Deep Learning-Based Emotion Recognition from Real-Time Videos

Wenbin Zhou, Justin Cheng, Xingyu Lei, Bedrich Benes(✉), and Nicoletta Adamo

Purdue University, West Lafayette, IN 47906, USA
{bbenes,nadamovi}@purdue.edu

Abstract. We introduce a novel framework for emotional state detection from facial expression targeted to learning environments. Our framework is based on a convolutional deep neural network that classifies people's emotions that are captured through a web-cam. For our classification outcome we adopt Russel's model of core affect in which any particular emotion can be placed in one of four quadrants: pleasant-active, pleasant-inactive, unpleasant-active, and unpleasant-inactive. We gathered data from various datasets that were normalized and used to train the deep learning model. We use the fully-connected layers of the VGG_S network which was trained on human facial expressions that were manually labeled. We have tested our application by splitting the data into 80:20 and re-training the model. The overall test accuracy of all detected emotions was 66%. We have a working application that is capable of reporting the user emotional state at about five frames per second on a standard laptop computer with a web-cam. The emotional state detector will be integrated into an affective pedagogical agent system where it will serve as a feedback to an intelligent animated educational tutor.

Keywords: Facial expression recognition · Deep learning · Education

1 Introduction

Facial expressions play a vital role in social communications between humans because the human face is the richest source of emotional cues [18]. We are capable of reading and understanding facial emotions because of thousands of year of evolution. We also react to facial expressions [13] and some of these reactions are even unconscious [14]. Emotions play an important role as a feedback in learning as they inform the teacher about the student's emotional state [3,31]. This is particularly important in on-line learning, where a fully automated system can be adapted to emotional state of the learner [15].

We introduce a novel deep-neural network-based emotion detection targeted to educational settings. Our approach uses a web-cam to capture images of the

National Science Foundation, # 10001364, Collaborative Research: Multimodal Affective Pedagogical Agents for Different Types of Learners.

M. Kurosu (Ed.): HCII 2020, LNCS 12182, pp. 321–332, 2020.
https://doi.org/10.1007/978-3-030-49062-1_22

learner and a convolutional neural networks to detect facial expressions in real time categorized by using Russell's classification model [55] and Fig. 1, which covers the majority of affective states a learner might experience during a learning episode. We also take the effort to deal with different scenes, viewing angles, and lighting conditions that may be encountered in practical use. We use transfer learning on the fully-connected layers of the VGG_S network which was trained on human facial expressions that were manually labeled. The overall test accuracy of the detected emotions was 66% and our system is capable of reporting the user emotional state at about five frames per seconds on a laptop computer. We plan to integrate the emotional state detector into an affective pedagogical agent system where it will serve as a feedback to an intelligent animated tutor.

2 Background and Related Work

Encoding and understanding emotions is particularly important in educational settings [3,31]. While face-to-face education with a capable, educated, and empathetic teacher is optimal, it is also not always possible. People have been looking at teaching without teachers ever since the invention of books and with the recent advances in technology, for example by using simulations [43,66]. We have also seen significant advances in distance learning platforms and systems [22,52]. However, while automation brings many advantages, such as reaching a wide population of learners or being available at locations where face-to-face education may not be possible, it also brings new challenges [2,9,50,61]. One of them is the standardized look-and-feel of the course. One layout does not fit all learners, the pace of the delivery should be managed, the tasks should vary depending on the level of the learner, and the content should be also calibrated to the individual needs of learners.

Affective Agents: Some of these challenges have been addressed by interactive pedagogical agents that have been found effective in enhancing distance learning [6,40,47,57]. Among them, animated educational agents play an important role [12,39], because they can be easily controlled and their behavior can be defined by techniques commonly used in computer animation, for example by providing adequate gestures [25]. Pedagogical agents with emotional capabilities can enhance interactions between the learner and the computer and can improve learning as shown by Kim et al. [30]. Several systems have been implemented, for example Lisetti and Nasoz [37] combined facial expression and physiological signals to recognize a learner's emotions. D'Mello and Graesser [15] introduced AutoTutor and they shown that learners display a variety of emotions during learning and they also shown that AutoTutor can be designed to detect emotions and respond to them. A virtual agent SimSensei [42] engages in interviews to elicit behaviors that can be automatically measured and analyzed. It uses a multimodal sensing system that captures a variety of signals that assess the user's affective state, as well as to inform the agent to provide feedback. The manipulation of the agents affective states significantly influences learning [68] and has a positive influence on learner self-efficacy [30].

However, an effective pedagogical agent needs to respond to learners emotions that need to be first detected. The communication should be based on real input from the learner, pedagogical agents should be empathetic [11,30] and they should provide emotional interactions with the learner [29]. Various means of emotion detection have been proposed, such as using eye-tracker [62], measuring body temperature [4], using visual context [8], or skin conductivity [51] but a vast body of work has been focusing on detecting emotions in speech [28,35,65].

Facial Expressions: While the above-mentioned previous work provides very good results, it may not be always applicable in educational context. Speech is often not required while communicating with educational agents, and approaches that require attached sensors may not be ideal for the learner. This leaves the detection of facial expressions and their analysis as a good option.

Various approaches have been proposed to detect facial expressions. Early works, such as the FACS [16], focus on facial parameterization, where the features are detected and encoded as a feature vector that is used to find a particular emotion. Recent approaches use active contours [46] or other automated methods to detect the features automatically. A large class of algorithms attempts to use geometry-based approaches, such as facial reconstruction [59] and others detect salient facial features [20,63]. Various emotions and their variations have been studied [45] and classified [24], and some focus on micro expressions [17]. Novel approaches use automated feature detection by using machine learning methods such as support vector machine [5,58], but they share the same sensibility to the facial detector as the above-mentioned approaches (see also a review [7]).

One of the key components of these approaches is a face tracking system [60] that should be capable of a robust detection of the face and its features even in varying light conditions and for different learners [56]. However, existing methods often require careful calibration, similar lighting conditions, and the calibration may not transfer to other persons. Such systems provide good results for head position or orientation tracking, but they may fail to detect subtle changes in mood that are important for emotion detection.

Deep Learning: Recent advances in deep learning [34] brought deep neural networks also to the field of emotion detection. Several approaches have been introduced for robust head rotation detection [53], detection of facial features [64], speech [19], or even emotions [44]. Among them, EmoNets [26] detects acted emotions from movies by simultaneously analyzing both video and audio streams. This approach builds on the previous work for CNN facial detection [33]. Our work is inspired by the work of Burket et al. [10] who introduced deep learning network called DeXpression for emotion detection from videos. In particular, they use the Cohn-Kanade database (CMU-Pittsburg AU coded database) [27] and the MMI Facial Expression [45].

3 Classification of Emotions

Most applications of emotion detection categorize images of facial expressions into seven types of human emotions: anger, disgust, fear, happiness, sadness, sur-

prise, and neutral. Such classification is too detailed in the context of students' emotions, for instance when learners are taking video courses in front of a computer the high number of emotions is not applicable in all scenarios. Therefore, we use a classification of emotions related and used in academic learning [48,49]. In particular, we use Russell's model of core affect [55] in which any particular emotion can be placed along two dimensions (see Fig. 1): 1) valence (ranging from unpleasant to pleasant), and 2) arousal (ranging from activation to deactivation). This model covers a sufficiently large range of emotions and is suitable for deep learning implementation.

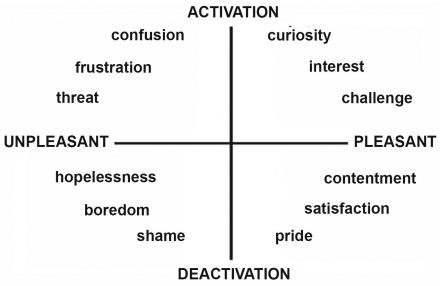

Fig. 1. Mapping of emotions from the discrete model to the 4-quadrant model (from Russel et al. [55]).

The two main axis of the Russel's divide the emotion space into four quadrants: 1) upper-left quadrant (active-unpleasant) includes affective states based on being exposed to instruction such as confusion or frustration, 2) upper-right quadrant (active-pleasant) includes curiosity and interest, 3) lower-right quadrant (inactive-pleasant) includes contentment and satisfaction, and 4) lower-left quadrant (inactive-pleasant) includes hopelessness and boredom.

Most of the existing image databases (some of them are discussed in Sect. 4.1) classify the images of facial expressions into the seven above-mentioned discrete emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral). We transform the datasets according to Russell's 4-quadrants classification model by grouping the images by using the following mapping:

- pleasant-active ⇐ happy, surprised,
- unpleasant-active ⇐ angry, fear, disgust,
- pleasant-inactive ⇐ neutral, and
- unpleasant-inactive ⇐ sad.

This grouping then assigns a unique label denoted by L to each image as:

$$L \in \{active - pleasant, active - unpleasant, \\ inactive - pleasant, inactive - unpleasant\}. \quad (1)$$

4 Methods

4.1 Input Images and Databases

Various databases of categorized (labeled) facial expressions with detected faces and facial features exist. We used images from the Cohn-Kanade database (CK+) [27], Japanese Female Facial Expression (JAFFE) [38], The Multimedia Understanding Facial Expression Database (MUG) [1], Indian Spontaneous

Expression Database (ISED) [23], Radboud Faces Database (RaFD) [32], Oulu-CASIA NIR&VIS facial expression database (OULU) [67], AffectNet [41], and The CMU multi-pose, illumination, and expression Face Database (CMU-PIE) [21].

Table 1. Databases used for training the deep neural network.

Input database	# of images	sad	happy	neutral	surprise	fear	anger	disgust
CK+ [27]	636	28	69	327	83	25	45	59
JAFFE [38]	213	31	31	30	30	32	30	29
MUG [1]	401	48	87	25	66	47	57	71
ISED [23]	478	48	227	0	73	0	0	80
RaFD [32]	7,035	1,005	1,005	1,005	1,005	1,005	1,005	1,005
Oulu [67]	5,760	480	480	2,880	480	480	480	480
AffectNet [41]	28,7401	25,959	134,915	75,374	14,590	6,878	25,382	4,303
CMU-PIE [21]	551,700	0	74,700	355,200	60,900	0	0	60,900

Table 1 shows the number of images and the subdivision of each dataset into categories (sad, happy, neutral, surprise, fear, anger, and disgust). Figure 2 shows the distributions of data per expression (top-left), per database (top-right), and the percentage distribution of each expression in the dataset (bottom-left). In total we had 853,624 images with 51% neutral faces, 25% happy, 3% sad, 8% disgust, 3% anger, 1% fear, and 9% surprise.

The lower right image in Fig. 2 shows the percentage of the coverage of each image by label L from Eq. (1). The total numbers were: active-pleasant: 288,741 images (12%), active-unpleasant 102,393 images (34%), inactive-pleasant 434,841 (51%), and inactive-unpleasant 27,599 (3%). The re-mapped categories were used as input to training the deep neural network in Sect. 4.2.

It is important to note that the actual classification of each image into its category varies in each databases and some are not even unique. Certain images may be classified by only one person while some are classified by various people, which brings more uncertainty. Moreover, some databases are in color and some are not. While it would be ideal to have uniform coverage of the expressions in all databases, the databases are unbalanced in both quality of images and the coverage of facial expressions (Fig. 2).

Also, certain expressions are easy to classify, but some may be classified as mixed and belonging to multiple categories. In this case, we either removed the image from experiments or put it into only one category. Interestingly, the most difficult expression to classify is neutral, because it does not represent any emotional charge and may be easily misinterpreted. This expression is actually the most covered in the dataset that should, in theory, improve its detection if correctly trained.

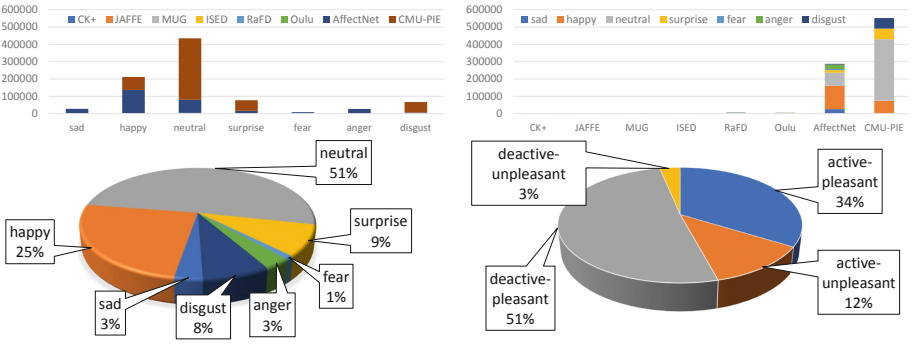

Fig. 2. Statistics of the used datasets used: contribution per database and expression (top row) and overall percentage of each expression (bottom left) and percentage of each contribution after remapping to Russel quadrants Eq. (1) (bottom right).

4.2 Deep Neural Network

We used deep neural network VGG_S Net [36] and Caffe. VGG_S Net is based on VGG Net that has proven successful in ImageNet Large Scale Visual Recognition Challenge [54] and the VGG_S is effective in facial detection.

Figure 3 shows the VGG_S neural network architecture. The network is a series of five convolutional layers, three fully-connected layers, eventually leading to a softmax classifier that outputs the probability value. We modified the output layer of the original net so that it generates the probability of the image to have a label from Eq. (1). The training input is a set of pairs $[image, L]$, where L is the label belonging to one of the four categories from Russel's diagram from Eq. (1). During the inference stage, the softmax layer outputs the probability of the input image of having the label L.

4.3 Training

We trained the network on images from datasets discussed in Sect. 4.1. We used data amplification by using Gaussian blur and applying variations of contrast, lighting, and subject position to the original images from each dataset to make our program more accurate in practical scenarios. The input images were preprocessed by using Haar-Cascade filter provided by OpenCV that crops the image by only including the face without significant background. This, in effect, that reduces the training times.

In order to have a balanced dataset, we would prefer to have similar number of images for each label from the categories in Eq. (1). Therefore, the lowest amount of images (inactive-unpleasant) dictated the size of the training set. We trained with 68,012 images, batch size 15 images, we used 80,000 iterations, and the average accuracy was set to 0.63 with 10,000 epochs. The training time

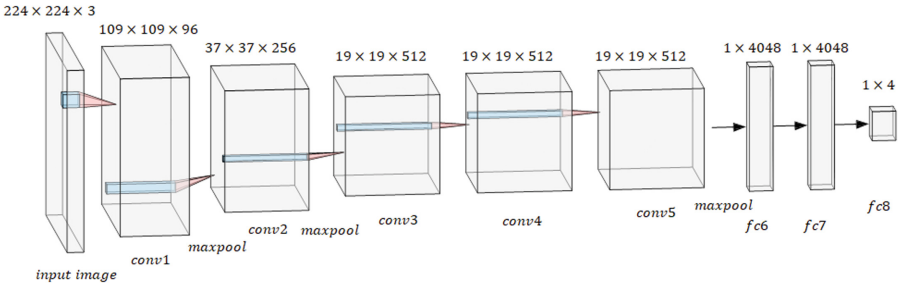

Fig. 3. Deep neural network architecture used in our framework.

was about 70 min on an desktop computer equipped with Intel Xeon(R) W-2145 CPU running at 3.7 GHz, with 32 GB of memory, and with NVidia RTX2080 GPU.

5 Results

Testing: We divided the dataset randomly into two groups in the ratio 80:20. We trained on 80% of the images, tested on the remaining 20%, and we repeated the experiment three times with random split of the input data each time.

Table 2. Average and standard deviation of the three runs of our testing.

	pleasant-active	pleasant-inactive	unpleasant-active	pleasant-inactive
pleasant-active	**(70.0, 1.6)**	(22.3, 4.1)	(2.0, 0.0)	(6.3, 2.1)
pleasant-inactive	(2.3, 1.2)	**(87.3, 0.5)**	(4.0, 0.8)	(6.0, 1.4)
unpleasant-active	(1.7, 0.5)	(41.7, 0.9)	**(44.0, 1.6)**	(8.3, 5.7)
pleasant-inactive	(5.0, 0.8)	(12.0, 3.7)	(9.3, 2.9)	**(62.0, 7.0)**

Table 2 shows the average and standard deviation of the confusion matrices from the three runs of our experiments and Fig. 4 show the confusion matrices of the individual runs. The main diagonal indicates that pleasant-active was detected 70% with standard deviation about 1.6% correctly and misdetected as pleasant-inactive in 22.3%, unpleasant-active in 2%, and unpleasant-active in 6.3% of cases. Similarly, pleasant-inactive was detected correctly for 87.3% of cases, unpleasant-active in 44% and the least precise was unpleasant-inactive with 62%. This is an expected result, because the lower part of the Russel's diagram (Fig. 1) includes passive expressions that are generally more difficult to detect. We achieved an overall accuracy of 66%.

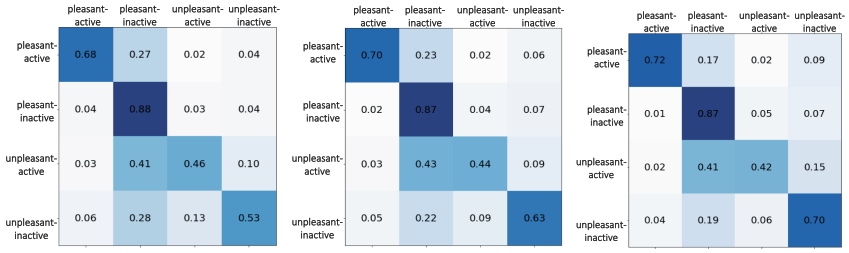

Fig. 4. Normalized confusion matrices for the results of our experiment.

Deployment: The trained deep neural network was extracted and used in real-time session that categorizes facial expressions into the four quadrants of Russel's diagram. We used a laptop computer with a web cam in resolution $1{,}920 \times 1{,}080$ equipped with CPU Intel Core i5-6300U at 2.4 GHz. We used the Caffe environment on Windows 10 and OpenCV to monitor the input image from the camera and detect face. Only the face was sent to the our trained network as the background was cropped out. The neural network classified the image and sent the result back to the application that displayed it as a label on the screen. An example in Fig. 5 shows several samples of real-time detection of facial expressions by using our system.

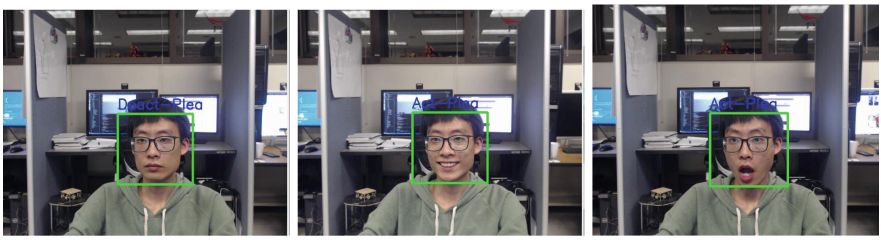

Fig. 5. Examples of real-time expression detection by using our system.

6 Conclusions

The purpose of this project is to develop a real-time facial emotion recognition algorithm that detects and classifies human emotions with the objective of using it as a classifier in online learning. Because of this requirement, our detector reports a probability of an emotion belonging to one of the four quadrants of Russel's diagram.

Our future goal is to integrate the recognition algorithm into a system of affective pedagogical agents that will respond to the students' detected emotions using different types of emotional intelligence. Our experiments show that the

overall test accuracy is sufficient for a practical use and we hope that the entire system will be able to enhance learning.

There are several possible avenues for future work. While our preliminary results show satisfactory precision on our tested data, it would be interesting to actually validate our system in a real-world scenario. We conducted a preliminary user study in which we asked 10 people to make certain facial expression and we validated the detection. However, this approach did not provide satisfactory results, because we did not find a way to verify that the people were actually in the desired emotional state and their expressions were genuine - some participants started to laugh each time the system detected emotion they were not expecting. Emotional state is a complicated. Happy people cannot force themselves to make sad faces and some of the expressions were difficult to achieve even for real actors. So while validation of our detector remains a future work, another future work is increasing the precision of the detection by expanding the training data set and tuning the parameters of the deep neural network.

Acknowledgments. This work has been funded in part by National Science Foundation grant # 1821894 - *Collaborative Research: Multimodal Affective Pedagogical Agents for Different Types of Learners.*

References

1. Aifanti, N., Papachristou, C., Delopoulos, A.: The MUG facial expression database. In: 11th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2010, pp. 1–4. IEEE (2010)
2. Allen, I.E., Seaman, J.: Staying the Course: Online Education in the United States. ERIC, Newburyport (2008)
3. Alsop, S., Watts, M.: Science education and affect. Int. J. Sci. Educ. **25**(9), 1043–1047 (2003)
4. Ark, W.S., Dryer, D.C., Lu, D.J.: The emotion mouse. In: HCI (1), pp. 818–823 (1999)
5. Bartlett, M.S., Littlewort, G., Fasel, I., Movellan, J.R.: Real time face detection and facial expression recognition: development and applications to human computer interaction. In: 2003 Conference on Computer Vision and Pattern Recognition Workshop, vol. 5, p. 53. IEEE (2003)
6. Baylor, A.L., Kim, Y.: Simulating instructional roles through pedagogical agents. Int. J. Artif. Intell. Educ. **15**(2), 95–115 (2005)
7. Bettadapura, V.: Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722 (2012)
8. Borth, D., Chen, T., Ji, R., Chang, S.F.: SentiBank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In: Proceedings of the 21st ACM International Conference on Multimedia, pp. 459–460 (2013)
9. Bower, B.L., Hardy, K.P.: From correspondence to cyberspace: changes and challenges in distance education. New Dir. Community Coll. **2004**(128), 5–12 (2004)
10. Burkert, P., Trier, F., Afzal, M.Z., Dengel, A., Liwicki, M.: DeXpression: deep convolutional neural network for expression recognition. arXiv preprint arXiv:1509.05371 (2015)

11. Castellano, G., et al.: Towards empathic virtual and robotic tutors. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 733–736. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_100
12. Craig, S.D., Gholson, B., Driscoll, D.M.: Animated pedagogical agents in multimedia educational environments: effects of agent properties, picture features and redundancy. J. Educ. Psychol. **94**(2), 428 (2002)
13. Dimberg, U.: Facial reactions to facial expressions. Psychophysiology **19**(6), 643–647 (1982)
14. Dimberg, U., Thunberg, M., Elmehed, K.: Unconscious facial reactions to emotional facial expressions. Psychol. Sci. **11**(1), 86–89 (2000)
15. D'Mello, S., Graesser, A.: Emotions during learning with autotutor. In: Adaptive Technologies for Training and Education, pp. 169–187 (2012)
16. Ekman, P.: Biological and cultural contributions to body and facial movement, pp. 34–84 (1977)
17. Ekman, P.: Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage, Revised edn. WW Norton & Company, New York (2009)
18. Ekman, P., Keltner, D.: Universal facial expressions of emotion. In: Segerstrale, U., Molnar, P. (eds.) Nonverbal Communication: Where Nature Meets Culture, pp. 27–46 (1997)
19. Fayek, H.M., Lech, M., Cavedon, L.: Evaluating deep learning architectures for Speech Emotion Recognition. Neural Netw. **92**, 60–68 (2017)
20. Gourier, N., Hall, D., Crowley, J.L.: Estimating face orientation from robust detection of salient facial features. In: ICPR International Workshop on Visual Observation of Deictic Gestures. Citeseer (2004)
21. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. Image Vis. Comput. **28**(5), 807–813 (2010)
22. Gunawardena, C.N., McIsaac, M.S.: Distance education. In: Handbook of Research on Educational Communications and Technology, pp. 361–401. Routledge (2013)
23. Happy, S., Patnaik, P., Routray, A., Guha, R.: The indian spontaneous expression database for emotion recognition. IEEE Trans. Affect. Comput. **8**(1), 131–142 (2015)
24. Izard, C.E.: Innate and universal facial expressions: evidence from developmental and cross-cultural research (1994)
25. Cheng, J., Zhou, W., Lei, X., Adamo, N., Benes, B.: The effects of body gestures and gender on viewer's perception of animated pedagogical agent's emotions. In: Kurosu, M. (ed.) HCII 2020. LNCS, vol. 12182, pp. 169–186. Springer, Cham (2020)
26. Kahou, S.E., et al.: EmoNets: multimodal deep learning approaches for emotion recognition in video. J. Multimodal User Interfaces **10**(2), 99–111 (2016). https://doi.org/10.1007/s12193-015-0195-2
27. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–53. IEEE (2000)
28. Kim, S., Georgiou, P.G., Lee, S., Narayanan, S.: Real-time emotion detection system using speech: multi-modal fusion of different timescale features. In: 2007 IEEE 9th Workshop on Multimedia Signal Processing, pp. 48–51. IEEE (2007)
29. Kim, Y., Baylor, A.L.: Pedagogical agents as social models to influence learner attitudes. Educ. Technol. **47**(1), 23–28 (2007)
30. Kim, Y., Baylor, A.L., Shen, E.: Pedagogical agents as learning companions: the impact of agent emotion and gender. J. Comput. Assist. Learn. **23**(3), 220–234 (2007)

31. Kirouac, G., Dore, F.Y.: Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. J. Nonverbal Behav. **9**(1), 3–7 (1985). https://doi.org/10.1007/BF00987555
32. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A.: Presentation and validation of the Radboud Faces Database. Cogn. Emot. **24**(8), 1377–1388 (2010)
33. Le, Q.V., Zou, W.Y., Yeung, S.Y., Ng, A.Y.: Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In: CVPR 2011, pp. 3361–3368. IEEE (2011)
34. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
35. Lee, C.M., Narayanan, S.S.: Toward detecting emotions in spoken dialogs. IEEE Trans. Speech Audio Process. **13**(2), 293–303 (2005)
36. Levi, G., Hassner, T.: Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 503–510 (2015)
37. Lisetti, C.L., Nasoz, F.: MAUI: a multimodal affective user interface. In: Proceedings of the Tenth ACM International Conference on Multimedia, pp. 161–170 (2002)
38. Lyons, M., Kamachi, M., Gyoba, J.: Japanese Female Facial Expression (JAFFE) Database, July 2017. https://figshare.com/articles/jaffe_desc_pdf/5245003
39. Martha, A.S.D., Santoso, H.B.: The design and impact of the pedagogical agent: a systematic literature review. J. Educ. Online **16**(1), n1 (2019)
40. Miles, M.B., Saxl, E.R., Lieberman, A.: What skills do educational "change agents" need? An empirical view. Curric. Inq. **18**(2), 157–193 (1988)
41. Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans. Affect. Comput. **10**(1), 18–31 (2017)
42. Morency, L.P., et al.: SimSensei demonstration: a perceptive virtual human interviewer for healthcare applications. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
43. Neri, L., et al.: Visuo-haptic simulations to improve students' understanding of friction concepts. In: IEEE Frontiers in Education, pp. 1–6. IEEE (2018)
44. Ng, H.W., Nguyen, V.D., Vonikakis, V., Winkler, S.: Deep learning for emotion recognition on small datasets using transfer learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 443–449 (2015)
45. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. In: 2005 IEEE International Conference on Multimedia and Expo, pp. 5–pp. IEEE (2005)
46. Pardàs, M., Bonafonte, A.: Facial animation parameters extraction and expression recognition using hidden Markov models. Sig. Process. Image Commun. **17**(9), 675–688 (2002)
47. Payr, S.: The virtual university's faculty: an overview of educational agents. Appl. Artif. Intell. **17**(1), 1–19 (2003)
48. Pekrun, R.: The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. Educ. Psychol. Rev. **18**(4), 315–341 (2006). https://doi.org/10.1007/s10648-006-9029-9
49. Pekrun, R., Stephens, E.J.: Achievement emotions: a control-value approach. Soc. Pers. Psychol. Compass **4**(4), 238–255 (2010)
50. Phipps, R., Merisotis, J., et al.: What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education (1999)

51. Picard, R.W., Scheirer, J.: The Galvactivator: a glove that senses and communicates skin conductivity. In: Proceedings of the 9th International Conference on HCI (2001)
52. Porter, L.R.: Creating the Virtual Classroom: Distance Learning with the Internet. Wiley, Hoboken (1997)
53. Rowley, H.A., Baluja, S., Kanade, T.: Rotation invariant neural network-based face detection. In: Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231), pp. 38–44. IEEE (1998)
54. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
55. Russell, J.A.: Core affect and the psychological construction of emotion. Psychol. Rev. **110**(1), 145 (2003)
56. Schneiderman, H., Kanade, T.: Probabilistic modeling of local appearance and spatial relationships for object recognition. In: Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231), pp. 45–51. IEEE (1998)
57. Schroeder, N.L., Adesope, O.O., Gilbert, R.B.: How effective are pedagogical agents for learning? A meta-analytic review. J. Educ. Comput. Res. **49**(1), 1–39 (2013)
58. Tian, Y.I., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 97–115 (2001)
59. Tie, Y., Guan, L.: A deformable 3-D facial expression model for dynamic human emotional state recognition. IEEE Trans. Circ. Syst. Video Technol. **23**(1), 142–157 (2012)
60. Viola, P., Jones, M., et al.: Robust real-time object detection. Int. J. Comput. Vis. **4**(34–47), 4 (2001)
61. Volery, T., Lord, D.: Critical success factors in online education. Int. J. Educ. Manag. **14**(5), 216–223 (2000)
62. Wang, H., Chignell, M., Ishizuka, M.: Empathic tutoring software agents using real-time eye tracking. In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications, pp. 73–78 (2006)
63. Wilson, P.I., Fernandez, J.: Facial feature detection using Haar classifiers. J. Comput. Sci. Coll. **21**(4), 127–133 (2006)
64. Yang, S., Luo, P., Loy, C.C., Tang, X.: From facial parts responses to face detection: a deep learning approach. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3676–3684 (2015)
65. Yu, F., Chang, E., Xu, Y.-Q., Shum, H.-Y.: Emotion detection from speech to enrich multimedia content. In: Shum, H.-Y., Liao, M., Chang, S.-F. (eds.) PCM 2001. LNCS, vol. 2195, pp. 550–557. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45453-5_71
66. Yuksel, T., et al.: Visuohaptic experiments: exploring the effects of visual and haptic feedback on students' learning of friction concepts. Comput. Appl. Eng. Educ. **27**(6), 1376–1401 (2019)
67. Zhao, G., Huang, X., Taini, M., Li, S.Z., PietikäInen, M.: Facial expression recognition from near-infrared videos. Image Vis. Comput. **29**(9), 607–619 (2011)
68. Zhou, L., Mohammed, A.S., Zhang, D.: Mobile personal information management agent: supporting natural language interface and application integration. Inf. Process. Manag. **48**(1), 23–31 (2012)