

Retraction

Retracted: Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation

Complexity

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] W. Wang, K. Xu, H. Niu, and X. Miao, "Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation," *Complexity*, vol. 2020, Article ID 4065207, 9 pages, 2020.

Research Article

Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation

Weiqing Wang,¹ Kunliang Xu,¹ Hongli Niu,¹ and Xiangrong Miao² 

¹School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

²Beijing Energy Conservation and Environmental Protection Center, Beijing 101199, China

Correspondence should be addressed to Xiangrong Miao; miaoxr@bjbeec.cn

Received 14 July 2020; Revised 26 August 2020; Accepted 28 August 2020; Published 11 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Weiqing Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online education has developed rapidly due to its irreplaceable convenience. Under the severe circumstances caused by COVID-19 recently, many schools around the world have delayed opening and adopted online education as one of the main teaching methods. However, the efficiency of online classes has long been questioned. Compared with traditional face-to-face classes, there is a lack of direct, timely, and effective communication and feedback between teachers and students in the online courses. Previous studies have shown that there is a close and stable relationship between a person's facial expressions and emotions generally. From the perspective of computer simulation, a framework combining a face expression recognition (FER) algorithm with online courses platforms is proposed in this work. The cameras in the devices are used to collect students' face images, and the facial expressions are analyzed and classified into 8 kinds of emotions by the FER algorithm. An online course containing 27 students conducted on Tencent Meeting is used to test the proposed method, and the result proved that this method performs robustly in different environments. This framework can also be applied to other similar scenarios such as online meetings.

1. Introduction

Facial expression is one of the most powerful, natural, and universal signals for human beings to convey their emotional states and intentions regardless of national borders, race, and gender [1, 2], and there were multitudinous related applications such as the health management [3], aided driving [4, 5], and others [6–9]. In the earlier researches on facial expressions of emotion, Ekman and Friesen argued that human beings perceive certain basic emotions in the same way regardless of their cultural background, and they defined the typical facial expressions into 6 categories: anger, disgust, fear, happiness, sadness, and surprise [10, 11]. Based on studies of Ekman and Friesen, Ekman and Heider [12–14], Matsumoto [15] provided enough proofs for another universal facial expression, contempt. Additionally, FER2013 [16], a large-scale and unconstrained database introduced in the ICML 2013 Challenges in Representation Learning,

labeled its facial images into anger, disgust, fear, happiness, sadness, surprise, and neutral, which has been widely used in designing facial expression recognition (FER) systems. In subsequent researches, although researchers introduced many models that can provide a wider range of emotions to deal with the complexity and subtlety of facial expressions [17–20], the classification that describes basic emotions discretely is still the most widely used method in FER due to its generality and intuitive definition of facial expressions [21], and Figure 1 displays the 8 basic facial expression phenotypes from datasets CK+ [22] and FER2013 [16].

For determining facial expressions, Ekman and Friesen [23] proposed a Facial Action Coding System (FACS) for determining facial expressions, which is based on a fact that expressions result from the change of facial parts. With the assistance of the computers, more advanced methods have been proposed during the last decades [24], and the feature points can be seen in Figure 2.



FIGURE 1: Basic facial expression phenotypes from CK+ and FER2013: from left to right are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise.

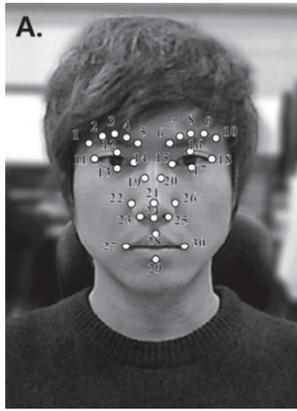


FIGURE 2: Facial analysis points used for facial expression [25].

With the development of artificial intelligence and deep learning, numerous FER algorithms have been proposed to deal with the expression information in facial representations, which has improved the accuracy of recognition gradually and achieved better performance than traditional methods [26, 27]. The tasks of FER can be mainly divided into two categories: static images (represented by photographs) [28–30] and dynamic sequence (represented by videos) [31–33] that take the dynamic relationship between the continuously changing images into account and therefore pose additional challenges than the former. In addition to the vision-based methods, other biometric techniques [34, 35] can also be adopted to assist the recognition of expression.

Sufficient labeled training databases that include as many variations of the populations and environments as possible are important for researchers to design and test a FER model or system; the existing databases are mainly divided into controlled and uncontrolled. On one hand, the controlled databases, represented by CK+ [22], Jaffe [36], and MMI [37], are collected from laboratory environments with sufficient light and simple backgrounds. Nowadays, because most actual scenes are complex and changeable due to factors such as lighting, FER in laboratory or controlled environments is generally considered to be of little practical significance and used mainly for the proof of concept of features extraction and classification methods. On the other hand, the uncontrolled databases, such as FER2013 [15] and AFEW [38], are collected from complex environments with vastly different backgrounds, occlusions, and illuminations; these scenes are more similar to the actual situations and have been increasingly used in more and more researches.

Limited by the hardware and insufficient processing capability, the majority of the traditional methods for FER

employed hand-craft features or shallow learning, such as local binary patterns (LBP) [28] and nonnegative matrix factorization (NMF) [39]. With the development of processing capabilities and computer simulation, all kinds of machine learning algorithms, such as Artificial Neural Networks (ANNs), Support Vector Machines (SVM), and Bayesian classifiers, were applied to FER, and the high accuracy has been verified in controlled environments so that the faces can be detected effectively. However, these methods were weak in generalization ability while this is the key to evaluate the practicality of a model [40]. Deep learning algorithms can solve this problem, and it is also robust in the uncontrolled environments. Recent works have shown that convolutional neural networks (CNNs), because of their effectiveness in feature extraction and classification tasks, performed well in addressing the computer vision problems especially in FER [41, 42], and numerous models based on CNN structure are proposed constantly and have achieved better results than previous methods. Simonyan and Zisserman [43] adopted an architecture of very small (3×3) convolution filters to conduct a comprehensive evaluation of networks with increasing depth and the two best-performing ConvNet models were available publicly to facilitate the further research in this field. By increasing the depth and width of the network while keeping the computational budget constant, Szegedy et al. [44] introduced a deep convolutional neural network architecture named “Inception” in which the utilization of the computing resources can be improved significantly, and Jahandad et al. [45] worked on 2 convolutional neural network architectures (Inception-v1 and Inception-v3) based on “Inception” and proved that these 2 models performed better than others, and Inception-v1 with 22-layer-deep network performed better than 42-layer-deep Inception-v3 network when facing low-resolution input images and 2D images of signatures; however, Inception-v3 outperformed in ImageNet challenge. The general trend of neural networks is to increase the depth of the network and the width of layer. In theory, the deeper the neural network models, the stronger the learning capabilities, but the more difficult to train. He et al. [46] proposed a residual learning framework to reduce training difficulty of deeper networks and proved thoroughly that these residual networks are easier to optimize while increasing accuracy from the considerably increased depth. In addition, a part of researchers proposed that the accuracy of recognition can be further improved by combining CNNs with recurrent neural networks (RNNs) in which the CNNs are adopted as the inputs to RNNs [47, 48].

During the past decades, online education has developed rapidly whether at universities or training institutions [49], which offers potential application opportunities for FER.

Significantly different from the traditional face-to-face courses, online courses are often considered of less constraining force and effective communication, which will inevitably lead to faculty's suspicions towards this novel educational method [50, 51] while there are several studies that argue that the students' learning outcomes achieved by online education may be comparable to traditional face-to-face courses [52, 53], except for the skills that require optimum precision and a greater degree of haptic awareness [54]. It is undeniable that the rapid growth of online education can effectively provide the convenience and flexibility for more students, so it also has broad development space in the future; therefore, how to ensure that students keep the same level of concentration and learning efficiency as the traditional courses during online education is critical to promote the further development of online education.

In brief, the main contribution of this paper is as follows. By combining the existing online education platforms with facial expression recognition model based on the architecture of convolutional neural network, this work proposed a framework that enables real-time monitoring of students' emotions in online courses and ensures that the feedback expressed by facial expression can be provided to teachers timely, so that they can flexibly adjust the teaching programs and ultimately improve the quality and efficiency of online education.

2. Proposed Framework

The framework mainly consists of two parts: online courses platforms, in this paper we took Tencent Meeting as an example for mode testing, and a deep learning model based on CNN, inspired by Kuo et al. [27], before which it is noted that the original images collected from online courses need to be preprocessed, including face detection, alignment, rotation, and resize, according to the different elements in the original images. Figure 3 exhibits the process of the FER, and the detailed steps of the proposed framework are as follows: first, the cameras built in the electronic devices are utilized to capture the facial images of the attending students. Second, the facial expression recognition algorithm trained by the standard facial expression database is employed to detect the faces and classify the facial expressions in terms of anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral. Third, the histogram of probability distribution about the expression is plotted and provided for the teacher so that the teaching plan can be adjusted timely.

2.1. Online Education Platforms. The advances in technological delivery modalities have spawned a large number of online education platforms and greatly improved the flexibility of education, enabling teachers to adopt diverse technical methods to assist teaching without worrying about the limit on the number of students in traditional classroom-based course, and students in different regions can communicate in real time without having to consider the traffic and other issues. The same teaching materials as traditional

classes can be uploaded to these platforms for students' reference. Currently, in platforms that have online teaching functions, such as DingTalk, Zoom, and Rain classroom, teachers can adopt the method of video meeting and take advantage of the camera built in devices to capture and recognize students' facial expressions in real time. The captured images will be preprocessed and then used as the input of CNN.

2.2. The Preprocessing Based on IntraFace. Effective preprocessing can reduce the interference of face-like objects in the background when detecting faces in an image and then standardize the face images according to the heuristic knowledge, which will effectively improve the efficiency of the deep learning model. We employed IntraFace [55], a publicly available software package that integrates algorithms for facial feature tracking, head pose estimation, facial attribute detection, etc., as the tool of preprocessing. As shown in Figure 4, IntraFace can also be used to detect multiple faces at the same time. The key features of each face including eyebrows, eyes, nose tip, and mouth can be recognized effectively, and the facial expression can be detected by rectangular outlines accordingly; these outlines are constructed by the feature points at the edge of every face, including the uppermost and the lowermost, which determined the vertical width, and the rightmost and the leftmost, which determined the horizontal width of the face image. In order to prevent the omission of facial information while reducing the noise of background, we enlarge the rectangular outlines by 1.05 times to cover more facial content. Furthermore, considering that the size of images input into the learning model is preset to 48×48 , the detected images will be rotated with nose tip as the center and resized appropriately to make it consistent with the input size.

2.3. The Learning Model Based on CNN. The architecture of the applied deep learning model based on CNN is illustrated in Figure 5, which referred to the research results proposed by Kuo et al. [27], and the prior performance of this model in FER over the other similar has also been proven. After a convolutional layer of 32 feature maps, the input layer is followed by 2 blocks, which consists of 2 convolutional layers and 1 max-pooling layer with 64 feature maps separately. And the size of kernels in the first convolutional layer is set to 3×3 , the second is 5×5 , the max-pooling layers both consist of a kernel of size 2×2 and stride 2, and as a consequence, the input image will be compressed to a quarter. And there are 2 following fully connected layers of 2048 and 1024 neurons, respectively, in which Rectified Linear Units (ReLUs) [56–59] are adopted as the activation function. In order to prevent overfitting, a Dropout is added after each of the 2 fully connected layers, which will release a part of neurons according to the presetting drop-probability; in this paper, the 2 values are both set to 0.5. The following output layer is composed of 8 units, and softmax [60] is adopted as the activation function to classify the expressions examined in terms of anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral.

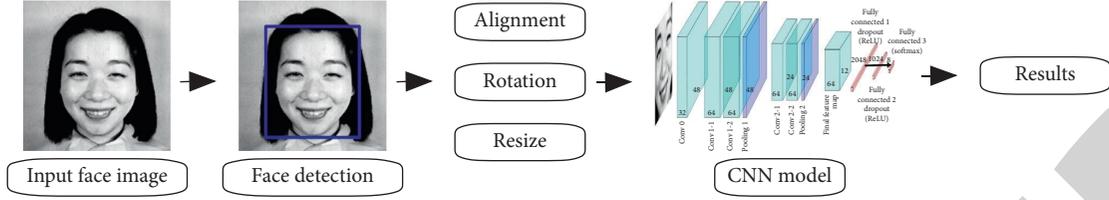


FIGURE 3: The process of FER.

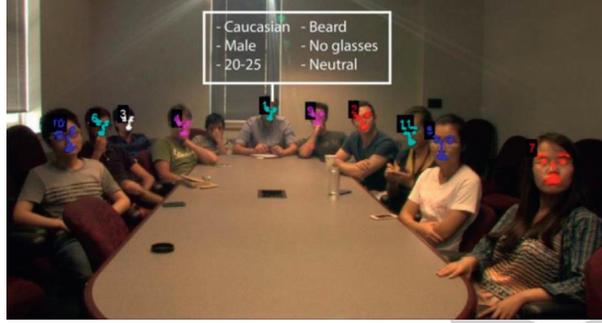


FIGURE 4: Automatic output of IntraFace to measure audience reaction [55].

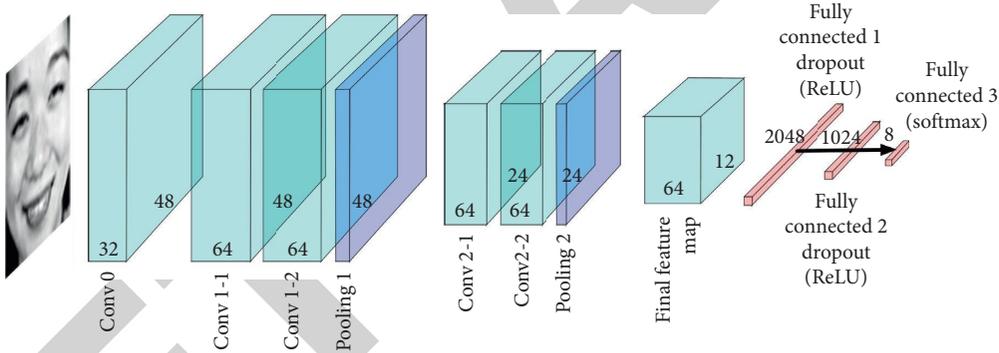


FIGURE 5: The structure of the applied CNN model.

The proposed model was trained by databases Jaffe, CK+, and FER2013 that include the above 8 basic expressions at the same time. Because the small FER databases usually only contain a few hundreds of images, obviously this is not enough for model training; we adopt online augmentation strategy with both horizontal flipping and random shifting to increase the images of training sets. More details about the CNN model are given in Table 1.

And in this model, the output size N of each convolutional layer can be formulated as

$$N = \frac{I - F + 2P}{S} + 1, \quad (1)$$

where I, F, P , and S denote the input size, kernel size, padding size, and stride size, respectively.

In each max-pooling layer, the padding size is 0, and the output size can also be expressed as

$$N = \frac{I - F}{S} + 1. \quad (2)$$

Rectified Linear Units (ReLU) are adopted as the activation function in the convolutional layers and max-pooling layers to avoid gradient explosion and ensure faster convergence speed during the back-propagation operation, which can be formulated as

$$f(x) = \begin{cases} 0, & x \leq 0; \\ x, & x > 0. \end{cases} \quad (3)$$

Softmax is used as the activation function in the output layer, the input of which is the matrix $z_i = w_i^T x + b_i$

TABLE 1: The parameters of the applied CNN model.

Layers	Kernel number	Kernel size	Steps	Image size
Input	0	0	None	$48 \times 48 \times 1$
Convolutional 0	32	1×1	1	$48 \times 48 \times 32$
Convolutional 1-1	64	3×3	1	$48 \times 48 \times 64$
Convolutional 1-2	64	5×5	1	$48 \times 48 \times 64$
Pooling 1	0	2×2	2	$24 \times 24 \times 64$
Convolutional 2-1	64	3×3	1	$24 \times 24 \times 64$
Convolutional 2-2	64	5×5	1	$24 \times 24 \times 64$
Pooling 2	0	2×2	2	$12 \times 12 \times 64$
Fully connected 1		Dropout = 0.5		$1 \times 1 \times 2048$
Fully connected 2		Dropout = 0.5		$1 \times 1 \times 1024$
Output	0	0	None	$1 \times 1 \times 8$

output from the fully connected layers. The formulation is given as

$$S_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}, \quad (4)$$

where K represents the output dimensions of the layer, meaning there are K kinds of results, and S_i represents the probability of result i , $i = 1, 2, \dots, K$.

And Softmax loss, which is used for gradient derivation and update, can be calculated by the following:

$$L = - \sum_{i=1}^K y_i \log S_i,$$

$$\frac{\partial L}{\partial z_k} = \frac{\partial L}{\partial S_i} \frac{\partial S_i}{\partial z_k},$$

$$\frac{\partial L}{\partial S_i} = \sum_{i=1}^K \left(-\frac{y_i}{S_i} \right),$$

$$\frac{\partial S_i}{\partial z_k} = \begin{cases} \frac{e^{z_i} \sum_{k=1}^K e^{z_k} - e^{z_i} e^{z_k}}{(\sum_{k=1}^K e^{z_k})^2} = S_i(1 - S_k), & i = k; \\ \frac{e^{z_i} \cdot e^{z_k}}{(\sum_{k=1}^K e^{z_k})^2} = -S_i S_k, & i \neq k, \end{cases} \quad (5)$$

$$\frac{\partial L}{\partial z_k} = \frac{\partial L}{\partial S_i} \frac{\partial S_i}{\partial z_k} = - \sum_{i=1}^K \frac{y_i}{S_i} \cdot \frac{\partial S_i}{\partial z_k}$$

$$= -S_k(1 - S_k) \cdot \frac{y_k}{S_k} + \sum_{i \neq k} \left(-\frac{y_i}{S_i} \right) \cdot (-S_i S_k)$$

$$= -y_k + y_k S_k + \sum_{i \neq k} y_i S_k = S_k - y_k,$$

where L denotes the loss function and y_i is the label variable, the value of which is 1 or 0 according to whether the output is consistent with the actual value.

3. Experiment and Results

In order to test the performance of the proposed framework in practical applications, we captured an image that includes 27 people from an online meeting held on Tencent Meeting and input it into the CNN model. This image is taken before the end of the meeting; the moderator was making a concluding speech in a pleasant atmosphere. In addition, everyone was told that the meeting was coming to an end, according to the experiment conducted by Tonguç and Ozkara [61], students' happiness will be significantly improved within a few minutes before the end of a lecture, so under the similar circumstance, it can be inferred that the emotions presented by most of the faces in this image are happy or neutral. Figure 6 shows the input (left) and output (right) images of the CNN model. It can be seen clearly from the result that all the faces were recognized and marked by the rectangular outlines, and the responding facial expressions were also labeled. In the total 27 faces, 10 were labeled "happy," 15 were labeled "neutral," and 2 were labeled "sad," noting that the 2nd image in the last line and the 3rd image in the 4th line from last, marked by red outlines, were not detected by the outlines precisely; the reason may be that the 2 face images are so incomplete that the features presented are too insufficient to recognize. Figure 7 shows the probability distribution histogram of emotions, from which we can observe the overall emotions intuitively and judge the emotional state of class accordingly. It is worth noting that the probability of happiness is significantly higher than that of neutral in this figure, while the faces labeled "happy" are less than "neutral" as exhibited in Figure 6. The difference can be explained as follows: there may be features of multiple expressions on a face at the same time; the expression presented on this face will be labeled according to the most likely expression decided by these features, but the overall expression of an image including multiple faces is decided by the sum of various expression features contained in each face. In some faces that are marked as "happy," the probability of happiness may be much higher than neutral, while in some faces that are marked as "neutral," the probability of happiness may be only slightly lower than neutral. Overall, the result of this experiment can provide favorable support for the performance of the model when applied to real environment.

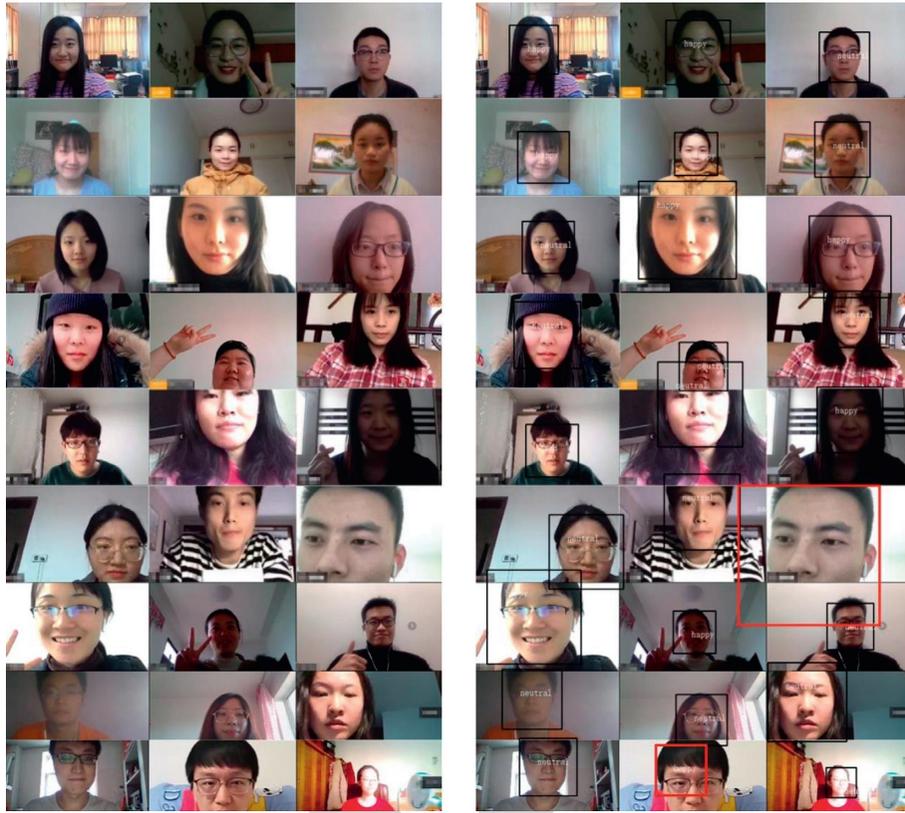


FIGURE 6: The input (a) and output (b) images of the FER model.

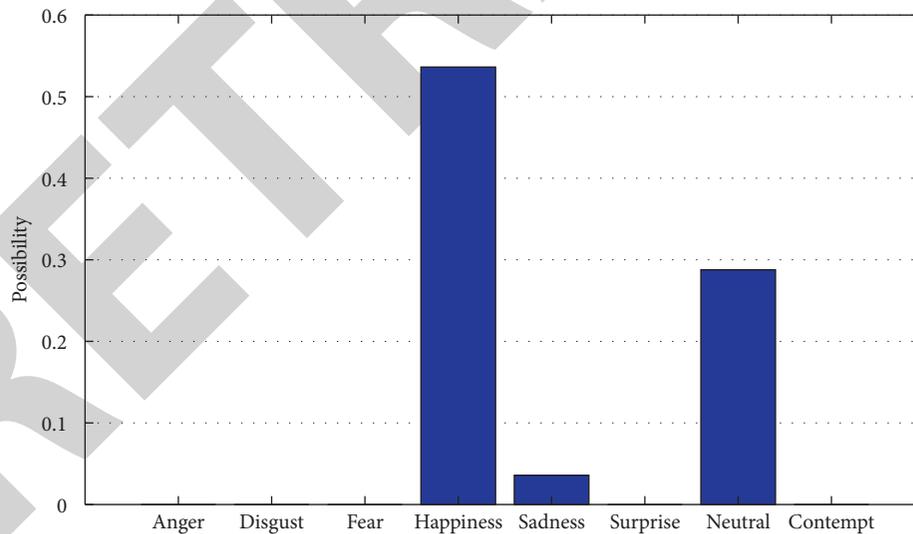


FIGURE 7: The probability distribution of emotions.

4. Conclusion and Discussion

In this study, by combining the online courses platforms and a compact deep learning model based on the architecture of CNN, we construct a framework to analyze students' emotions according to their facial expressions from the perspective of computer simulation. The overall result can be

presented in a histogram intuitively, and teachers can adjust their teaching strategies accordingly to improve the efficiency of online teaching.

With reference to the studies of Ekman et al. and FER2013, the emotions were classified into anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral in the proposed framework. To verify the applicability of this

framework in a real environment, we captured an image including the facial images of all participants at one time in a real online meeting; there were 12 participants in this meeting, and the captured time was determined at the end of meeting. A total of 12 faces were captured, of which 11 were effectively recognizable faces that contain enough feature points. By inputting this image into the applied CNN model, we obtained the emotional tags for each valid face and got the overall emotion at that time. It has been proved that the framework has good applicability in practical activities and plays a positive role in solving the problems, such as the lack of binding force on students, and teachers cannot get timely feedback. Ultimately, it will contribute to improving the quality of online education.

Despite the above benefits, there is still much room for improvement in this framework and its applications. From the perspective of technologies, with the development of computer simulation, algorithms with better performance and shorter operation time, including preprocessing and deep learning models, will be continuously developed over time. For instance, the image preprocessing contains face detection, alignment, rotation, and resize, but when facing problems, such as backlight, shadows, and facial incompleteness, caused by complex environments, these current methods are always powerless, and these shortcomings may be solved in the future. What is more, although the CNN model in the proposed framework currently performs well, it will be replaced by models with higher learning capabilities and higher classification accuracy in the future. In order to ensure the competitiveness of the framework in a longer period, it needs to be adjusted and maintained regularly, and more advanced algorithms and technologies should be adopted to update it.

In addition, with a large number of participants in the online courses, we have no way to ensure that everyone keeps the high level of concentration, and then students' expressions may not fully represent their emotions due to these subjective factors. Taking measures like setting thresholds can filter out some invalid information and highlight the main emotions in the image. Finally, the teaching efficiency can be improved as a result.

Data Availability

The data used in this manuscript can be accessed by readers via the authors' BaiduPan at <https://pan.baidu.com/s/1dbKUfeKp5joeYh4wSOU7Qw> with the extraction code "qin6."

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Study on Influence of Chinese Stock Market under Economic Uncertainty (No. FRF-DF-20-11).

References

- [1] C. Darwin and P. Prodger, *The Expression of the Emotions in Man and Animals*, Oxford University Press, Oxford, MA, USA, 1998.
- [2] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001.
- [3] R. Donoso, C. San Martín, and G. Hermosilla, "Reduced isothermal feature set for long wave infrared (LWIR) face recognition," *Infrared Physics & Technology*, vol. 83, pp. 114–123, 2017.
- [4] T. Liu, H. Liu, Z. Chen et al., "FBRDLR: fast blind reconstruction approach with dictionary learning regularization for infrared microscopy spectra," *Infrared Physics & Technology*, vol. 90, pp. 101–109, 2018.
- [5] Z. Huang, H. Fang, Q. Li et al., "Optical remote sensing image enhancement with weak structure preservation via spatially adaptive gamma correction," *Infrared Physics & Technology*, vol. 94, pp. 38–47, 2018.
- [6] Y. Bi, M. Lv, Y. Wei, N. Guan, and W. Yi, "Multi-feature fusion for thermal face recognition," *Infrared Physics & Technology*, vol. 77, pp. 366–374, 2016.
- [7] H. Liu, Z. Zhang, S. Liu, J. Shu, T. Liu, and T. Zhang, "Blind spectrum reconstruction algorithm with L_0 -sparse representation," *Measurement Science and Technology*, vol. 26, no. 8, pp. 085501–085507, 2015.
- [8] H. Wu, Y. Liu, L. Qiu, and Y. Liu, "Online judge system and its applications in C language teaching," in *Proceedings of the International Symposium on Educational Technology (ISET)*, pp. 57–60, Beijing, China, July 2016.
- [9] T. Liu, Z. Chen, H. Liu, Z. Zhang, and Y. Chen, "Multi-modal hand gesture designing in multi-screen touchable teaching system for human-computer interaction," in *Proceedings of the Second International Conference on Advances in Image Processing*, pp. 100–109, Chengdu China, June 2018.
- [10] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.
- [11] P. Ekman, "Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique," *Psychological Bulletin*, vol. 115, no. 2, pp. 268–287, 1994.
- [12] P. Ekman and W. V. Friesen, "A new pan-cultural facial expression of emotion," *Motivation and Emotion*, vol. 10, no. 2, pp. 159–168, 1986.
- [13] P. Ekman and W. V. Friesen, "Who knows what about contempt: a reply to Izard and Haynes," *Motivation and Emotion*, vol. 12, no. 1, pp. 17–22, 1988.
- [14] P. Ekman and K. G. Heider, "The universality of a contempt expression: a replication," *Motivation and Emotion*, vol. 12, no. 3, pp. 303–308, 1988.
- [15] D. Matsumoto, "More evidence for the universality of a contempt expression," *Motivation and Emotion*, vol. 16, no. 4, pp. 363–368, 1992.
- [16] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," *Neural Information Processing*, Springer, Berlin, Germany, pp. 117–124, 2013.
- [17] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.
- [18] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: a survey of registration, representation, and

- recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1113–1133, 2015.
- [19] B. Martinez and M. F. Valstar, "Advances, challenges, and opportunities in automatic facial expression recognition," in *Advances in Face Detection and Facial Image Analysis*, pp. 63–100, Springer, Cham, Switzerland, 2016.
- [20] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: current trends and future directions," *Image and Vision Computing*, vol. 31, no. 2, pp. 120–136, 2013.
- [21] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, In press.
- [22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 94–101, San Francisco, CA, USA, July 2010.
- [23] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environmental Psychology and Nonverbal Behavior*, vol. 1, no. 1, pp. 56–75, 1976.
- [24] S. Bayrakdar, D. Akgün, and İ. Yücedağ, "A survey on automatic analysis of facial expressions," *Sakarya University Journal of Science*, vol. 20, no. 2, pp. 383–398, 2016.
- [25] Y. Kim, S. Kang, S. Lee et al., "Efficiently detecting outlying behavior in video-game players," *PeerJ*, vol. 3, p. e1502, 2015.
- [26] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *The Visual Computer*, vol. 36, no. 2, pp. 405–412, 2020.
- [27] C. M. Kuo, S. H. Lai, and M. Sarkis, "A compact deep learning model for robust facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2202–2208, Salt Lake City, UT, USA, June 2018.
- [28] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.
- [29] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1805–1812, Columbus, OH, USA, June 2014.
- [30] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proceedings of the IEEE Winter Conference Applications of Computer Vision (WACV)*, pp. 1–10, Lake Placid, NY, USA, March 2016.
- [31] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.
- [32] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the IEEE International Conference Computer Vision (ICCV)*, pp. 2983–2991, Santiago, Chile, December 2015.
- [33] X. Zhao, X. Liang, L. Liu et al., "Peak-piloted deep network for facial expression recognition," *Computer Vision—ECCV 2016*, Springer, Cham, Switzerland, pp. 425–442, 2016.
- [34] I. Bhardwaj, N. D. Londhe, and S. K. Kopparapu, "A spoof resistant multibiometric system based on the physiological and behavioral characteristics of fingerprint," *Pattern Recognition*, vol. 62, pp. 214–224, 2017.
- [35] J. Zhao, R. Ma, and X. Zhang, "Speech emotion recognition based on decision tree and improved SVM mixed model," *Transactions of Beijing Institute of Technology*, vol. 37, no. 4, pp. 386–395, 2017.
- [36] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "The Japanese female facial expression (Jaffe) database," in *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, pp. 14–16, Nara, Japan, April 1998.
- [37] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Amsterdam, Netherland, July 2005.
- [38] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Collecting large, richly annotated facial-expression databases from movies," *IEEE Multimedia*, vol. 19, no. 3, pp. 34–41, 2012.
- [39] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 1, pp. 38–52, 2011.
- [40] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon, "Emotion recognition in the wild challenge 2014: baseline, data and protocol," in *Proceedings of the 16th International Conference on Multimodal Interaction*, pp. 461–466, ACM, Istanbul Turkey, November 2014.
- [41] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, "Attention mechanism-based CNN for facial expression recognition," *Neurocomputing*, vol. 411, pp. 340–350, 2020.
- [42] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, pp. 82–92, 2019.
- [43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [44] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.
- [45] A. Jahandad, S. M. Sam, K. Kamardin, N. N. Amir Sjarif, and N. Mohamed, "Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3," *Procedia Computer Science*, vol. 161, pp. 475–483, 2019.
- [46] 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, Las Vegas, NV, USA, June 2016.
- [47] I. Banerjee, Y. Ling, M. C. Chen et al., "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," *Artificial Intelligence in Medicine*, vol. 97, pp. 79–88, 2019.
- [48] H. Hasan, B. Huang, and G. Tian, "Facial expression recognition based on deep convolution long short-term memory networks of double-channel weighted mixture," *Pattern Recognition Letters*, vol. 131, pp. 128–134, 2020.
- [49] I. Allen and J. Seaman, "Digital compass learning: distance education enrollment report 2017," Babson Survey Research Group, Babson Park, MA, USA, 2017.
- [50] P. Shea, T. Bidjerano, and J. Vickers, "Faculty attitudes toward online learning: failures and successes," SUNY Research Network, Albany, NY, USA, 2016.

- [51] D. Lederman, *Conflicted Views of Technology: A Survey of Faculty Attitudes*, Inside Higher Education, Washington, WA, USA, 2018.
- [52] C. L. Cason and J. Stiller, "Performance outcomes of an online first aid and CPR course for laypersons," *Health Education Journal*, vol. 70, no. 4, pp. 458–467, 2010.
- [53] S. Maloney, M. Storr, S. Paynter, P. Morgan, and D. Ilic, "Investigating the efficacy of practical skill teaching: a pilot-study comparing three educational methods," *Advances in Health Sciences Education*, vol. 18, no. 1, pp. 71–80, 2013.
- [54] E. Dolan, E. Hancock, and A. Wareing, "An evaluation of online learning to teach practical competencies in undergraduate health science students," *The Internet and Higher Education*, vol. 24, pp. 21–25, 2015.
- [55] F. D. I. Torre, W. Chu, and X. Xiong, "IntraFace," in *Proceedings of the IEEE International Conference on Automatic Face And Gesture Recognition*, Lille, France, May 2015.
- [56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, Nevada, December 2012.
- [57] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognition*, vol. 92, pp. 64–81, 2019.
- [58] L. Zhang, P. Shen, X. Peng et al., "Simultaneous enhancement and noise reduction of a single low-light image," *IET Image Processing*, vol. 10, no. 11, pp. 840–847, 2016.
- [59] Q. Zhu, J. Zhang, W. Wei et al., "A neuro-heuristic approach for recognition of lung diseases from X-ray images," *Expert Systems with Applications*, vol. 126, pp. 218–232, 2019.
- [60] C. S. Chin, J. Si, A. S. Clare, and M. Ma, "Intelligent image recognition system for marine fouling using Softmax transfer learning and deep convolutional neural networks," *Complexity*, vol. 2017, Article ID 5730419, 9 pages, 2017.
- [61] G. Tonguç and B. O. Ozkara, "Automatic recognition of student emotions from facial expressions during a lecture," *Computers & Education*, vol. 148, Article ID 103797, 2020.