Multimedia Tools and Applications (2021) 80:24287–24301
https://doi.org/10.1007/s11042-021-10836-w

Action unit classification for facial expression recognition using active learning and SVM

Li Yao[1] · Yan Wan[1] · Hongjie Ni[1] · Bugao Xu[2]

Received: 3 June 2019 / Revised: 7 December 2020 / Accepted: 10 March 2021 /
Published online: 4 April 2021

Abstract
Automatic facial expression analysis remains challenging due to its low recognition accuracy and poor robustness. In this study, we utilized active learning and support vector machine (SVM) algorithms to classify facial action units (AU) for human facial expression recognition. Active learning was used to detect the targeted facial expression AUs, while an SVM was utilized to classify different AUs and ultimately map them to their corresponding facial expressions. Active learning reduces the number of non-support vectors in the training sample set and shortens the labeling and training times without affecting the performance of the classifier, thereby reducing the cost of labeling samples and improving the training speed. Experimental results show that the proposed algorithm can effectively suppress correlated noise and achieve higher recognition rates than principal component analysis and a human observer on seven different facial expressions.

Keywords Action unit · Facial expression recognition · Active learning · Support vector machine

1 Introduction

Action units (AUs) are the fundamental actions of individual muscles or groups of muscles. Ekman and Friesen analyzed the relationship between AU movement and facial expressions in the Emotion Facial Action Coding System (EMFACS) [3, 7] and claimed that all AUs are external representations of muscle movements. The measurements in EMFACS are AUs, not muscles, for two reasons. First, for a few appearances, two or more muscles are combined into a single AU because the changes in appearance they produce cannot be distinguished. Second,

✉ Yan Wan
winniewan@dhu.edu.cn

[1] School of Computer Science and Technology, Donghua University, Shanghai, China

[2] Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA

the appearance changes produced by one muscle are sometimes separated into two or more AUs to represent the relatively independent actions of different parts of the muscle.

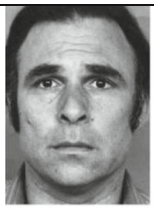
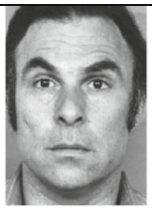
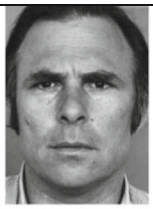
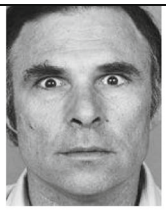
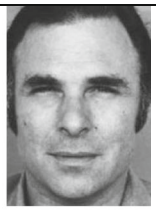
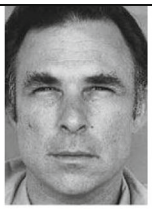
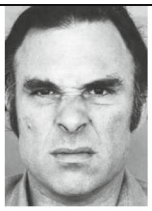
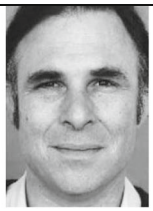
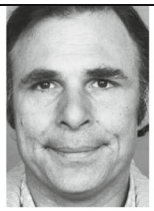
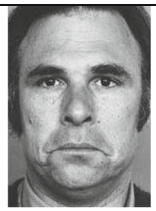
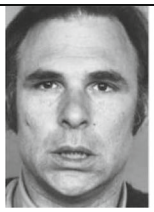
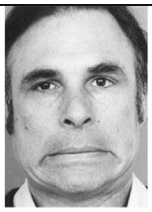
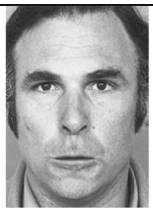
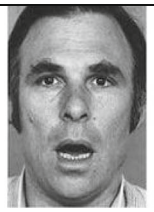
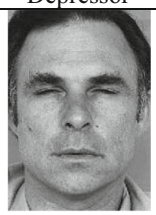
Our algorithm is based on a simplified EMFACS model in which 15 AUs are related to facial expressions, as shown in Table 1.

Facial expression is one of the most powerful, natural, and direct means of expressing emotion and intention. The automatic analysis of facial expression is an interesting and challenging problem with important applications in the fields of human–computer interaction and data-driven animation. Automatic facial expression recognition has attracted much attention in recent years owing to its wide range of applications.

(1) Human–computer interaction

Human–computer interaction is the most direct application of facial expression recognition, where a computer can recognize different expressions to guess the inner feelings of human beings and present different reactions according to the different expressions of human faces.

Table 1 Action units related to facial expressions.

				
（a）AU1 Inner Brow Raiser	（b）AU2 Outer Brow Raiser	（c）AU4 Brow Lowerer	（d）AU5 Upper Lid Raiser	（e）AU6 Cheek Raiser
				
（f）AU7 Lid Tighterner	（g）AU9 Nose Wrinkler	（h）AU12 Lip Corner Puller	（i）AU14 Dimpler	（j）AU15 Lip Corner Depressor
				
（k）AU16 Lower Lip Depressor	（l）AU20 Lip Stretcher	（m）AU23 Lip Tightener	（n）AU26 Jaw Drop	（o）AU45 Blinking

For example, if a machine finds a human with a sad expression, it can take a soothing action or send out a comforting message according to an established procedure.

(2) Medical application

Patients' expressions often express a variety of emotional information about disease, diagnosis, and treatment. For some critically and mentally ill patients, nursing robots with expression analysis can be used. When a patient's illness or mood fluctuates greatly, a robot can quickly predict the patient's current state and take relevant preventive measures to better avoid accidents. Alternatively, if a camera finds that a patient has a negative expression, such as pain, the machine can send out a command or trigger an action, and nursing staff can quickly reach the patient.

(3) Automotive safety

Expression can play an important role in preventing driving with fatigue. Fatigued driving statistically ranks first among the causes of traffic accidents in China. In view of this, electronic equipment for facial expression recognition based on facial movement units can analyze a driver's mental state in real time and take necessary measures, such as playing an alarm or stimulating music, to prevent possible traffic accidents.

(4) Animation synthesis

Expression animation is an important branch of computer animation research. If a computer can synthesize human expression, it will be beneficial to human understanding of robots.

2 Related work

In recent years, domestic and foreign AU detection research has mainly focused on the following three aspects: (1) extracting effective face features, (2) the relationship between different AUs, and (3) extracting time series information from AU images.

An AU detection object is a face object, and extracting effective face features is the first key step. In the early research on AU detection, the face was divided into several regions, where geometric or texture features were extracted from each region, combined into a single feature, and classified with a classifier. Cheng et al. [5] presented a location-aware music recommender system called Venue-Music, used hidden variables to describe users' music preferences, constructed a latent variable model to infer users' music preferences in different situations, and realized personalized music recommendations. Fabian et al. [3] combined the geometric and texture features near key points to enhance the feature representation of each face region. Zhao et al. [27] used joint region learning to detect AUs, selected 49 key points near the eyes, nose, and mouth, and extracted SIFT features to represent each region. The core of these methods is extracting the artificial features of key areas of the face to obtain a feature representation suitable for AU detection. In recent years, the extraction of artificial features has been replaced by the feature representation of deep learning. Gadi et al. [9] used a seven-layer convolutional neural network to complete the strength calculation of AU detection. Zhao [28] further subdivided the face into 8×8 regions and used deep learning to obtain the features

of all regions. Since 2015, feature extraction of faces in AU detection has mostly been realized by deep learning methods. According to Chen [4], an image will show some spatial differences under different angles of sensor acquisition.

Feature extraction is an important step in micro-expression recognition. The performance of micro-expression recognition has been improved by the spatiotemporal complete local quantization pattern feature [11], spatiotemporal local binary pattern integral graph feature [10], and six intersection local binary pattern features [23]. In addition, the independent tensor color space proposed by Wang et al. [21] combines some features of LBP-TOP and obtains a good micro-expression recognition effect of RGB space. The gradient feature [17] can also be used to describe the dynamic changes of the face. In recent years, based on its characteristics, optical flow has gradually become the main focus of micro-expression action recognition. Xu [25] used an optical flow field to observe the changes in facial micro-motion and proposed facial dynamic characteristics. Liu et al. [12] divided the face into different regions and calculated the average optical flow of local regions. They chose a support vector machine (SVM) classifier to realize micro-motion recognition technology with high recognition accuracy and proposed the main direction average optical flow. In December 2018, the HCP lab [13] proposed an improved BB-FCN model based on a cascaded backbone branched fully collaborative network (FCN) for facial landmark localization. The model is two-stage and explores in a coarse-to-fine manner, which can give a full response map and is applicable to uncontrolled environments.

AU detection has been studied for multiple decades, and its goal is to recognize and predict AU labels in each frame of a facial expression video. Automatic detection of AUs has a wide range of applications, such as human–machine interfaces, affective computing, and driving monitoring [2].

According to the physiological distribution of muscles, the movement rules of muscles can be grouped into AUs according to the relevant characteristics. The face can be divided into several major areas, such as left eye, right eye, nose, and mouth [14].

An AU is a subtle movement of the facial muscles. AUs can be combined to represent all possible facial expressions (e.g., frowning, sipping mouth, etc.), and they are the cornerstone of facial expression. Face AU recognition is a multi-label classification problem. The multi-label constraints can be limited to a finer granularity to achieve higher accuracy.

Automatic facial expression analysis is an interesting and challenging area that has an impact on many applications, including human–computer interaction and data-driven animation [19]. Ekman [6] proposed basic emotions for the modern theory of expression analysis. Ekman and Friesen built the EMFACS [8] in 1983, revealing the relationship between facial muscle changes and emotional states. They classified the muscle groups AUs according to the physiological distribution and characteristics of relevant muscle movements, and they summarized seven universal facial expressions that represent emotion: happiness, sadness, anger, fear, surprise, disgust, and contempt [16].

Based on this theory, much progress has been made in automatic facial expression recognition. Shreve adopted mathematical methods to detect changes in the optical flow field generated by non-rigid muscle movement of the face when producing an expression [20]. For classifying micro-expressions, Polikovsky divided the face into twelve areas, each having its own timing characteristics, namely, 3D-gradient orientation histogram descriptors [18]. In the continuous expression frame, the changes between these characteristics were utilized to determine AUs and identify different expressions. However, the twelve facial regions were selected manually, and the face position did not change significantly in the video. Fu calculated discriminant features based on discriminant tensor analysis (DTSA) and then used

the extreme learning machine (ELM) classifier to recognize facial emotions when subjects in an image were sparse and met other requirements [22]. Abdul-Majjed used an SVM to classify seven facial expressions [1] and achieved 75.8% recognition accuracy.

The SVM classifier is illustrated by two samples in Fig. 1. The red and blue dots in the figure represent two types of training samples. H is the classification line that separates the two types correctly. H1 and H2 are parallel to the classification line H and are the samples closest to the classification line. The distance between H1 and H2 is called the category gap. The optimal classification line is not only to separate the two categories correctly but also to maximize the gap between the two categories. When extended to high-dimensional space, the optimal classification line becomes the optimal classification surface.

The classification principle of the SVM can be summarized as finding a classification hyperplane such that the two kinds of sample points in the training sample can be separated and as far away from the plane as possible. For the linear non-separable problem, the data in the low-dimensional input space is mapped to the high-dimensional space through the kernel function such that the linear non-separable problem in the original low-dimensional space can be transformed into a linear separable problem in the high-dimensional space.

The kernel functions of the SVM are usually seen as follows:

(1) Polynomial kernel function:

$$K\left(x_i,x_j\right) = \left(\gamma x_i^T x_j + r\right)^d, \gamma > 0.$$

The SVM based on the kernel function is a d-order polynomial classifier. When d = 1, we obtain a linear support vector machine.

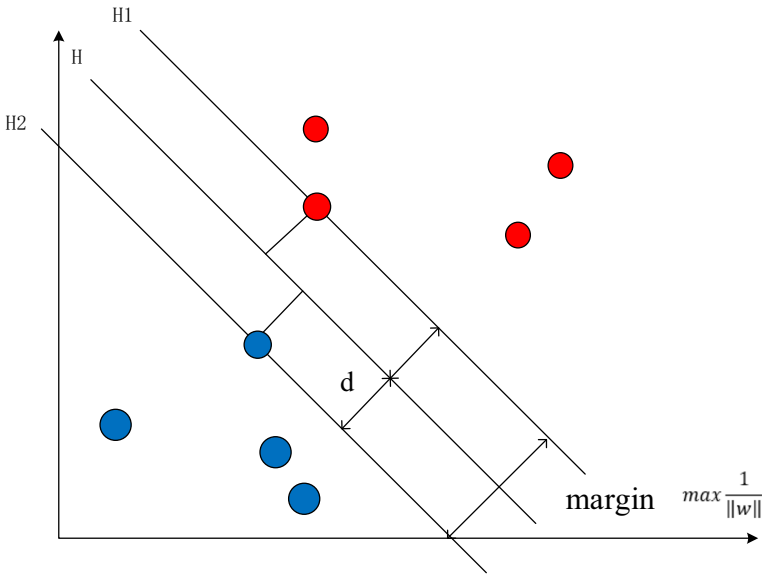


Fig. 1 Support vector machine

(2) Radial basis function:

$$K(x_i,x_j) = \exp\left(-\gamma\left\|x_i-x_j\right\|^2\right), \gamma > 0.$$

(3) Sigmoid kernel function:

$$K(x_i,x_j) = tanh\left(\gamma x_i^T x_j + \gamma\right).$$

Among them, there is no definite method to select the kernel function in an SVM, and it mostly depends on experience.

The basic idea of the combination of active learning and SVM is as follows. First, none of the candidate samples are labeled with categories. Based on the prior knowledge, the initial training sample set can be constructed by randomly selecting a small number of samples from the candidate sample set and labelling their categories to ensure that it contains at least one positive sample and one negative sample. The initial training sample set is used to train an SVM classifier. Under the classifier, an algorithm is used to select the sample that is most conducive to the performance of the classifier from the candidate sample set, label the category, add it to the training sample set, retrain the classifier, and repeat the process until the candidate sample set is empty or reaches a certain index.

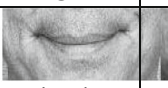
SVMs have been widely used in machine learning and pattern recognition because they can effectively solve non-linear problems for a global optimal solution and have strong generalization ability without over-learning. However, to obtain a high recognition rate, a large number of training samples must be used to obtain the necessary information for classification. Because unmarked samples are abundant and manual marking is expensive, SVM training can be extremely time-consuming. Active learning can effectively reduce sample complexity to shorten the training time and improve the accuracy of expression classification.

In this study, based on the feature points in the current expression state, the relative motion of feature points was extracted by comparing the positions of feature points in the neutral expression (no expression) state, and a two-level classifier based on an active learning support vector machine was designed. The first classifier is used to recognize the AU, and the second classifier is based on the feature of the AU to recognize the facial expression. Taking the probability of AU output by the first classifier as the expression feature weight can effectively prevent the second classifier from over-fitting when distinguishing expression categories.

The first classifier is used for the classification of AUs. There are 14 AUs related to facial expressions (see Table 2). According to the regional characteristics of the AU, the eyebrow and eye region, nose region, and mouth region are selected as three separate regions. In the corresponding neutral expression state, the change of feature points relative to the current feature points is used as the feature input, and the SVM based on active learning is divided into three regions, corresponding to the three AU recognition models established. As shown in the first-level classifier in Fig. 2, the eyebrow and eye region model is used to identify AU1, AU2, AU4, AU7, and AU45; the nose region model is used to identify AU9; and the mouth region model is used to identify AU12, AU14–16, AU20, AU23, and AU26.

The second-level classifier distinguishes expression categories according to the combination of AUs obtained by the first-level classifier. An SVM based on active learning is used to model 7 kinds of expressions (6 basic expressions and a neutral expression). The input of each

Table 2 AUs in this study

AU1	AU2	AU4	AU5	AU6
 Inner Brow Raiser	 Outer Brow Raiser	 Brow Lowerer	 Upper Lid Raiser	 Cheek Raiser
AU7	**AU9**	**AU12**	**AU14**	**AU15**
 Lid Tighterner	 Nose Wrinkler	 Lip Corner Puller	 Dimpler	 Lip Corner Depressor
AU16	**AU23**	**AU26**	**AU28**	**AU45**
 Lower Lip Depressor	 Lip Tightener	 Jaw Drop	 Suck Lips	 Blinking

model is composed of 14-dimensional vectors of the output of the first-level classifier (each component is the occurrence probability of the AU related to the expression); the output corresponds to the probability of the expression of the classifier. Finally, the classifier with the highest output probability is used as the expression category of the current face image.

3 AU recognition with active learning and SVM

Based on the feature points extracted from the gradient histogram, this paper proposes a combination of active learning and SVM for the extraction of AUs and facial expression classification, as shown in Fig. 3.

Machine learning can make a classifier produce the least error probability through the learning of a limited number of samples. To ensure the learning outcome, it is necessary to

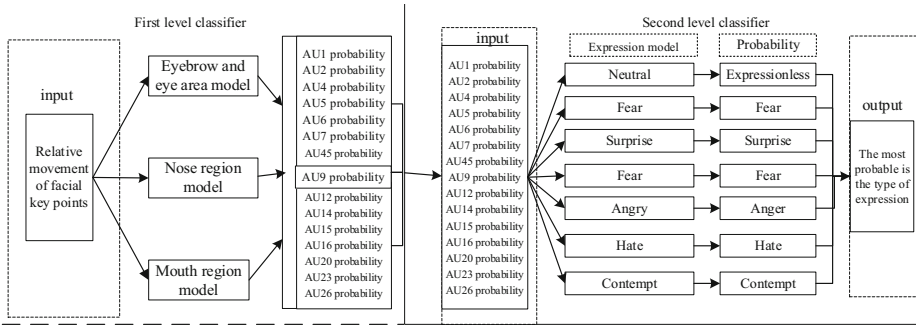


Fig. 2 Expression classification flowchart

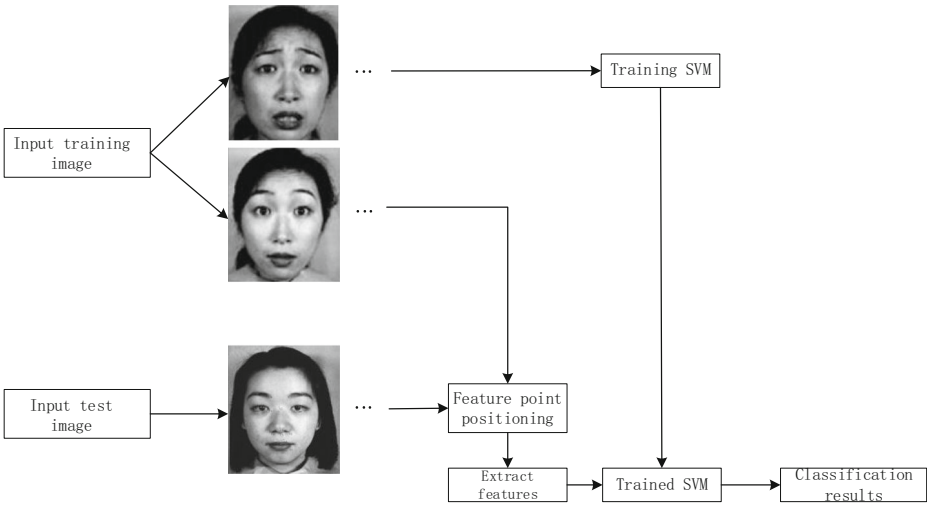


Fig. 3 Graph of facial expression recognition model

increase the number of training samples. However, labeling samples is usually a very expensive procedure. Active learning can help to mitigate the problem by selecting the data instances marked by experts through a strategy to obtain the needed amount of data for the machine learning task.

Combined with an SVM, active learning is effective for AU identification. Active learning uses the existing model to acquire new knowledge by simulating the process of human learning. Based on continuously accumulated information, the existing model can be corrected to become more accurate. An active learning algorithm A consists of five elements:

$$A = (C, L, S, Q, U) \tag{1}$$

where C is a classifier or set of classifiers, L is a group of labeled training samples, S is a supervisor that correctly labels samples in the set of all unlabeled samples (U), and Q is a query function for obtaining information on samples in U.

Initially, the training sample set L is empty. A small part of U is selected by S to be marked as labeled samples and added to L. Thus, an initial classifier model is set up. After that, a certain unlabeled sample in U is selected according to a query criterion Q and added to L. This iteration continues until the stop standard is reached, as illustrated in Fig. 4.

The active learning algorithm is an iterative process. The classifier is trained by adopting iterative feedback samples that continuously improve the accuracy. Taking AU2 (outside eyebrow lift) and AU4 (eyebrow droop) as examples, the training process usually includes four steps (Fig. 5):

- The AU2 and AU4 classifiers are initialized, and an SVM with the RBF kernel is used to generate a predictive result.
- Fragments of the AUs are extracted and labeled.
- Fragments are labeled by the EMFACS training coder.
- The new samples are used to train the AU classifier.

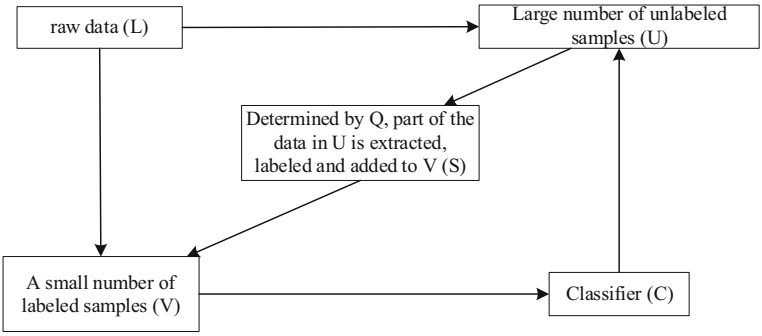


Fig. 4 Active learning flowchart

After the AU2 and AU4 classifiers are trained, they are used to predict each existing facial behavior combined with RBF core SVMs. In our experiment, the smiling expression was initialized with 8392 pictures and the AU2 and AU4 classifiers, in which digital 0–100 was utilized to determine the classifier's level. The sigmoid function was applied to calibrate the output of the classifier.

In the experiment, in the image excluding the invalid information area, the strong classifier formed by the weak classifier was used to match the features of the area including the target object (that is, the face detection area), and these face matching blocks were collected to filter out the noise. Then, the search window was expanded by 10%, iterative matching was conducted, and all results were stored in the element sequence. The effective face candidate region was detected and marked with a blue rectangle to obtain the width and height of the region. In general, the Viola Jones face detection algorithm is the best in terms of speed and accuracy and meets the requirements for real-time video flow detection. In the experiment, the

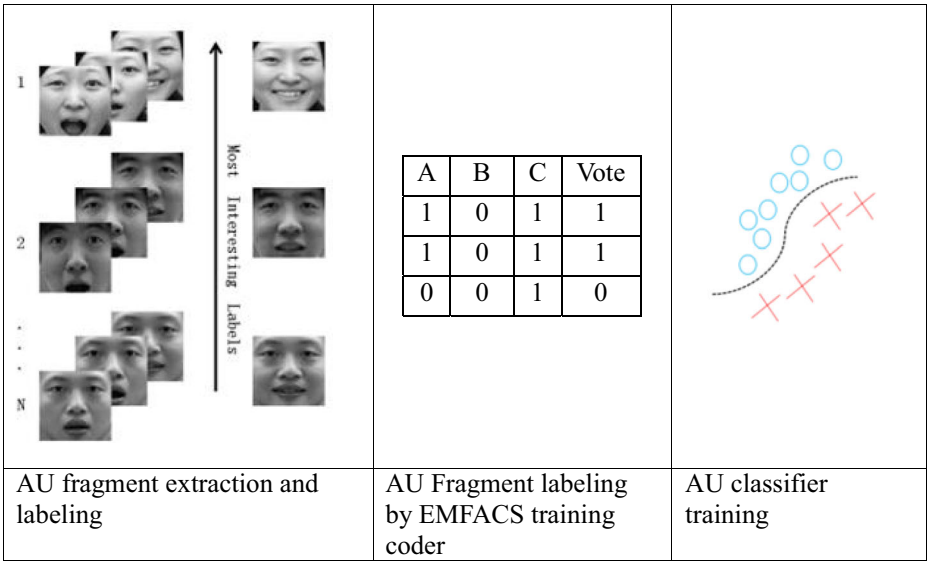


Fig. 5 Positive sample flow generated by active learning for an AU

traditional face detection method based on LBP features was compared with the Viola Jones-based method [24], as shown in Table 3. Therefore, the method based on Viola Jones was used in this study.

The training set used in this study was a subset of CAS-PEAL [26] (including neutral expressions and several common expressions). CAS-PEAL, collated by the Institute of Computing Technology of the Chinese Academy of Sciences in 2003, has a total of 99,450 face images of 1040 volunteers.

Experiments showed that the reduced number of non-support vector samples has little or no effect on the performance of the classifier. In this study, the active learning-based SVM reduced the number of non-support vectors in the training sample set and shortened the labeling time and training time without affecting the performance of the classifier, thereby reducing the cost of labeling samples and improving the training speed.

Active learning was adopted to obtain a large number of **positive** samples for each AU. Without active learning, the amount of data containing AU2 and AU4 can only be reduced by 2% and the amount of smiling data by 20%. On the contrary, with active learning, 30% more images in the dataset can be used by the AU2 and AU4 classifiers. Therefore, more positive sample training sets and negative sample training sets can be obtained by active learning.

The RBF kernel function was used in the SVM as an AU classifier for a given specific training set T:

$$T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{2}$$

$$k_{RBF}(x_i, x_j) \leq \varphi(x_i), \varphi(x_j) \geq exp\left(-\gamma(x_i - x_j)^T (x_i - x_j)\right) \tag{3}$$

where γ is a hyper function, and φ is the approximate mapping function.

Because the training samples are projected onto a three-dimensional space, a scale problem exists at the same time when an RBF kernel is used by the SVM. Thus, the Nystrom method was used to randomly select a subset of training samples. Specifically, Ns samples were randomly selected from the training dataset, and sample X was mapped by

$$(\widetilde{\varphi}(x))_i = exp\left(-\gamma\left(x - x_i\right)^T (x - x_i)/\sqrt{(s)_i}\right) \ (\text{for } i = 1 \ldots Ns) \tag{4}$$

where s is the eigenvalue matrix of the Ns sample kernel. The normalization process was $k_{RBF}(x_i, x_j) \leq \widetilde{\varphi}(x_i, x_j)$, $\widetilde{\varphi} > Ns$ for all samples in the subset of samples (i, j). This mapping was applied to all data in this article, and then a linear SVM was learned in three-dimensional space. In this case, the time-consuming classification mainly depends on the calculation of the feature vector, which is directly proportional to Ns. We can reduce the system response time by reducing Ns or obtaining an approximation closer to the RBF kernel by increasing Ns. By training the classifier for each AU, the AU detection module required by our algorithm is established.

Table 3 Comparison between LBP and Viola Jones-based methods

	Accuracy based on LBP	Accuracy based on Viola Jones
CK+face data set	92.26%	97.69%
LFW face data set	94.67%	98.88%

Table 4 Mapping between emotions and AUs

Emotion	Bonus movement	Subtraction action	Action unit
Joy	Smile	Brow Raise Brow Furrow	AU6+AU12
Sadness	Brow Furrow Lip Suck Eye Widen	Brow Raise Smile Lip Press Mouth Open	AU1+AU4+AU15
Surprise	Inner Brow Raise Jaw Drop Eye Widen	Smile Brow Furrow	AU1+AU2+ AU5**D**+AU26
Fear	Inner Brow Raise Brow Raise Eye Widen Lip Stretch	Brow Furrow	AU1+AU2+ AU4+AU5+ AU7+AU20+ AU26
Anger	Brow Furrow Eye Widen Chin Raise	Inner Brow Raise Brow Raise Smile	AU4+AU5+ AU7+AU23
Hate	Inner Brow Raise Brow Furrow Lip Corner Depressor	Brow Raise Eye Widen Mouth Open Lip Suck Smile	AU9+AU15+ AU16
Disdain	Brow Furrow Smirk	Smile	AU12**U**+AU14**U**

3.1 Emotional facial action coding system (EMFACS)

EMFACS, proposed by Friesen and Ekman in 1983, suggested combining AUs to form facial expressions. Based on observations and experimental data, we present the mapping of AUs to emotions in Table 4. Note that the letters **U**/**D** in the table indicate the directions of the muscle movement, where **U** means upwards and **D** means downwards.

3.2 Experimental results and analysis

In the experiment, the pictures used to analyze the recognition rates were taken from the CK face expression database [15] established by the Robot Research Center and the Department of Psychology at Carnegie Mellon University, USA. A total of 583 pictures were selected for a preliminary study to verify our facial expression recognition algorithm. Figure 6 shows the seven basic facial expressions of a volunteer. From left to right, they are neutral, fear, surprise, sadness, anger, hate, and joy.

Fig. 6 Seven basic facial expressions

Table 5 Recognition rates for female expressions in the CK data set

Emotion	Sample number	Recognized number	Recognition rate
Joy	80	80	100%
Sadness	36	33	91.67%
Surprise	80	76	95%
Fear	36	34	94.44%
Anger	54	50	92.59%
Hate	72	65	90.28%
Neutral	30	27	90%

Tables 5 and 6 list the numbers and rates of male and female facial expressions that were correctly recognized by our algorithm from the CK dataset. The average recognition rates for female and male expressions were 94.07% and 90.77%, respectively. The recognition rate for female facial expressions was higher than that for male facial expressions, indicating that women have richer and more distinguishable expressions than men. Regardless of gender in the samples, the hate and neutral facial expressions seem to be more difficult to recognize than the joy and surprise expressions.

We conducted three experiments using state-of-the-art methods, Keras CNN, Principal Component Analysis (PCA) and ResNet18 Model, and the Fer2013 facial expression dataset. The Fer2013 dataset consists of 35,886 facial expression pictures, including 3589 public-test pictures and 3589 private-test pictures. Each picture is a grayscale image with a fixed size of 48×48. There are seven kinds of expressions labeled by seven numbers: $0-$anger, $1-$ disgust, $2-$fear, $3-$happy, $4-$sad, $5-$surprise, and $6-$normal. Table 7 lists the recognition rates of the three algorithms on the Fer2013 dataset. The classification accuracies of Keras CNN are 47.49% on the public-test set and 65% on the private-test set, which is the lowest among the three methods. PCA has a slightly higher accuracy than ResNet18 Model, and thus it is selected in the following comparison tests.

Table 8 lists the recognition rates of the same facial pictures with the PCA method, a human observer, and our algorithm (active learning + SVM). Of the seven facial expressions, five expressions (joy, sadness, anger, hate, and neutral) were recognized correctly by our algorithm with higher rates than PCA and the human observer. Some expressions (anger, hate, and neutral) seemed to be difficult to recognize by the PCA and observer, but our algorithm achieved over 86% recognition rates on those expressions. On average, the PCA and observer had similar recognition rates on the seven facial expressions, 85% and 84%, respectively, while our algorithm reached a significantly

Table 6 Recognition rates for male expressions in the CK data set

Emotion	Sample number	Recognized number	Recognition rate
Joy	38	38	100%
Sadness	22	20	90.90%
Surprise	54	51	94.44%
Fear	14	12	85.71%
Anger	32	29	87.88%
Hate	29	22	82.76%
Neutral	6	5	83.33%

Table 7 Recognition rates of Keras CNN, PCA and ResNet18 Model

Method	Accuracy on public-test set	Accuracy on private-test set
Keras CNN	47.49%	65.00%
PCA	71.50%	73.11%
ResNet18	71.20%	72.89%

higher rate of 93%, although it did have lower recognition rates on the surprise and fear expressions than PCA, the observer, and both.

4 Conclusion

In the traditional SVM training method, the SVM requires a large number of manually labeled samples to train the classifier, which is not only expensive but also affects the training speed of the classifier. Reducing the number of non-support vector samples has little or no effect on the performance of the classifier. The SVM based on active learning reduces the number of non-support vectors in the training sample set and shortens the manual marking time and training time without affecting the performance of the classifier, thus reducing the cost of labeled samples and improving the training speed. In this study, we utilized an active learning and SVM algorithm to extract relevant AUs from facial images and to reconstruct the relationship between the AUs and emotions for facial expression recognition. The proposed algorithm improves both the robustness and accuracy of recognizing various facial expressions, providing a new concept for facial expression recognition for actual scenes. With our algorithm, female facial expressions could be recognized at higher rates than the corresponding male facial expressions, but different facial expressions, regardless of being female or male, had different recognition rates ranging from 90% to 100% for females and from 83% to 100% for males. Compared to PCA and a human observer, our algorithm achieved a significantly higher average recognition rate on the seven considered facial expressions.

Table 8 Comparison of recognition rates between PCA, human observer, and our algorithm

Emotion	PCA	Observer	Our algorithm	Effectiveness
Joy	98%	98%	100%	↗
Sadness	72%	74%	91%	↗
Surprise	98%	98%	95%	↘
Fear	97%	76%	92%	↘
Anger	86%	73%	95%	↗
Hate	72%	79%	86%	↗
Neutral	69%	88%	90%	↗
Average	85%	84%	93%	

Note: a green arrow represents an increase, and a red arrow represents a decrease

References

1. Abdul-Majjed IO (2017) Emotion recognition system based on facial expressions using svm. Recent Developments in Intelligent Computing, Communication and Devices. pp. 31–35.
2. AU R-CNN (2019) Encoding expert prior knowledge into R-CNN for action unit detection. Ma C., Chen L., Yong J. H. Neurocomputing
3. Benitez-Quiroz CF, Srinivasan R, Martinez AM (2016) EmotioNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild [C]// the IEEE conference on computer vision and pattern recognition (CVPR). IEEE:5562–5570
4. Chen L (2014) A fair comparison should be based on the same protocol–comments on "trainable convolution filters and their application to face recognition"[J]. IEEE Trans Pattern Anal Mach Intell 36(3):622–623
5. Cheng ZY, Shen JL (2016) On effective location-aware music recommendation[J]. ACM Trans Inf Syst 34(2):1–13
6. Ekman P (1978) A technique for the measurement of facial action. Palo alto [J]. Facial Action Coding System
7. Friesen W, Ekman P (1983) EMFACS-7: emotional facial action coding system. Unpublished manual, University of California, California
8. Friesen W, Ekman P (1983) EMFACS-7: emotional facial action coding system. Unpublished manual, University of California, California
9. Gudi A, Tasli HE, Den Uyl TM et al (2015) Deep learning based FACS unit occurrence and intensity estimation[C]//2015 11th IEEE international conference and workshops on automatic face and gesture recognition(FG). IEEE 6:1–5
10. Huang X, Wang SJ, Zhao G et al. (2015) Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection[C]// second workshop on computer vision for affective computing at Iccv. IEEE Computer Society
11. Huang X, Zhao G, Hong X et al. (2016) Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns[J]. Neurocomputing, 2016, 175 (JAN.29PT.A):564–578.
12. Liu Y, Zhang J, Yan W et al (2016) A Main directional mean optical flow feature for spontaneous micro-expression recognition [J]. IEEE Trans Affect Comput 7(4):299–310
13. Liu L, Li G, Xie Y, Yu Y, Wang Q, Lin L (2019) Facial landmark machines: a backbone-branches architecture with progressive representation learning [J]. IEEE Transactions on Multimedia 21(9):2248–2262
14. Lu WL Facial expression recognition based on emotional geometry and support vector machines [D]. Fudan University
15. Lucey P, Cohn JF, Kanade T et al (2010) The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. Computer Vision and Pattern Recognition Workshops IEEE 36(1):94–101
16. Merghani W, Davison AK, Yap MHA (n.d.) Review on Facial Micro-Expressions Analysis: Datasets, Features and Metrics. PREPRINT SUBMITTED TO IEEE JOURNAL
17. Polikovsky S, Kameda Y, Ohta Y et al (2013) Facial micro-expression detection in hi-speed video based on facial action coding system (FACS)[J]. IEICE Trans Inf Syst 96(1):81–92
18. Polikovsky S, Kameda S, Ohta Y (2013) Facial micro-expression detection in hi-speed video based on facial action coding system (FACS) [J]. IEICE Transactions on Information & Systems 96(1):81–92
19. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput 27:803–816
20. Shreve M, Godavarthy S, Manohar V et al. (2009) Towards macro-and micro-expression spotting in video using strain patterns, in: Applications of Computer Vision (WACV), pp. 1–6.

21. Wang SJ, Yan WJ, Li X et al. (2014) Micro-expression recognition using dynamic textures on tensor independent color space[C]// international conference on pattern recognition. IEEE
22. Wang SJ, Chen HL, Yan WJ, Chen YH, Fu X (2014) Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. Neural Process Lett 39(1): 25–43
23. Wang Y, See J, Phan W et al (2015) LBP with six intersection points: reducing redundant information in LBP-TOP for micro-expression recognition[C]// Accv. Springer, Cham
24. Xian W, Yan Z, Xin M et al (2012) Face recognition algorithm based on improved LBP [J]. Opt Eng 39(7): 109–114
25. Xu F, Zhang J, Wang JZ (2017) Microexpression identification and categorization using a facial dynamics map[J]. IEEE Trans Affect Comput 8(2):254–267
26. Zhang X, Shan S, Cao B et al (2005) Cas-peal: a large-scale Chinese face database and some primary evaluations. Journal of Computer Aided Design & Computer Graphics 17(1):9–17
27. Zhao K, Chu WS, De la Torre F. et al (2015) Joint patch and multi-label learning for facial action unit detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2207–2216
28. Zhao K, Chu WS, Zhang H (2016) Deep Region and Multi-label Learning for Facial Action Unit Detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3391–3399.

Li Yao received his Ph.D. degree in 2002 from the Zhejiang University and then joined the School of Computer and Science of the Donghua University. His research interests include machine learning, artificial intelligence, and 3D imaging and modeling.

Yan Wan received her Ph.D. degree in 2001 from the Shanghai Jiaotong University and joined the faculty of the Donghua University in 2002. Since 2014, she has been a professor in School of Computer and Science, University of North Texas. Her research interests include machine learning, artificial intelligence, and image processing and pattern recognition.

Hongjie Ni received his Master degree in 2019 from the Donghua University. His research interests include machine learning, and image processing.

Bugao Xu received his Ph.D. degree in 1992 from the University of Maryland at College Park and joined the faculty of the University of Texas at Austin in 1993. Since 2016, he has been a professor in Department of Merchandising and Digital Retailing and Department of Computer Science and Engineering, University of North Texas. His research interests include high-speed imaging systems, image and video processing, and 3D imaging and modeling.