2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta) | 979-8-3503-4655-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00002

Proceedings

2022 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Autonomous & Trusted Vehicles, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse

(SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse)

Haikou, China
15-18 December 2022

IEEE COMPUTER SOCIETY CONFERENCE PUBLISHING SERVICES

Los Alamitos, California
Washington • Tokyo

IEEE COMPUTER SOCIETY

2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta) | 979-8-3503-4655-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00003

ISBN-13: 979-8-3503-4655-8
BMS Part# CFP2275H-ART

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1 800 272 6657
Fax: + 1 714 821 4641
http://computer.org/cps
cps@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1 732 981 0060
Fax: + 1 732 981 9667
http://shop.ieee.org/store/
customer-service@ieee.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81 3 3408 3118
Fax: + 81 3 3408 3553
tokyo.ofc@computer.org

Editorial production by Cristina Ceballos
Cover art production by Hector Torres

IEEE Computer Society
***Conference Publishing Services* (CPS)**
http://www.computer.org/cps

2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta) | 979-8-3503-4655-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00004

2022 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Autonomous & Trusted Vehicles, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse (SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse)

SmartWorld-UIC-ScalCom-DigitalTwin-PriComp-Metaverse 2022

Table of Contents

The 19 th IEEE International Conference on Ubiquitous Intelligence and Computing (UIC 2022)

Regular Research Papers

Intelligent/Smart Object & Interaction I

Intelligent/Smart Object & Interaction II

Intelligent/Smart Object & Interaction III

Intelligent/Smart Object & Interaction IV

Intelligent/Smart Object & Interaction V

Intelligent/Smart Environment & Application I

Intelligent/Smart Environment & Application II

Intelligent/Smart Environment & Application III

Intelligent/Smart Environment & Application IV

Intelligent/Smart Environment & Application V

Intelligent/Smart Environment & Application VI

Intelligent/Smart Environment & Application VII

Intelligent/Smart Environment & Application VIII

Intelligent/Smart Environment & Application IX

Intelligent/Smart Environment & Application X

Intelligent/Smart Environment & Application XI

Intelligent/Smart Environment & Application XII

Intelligent/Smart Systems & Services I

Intelligent/Smart Systems & Services II

Intelligent/Smart Systems & Services III

Intelligent/Smart Systems & Services IV

Intelligent/Smart Systems & Services V

Intelligent/Smart Systems & Services VI

Personalization and Social Aspects I

Personalization and Social Aspects II

Personalization and Social Aspects III

Ubiquitous Intelligence with Blockchain Technology

Research Papers

Intelligent/Smart Object & Interaction I

Intelligent/Smart Object & Interaction II

Intelligent/Smart Object & Interaction III

Intelligent/Smart Object & Interaction IV

Intelligent/Smart Object & Interaction V

Intelligent/Smart Environment & Application I

Intelligent/Smart Environment & Application II

Intelligent/Smart Environment & Application III

Intelligent/Smart Environment & Application IV

Intelligent/Smart Environment & Application V

Intelligent/Smart Environment & Application VI

Intelligent/Smart Environment & Application VII

Intelligent/Smart Environment & Application VIII

Intelligent/Smart Environment & Application IX

Intelligent/Smart Environment & Application X

Intelligent/Smart Environment & Application XI

Intelligent/Smart Environment & Application XII

Intelligent/Smart Systems & Services I

Intelligent/Smart Systems & Services II

Intelligent/Smart Systems & Services III

Intelligent/Smart Systems & Services IV

Intelligent/Smart Systems & Services V

Intelligent/Smart Systems & Services VI

Intelligent/Smart Systems & Services VII

Personalization and Social Aspects I

Personalization and Social Aspects II

Personalization and Social Aspects III

Personalization and Social Aspects IV

Personalization and Social Aspects V

Ubiquitous Intelligence with Blockchain Technology

The 2022 IEEE International Conference on ATC (ATC 2022)

Autonomous and Trust Vehicles

Communications and Networking

The 22nd IEEE International Conference on Scalable Computing and Communications (ScalCom 2022)

Cloud and Fog Computing

Blockchain, Modelling and Simulations

The 2022 IEEE International Conference on Digital Twin (DigitalTwin 2022)

Modeling & Simulation

Technologies & Applications I

Technologies & Applications II

Evaluation & Optimization, Security & Privacy

The 2022 IEEE International Conference on Privacy Computing (PriComp 2022)

Privacy Computing I

Privacy Computing II

Privacy Computing III

The 2022 IEEE International Conference on Metaverse (Metaverse 2022)

Metaverse Computing and Communications

Metaverse Interaction

Metaverse Security

Metaverse Applications

2022 IEEE SmartWorld - No show papers

2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Metaverse)

2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta) | 979-8-3503-4655-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00157

Engagement Detection of Online Learners Based on Key Frames

1st Zhurong Zhou
SOUTHWEST UNIVERSITY
COLLEGE OF COMPUTER & INFORMATION SCIENCE
Chongqing, CHINA
zhouzr@swu.edu.cn

2nd Danqing Pu
SOUTHWEST UNIVERSITY
COLLEGE OF COMPUTER & INFORMATION SCIENCE
Chongqing, CHINA
pdq737@email.swu.edu.cn, corresponding author

3rd Fan Yang
SOUTHWEST UNIVERSITY
COLLEGE OF COMPUTER & INFORMATION SCIENCE
Chongqing, CHINA
fanyang0329@gmail.com

***Abstract*—In recent years, more and more students are willing to choose courses on online learning platforms, such as MOOC. However, the online learning platform cannot capture the learning status of students, and cannot give real-time feedback when students' attention is out of learning. The lack of feedback mechanisms mainly leads to a staggering dropout rate of MOOC. Therefore, engagement detection is very important for online learning platforms. Nowadays, many researchers have focused on engagement detection which are based on pictures. This method of selecting pictures is through random selection or repeated selection after short interval of time. However, both of these methods may lead to the problem of inaccurate samples or heavy identification workload. In this paper, we propose to introduce Key Frames that can more accurately represent the engagement level of the online learning personal video in the sample selection stage. Experimental results show that our proposed method has better results.**

***Index Terms*—engagement detection, Key Frame, Local Directional Pattern, expression similarity, online learning**

I. Introduction

There are growing evidences prove that students' enthusiasm and engagement are closely related to students' academic performance [1], [2]. Scholars have different conceptual definitions of engagement. Chapman [3] offers a possible definition: students' cognitive investment in, active participation in and emotional commitment to their learning. The Australian Council of Educational Research(ACER) proposes another: students' participation with activities and conditions may generate high quality learning [4]. We accept both definitions, because they both highlight the importance of high engagement for student learning.

In the past few years, online learning courses have grown significantly and are still growing [5]. More and more students are willing to choose online learning courses because it saves students' costs and has no time and space limitations. With the upgrading of software and hardware equipment and the rapid expansion of online teaching resources, online education has an unprecedented momentum of development. In the next few years, the number of online learning users will continue to grow. However, many existing online learning platforms lack a feedback mechanism for students' learning status. Lacking of feedback mechanism is also one of the important reasons for the high dropout rate of online learning platforms, such as MOOC [6].

In the traditional way of education, teachers and students communicate face to face. Teachers' friendly smiles, humorous language and elegant gestures can all increase students' interest and participation in the classroom. When students encounter difficult or unintelligible content in learning, teachers can help students solve problems one-on-one to avoid the accumulation of problems which leads to negative learning emotions. In addition, teachers can observe students' attention in class from their facial expressions and body movements in time. When a student is drowsy, talking or in a daze, the teacher can immediately take activities to attract and encourage students to be more involved in the classroom.

In the online learning method, students mainly study independently. The course is composed of static content and students can find answers to questions through the Internet or other forms of communication (such as mailing teachers or classroom forums). These exchanges usually take longer to receive answers. Due to the delayed feedback, students' problems cannot be solved in time, and the content of the course is not fully grasped. Over time, students' interest in the course is reduced and the students can't concentrate on the course. Then a large number of students gave up the course, leading to a high dropout rate on online learning platforms.

One of the ways to bridge the gap between offline

979-8-3503-4655-8/22/$31.00 ©2022 IEEE

DOI 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00157

teaching and online teaching is to add feedback mechanisms about student status to online learning platforms. The current sensor equipment and computer technology are sufficient to realize the real-time feedback mechanism of the online learning system. The online learning platform can obtain the behavioral state of learners through the camera, and analyze the level of engagement and make timely interventions for students who are not sufficiently focused.

Because facial expressions are directly related to the participation of online learners, and the current facial expression recognition technology is relatively successful. In this paper, we mainly identify the concentration of students through the analysis of facial expressions. Our goal is to apply engagement detection to different online learning environments, to ensure the active participation of students, improve student learning effects, and reduce the dropout rate of online learning systems.

In the previous literature, most authors randomly select a frame from the video for recognition, but a person's expression can change in a short period of time, so randomly extracting a frame is not accurate enough, and some authors selects a frame through a time window, but the size of the time window is difficult to grasp, too large is easy to be inaccurate, and too small increases the workload of recognition. Therefore, in this paper, we propose a model that can select frames based on the similarity of expression features. When the expression changes significantly, more frames are selected, and when the expression changes not obviously, fewer frames are selected. We extract the features of different facial expressions using Local Directional Pattern (LDP) [7], use Principal Component Analysis (PCA) a to reduce the dimensionality of the feature data, and use the Cosine Similarity algorithm to calculate the similarity between expressions in each frame of the video. We select a frame from each video clips where each frame is similar, then take the selected all frames as the Key Frames of the video, use the engagement recognition model (CNN and SVM) to identify the engagement level of the Key Frame, and use the statistical analysis method to determine the overall engagement level of the video.

The rest of this paper is arranged as follows. In section 2, we describe existing related work in the field of engagement detection. Section 3 explains the proposed architecture of engagement detection. Section 4 discusses the experimental results, and Section 5 draws conclusions.

II. Literature Review

There is a large amount of relevant literature that focuses on the engagement detection of learners in online and traditional teaching environments [8], [9]. The research methods are mainly divided into three categories: self-report physiological and engagement tracing, and automated measurements.

Self-report is a questionnaire where students report their own level of engagement, distraction, excitement or boredom [10]. self-report method does not necessarily directly indicate the degree of learner's engagement, but take the engagement as a descriptive latent variable, and use factor analysis to determine the degree of concentration [11]. Self report is favored by many researchers because it can easily obtain useful information, but it also has many shortcomings. For example, students may hide their true thoughts in the questionnaire [12], and different students have different definitions of engagement.

Engagement tracing. The performance of the learner is evaluated based on the learner's reaction time and accuracy to some practice questions and test questions [13]. For example, the very short response time to simple questions indicates that the learner did not participate, but gave random answers effortlessly.

Automated measurements. The methods in the automatic category are divided into log file analysis, sensor data analysis, and computer vision based methods. In the log file analysis, concentration detection is performed based on the learner's behavior stored in the log file [14], mainly using machine learning and data mining approaches. Cocea and Weibelzahl [15] monitor learners' behavior (such as average time spent on pages, number of tests attended, number of correctly answered tests) to predict disengagement. In sensor data analysis category, physiological and neural sensor readings are used to detect engagement. Chaouachi [16] believes that engagement is usually equal to the level of arousal or alertness and the level of EEG, blood pressure, heart rate or galvanic skin response.

The third method of automatic engagement recognition is based on computer vision, which is also the method used in this paper. Computer vision-based methods provide a variety of ways to measure learner engagement by investigating clues from gestures and postures [17], [18], eye movements [19] , and facial expressions [20]. Gravesgaard [21] analyzed the gestures and postures in the computer-mediated tutorial dialogue, and studied the relationship between the learner's posture, gestures, dialogue and the tutor. Aslan [22] uses an eye tracker to detect the user's line of sight, and combines this information with statistical facial features and depth information. Ekman and Friesen [23] proposed the Facial Action Coding System (FACS), which is a widely used method for describing facial muscle action units (AU) and corresponding expressions.

At present, most researches are mainly focused on identifying the level of engagement from pictures [8], [24]. But the expression at a certain moment is not enough to represent the concentration of the current time period. For example, if the eyes are closed at a certain moment, it will be considered as sleeping, but it may actually be blinking. The aim of this study is to evaluate the engagement of an online learning video.

Fig. 1. An example of a Key Frames.

$$\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}\ \begin{bmatrix}-3 & 5 & 5\\ -3 & 0 & 5\\ -3 & -3 & -3\end{bmatrix}\ \begin{bmatrix}5 & 5 & 5\\ -3 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\ \begin{bmatrix}5 & 5 & -3\\ 5 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}$$

East M_0 North east M_1 North M_2 North west M_3

$$\begin{bmatrix}5 & -3 & -3\\ 5 & 0 & -3\\ 5 & -3 & -3\end{bmatrix}\ \begin{bmatrix}-3 & -3 & -3\\ 5 & 0 & -3\\ 5 & 5 & -3\end{bmatrix}\ \begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & -3\\ 5 & 5 & 5\end{bmatrix}\ \begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & 5\\ -3 & 5 & 5\end{bmatrix}$$

West M_4 South west M_5 South M_6 South east M_7

Fig. 3. Kirsch edge masks in all eight directions.

III. The Proposed Approach

Definition 1 Key Frame: We calculate the facial expression similarity of each frame, and select the frame with large difference as the Key Frame. Each Key Frame can represent the expression of the learner in a video clip. An example of a Key Frames is shown in Fig. 1.

The framework we proposed for engagement detection of online learners is shown in Fig. 2. Our work is mainly divided into three parts: A. Find Key Frames, B. Classification, C. Statistical Analysis.

In part A, we first extract the face of each frame in the video, use LDP to obtain the facial features, and then use the Cosine Similarity algorithm to calculate the similarity between the facial expressions, and finally extract the frame with large differences as the Key Frame. In Part B, we used CNN and SVM models to identify the engagement level of Key Frames, which are divided into engaged and disengaged. In part C, we use the engagement of the Key Frame as the engagement of the left and right intervals of the Key Frame, and perform statistical analysis to obtain the overall engagement of the video.

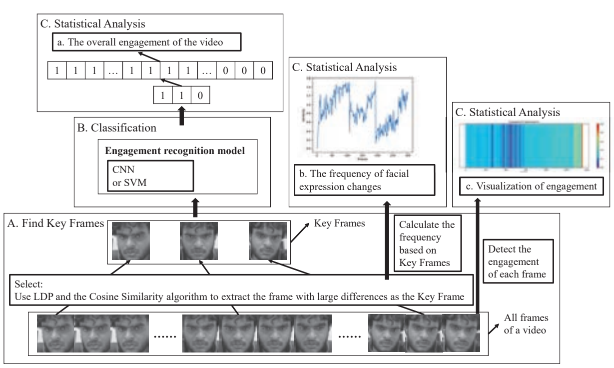

Fig. 2. The framework of the method we proposed.

A. Find Key Frames

First, we extract the face of the same person from each frame of the video, and then generate the Local Directional Pattern (LDP) codes for the whole face. LDP is a robust and person independent feature extraction technique [7], it calculates edge response values in different directions and uses these values to encode image textures. Because the edge response is less sensitive to light and noise than the intensity value, the resulting LDP features describe local primitives in a more stable manner, including different types of curves, corners, and intersections, and retain more information. LDP assigns an eight-bit binary code to each pixel of the input image. Each pixel of the image has a local 3×3 neighborhood pixel gray value, and the pixel is in the center of the neighborhood. Use Kirsch mask [25] to calculate the 8 directions of the pixel to get the eight-directional edge response values$\{\mathrm{m_i}\}$, $i = 0, \cdots, 7$. These masks are shown in the Fig. 3. The response value is not equally important in all directions, because the presence of corners or edges in a specific direction shows a high response value. Therefore, we need to find the most prominent k directions to generate LDP. Set these k directions to 1,we empirically set k=3 and the remaining directions to 0. Finally, LDP can be obtained by the following formulas:

$$LDP_k = \sum_{i=0}^{7} b_i\left(m_i - m_k\right) \times 2^i \tag{1}$$

$$b_i(a) = \begin{cases} 1, & a \geq 0, \\ 0, & a < 0, \end{cases} \tag{2}$$

Where m_i is the most prominent k directions response. The LDP bit positions is shown in Fig. 4.

b_3	b_2	b_1
b_4	P	b_0
b_5	b_6	b_7

Fig. 4. The LDP binary bit positions.

After computing all the LDP code for each pixel, we divide the entire image into 4 regions: left eye, right eye, nose and mouth (see Fig. 5), and divide the 4 regions into 9, 9, 9, 18 blocks respectively then calculate the LDP histograms of each blocks respectively. The input block I of size M×N is represented by an LDP histogram H, using

$$B(i) = \sum_{r=1}^{M}\sum_{c=1}^{N} f\left(LDP_k(r,c), i\right) \quad i = 0, \cdots, l-1 \tag{3}$$

$$f(a,i)=\begin{cases}1, & a=i,\\ 0, & a\neq i,\end{cases} \tag{4}$$

Where i is the LDP code value, l is the number of histogram bins, we set l = 32. Connect the histogram of all blocks in each area as the histogram of the current area, represented by H^k (k=1,2,3,4). Finally, create a histogram of the entire image by placing all the histograms of the four regions side by side, as shown below:

$$H=\left(H^1H^2H^3H^4\right) \tag{5}$$

H is a high-dimensional data, in order to better analyze it, we use PCA for dimensionality reduction. In this paper, we propose Cosine Similarity algorithm to calculate the expression similarity between different faces. The formula for Cosine Similarity algorithm is given below, $A=(x_1,x_2,\ldots,x_n)$, $B=(y_1,y_2,\ldots,y_n)$ is the feature vector of two faces, then cosine similarity can be defined as,

$$\cos\theta=\frac{A\cdot B}{|A|\cdot|B|}=\frac{\sum_{i=1}^{n}(x_i\cdot y_i)}{\sqrt{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i^2}} \tag{6}$$

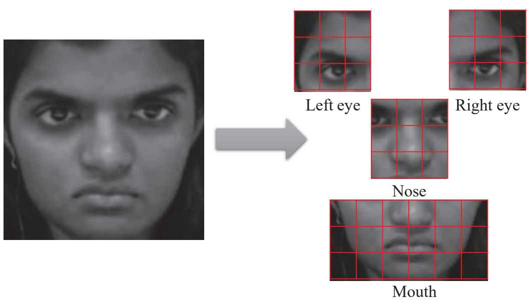

Fig. 5. Divide the entire face into 4 regions.

We process each frame of the face of the video to obtain the feature vector of each frame, then take the first frame of the video clip as the reference, and calculate the similarity in sequence with each frame after the time of the reference face. When the similarity result is less than -0.5, we think that the expression of the current frame has changed a lot compared with the reference face. Using this face as a benchmark, continue to calculate backwards until the end of the last frame of the video. Finally, all reference faces are selected as the Key Frames of this video for subsequent processing.

B. Classification

Convolutional Neural Network: We use the convolutional neural network (CNN) model as our engagement classification model. The model contains three convolutional layers (Conv.), followed by one max pooling layer and two average pooling layers with stride 2, and two fully connected (FC) layers, respectively. A rectified linear unit (ReLU) activation function is applied after all Conv. and FC layers. The last step of the CNN model includes a softmax layer, which consists of two neurons indicating engaged and disengaged.

Support Vector Machine: We also use the histogram features obtained in the previous part and linear support vector machine (SVM) to train a method as another engagement detection model. SVM is a machine learning technology that can be used to classify facial expressions. In this article, we use it to perform engagement detection, and also obtain two classifications, engaged and disengaged classes.

C. Statistical Analysis

1) The overall engagement of the video: Use our classification model to identify the selected Key Frames. The result of recognition is used as the type of all frames between two moments, the middle moment between the current Key Frame and the previous Key Frame, and the middle moment between the current Key Frame and the next Key Frame. The analysis process is shown in Fig. 6. We use a tuple index $=(KF_1,KF_2,\cdots,KF_m)$ to record the position of the selected Key Frame, where m is the number of Key Frames, the intermediate moments between Key Frames, starting and ending position are recorded with position $=\left(KF_1,\lceil\frac{KF_1+KF_2}{2}\rceil,\cdots,\lceil\frac{KF_{m-1}+KF_m}{2}\rceil,n-1\right)$, where KF_1 is the starting position and the value is equal to 0, and n is the number of frames the video contains, results $=(R_1,R_2,\cdots,R_m)$ record the recognition results of the Key Frames, the result value is 0 (disengaged) or 1(engaged). The engagement level of the video is calculated by the following formula:

$$V=\frac{\sum_{i=1}^{L}\left(\text{position}_i-\text{position}_{i-1}\right)\times\text{results}_{i-1}}{n} \tag{7}$$

Where L is the length of tuple position, and V is the overall engagement level of online learning video.

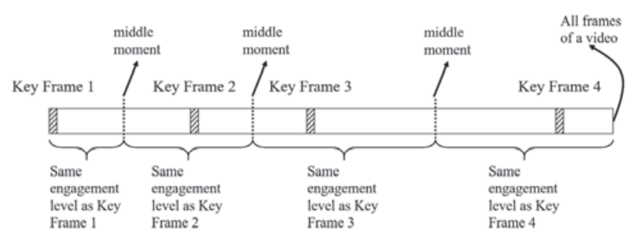

Fig. 6. The analysis process of video engagement.

2) The frequency of facial expression changes: In a class, the expressions of the students who focused on listening to the class are very varied, indicating that the content of the teacher's teaching materials is very exciting.

3) Visualization of engagement: In order to better use the engagement recognition results in the classroom feedback system, we display the results visually, and teachers can judge the change trend of students' engagement in real time through color changes.

IV. EXPERIMENTS AND RESULTS

A. Evaluation Metrics

In this paper, we use two performance metrics including classification accuracy and F1 measure. Classification accuracy specifies the number of positive (engaged) and negative (disengaged) samples which are correctly classified and are divided by all samples.

$$\text{Accurancy } = \frac{TP + TN}{TF + FP + TN + FN} \tag{8}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively. F1 measure is calculated by following formula:

$$\mathrm{F1} = 2 \times \frac{p \times r}{p + r} \tag{9}$$

Where p is precision defined as $\frac{TP}{TP+FP}$ and r is recall defined as $\frac{TP}{TP+FN}$.

B. Implementation Details

In order to evaluate the engagement of online learners, our experiment uses the Dataset for Affective States in E-Environments (DAiSEE [26]). This dataset includes 112 individuals, where 80 male and 32 female. And it contains 9068 video clips in 6 different locations (such as dormitories, crowded laboratory spaces, libraries, etc.) and 3 different lighting settings (bright, dark and neutral); each video clip is 10 seconds long. This dataset consists of labels for four affective states related to user engagement, viz., engagement, frustration, confusion, and boredom, and each of the affective states is defined at four levels: (1) very low (2) low (3) high and (4) very high. The advantage of the aforementioned 4-level annotation of the emotional state related to the intensity value is that it can be changed to any other n-level engagement (for example, 2-levels or 3-levels or more) when required by a given application.

In our experiment, the engaged face images with the intensity values lower than 1 are given the label disengaged; or The face images with labels bored, confused, and frustratedand intensity values higher than 2 are given the label disengaged; other faces are labeled as engaged. Some sample faces of the DAiSEE dataset after the new label is assigned are shown in Fig. 7.

In the selection of Key Frames, first, we detect the face from each frame of the video and set the size to 200*200 pixels, then divide the face into 4 blocks, obtain the LDP feature code of each region, and create a 32-bit histogram,

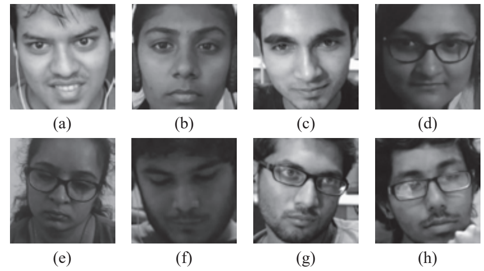

Fig. 7. Some sample faces of the DAiSEE dataset: (a)-(d) Faces with labels engaged; (e)-(h) Faces with labels disengaged.

so that 4 regions get a total of 32*(9+9+9+18) =1140 high-level data, and the dimension of this data is 200 after using PCA. The Cosine Similarity algorithm is used to calculate the expression similarity. On average, 3 Key Frames are selected for each video, then scale the face to a common size of 48×48 pixels.

C. Results

1) The overall engagement of the video: The results of our model on the videos is shown in table 1.In addition, we randomly select a face as a representative from all frames of the video, use CNN and SVM models to predict, and then use the same statistical analysis method to get the concentration level of each video. Repeat ten times, and the average effect obtained is shown in Table 2.

TABLE I
RECOGNITION PERFORMANCE OF OUR PROPOSED METHOD

Classification model	Accurancy	F1
SVM	0.7384	0.8420
CNN	0.7848	0.8595

TABLE II
RECOGNITION PERFORMANCE OF RANDOMLY SELECTING FACES

Classification model	Accurancy	F1
SVM	0.6805	0.8045
CNN	0.7301	0.8089

2) The frequency of facial expression changes: The frequency of facial expression changes sample is shown in Fig. 8. In Fig. 8(a), we can see that the student's expression changes very drastically, reaching the threshold twice within a period of time and extracting 3 Key Frames, while Fig. 8(c) changes relatively smoothly.

3) Visualization of engagement: We identified the engagement of each frame of a video, and visually displayed it to form a spectrogram of the engagement. We can see the engagement spectrogram in Fig. 9. As time progresses, the engagement of each frame is mapped to color and displayed in the spectrogram. The closer the recognition result is to

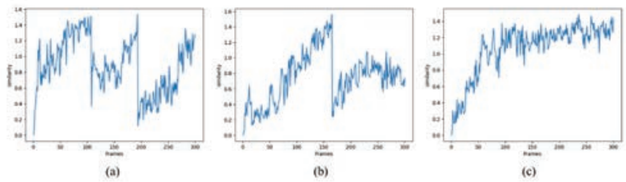

Fig. 8. The frequency of facial expression changes.

0, the more engaged it is, and the closer the color in the spectrogram is to blue; the closer the recognition result is to 1, the less engaged it is, and the closer the color in the spectrogram is to red.

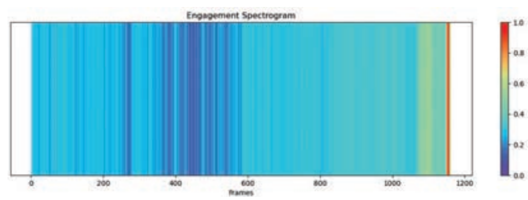

Fig. 9. The engagement spectrogram.

V. Conclusion

In this paper, we propose a model to identify the engagement level of online learners, and solve the problem of how to select Key Frames in the video in engagement recognition. It can be seen from the experimental results that the model we proposed is better than the traditional method. In the future, we can analyze the engagement of students in a certain course for one semester to predict the students' final exam results; analyze the engagement of all students different teaching processes to judge whether the teacher's teaching materials and methods can capture the attention of students. Our method can also be well applied to offline learning environments, and accurately detect the engagement of the face captured by the camera.

References

[1] Kahu, E.R.: Framing student engagement in higher education. Studies in higher education 38 (2013) 758–7

[2] Yin H. What motivates Chinese undergraduates to engage in learning? Insights from a psychological approach to student engagement research[J]. Higher Education, 2018, 76(5): 827-847.

[3] Chapman E. Alternative approaches to assessing student engagement rates[J]. Practical Assessment, Research, and Evaluation, 2002, 8(1): 13.

[4] Coates H. Attracting, engaging and retaining: New conversations about learning. Australasian student engagement report[J]. Camberwell, Vic: Australian Council for Educational Research Ltd, 2008.

[5] Booth B M, Ali A M, Narayanan S S, et al. Toward active and unobtrusive engagement assessment of distance learners[C]//2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2017: 470-476.

[6] Lan M, Hew K F. Examining learning engagement in MOOCs: A self-determination theoretical perspective using mixed method[J]. International Journal of Educational Technology in Higher Education, 2020, 17(1): 1-24.

[7] Jabid T, Kabir M H, Chae O. Robust facial expression recognition based on local directional pattern[J]. ETRI journal, 2010, 32(5): 784-794.

[8] Whitehill J, Serpell Z, Lin Y C, et al. The faces of engagement: Automatic recognition of student engagementfrom facial expressions[J]. IEEE Transactions on Affective Computing, 2014, 5(1): 86-98.

[9] TS A, Guddeti R M R. Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks[J]. Education and information technologies, 2020, 25(2): 1387-1415.

[10] D'Mello S K, Craig S D, Sullins J, et al. Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue[J]. International Journal of Artificial Intelligence in Education, 2006, 16(1): 3-28.

[11] Matthews G, Campbell S E, Falconer S, et al. Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry[J]. Emotion, 2002, 2(4): 315.

[12] D'Mello S, Lehman B, Pekrun R, et al. Confusion can be beneficial for learning[J]. Learning and Instruction, 2014, 29: 153-170.

[13] Joseph E. Engagement tracing: using response times to model student disengagement[J]. Artificial intelligence in education: Supporting learning through intelligent and socially informed technology, 2005, 125: 88.

[14] Sundar P, Kumar S. Disengagement detection in online learning using log file analysis[J]. International journal of computer technology and applications, 2016, 9(27): 195-301.

[15] Cocea M, Weibelzahl S. Log file analysis for disengagement detection in e-Learning environments[J]. User Modeling and User-Adapted Interaction, 2009, 19(4): 341-385.

[16] Chaouachi M, Pierre C, Jraidi I, et al. Affect and mental engagement: Towards adaptability for intelligent[C]//Twenty-Third International FLAIRS Conference. 2010.

[17] Tofighi G, Gu H, Raahemifar K. Vision-based engagement detection in virtual reality[C]//2016 Digital Media Industry Academic Forum (DMIAF). IEEE, 2016: 202-206.

[18] Chang C, Zhang C, Chen L, et al. An ensemble model using face and body tracking for engagement detection[C]//Proceedings of the 20th ACM international conference on multimodal interaction. 2018: 616-622.

[19] Miller B W. Using reading times and eye-movements to measure cognitive engagement[J]. Educational psychologist, 2015, 50(1): 31-42.

[20] Mohamad Nezami O, Dras M, Hamey L, et al. Automatic recognition of student engagement using deep learning and facial expression[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2020: 273-289.

[21] Grafsgaard J F, Fulton R M, Boyer K E, et al. Multimodal analysis of the implicit affective channel in computer-mediated textual communication[C]//Proceedings of the 14th ACM international conference on Multimodal interaction. 2012: 145-152.

[22] Aslan S, Cataltepe Z, Diner I, et al. Learner engagement measurement and classification in 1: 1 learning[C]//2014 13th International Conference on Machine Learning and Applications. IEEE, 2014: 545-552.

[23] Ekman P, Friesen W V. Facial action coding system[J]. Environmental Psychology Nonverbal Behavior, 1978.

[24] Upadhyay H, Kamat Y, Phansekar S, et al. User Engagement Recognition Using Transfer Learning and Multi-task Classification[M]//Intelligent Data Communication Technologies and Internet of Things. Springer, Singapore, 2021: 411-420.

[25] Pratt W K. Digital image processing: PIKS Scientific inside[M]. Hoboken, New Jersey: Wiley-interscience, 2007.

[26] Gupta A, D'Cunha A, Awasthi K, et al. Daisee: Towards user engagement recognition in the wild[J]. arXiv preprint arXiv:1609.01885, 2016.

报告编号：202309-229

检 索 报 告

项目名称：论文被 EI 收录情况证明

委 托 人：西南大学外国语学院 蒲丹青

日　　期：2023 年 9 月 21 日

认证单位：教育部科技查新工作站 N08

二〇二〇年制

<table>
<tr><td>检索项目名称</td><td colspan="4">委托人提交论文被 EI 收录情况</td></tr>
<tr><td rowspan="2">查新机构</td><td>名　称</td><td>教育部科技查新工作站 N08</td><td>邮　编</td><td>400715</td></tr>
<tr><td>地　址</td><td>重庆市北碚区西南大学图书馆</td><td>电　话</td><td>023-68253283</td></tr>
<tr><td>联系人</td><td colspan="4">蒲丹青（联系电话：15310292745）</td></tr>
<tr><td>委托文献目录</td><td colspan="4">1. Engagement Detection of Online Learners Based on Key Frames
Authors: Zhou, Zhurong (1); Pu, Danqing (1); Yang, Fan (1)
Source title: Proceedings - 2022 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Autonomous and Trusted Vehicles, Scalable Computing and Communications, Digital Twin, Privacy Computing, Metaverse, SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse 2022
Part number: 1 of 1
Issue title: Proceedings - 2022 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Autonomous and Trusted Vehicles, Scalable Computing and Communications, Digital Twin, Privacy Computing, Metaverse, SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse 2022
Issue date: 2022
Publication year: 2022</td></tr>
<tr><td>检索的数据库范围</td><td colspan="4">EI Compendex　(1969—今)</td></tr>
<tr><td>检索要点</td><td colspan="4">论文被 EI 收录情况</td></tr>
<tr><td>检索结论</td><td colspan="4">经检索，委托人提交的 1 篇论文被 EI 收录。检索结果详细情况见附件。

<table>
<tr><th>序号</th><th>作者排名</th><th>EI 检索号</th><th>文献类型</th><th>出版时</th><th>语种</th></tr>
<tr><td>1</td><td>2</td><td>20233414586538</td><td>Conference article (CA)</td><td>2022 年</td><td>英文</td></tr>
</table>

检索人（签名）：代洪波　代洪波
职称：副研究馆员
教育部科技查新工作站 N08
2023 年 9 月 21 日</td></tr>
</table>

附件：1. Engagement Detection of Online Learners Based on Key Frames

Accession number: 20233414586538

Authors: Zhou, Zhurong (1); Pu, Danqing (1); Yang, Fan (1)

Author affiliation: (1) Southwest University, College of Computer & Information Science, Chongqing, China

Corresponding author: Zhou, Zhurong(zhouzr@swu.edu.cn)

Source title: Proceedings - 2022 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Autonomous and Trusted Vehicles, Scalable Computing and Communications, Digital Twin, Privacy Computing, Metaverse, SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse 2022

Abbreviated source title: Proc. - IEEE SmartWorld, Ubiquitous Intell, Comput., Auton. Trust. Veh., Scalable Comput. Commun., Digit. Twin, Priv. Comput., Metaverse, SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse 2023
Part number: 1 of 1
Issue title: Proceedings - 2022 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Autonomous and Trusted Vehicles, Scalable Computing and Communications, Digital Twin, Privacy Computing, Metaverse, SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse 2022
Issue date: 2022
Publication year: 2022
Pages: 1065-1070
Language: English
ISBN-13: 9798350346558
Document type: Conference article (CA)
Conference name: 2022 IEEE SmartWorld, 19th IEEE International Conference on Ubiquitous Intelligence and Computing, 2022 IEEE International Conference on Autonomous and Trusted Vehicles Conference, 22nd IEEE International Conference on Scalable Computing and Communications, 2022 IEEE International Conference on Digital Twin, 8th IEEE International Conference on Privacy Computing and 2022 IEEE International Conference on Metaverse, SmartWorld/UIC/ATC/ScalCom/DigitalTwin/PriComp/Metaverse 2022
Conference date: December 15, 2022 - December 18, 2022
Conference location: Haikou, China
Conference code: 191151
Sponsor: IEEE; IEEE Computational Intelligence Society
Publisher: Institute of Electrical and Electronics Engineers Inc.
Abstract: In recent years, more and more students are willing to choose courses on online learning platforms, such as MOOC. However, the online learning platform cannot capture the learning status of students, and cannot give real-time feedback when students' attention is out of learning. The lack of feedback mechanisms mainly leads to a staggering dropout rate of MOOC. Therefore, engagement detection is very important for online learning platforms. Nowadays, many researchers have focused on engagement detection which are based on pictures. This method of selecting pictures is through random selection or repeated selection after short interval of time. However, both of these methods may lead to the problem of inaccurate samples or heavy identification workload. In this paper, we propose to introduce Key Frames that can more accurately represent the engagement level of the online learning personal video in the sample selection stage. Experimental results show that our proposed method has better results.

Number of references: 26
Main heading: Students
Controlled terms: E-learning - Learning systems
Uncontrolled terms: Engagement detection - Expression similarity - Feedback mechanisms - Key-frames - Learning platform - Local directional pattern - Online learning - Random selection - Real-time feedback - Short-interval
DOI: 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00157
Compendex references: YES
Database: Compendex
Data Provider: Engineering Village