

Using Facial Expression to Detect Emotion in E-learning System: A Deep Learning Method

Ai Sun¹, Ying-Jian Li²(✉), Yueh-Min Huang¹, and Qiong Li²

¹ Department of Engineering Science,
National Cheng Kung University, Tainan, Taiwan
sun_ai@hotmail.com, huang@mail.ncku.edu.tw

² School of Computer Science and Technology,
Harbin Institute of Technology, Harbin, China
hitlyj@stu.hit.edu.cn, qiongli@hit.edu.cn

Abstract. E-learning system is becoming more and more popular among students nowadays. However, the emotion of students is usually neglected in e-learning system. This study is mainly concerned about using facial expression to detect emotion in e-learning system. A deep learning method called convolutional neural network (CNN) is used in our research. Firstly, CNN is introduced to detect emotion in e-learning system based on using facial expression in this paper. Secondly, the training process and testing process of CNN are described. To learn about the accuracy of CNN in emotion detection, three databases (CK+, JAFFE and NVIE) are chosen to train and test the model. 10-fold cross validation method is used to calculate the accuracy. Thirdly, we introduce how to apply the trained CNN to e-learning system, and the design of e-learning system with emotion detection module is given. At last, we propose the design of an experiment to evaluate the performance of this method in real e-learning system.

Keywords: E-learning · Emotion detection · Facial expression · Deep learning

1 Introduction

In recent years, with the development of e-learning technology, more and more students choose to learn knowledge using the e-learning system. With e-learning system, people can learn what they need at any time and at anywhere, as long as there is network around them. As a result, learning through the e-learning system is becoming more and more popular.

One of the differences between e-learning and face-to-face education (traditional education) is the lack of affective factors. It is well known that emotion plays an important role in learning process [1, 2]. For example, the learning efficiency could be higher if the learner is in happy state and be lower if the learner is sad. In face-to-face education, instructors can observe the motion of students and adjust the teaching strategies in time. However, in e-learning system, teachers are usually separated from

students. In this case, it is difficult for the teachers to feel the emotion of students and detect the problems especially incurred by the confusion or sadness.

Due to the important influence of emotions on different areas, lots of approaches are proposed to detect emotions. All these approaches can be divided into two kinds: (1) using off-the-shelf software (FaceReader [3, 4], Xpress Engine or other alternative products) [5]; (2) using other technologies instead of software [6–9].

Moridis and Economides [5] used FaceReader to observe the emotions of students. The best advantage of using these off-the-shelf products is that the researchers don't need to learn about how to implement a classifier because the software does that automatically. However, the shortage is the recognition rate is not high enough. Brodny et al. [10] tested the performance of different softwares, the recognition rates of FaceReader on CK+ database [11] and MMI database [12] were 77.59% and 56.10%, while Xpress Engine achieved 87.60% and 45.12%, respectively.

Some researchers were not satisfied with the result of off-the-shelf software and proposed various kinds of approaches to classify emotions. Some common approaches are speech-based method [6], text-based method [7], facial-expression-based method [8, 9] and multi-modal-based method [13]. Chi-Chun et al. [6] applied decision tree, a kind of machine learning model, to detect emotions in speech, it was found the testing performance was improved by 3.37%. The emotions in text were evaluated by Chan and Chong [7] with a sentiment analysis engine. Decision tree approach was applied to classify emotions on CK+ database using facial expression by Salmam et al. [8] and the accuracy rate was 90%. Lee et al. [9] used SVM, another machine learning model, to evaluate the emotions on CK+ database and JAFFE database [14]. The recognized rates were 94.39% and 92.22%, respectively. Han and Wang [13] used multi-modal signals to detect emotions. As we can see, different approaches may result in various performances. The performance can be different when using different machine learning models. This suggests that the recognition rate can be improved if we use reasonable machine learning models. Motivated by this finding, we want to try more effective approaches to detect emotions.

In the past few years, deep learning has become a popular research area. For classification tasks, the basic of using general machine learning methods is feature extraction. Different features would result in different accuracy rates. However, feature extraction is complicated work. How many features should be extracted? What kinds of features are the most effective for classification? These are important questions that should be noticed. Fortunately, deep learning technology can learn the features automatically instead of extracting features manually. Lots of work using deep learning has been done till now and excellent result has been achieved in various areas [15–17]. In speech recognition domain, Dahl et al. [15] applied deep learning technology in large vocabulary speech recognition and the accuracy rate was improved by 9.2%. In machine translation area, Devlin et al. [16] proposed a deep learning method to improve the recognition rate of sentences and was regarded as the best paper in ACL in 2014. In digital image processing area, Convolutional Neural Networks (CNN), a kind of deep learning model, is widely accepted due to its wonderful performance. Sun et al. [17], applied CNN to implement face recognition and the accuracy was 99.53% (even better than that of human beings). So what if we apply deep learning technology (CNN) to detect emotion in e-learning system is the research purpose for this paper.

According to the analysis above, the main ideas of this research are as follows:

- (1) Introduce CNN to detect emotions based on using facial expression in e-learning system.
- (2) Design the framework to detect emotions using facial expression in e-learning system.
- (3) Design the experiment to test the performance of CNN in real e-learning system.

2 Related Works

In this section, some works that has been done by pioneers in facial expression recognition are going to be introduced.

2.1 Common Approaches in Facial Expression Recognition

Ekman et al. [18] proposed that facial expressions could be divided into six basic emotions: happiness, surprise, sadness, fear, anger, and disgust. Together with natural, these seven emotions are usually used in facial expression recognition. Three key crucial parts were included in facial expression recognition: (1) face detection; (2) feature extraction; (3) facial expression (emotion) classification. The frame work of emotion detection using facial expression is showed in Fig. 1.

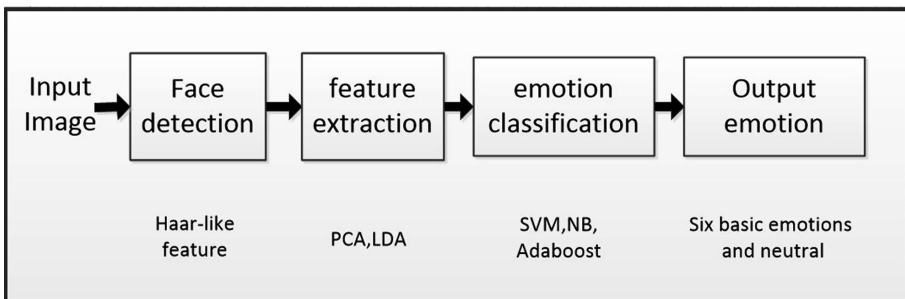


Fig. 1. The frame work of emotion detection using facial expression

Even in nowadays, lots of work has been done in facial expression recognition followed these three crucial parts. Face detection is a preprocess step of facial expression recognition. To describe the variance of facial expression, the FACS system was proposed by Friesen and Ekman [20]. They defined 44 Action Units (AU) and group of AU were used to describe a certain facial expression. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two usually used methods to extract features. PCA prefers the best features that can maintain the information of data while LDA chooses the features that can benefit classification well. There are various kinds of classifiers that can be used to complete facial expression classification task. SVM was applied on CK+ database by Lee et al. [9]. Eskill and Benli

[19] used SVM, Naive Bayesian (NB) and Adaboost methods to classify emotions using facial expression.

2.2 Deep Learning in Emotion Facial Expression Recognition

Because of the excellent performance in different areas, deep learning technology was employed to detect facial expressions by some researchers [21–24]. Soleymani et al. [21] applied Long-short-term-memory recurrent neural networks (LSTM-RNN) and conditional random fields (CCRF) to classify emotions using facial expression, as well as electroencephalogram (EEG) signals. The best result was 0.043 ± 0.025 measured by RMSE using facial features. Lv et al. [22] proposed to classify emotions using only facial components. In their work, deep belief network (DBN) was trained to learn important components first. After that, tanh function was applied to divide emotions into seven classes. In order to achieve the same purpose, two kinds of deep learning structures, deep neural network (DNN) and CNN, were implemented by Jung et al. [23]. There are two main differences between DNN and CNN: (1) CNN has Convolution kernel; (2) There are less parameters in CNN. As a result, CNN is easier to train compared with DNN. Liu et al. [24] proposed Deep Networks (AUDN) to achieve facial expression recognition.

2.3 Emotion Detection in E-learning System

In this subsection, some methods to classify emotions of students in e-learning system are introduced.

In order to detect the emotions of students in e-learning system, different kinds of signal were used, such as text [25, 26], facial [27, 28] and so on. Ortigosa et al. [25] presented Sentbuk, a Facebook application, to achieve emotion analysis with a accuracy of 83.27%. With Sentbuk, emotional change can also be detected. They described the usefulness of their work in e-learning system at the end of the paper. Qin et al. [26] paid special attention to negative emotion. Interactive text was used to detect negative emotion, after which music was selected to regulate the emotion of student. Sun et al. [27] trained a SVM classifier using images from CK database. 265 facial images of 42 people were used as testing sets and the accuracy was 84.55%. Most e-learning systems are mainly focus on single user detection. However, Ashwin et al. [28] proposed multi-user face detection based on e-learning system using SVM. It was the first work on multi-user face detection for e-learning Systems. The proposed method was tested on three databases: LFW, Fddb and YFD, and the accuracy were between 89% and 100%. To speed up the processing, Central Processing Unit (CPU) and Graphics Processing Unit (GPU) were used.

3 Method Description

In this section, the design of CNN is introduced first. Then, the way how to train and test the CNN is described. The framework of emotion detection in e-learning system is also given in this part. Finally, a rough design of testing the performance of CNN in e-learning is proposed.

3.1 Design of CNN

CNN has been used in many tasks by researchers. The different CNNs are with different parameters, such as the depth, the size of convolution kernel, the choice of activation function and so on. The structure of our CNN is shown in Fig. 2. There are three convolution layers with RELU (a kind of activation function), three pooling layers and two fully connected layers in our CNN. The convolution kernels are set at size of $5 * 5$. In each pooling layer, we choose max pooling as pooling method. The size of pooling window and pooling stride are set at $3 * 3$ and 2, respectively. The number of units in the first fully connected layer is set as 1024. In the last layer, there are 7 units with softmax function and each one is stand for a certain kind of emotion. The output is the probability of the input image's emotion.

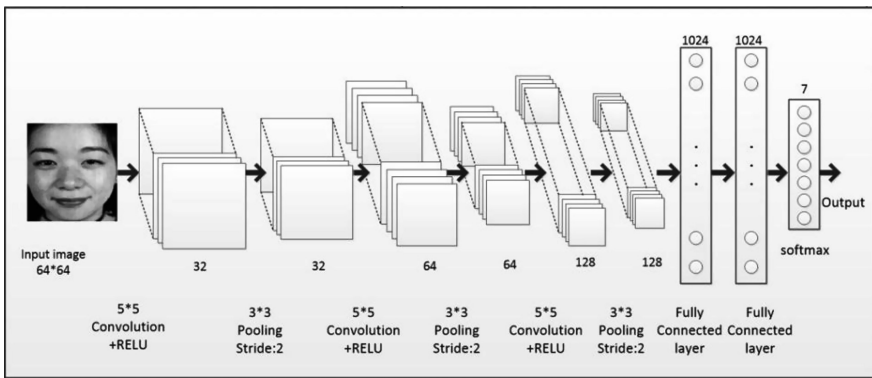


Fig. 2. Structure of our CNN

3.2 Training and Testing Framework Design

The same as other machine learning method, the use of CNN is divided into two steps: training step and testing step. The two steps are described in this subsection.

Training Step

In the training step of CNN, we decide to use stochastic gradient descent (SGD) algorithm. The advantage of SGD is SGD can reduce the over fitting phenomenon. To speed up the training process, the following methods can be used:

- (1) Use a batch of images to regulate the weight in CNN every time instead of a single one.
- (2) Use GPU. GPU is faster than CPU in image processing. There are many deep learning frameworks supporting GPU, such as Tensorflow, Caffe and Theano.
- (3) Detect the location of face and cut it out of the image before training. Using OpenCV library can implement this idea. The faces for training are put into training data set and others are put in testing data set.

The training process is shown in Fig. 3. CNN training needs lots of images with emotion labels. We aim at training a CNN that can be generally used to all kinds of faces. As a result, the training images should include both eastern faces and western faces. In our research, CK+ database, JAFFE database and NVIE database are chosen.

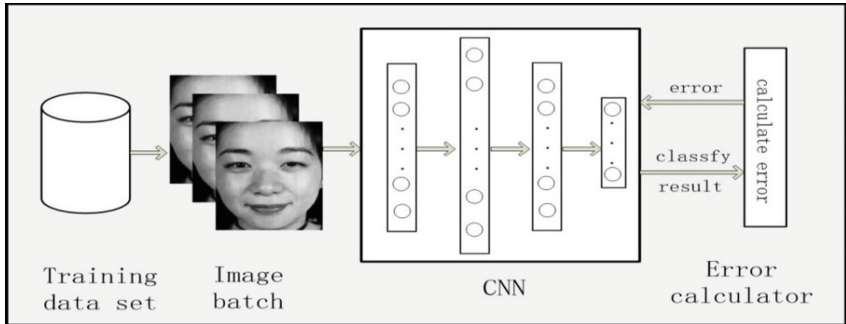


Fig. 3. The training process of CNN using SGD

CK+ database consist of 593 sequences from 123 subjects. There are eight kinds of emotions in this database: neutral, happy, surprise, angry, disgust, fear, sadness and contempt. In our research, however, we are going to use all the images but those with contempt emotion.

JAFFE database contains 213 facial expression images from 10 Japanese women. The extension of each image is “tiff” and the resolution ratio is $256 * 256$. There are seven emotion labels in this database: happy, surprise, angry, disgust, fear, sadness and neutral.

NVIE database [29] consists of natural visible and infrared facial expressions. In this database, there are lots of images from about 100 subjects under front, left and right illumination. The emotion is activated by eliciting videos. The number of emotion labels is 6. They are happy, surprise, angry, disgust, fear and sadness.

Testing Step

In order to learn about the performance of CNN on the data set, 10-fold cross validation is designed. The testing process of each image is shown in Fig. 4.

The classification result is compared with the label of the image and is recorded if the result is correct. The accuracy is the ratio of correct-classified number and testing set size. In 10-fold cross validation, the average of ten accuracy rates is set as the final accuracy rate.

3.3 Application in E-learning Systems

In this subsection, we introduce how to apply the trained CNN to detect emotion in e-learning system.

The e-learning system with emotion detector can be divided into two modules, emotion detection module and teaching strategy regulation module. At the beginning, the teacher teaches the students with a certain strategy. The emotion state of students

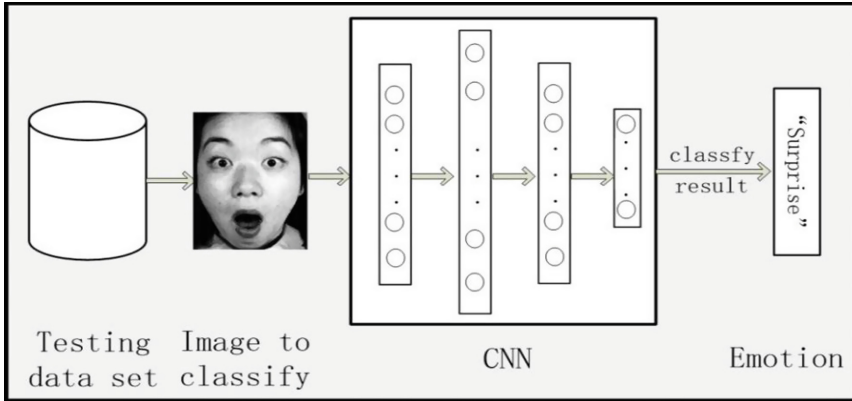


Fig. 4. The testing process of CNN

will be detected by emotion detection module. The detected emotion returns to teacher as feedback. The teacher regulates his teaching strategy and content according to the emotion of most students. The whole system is shown in Fig. 5.

In emotion detection module, the learners' facial expressions are shot by a camera first and we can get the image sequence. Second, we use OpenCV library to detect and cut the face area out of each image. All of the face areas form the face sequence. After that, the faces in face sequence are fed to CNN one by one. The emotion of each student was obtained.

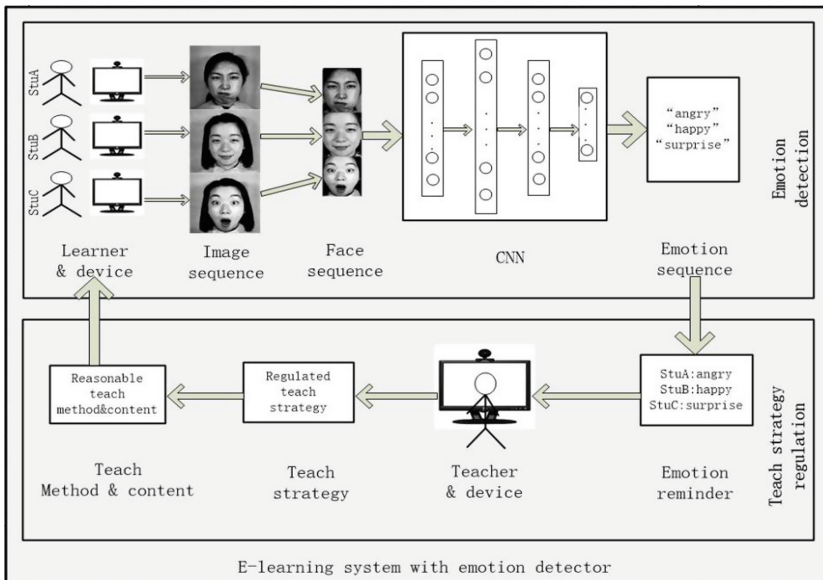


Fig. 5. E-learning system with emotion detector

In teaching strategy regulation module, the emotion sequence obtained from emotion detection module is input into the emotion reminder. The emotion reminder is a device that can remind the teacher of the students’ emotion state. The teaching strategy is regulated by the teacher according to the emotion of students. After teaching strategy regulation, new emotion may arise among students. The new emotion can be detected and sent to teacher and the teaching strategy may be changed further. As a result, this is a dynamic process. The advantage of this kind of e-learning system is that it can improve the efficiency of learning.

In our research, we are concerned the accuracy rate of the CNN in emotion classification in e-learning system. In last subsection, we designed 10-fold cross validation to test the performance of CNN on the testing dataset. However, how the CNN will perform in real e-learning system is not tested. The rough design of this idea is given as follows:

- (1) Choose 50 college students to learning a course via e-learning system.
- (2) For each student (for example, StuA), the testing process can be described by Fig. 6. StuA learn the course via computer or other electrical device. In the process of learning, his facial expression is recorded by a camera every 20 s. On one hand, the face area is cut out of each photo and is fed to CNN. We can get the classification result by CNN. On the other hand, the recorded photos are classified by StuA manually. We can get the classification result by human. The two results are compared and the number of correctly-classified photos is recorded.

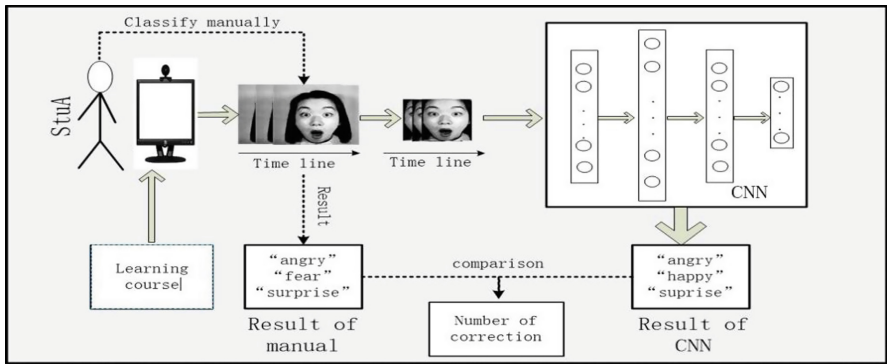


Fig. 6. Testing in real e-learning system

- (3) Calculate the accuracy using the following formula.

$$accuracy = \frac{N_{correct}}{N_{all}},$$

where $N_{correct}$ is the sum of correctly- classified photos for all students and N_{all} is the number of all recorded photos.

4 Conclusion and Future Work

E-learning systems are becoming more and more popular nowadays. However, emotion, an important factor in learning process, is often ignored in e-learning system. In this paper, we are mainly concerned about the emotion detection in e-learning system using facial

Due to the excellent performance of CNN, we introduced this model to detect emotion in e-learning system. First, we characterize the structure of our CNN. After that, the training and testing process of CNN was described. To train the CNN, three facial expression databases (CK+, JAFFE and NVIE) were selected. The working process of CNN in e-learning system was showed. At last, the rough design to test the accuracy in e-learning system is offered.

There are some questions left in this paper: (1) How does the proposed CNN perform on the testing data set and in e-learning system? (2) How long does it take to train this CNN? In near future, we are going to focus on these questions. For the first one, we are going to implement the CNN. The accuracy of CNN should be tested on testing data set given in this paper. Also, we will carry out the experiment in e-learning system according to the design from Subsect. 3.3. For the second question, CNN training may take a long time if we use CPU only. As a result, we are going to use GPU to speed up the training process and record the training time.

References

1. O'Regan, K.: Emotion and e-learning. *J. Asynchronous Learn. Netw.* **7**(3), 78–92 (2003)
2. Woolf, B.P.: Building Intelligent Interactive Tutors: Student Centered Strategies for Revolutionizing E-Learning, vol. 59. Morgan Kaufmann Inc., Burlington (2010). 337–379
3. Yi, M.J.D., Kuilenburg, H.V.: The FaceReader: online facial expression recognition. In: *Proceedings of the Measuring Behaviour*, vol. 30, pp. 589–590 (2005)
4. Yu, C.Y., Ko, C.H.: Applying FaceReader to recognize consumer emotions in graphic styles ☆. *Procedia Cirp* **60**, 104–109 (2017)
5. Moridis, C.N., Economides, A.A.: Affective learning: empathetic agents with emotional facial and tone of voice expressions. *IEEE Trans. Affect. Comput.* **3**(3), 260–272 (2012)
6. Chi-Chun, L., Emily, M., Carlos, B., Sungbok, L., Shrikanth, N.: Emotion recognition using a hierarchical binary decision tree approach. *Speech Commun.* **53**(9–10), 1162–1171 (2011)
7. Chan, S.W.K., Chong, M.W.C.: Sentiment analysis in financial texts. *Decis. Support Syst.* **94**, 53–64 (2017)
8. Salmam, F.Z., Madani, A., Kissi, M.: Facial expression recognition using decision trees. In: *IEEE International Conference on Computer Graphics, Imaging and Visualization*, pp. 125–130 (2016)
9. Lee, S.H., Kostas, P.K.N., Yong, M.R.: Intra-class variation reduction using training expression images for sparse representation based facial expression recognition. *IEEE Trans. Affect. Comput.* **5**(3), 340–351 (2014)
10. Brodny, G., Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., Wróbel, M.R.: Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions. In: *IEEE International Conference on Human System Interactions*, pp. 397–404 (2016)

11. 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: *IEEE Computer Vision and Pattern Recognition Workshops*, vol. 36, pp. 94–101 (2010)
12. Valstar, M.F., Pantic, M.: Induced disgust happiness and surprise: an addition to the MMI facial expression database. In: *Proceedings of the International Workshop on Emotion Corpora for Research on Emotion & Affect*, pp. 65–70 (2010)
13. Han, Z.Y., Wang, J., Bohai University: Emotion visualization method for speech and facial expression signals. *Electron. Des. Eng.* **24**, 146–149 (2016)
14. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: *1998 IEEE International Conference on Automatic Face and Gesture Recognition Proceedings*, pp. 200–205 (1998)
15. Dahl, G.E., Yu, D., Deng, L., Acero, A.: Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **20**(1), 30–42 (2012)
16. Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R.M., Makhoul, J.: Fast and robust neural network joint models for statistical machine translation. In *ACL (1)*, pp. 1370–1380 (2014)
17. Sun, Y., Liang, D., Wang, X., Tang, X.: Deepid3: face recognition with very deep neural networks. *arXiv preprint [arXiv:1502.00873](https://arxiv.org/abs/1502.00873)* (2015)
18. Ekman, P., Rolls, E.T., Perrett, D.I., Ellis, H.D.: Facial expressions of emotion: an old controversy and new findings. *Philos. Trans. R. Soc. B: Biol. Sci.* **335**(1273), 63–69 (1992)
19. Eskil, M.T., Benli, K.S.: Facial expression recognition based on anatomy. *Comput. Vis. Image Underst.* **119**, 1–14 (2014)
20. Friesen, E., Ekman, P.: *Facial action coding system: a technique for the measurement of facial movement*, Palo Alto (1978)
21. Soleymani, M., Asghari-Esfeden, S., Fu, Y., Pantic, M.: Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans. Affect. Comput.* **7**(1), 17–28 (2016)
22. Lv, Y., Feng, Z., Xu, C.: Facial expression recognition via deep learning. In: *IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 303–308 (2014)
23. Jung, H., Lee, S., Park, S., Kim, B., Kim, J., Lee, I., Ahn, C.: Development of deep learning-based facial expression recognition system. In: *IEEE Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pp. 1–4 (2015)
24. Liu, M., Li, S., Shan, S., Chen, X.: Au-inspired deep networks for facial expression feature learning. *Neurocomputing* **159**, 126–136 (2015)
25. Ortigosa, A., Martín, J.M., Carro, R.M.: Sentiment analysis in Facebook and its application to e-learning. *Comput. Hum. Behav.* **31**, 527–541 (2014)
26. Qin, J., Zheng, Q., Li, H.: A study of learner-oriented negative emotion compensation in e-learning. *Educ. Technol. Soc.* **17**(4), 420–431 (2014)
27. Sun, J.M., Pei, X.S., Zhou, S.S.: Facial emotion recognition in modern distant education system using SVM. In: *IEEE International Conference on Machine Learning and Cybernetics*, vol. 6, pp. 3545–3548 (2008)
28. Ashwin, T.S., Jose, J., Raghu, G., Reddy, G.R.M.: An e-learning system with multifacial emotion recognition using supervised machine learning. In: *IEEE International Conference on Technology for Education*, pp. 23–26 (2015)
29. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., et al.: A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Trans. Multimedia* **12**(7), 682–691 (2010)