and Synchronization (INDS), 2011, pp.1–5.

[16] Q. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. PAMI, 29(6):915–928, 2007.

[17] A. Klaser, M. Marszałek, and C. Schmid. A spatio-temporal descriptor based on 3d-gradients. In BMVC, pages 275–1, 2008

[18] M. Jeong, Driver's Facial Expression Recognition in Real-Time for Safe Driving. Sensors 2018, 18, pp. 4270.

[19] P. Scovanner, S. Ali, and M. Shah. A 3-dimensional sift descriptor and its application to action recognition. In ACM MM, 2007.

[20] H. Jung, S. Lee, and J. Yim, Joint fine-tuning in deep neural networks for facial expression recognition, in Proceedings of the International Conference on Computer Vision, 2015, pp. 2983-2991.

[21] M. Liu, R. Wang, S. Li, S. Shan and Q. Chen. Deeply learning deformable facial action parts model for dynamic expression analysis, Asian Conference Computer Vision, 2014, pp.143-157..

[22] H. Yao, A. Dhall and J. Thiran. Detecting emotional stress from facial expressions for driving safety. 2014 IEEE International Conference on Image Processing, Paris, 2014, pp. 5961-5965.

[23] K. Zhang and Z. Zhang and Z. Li and Q. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. Xiao IEEE Signal Processing Letters. 2016, vol 23, pp. 1499-1503.

[24] Lucey, P. Cohn, J.F. Kanade, T. Saragih, J. Ambadar, Z. Matthews, I. The extended cohn-kanade dataset (ck++):A complete dataset for action unit and emotion-specified expression. In Proceedings of the IEEE computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, 2018, pp. 94–101.

[25] M. Jeong, Driver's Facial Expression Recognition in Real-Time for Safe Driving. Sensors 2018, 18, pp. 4270.

[26] J.Ouyang, S. Kawaai, E.G.H.Goh, S. Shen, W. Ding. Audio-visual emotion recognition using deep transfer learning and multiple temporal models, in Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017, pp.577-582.

[27] Hasani, B. Mahoor, M.H. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 2278–2288.D.F.

[28] Q. LI., Q. LIN, M. JIANG, Facial Expression Recognition with Cross-connect LeNet-5 Network. Acta Automatica Sinica, 2018, 44(1), pp.176-182.