International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 38, No. 3, June, 2021, pp. 865-873
Journal homepage: http://iieta.org/journals/ts

Online Learning Behavior Analysis Based on Image Emotion Recognition

Shunye Wang

College of Electronic and Information Engineering, Langfang Normal University, Langfang 065000, China

Corresponding Author Email: wangshunye@lfnu.edu.cn

https://doi.org/10.18280/ts.380333

Received: 2 January 2021
Accepted: 16 April 2021

Keywords:
image emotion recognition, online learning, learning behavior analysis, learning emotion recognition

ABSTRACT

There are many limitations of the current online learning platforms in the monitoring of learning behaviors and the evaluation of teaching effects. For example, the platforms cannot sense and correct the changes in the learning states and emotions of the students. To overcome the limitations, this paper tries to analyze online learning behaviors based on image emotion recognition. Firstly, the flow of image emotion recognition was detailed to facilitate the analysis of online learning behaviors, and the key frames were extracted from human face images, using improved local binary pattern (LBP) and wavelet transform. Next, the authors constructed the structure for the system of online learning behavior analysis, proposed a learning emotion recognition method based on facial expressions, and established an image emotion classification model for online learning, based on the attention mechanism. Experimental results show that the proposed algorithm is effective in analyzing online learning behaviors based on image emotion recognition.

1. INTRODUCTION

Online learning is an emerging way of learning in the Internet era. Compared with traditional classroom learning, online learning boasts incomparable advantages, such as diverse channels of knowledge acquisition, individualized learning plans, and a growing number of students [1-3]. However, there are many limitations of the current online learning platforms in the monitoring of learning behaviors and the evaluation of teaching effects [4-6]. For example, the platforms cannot sense and correct the changes in the learning states and emotions of the students, because face-to-face teacher instruction is removed from online learning. To ensure the effect of online learning, it is particularly important to recognize the emotions of students, and effectively monitor the learning behaviors during online learning.

Existing research shows that the performance of convolutional neural network (CNN) in data processing is greatly affected by whether the emotional image data are labeled [7-10]. Based on semi-supervised dynamic learning, Chaabi et al. [11] constructed a largescale dataset of image emotions, which effectively makes up for the gap between image features and human emotions, and formulated a relational learning network that effectively segments the foreground and background in original images.

Traditionally, image emotion recognition is grounded on statistics. The traditional models rely heavily on artificial visual features, which take lots of time and labor to construct, and incur a high cost of labeling the target dataset [12-16]. To solve the small sample problem of image emotion recognition, Narula et al. [17] formulated a two-layer transfer CNN capable of extracting universal low-level image features and high-level semantic features, and thereby effectively solved the matching errors caused by the distribution difference between regions of interest (ROIs). Drawing on the salient feature map extracted by Itti’s visual attention model, Padhy et al. [18] computed the weighted histogram of each block in the map, designed an eigenvector composed of image emotions like color, texture, and expression, and applied the eigenvector to the emotion-based film recommendation system; the application led to excellent recommendation results.

To improve the quality of online teaching and education, researchers have implemented data mining on the learning behavior records and behavior features of students in online learning [19-22]. Wang and Zhang [23] explored deep into the correlations between the purposes, interests, learning types, and behavior features of online learning students, evaluated how these factors influence the learning effect, and proposed an improvement scheme for personalized learning management system. After analyzing student behaviors in online learning, Purwoningsih et al. [24] introduced the analysis results to predict the learning effect and provide a reference for teaching interventions, and observed that the strategy greatly improved the effect of personalized learning.

The previous application of image emotion recognition in online learning behavior analysis faces several challenges: unrobust recognition effects, and high cost of labeling salient areas of image emotions [25-28]. Besides, there is a severe lacking of online learning behavior analysis that combines facial organ state detection with emotion perception analysis. To solve the problem, this paper tries to analyze online learning behaviors based on image emotion recognition. Section 2 details the flow of image emotion recognition to facilitate the analysis of online learning behaviors, and extracts the key frames from human face images, using improved local binary pattern (LBP) and wavelet transform. Section 3 provides the basic structure for the system of online learning behavior analysis, and presents a learning emotion recognition method based on facial expressions. Section 4 establishes an image emotion classification model for online learning, based on the attention mechanism. Through experiments, the proposed algorithm was proved effective in analyzing online learning behaviors based on image emotion recognition.

2. EXTRACTION OF KEY FRAMES

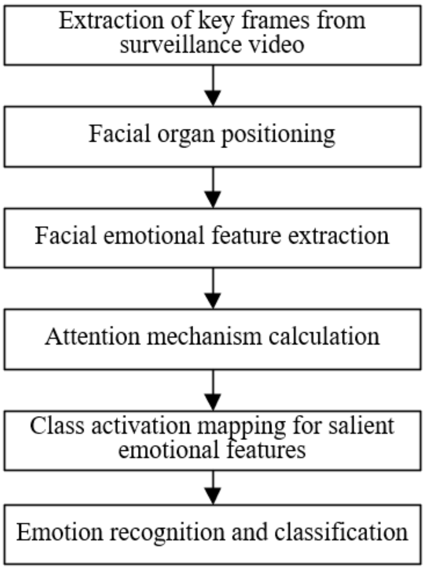

Extraction of key frames from surveillance video
Facial organ positioning
Facial emotional feature extraction
Attention mechanism calculation
Class activation mapping for salient emotional features
Emotion recognition and classification

Figure 1. Flow of image emotion recognition based on online learning behavior analysis

The image emotion recognition of surveillance video on online learning students mainly covers the following tasks: recognizing the static emotional features based on the facial emotional changes in the key frames of the video, and recognizing the dynamic emotional features based on the dynamic changes of facial expressions in a series of video images. Figure 1 explains the flow of image emotion recognition based on online learning behavior analysis. It can be seen that any emotion recognition method must extract features from facial expressions. This paper relies on improved LBP and wavelet transform to extract the key frames from facial expression images.

The LBP can characterize the relationship between a pixel and its neighboring pixels in the surveillance video image. In this paper, the improved uniform mode LBP operator $\Delta_{un\text{-}LBP}$ is adopted to extract the face features from each image in the surveillance video on online learning students. Let h_D be the grayscale of the central pixel in the image; $h_O(o=0, 1, L, O\text{-}1)$ be the grayscales of the other pixels in the circular neighborhood of the central pixel. Then, $\Delta_{un\text{-}LBP}$ can be expressed as:

$$\Delta_{un-LBP}=\begin{cases}\sum_{i=0}^{O-1} r\left(h_i-h_D\right), & if \quad V\left(\Delta_{LBP}\right)\le 2\\ O+1, otherwise\end{cases} \tag{1}$$

where, $V(\Delta_{un\text{-}LBP})$ can be calculated by:

$$V\left(\Delta_{LBP}\right)=\left|r\left(h_{o-1}-h_D\right)-r\left(h_i-h_D\right)\right| \\ +\sum_{i=0}^{O-1}\left|r\left(h_i-h_D\right)-r\left(h_{i-1}-h_D\right)\right| \tag{2}$$

The binary function $r(a)$ can be expressed as:

$$r\left(a\right)=\begin{cases}1, a\ge 0\\ 0, a<0\end{cases} \tag{3}$$

The two-dimensional (2D) Gabor wavelet was selected to process the texture of the images in the surveillance video on online learning students, aiming to extract the facial emotional features from the spatial domain and the time domain simultaneously. Let ||*|| be the remainder operation; λ and u be the direction and scale of Gabor filter, respectively; $l_{v,u}=[l_u cos\theta_v]$ and l_{max} be the central frequency and maximum central frequency, respectively, with $l_u=l_{max}/g$ and $\theta_v=v\pi/8$; $c=(a, b)$ be the coordinates of a pixel in an image; $\delta_{\lambda\text{-}u}$ be the kernel function of the 2D Gabor wavelet. Then, the wavelet transform of a grayscale image $GR(a, b)$ in the surveillance video can be expressed as:

$$P_{\lambda\text{-}u}\left(a,b\right)=\left\|GR\left(a,b\right)*\delta_{\lambda\text{-}u}\right\| \tag{4}$$

where, $\delta_{\lambda\text{-}u}$ can be calculated by:

$$\delta_{\lambda\text{-}u}=\frac{\left\|l_{\lambda\text{-}u}\right\|^2}{\varepsilon^2}e^{-\frac{\left\|l_{\lambda\text{-}u}\right\|^2\left\|c\right\|^2}{2\varepsilon^2}}\left(e^{il_{\lambda\text{-}u}c}-e^{\frac{\varepsilon^2}{2}}\right) \tag{5}$$

Let $FI_{SB}(a, b)$ and $FI_{XB}(a, b)$ be the real part filter and imaginary part filter of Gabor wavelet, respectively; $(FI*GR)$ be the convolution operation of $GR(a, b)$ and wavelet filter. After convolutional filtering of $GR(a, b)$ by $FI_{SB}(a, b)$ and $FI_{XB}(a, b)$ in eight directions and on five scales, the resulting real and imaginary parts of $GR(a, b)$ need to be computed by formula (6), in order to obtain the facial feature map after wavelet transform:

$$FEF\left(x,y\right)=\sqrt{\left(FI_{SB}\left(x,y\right)*GR\left(x,y\right)\right)^2+\left(FI_{XB}\left(x,y\right)*GR\left(x,y\right)\right)^2} \tag{6}$$

Finally, the grayscales of N facial expression feature maps were stretched to produce the one-dimensional (1D) Gabor eigenvector of the original facial expression image.

To effectively extract the emotional expressions of online learning students, it is necessary to associate the emotional information of the voice in the surveillance video with the emotional information of the facial expression image. Suppose there are m frames in each second of the surveillance video on online learning students. Then, each frame of the voice signal can be processed with Hamming window at the speed of m frames per second. Let FL and SH be the frame length and frame shift of the voice signal, respectively; SF=3FL is the sampling frequency. Then, the frame length can be calculated by:

$$FL\times m-\left(m-1\right)\times SH=SF \tag{7}$$

Formula 7 is equivalent to:

$$FL=3SF/\left(2m+1\right) \tag{8}$$

Let $\{g(1), g(2), \ldots, g(M)\}$ be the voice series processed by Hamming window, with M being the total number of frames.

Then, the time series $g(i)$ of the voice signal in frame i after Hamming window processing can be expressed as:

$$g(i)=\{v_i(j), j=1,2,\cdots,FL\} \tag{9}$$

The mean $r(i)$ of the absolute values of all amplitudes $v_i(j)$ in $g(i)$ can be computed by:

$$r(i)=\frac{\sum_{j=1}^{FL}|v_i(j)|}{FL} \tag{10}$$

Then, a series can be constructed based on $r(i)$:

$$a=\{r(i), i=1,2,...,M\} \tag{11}$$

Next, the voice frames and image frames in the surveillance video were paired on the time axis, and series a was sorted again by size. Based on the new series a', the image frame series $IF(i)$ was re-sorted into:

$$\begin{aligned} y&=\{I(i), i=1,2,...,N\} \\ b&=\{IF(i), i=1,2,...,M\} \end{aligned} \tag{12}$$

Considering the sheer size of the surveillance video, the image frames corresponding to the top 25% of voice frames in terms of amplitude were selected for extracting facial expression eigenvectors, in order to reduce the time cost of image frame processing. Suppose the LBP eigenvector LBP_i of frame i has θ dimensions. Then, we have:

$$LBP_i=\{\Delta_{lbp-i1},\Delta_{lbp-i2},...,\Delta_{lbp-i\theta}\} \tag{13}$$

For an m-frame surveillance video on online learning students, the series of LBP eigenvectors can be expressed as:

$$c=\{LBP_1, LBP_2,..., LBP_m\} \tag{14}$$

The Euclidean distance between LBP_i and the other (m-1) frames can be calculated by:

$$q_{i,j}=\sqrt{\left(\Delta_{lbp-i1}-\Delta_{lbp-j1}\right)^2+\left(\Delta_{lbp-i2}-\Delta_{lbp-j2}\right)^2+...+\left(\Delta_{lbp-i\theta}-\Delta_{lbp-j\theta}\right)^2} \tag{15}$$

The corresponding 1D vector can be described by:

$$Q_i=\left[q_{i,1}, q_{i,2},..., q_{i,M-1}\right] \tag{16}$$

The mean Q'_i of Q_i can be calculated by:

$$Q'_i=\sum_{j=1}^{M-1} q_{i,j} \tag{17}$$

The selected m frames of facial expressions demonstrate the emotions well. If a frame is highly similar as the previous and subsequent frames, the mean Q'_i will be small. In this case, the frame must be very similar to other frames. That is, the frame is relatively good at emotion expression in the frame series. This paper chooses the l frames with the minimum Q'_i and the best emotional expression effect as the targets of CNN-based emotion recognition.

3. LEARNING EMOTION RECOGNITION BASED ON FACIAL EXPRESSIONS

The surveillance video on online learning students is a rich data source for online learning behavior analysis. After extracting the key frames from the video, the online learning behavior analysis system was established. As shown in Figure 2, the system supports three functions: identity recognition, expression state recognition, and emotional perception. As can be seen from the figure, expression state recognition and emotional perception are the foundation work of accurate and thorough evaluation of learning behaviors. When a student participates in teaching activities, both positive and negative learning moods are closely related to the mouth features on his/her face. Therefore, the extraction of mouth features is an important aspect of recognizing facial expression changes from the surveillance video on online learning students.

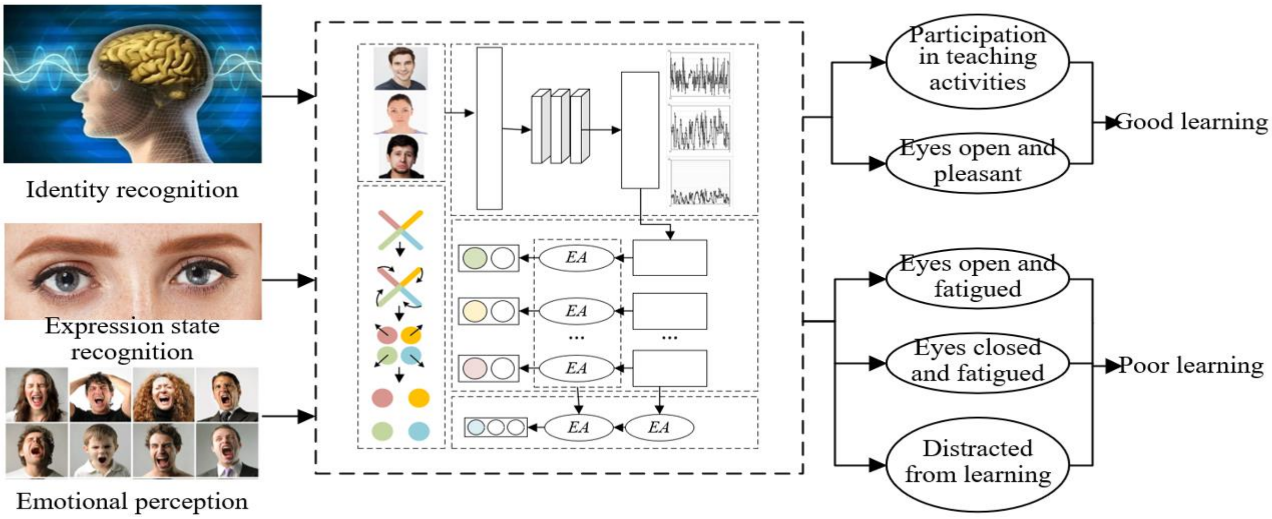


Figure 2. Structure of online learning behavior analysis system

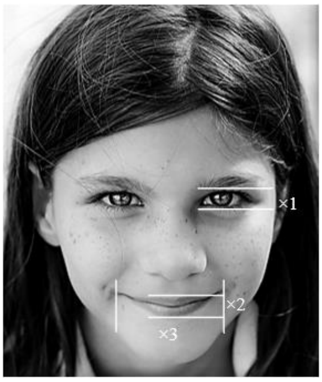


Figure 3. Characteristic parameters of facial organs

With the aid of the CNN, this paper positions four feature points on the face. The four points are respectively above, below, on the left to, and on the right to the center of the mouth. Firstly, the positions of the points were projected, and the distances were measured. Let a_1 be the degree of opening of the eyes; a_2 and a_3 be the vertical and horizontal distances, respectively; X and Y be the height and width of the mouth, respectively; C be the aspect ratio, i.e., the radian, of the mouth: C=A/B. Figure 3 shows the characteristic parameters of facial organs. Three emotions were defined as the target classes of emotion recognition for online learning behavior analysis: calm $MOOD_c$, pleasant $MOOD_j$, and fatigued $MOOD_f$.

$$\begin{cases} MOOD_c = 1 & 0.6 < C < 1.0 \\ MOOD_j = 1 & 0.4 < C < 0.6 \\ MOOD_f = 1 & 0.15 < C < 0.4 \end{cases} \tag{18}$$

Formula 18 was combined with Euclidean distance into a criterion for identifying emotion recognition targets, making it is more accurate to judge facial expressions. Let A=(a_1, a_2, a_3) be a 1D vector composed of three parameters. Then, the eigenvector of facial expressions can be given by:

$$\begin{aligned} & A_i = \left(a_{i1}, a_{i2}, a_{i3}\right), \\ & i = 1\left(MOOD_c\right), 2\left(MOOD_j\right), 3\left(MOOD_f\right) \end{aligned} \tag{19}$$

The difference vector between the three emotional features and the calm face:

$$Q_i A_i = \left(a_{i1}, a_{i2}, a_{i3}\right) - \left(a_{11}, a_{12}, a_{13}\right) \tag{20}$$

Let a'_I be the mean of a_i. Then, the Euclidean distance criterion can be established as:

$$q = \sqrt{\left(a_i - a'_i\right)^2} \tag{21}$$

In descending order of distance, the three moods can be ranked as $MOOD_c$, $MOOD_j$, and $MOOD_f$.

4. ATTENTION MECHANISM-BASED IMAGE EMOTION CLASSIFICATION

For online learning behavior analysis, image emotion recognition is premised on a large labeled dataset of image emotions of online learners. Nevertheless, it is extremely difficult to label a large dataset of image emotions, owing to the subjectiveness of image emotions. This paper constructs an image emotion classification model based on the attention mechanism (Figure 4). Specifically, the model consists of a CNN to acquire facial emotional features, a non-extreme channel attention mechanism that suppresses the negative effects of labels, and a class-specific activation map mechanism targeting the salient features of each type of emotions.

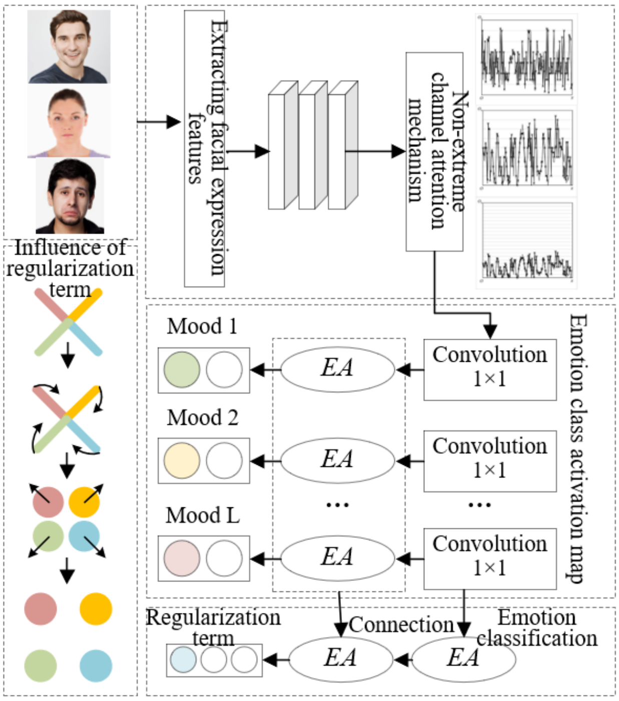

Extracting facial expression features
Non-extreme channel attention mechanism
Influence of regularization term
Mood 1
EA
Convolution 1×1
Mood 2
EA
Convolution 1×1
...
...
Mood L
EA
Convolution 1×1
Emotion class activation map
Regularization term
Connection
Emotion classification
EA
EA

Figure 4. Attention mechanism-based image emotion classification model

To begin with, the traditional CNN was implemented to extract facial expression features. On this basis, a feature matrix was derived for each facial expression image in the surveillance video. Let x=1, 2, 3, …, q and y = 1, 2, 3, …, q be the coordinates of the feature matrix; q be the length and width of the feature matrix; n be the number of convolution kernels; TZ be the number of classes in the emotional dataset. Then, the features extracted from the m-th image samples can be defined as $A^m_{x,y,n} \in \mathbb{R}^{TZ}$. As a nonlinear activation function, sigmoid function can be described as $g_{SIG}(a)=1/(1+e^{-a})$. Let ω and f be the weight and bias of the attention mechanism, respectively. The feature matrix can be converted into a 1D distribution of attention, using sigmoid function:

$$DV_n = g_{SIG}\left(\omega^T \cdot A_{x,y,n} + f\right) \tag{22}$$

Then, the attention distribution value was normalized to reduce the feature values of noisy labels:

$$FB_n = \frac{e^{DV_n}}{\sum_n e^{DV_n}} \tag{23}$$

The attention distribution value FB_n satisfies $\sum FB_n$=1. Let ⊗ be the multiplication of the elements of the matrix. By multiplying the normalized FB_n with the feature matrix of the previous facial expression images in the surveillance video, the attention feature map can be derived as:

$$SF^{m}_{x,y,n} = A^{m}_{x,y,n} \otimes FB^{m}_{n} \tag{24}$$

The softmax function can be described as $g_{SM}=(e^a/\sum e^a)$. To compare with the traditional attention mechanism, the channel attention mechanism FB_n in the field of image processing can be established as:

$$FB_n = g_{SM}\left(\omega^T \cdot A_{x,y,n} + f\right) \tag{25}$$

The corresponding spatial attention mechanism can be expressed as:

$$FB_{x,y,n} = g_{SM}\left(\omega^T \cdot A_{x,y,n} + f\right) \tag{26}$$

For comparison, the traditional attention mechanism can be expressed as:

$$FB_{x,y,n} = \frac{e^{A_{x,y,n}}}{\sum_i \sum_j e^{A_{x,y,n}}} \tag{27}$$

Let $IN_n^{ED}{}_l$ be the input eigenvector of the emotion detector ED_l after global average pooling; n_{SE} be the dimensionality of a specific emotion vector; l be the number of classes of the emotion dataset. Based on $IN_n^{ED}{}_l$ and $SF^m_{x,y,n}$, the emotion activation map $EA^{EDl}_{x,y,n}$ can be established as:

$$EA^{ED_l}_{x,y,n} = \sum_n IN^{ED_l}_n \cdot SF^m_{x,y,n} \tag{28}$$

Let L be the number of classes for the given emotions. Thus, a total of L emotion detectors were designed to derive L class activation maps $EA^{EDl}_{x,y,n}$ for salient features. Taking $EA^{EDl}_{x,y,n}$ as local features, an attention feature matrix, i.e., global feature, $SF^m_{x,y,n}$ could be established. Let ○ be the connection between features. Then, we have:

$$L_{x,y,n} = \left[SF^m_{x,y,n} \circ EA^{\gamma_1}_{x,y,n} \circ EA^{\gamma_2}_{x,y,n} \circ ...EA^{\gamma_l}_{x,y,n}\right] \tag{29}$$

To classify the emotions for online learning behavior analysis, the pretrained densely connected CNN was selected as the modeling basis. For the final emotion classification, the number of image samples was denoted as M, the input image as ST_i, the corresponding label as B_i. Then, the supervised learning method can be denoted as $\{ST_i,B_i\}^M_{i=1}$. Further, softmax function $g_{SM}=(e^a/\sum e^a)$ was taken as the final loss function for emotion classification. Through global average pooling, the final image emotion vector $o_i=GAP(\gamma_i, j, n)$ was obtained. On this basis, the minimize cross entropy loss function can be established as:

$$LOSS_{CLA} = -\sum\nolimits_{i=1}^{M} B_i log \frac{e^{\omega^T o_i}}{\sum_\gamma e^{\omega^T o_i}} \tag{30}$$

Suppose the M-th image belongs to the emotion class b_i, $i\in\Sigma L$. If B_i=1, the target image has the emotion of the current emotion detector; if B_i=0, the target image does not have the emotion of the current emotion detector. Suppose B_i is a one-hot code. Then, the loss function of the binary classifier for a specific class activation map can be established by:

$$LOSS_{TC} = -\sum\nolimits_\beta \sum\nolimits_{i=1}^{M} b_i log b_i' + \left(1-b_i\right)\left(1-log b_i'\right) \tag{31}$$

where, b'_l can be calculated by:

$$b_i' = \frac{1}{\left(1+e^{-g_i}\right)} \tag{32}$$

To further improve the emotion classification effect, this paper treats emotion classification as a linear problem, and introduces two regularization terms for the output emotion classes: the central loss to reduce inter-class distance, and the triplet loss to increase the inter-class distance. Let γ_{Bi} be the class center of eigenvector o_i. Then, the central loss $LOSS_{CL}$ can be described as:

$$LOSS_{CL} = \frac{1}{2}\sum\nolimits_{i=1}^{M}\left\|o_i - \gamma_{B_i}\right\|_2^2 \tag{33}$$

where, γ_{Bi} can be updated by:

$$\gamma_{B_i} = \frac{B_i e^{PA^T o_i}}{o \cdot \sum_\gamma e^{PA^T o_i}} \tag{34}$$

Only if $\sum^o_{j=1}B_{ij}=1$, could γ_{Bi} be updated based on o_i. Suppose φ is a hyperparameter. Then, the triplet loss $LOSS_{TL}$ can be described as:

$$LOSS_{TL} = \sum\nolimits_{i=1}^{M}\left[\left\|h\left(o_i^e\right)-h\left(o_i^r\right)\right\|_2^2 - \left\|h\left(o_i^e\right)-h\left(o_i^m\right)\right\|_2^2 + \phi\right] \tag{35}$$

The other hyperparameter $h(*)$ can be calculated by:

$$h\left(*\right) = \left\|\frac{e^{PA^T o_i}}{\sum_\gamma e^{PA^T o_i}}\right\|_2^2 \tag{36}$$

where, o^e_i and o^r_i correspond to the same emotion class; o^e_i and o^m_i correspond to different emotion classes. Let μ and η be the weight coefficients that balance the contributions of $LOSS_{CL}$ and $LOSS_{TL}$. The final emotion classification loss function can be established by:

$$\begin{aligned} LOSS &= LOSS_{CLA}\left(ST, B\right) + LOSS_{TC}\left(ST, b\right) \\ &+ \mu LOSS_{CL}\left(o\right) + \eta LOSS_{TL}\left(o\right) \end{aligned} \tag{37}$$

5. EXPERIMENTS AND RESULTS ANALYSIS

Figure 5 shows the LBP-based facial expression histogram. Based on the LBP, the calculation of mean Euclidean distance helps to extract the key frames from the surveillance video on online learning students. The extracted features lay a reasonable basis for subsequent emotion and mood recognition, and thereby improve the recognition efficiency of online learning behaviors. Contrastive experiments were designed to verify the effectiveness of the proposed key frame extraction method. Table 1 compares the emotion recognition results of different key frame extraction methods; Figure 6 compares the emotion recognition histograms of different key frame extraction methods.

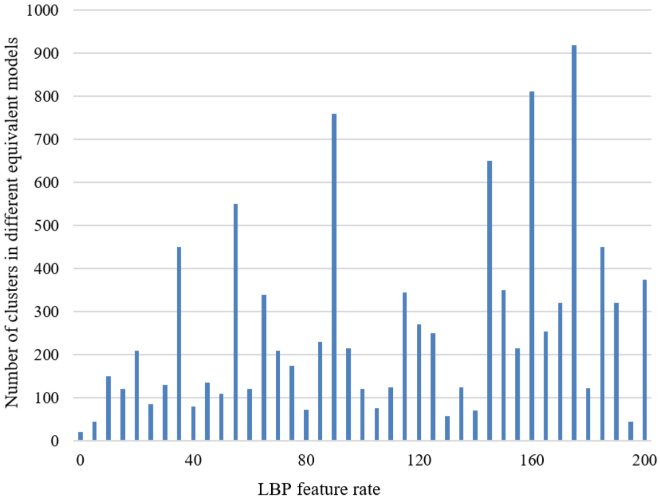


Figure 5. LBP-based facial expression histogram

As shown in Table 1 and Figure 6, the proposed key frame extraction method for online learning images recognized single static key frames slightly better than the contrastive extraction methods, which are respectively based on voice amplitude (+8.14%), clustering algorithm (+8.09%), and peak frame (+7.61%). This confirms the effectiveness of the proposed key frame extraction method in the recognition of online learning emotions. Besides, our emotion recognition approach, which is based on the mean facial expression features of the key frames from several images, outperformed the image emotion recognition for a single static key frame (+6.04%). Our algorithm can take the average of the frames with the best emotional expression effect. Its emotion recognition effect was much better than the contrastive methods. The advantage in recognition rate was 10.54%, 15.23%, and 8.66%, respectively. Therefore, image emotion recognition can be improved by the proposed approach, which solves the mean facial expression features from the key frames of several images.

The next is to verify the classification effectiveness of the regularization terms in the attention mechanism-based image emotion classification model. For this purpose, the eigenvector on the last layer of the proposed model was subject to dimensionality reduction. Then, the 2D scatterplots of online video image sample distribution were visualized on a test set containing a huge number of online learning image emotions. Figure 7 presents the 2D scatterplots on the test set after 20, 40, 60, and 80 training cycles. It can be inferred from the figure that, with the growing number of training cycles, the online video image samples of the same emotion class gradually approached each other, while those of different emotion classes departed from each other. This reflects the good effect of our model in emotion recognition.

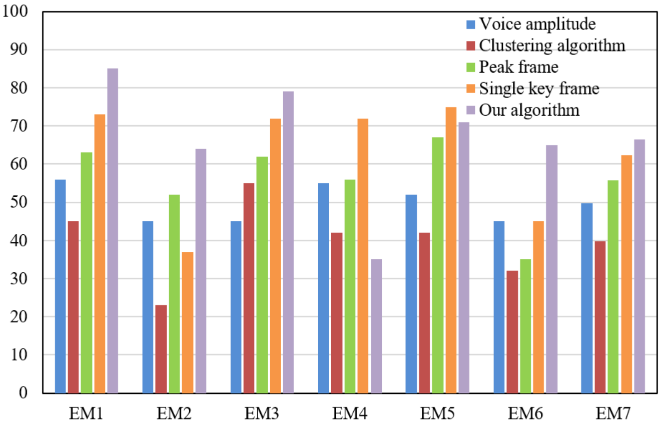
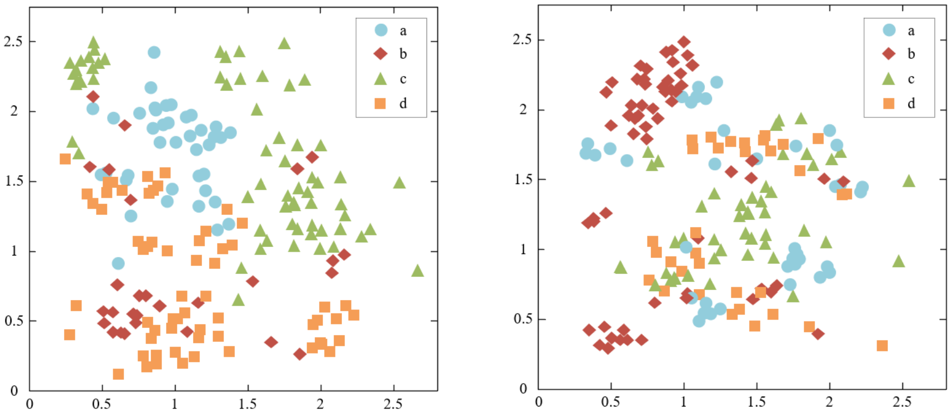


Figure 6. Emotion recognition histograms of different key frame extraction methods

Table 1. Emotion recognition results of different key frame extraction methods

Emotion class number	Voice amplitude	Clustering algorithm	Peak frame	Single key frame	Our algorithm
EM1	72.86	61.51	68.26	78.14	79.37
EM2	72.47	55.75	75.03	81.03	78.42
EM3	65.73	54.82	62.47	60.56	72.35
EM4	81.25	68.73	85.91	92.08	97.07
EM5	73.21	58.74	81.34	83.74	90.49
EM6	74.62	62.98	72.41	77.54	82.31



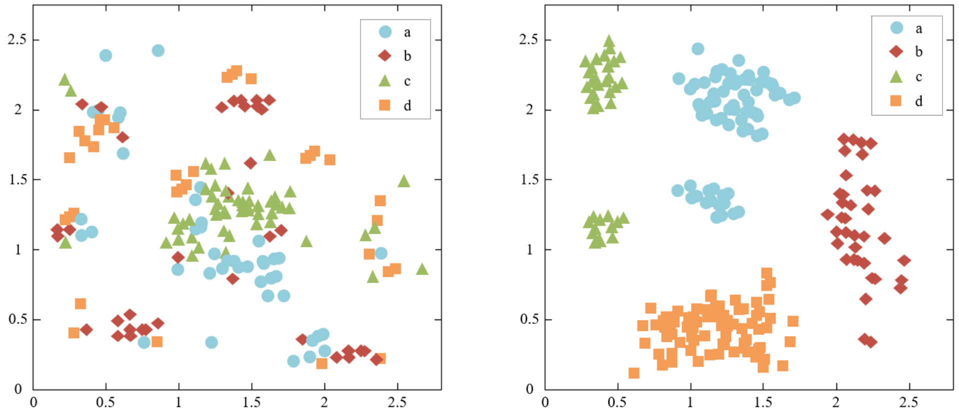


Figure 7. 2D scatterplots on the test set after different training cycles

Table 2. Experimental results on the selection of weight coefficients

η μ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	72.67	75.95	75.32	75.82	74.55	74.32	75.02	76.53	74.31	74.25	75.61
0.1	75.95	76.63	72.04	76.31	75.31	76.31	75.39	76.71	74.03	76.37	75.61
0.2	75.41	74.49	72.04	74.86	75.95	75.42	76.74	76.39	75.96	75.21	76.33
0.3	72.03	75.31	79.69	78.55	74.37	80.04	75.35	76.54	74.49	72.33	76.43
0.4	76.22	46.24	76.33	72.73	75.02	76.39	74.28	77.92	71.53	74.98	74.05
0.5	71.47	75.57	74.27	76.84	74.81	77.17	76.01	78.21	73.09	75.49	73.57
0.6	75.35	76.02	75.63	76.53	76.37	76.42	75.29	76.25	74.23	78.21	74.28
0.7	76.84	74.97	78.58	74.27	76.52	77.95	76.37	75.37	75.92	77.32	75.61
0.8	74.12	75.08	77.57	76.71	74.39	75.08	78.25	74.09	76.55	75.46	68.42
0.9	78.68	75.24	75.31	74.52	73.27	74.23	77.64	75.32	75.03	75.45	73.23
1.0	77.29	75.61	74.20	77.35	73.52	73.51	75.32	76.21	71.43	72.31	74.25

Next, a grid search was carried out with the interval of [0, 1] and the step length of 0.1. The purpose is to optimize the values of the weight coefficients μ and η, which balance the contributions of $LOSS_{CL}$ and $LOSS_{TL}$. The attention mechanism-based image emotion classification model is denoted as "Based". The model consists of a CNN, a non-extreme channel attention mechanism, a class-specific activation map mechanism, and two regularization terms (central loss and triplet loss). As shown in Table 2, the highest emotion recognition ate of our method was observed (80.04%) at μ=0.3, and η=0.5.

Table 3. Results of online learning behavior analysis

Test number	Eyes open and pleasant	Eyes open and calm	Eyes open and fatigued	Eyes semi-closed and pleasant	Eyes semi-closed and fatigued	Eyes closed and calm	Eyes closed and fatigued	Participating in teaching activities	Distracted from learning	Learning state
1	1	2	0	0	0	0	0	0	0	Good
2	0	2	1	1	0	1	1	0	1	Poor
3	0	1	0	1	1	0	0	0	0	Good
4	0	0	0	0	0	0	0	0	0	General
5	0	0	1	2	1	0	0	0	0	General
6	1	0	2	0	0	1	0	0	0	Poor
7	0	0	0	1	0	0	0	0	0	Good
8	2	1	0	3	0	1	0	0	1	General
9	0	0	1	2	1	1	0	0	0	Poor
10	4	2	0	0	0	0	1	0	0	Good

During online learning, the facial expressions of students can be judged by the states of eyes and mouths. For the online learning experiments, nine classes of learning states were defined, namely, eyes open and pleasant, eyes open and calm, eyes open and fatigued, eyes semi-closed and pleasant, eyes semi-closed and fatigued, eyes closed and calm, eyes closed and fatigued, participating in teaching activities, and distracted from learning (Table 3). The students are in a good learning state, if they are found to have one of the following three emotions: eyes open and pleasant, eyes open and calm, and participating in teaching activities; the students are in a general learning state, if they are found to have one of the following three emotions: eyes semi-closed and pleasant, eyes semi-closed and fatigued, and eyes closed and calm; the students are in a poor learning state, if they are found to have one of the following three emotions: eyes open and fatigued, eyes closed and fatigued, and distracted from learning. The experimental results show that the learning behaviors and changing learning ability of online learning students could be detected based on the states of student eyes and mouths, and facial emotional features. This strategy can visually display the quality of learning states, and facilitate the evaluation of learning behaviors.

6. CONCLUSIONS

This paper analyzes online learning behaviors based on image emotion recognition. After detailing the flow of image emotion recognition for online learning behavior analysis, this paper extracts the key frames from facial expression images through improved LBP and wavelet transform. Then, the mean expression feature was solved form several extracted key frames. The data and histogram evidences of experiments show that the proposed method can improve the effect of image emotion recognition, compared with different key frame extraction methods. After that, the authors established the structure of online learning behavior analysis system, proposed a learning emotion recognition method based on facial expressions, and constructed an emotional classification model for online learning images based on the attention mechanism. Finally, the authors drew the 2D scatterplots on the test set with different training cycles, and presented the results of online learning behavior analysis. The results demonstrate the classification effectiveness of the regularization terms in the attention mechanism-based image emotion classification model, and the applicability of our model to online learning behavior analysis.

ACKNOWLEDGEMENT

This paper was supported by Youth Artificial Intelligence Education Innovation Center in Langfang Normal University (Grant No. XPT202103) and Key Programs of Scientific Research Foundation of Hebei Educational Committee (Grant No. ZD2019056).

REFERENCES

[1] Xie, S.T., Chen, Q., Liu, K.H., Kong, Q.Z., Cao, X.J. (2021). Learning behavior analysis using clustering and evolutionary error correcting output code algorithms in small private online courses. Scientific Programming. https://doi.org/10.1155/2021/9977977

[2] Corchs, S., Fersini, E., Gasparini, F. (2019). Ensemble learning on visual and textual data for social image emotion classification. International Journal of Machine Learning and Cybernetics, 10(8): 2057-2070. https://doi.org/10.1007/s13042-017-0734-0

[3] Takanashi, T., Hirai, K., Horiuchi, T. (2019). Construction of facial emotion database through subjective experiments and its application to deep learning-based facial image processing. Electronic Imaging, 2019(11): 267-1-267-7(7). https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-267

[4] Islam, M.R., Ahmad, M. (2019). Virtual image from EEG to recognize appropriate emotion using convolutional neural network. 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, pp. 1-4. https://doi.org/10.1109/ICASERT.2019.8934760

[5] Rao, T., Li, X., Zhang, H., Xu, M. (2019). Multi-level region-based convolutional neural network for image emotion classification. Neurocomputing, 333: 429-439. https://doi.org/10.1016/j.neucom.2018.12.053

[6] Huang, W., Zhu, S., Yao, X. (2021). Destination image recognition and emotion analysis: evidence from user-generated content of online travel communities. The Computer Journal, 64(3): 296-304. https://doi.org/10.1093/comjnl/bxaa064

[7] Yang, K., Wang, C., Sarsenbayeva, Z., Tag, B., Dingler, T., Wadley, G., Goncalves, J. (2020). Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. The Visual Computer, 37: 1447-1466. https://doi.org/10.1007/s00371-020-01881-x

[8] Özcan, T., Baştürk, A. (2019). Static image-based emotion recognition using convolutional neural network. In 2019 27th signal processing and communications applications conference (SIU), Sivas, Turkey, pp. 1-4. https://doi.org/10.1109/SIU.2019.8806408

[9] Deepak, N.A., Shobha, N.S. (2021). Analysis of learner's behavior using latent Dirichlet allocation in online learning environment. In Computational Methods and Data Engineering, pp. 231-242. https://doi.org/10.1007/978-981-15-7907-3_18

[10] Hao, Y., Leng, C., Zheng, H., Zhang, H. (2021). Research on online learning behavior analysis based on big data architecture. 2021 2nd International Conference on Computers, Information Processing and Advanced Education, pp. 519-523. https://doi.org/10.1145/3456887.3457004

[11] Chaabi, Y., Messoussi, R., Hilaire, V., Lekdioui, K., Ruichek, Y., Touahni, R. (2014). An automatic system for the determination of learner's sociological behavior from textual asynchronous conversations analysis in online collaborative learning. 2014 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14), Rabat, Morocco, pp. 1-7. https://doi.org/10.1109/SITA.2014.6847305

[12] El Haddioui, I., Khaldi, M. (2012). Learner behavior analysis on an online learning platform. International Journal of Emerging Technologies in Learning (iJET): 7(2), 22-25. http://dx.doi.org/10.3991/ijet.v7i2.1932

[13] Kim, J.H., Poulose, A., Han, D.S. (2021). The extensive usage of the facial image threshing machine for facial emotion recognition performance. Sensors, 21(6): 2026. https://doi.org/10.3390/s21062026

[14] Bum, J., Choo, H., Whang, J. J. (2021). Image-based lifelogging: user emotion perspective. CMC-Computers Materials & Continua, 67(2): 1963-1977. http://dx.doi.org/10.32604/cmc.2021.014931

[15] Kumar, P., Raman, B. (2021). Domain adaptation based technique for image emotion recognition using image captions. Computer Vision and Image Processing: 5th International Conference, CVIP 2020, Prayagraj, India, pp. 394-406.

[16] Rao, T., Li, X., Xu, M. (2020). Learning multi-level deep representations for image emotion classification. Neural Processing Letters, 51(3): 2043-2061. https://doi.org/10.1007/s11063-019-10033-9

[17] Narula, V., Feng, K., Chaspari, T. (2020). Preserving privacy in image-based emotion recognition through user anonymization. In Proceedings of the 2020 International

Conference on Multimodal Interaction, pp. 452-460. https://doi.org/10.1145/3382507.3418833

[18] Padhy, N., Singh, S.K., Kumari, A., Kumar, A. (2020). A literature review on image and emotion recognition: proposed model. Smart Intelligent Computing and Applications, pp. 341-354. https://doi.org/10.1007/978-981-32-9690-9_34

[19] Nagahama, T., Morita, Y. (2017). An analysis of students' learning behaviors using variable-speed playback functionality on online educational platforms. International Conference on Human-Computer Interaction, pp. 154-159. https://doi.org/10.1007/978-3-319-58753-0_24

[20] Nasir, M., Dutta, P., Nandi, A. (2020). Recognition of changes in human emotion from face image sequence using triangulation induced Barycentre-Orthocentre paired distance signature. 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, pp. 101-106. https://doi.org/10.1109/ComPE49325.2020.9200068

[21] Joseph, A., Geetha, P. (2020). Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow. The Visual Computer, 36(3): 529-539. https://doi.org/10.1007/s00371-019-01628-3

[22] Ruan, C., Yang, L. (2021). Analysis of online learning behavior characteristics and influencing factors based on link prediction. 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, pp. 1059-1062. https://doi.org/10.1109/IPEC51340.2021.9421205

[23] Wang, J., Zhang, Y. (2019). Clustering study of student groups based on analysis of online learning behavior. Proceedings of the 2019 International Conference on Modern Educational Technology, pp. 115-119. https://doi.org/10.1145/3341042.3341065

[24] Purwoningsih, T., Santoso, H.B., Hasibuan, Z.A. (2019). Online Learners' behaviors detection using exploratory data analysis and machine learning approach. 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, pp. 1-8. https://doi.org/10.1109/ICIC47613.2019.8985918

[25] Shin, D., He, S., Lee, G.M., Whinston, A.B. (2015). Sharing behavior in online social media: An empirical analysis with deep learning. Workshop on E-Business, pp. 222-227. https://doi.org/10.1007/978-3-319-45408-5_26

[26] Hossain, M.A., Assiri, B. (2020). Emotion specific human face authentication based on infrared thermal image. 2020 2nd International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, pp. 1-6. https://doi.org/10.1109/ICCIS49240.2020.9257683

[27] Uchida, M., Akaho, R., Ogawa-Ochiai, K., Tsumura, N. (2019). Image-based measurement of changes to skin texture using piloerection for emotion estimation. Artificial Life and Robotics, 24(1): 12-18. https://doi.org/10.1007/s10015-018-0435-0

[28] Zou, X. (2021). Analysis of consumer online resale behavior measurement based on machine learning and BP neural network. Journal of Intelligent & Fuzzy Systems, 40(2): 2121-2132. https://doi.org/10.3233/JIFS-189212