Research on Facial Expression Recognition Based on Active Learning

Yan Wan¹, Hongjie Ni¹*, Xiaodong Peng¹, Yining Guo¹, Li Yao¹ and Xiangyang Feng¹
¹School of Computer Science and Technology, Donghua University, No. 2999 Renmin North Road, Songjiang District, Shanghai, China
617475jordan@sina.com

Abstract: In order to solve the problems of low recognition accuracy and poor robustness in traditional facial expression recognition research methods, this paper presents a study of methods based on active learning. The method is based on active learning and combines support vector machine algorithms to construct a learning network similar to the human visual system. Active learning learns the corresponding facial expression action unit through training, and utilizes support vector machines to classify different action units and ultimately maps to corresponding facial expressions, thereby realizing recognition of facial expressions. Experimental results show that this method can effectively suppress correlated noise and obtain relative information. It not only has good robustness, but also improves the recognition rate of facial expressions and can meet actual needs.

1. Introduction
In the field of facial expression recognition, the Ekman [1] is the main force in foreign teams. Since most of the team's research results are in the state of confidentiality, the results of the team can only be used for theoretical research. The method of American Shreve’s team of recognizing facial expressions mainly adopts mathematics to detect changes in the optical flow field [2] generated by the non-rigid muscle movement of the face when the face is producing an expression. Polikovsky [3] utilizes coding to divide the face into twelve parts. Each part has its own different timing characteristics. In the continuous expression frame, the changes between these characteristics are utilized to find out the area where the expression occurs and thus the different expressions are identified. Unlike Polikovsky, Zhao Guoying at the University of Oulu in Finland focuses on the overall face features, namely the extraction of texture features. In China, Fu Xiaolan [4], with the help of the National Natural Science Foundation project “Study on the expression of micro-expressions for automatic lies recognition”, puts to use machine vision technology and cognitive psychology theory system to study the factors that influence expression recognition and recognize the basic characteristics of expression.

In some cases, unmarked data is abundant, but manual marking is expensive. In this scenario, the learning algorithm can actively query the user/teacher tag. This iterative supervision learning is called active learning. Active learning is a special semi-automatic machine learning algorithm in the learning algorithm. It can interactively query the user (or other information source) to obtain the output required by the new data point. In the statistical literature, it is sometimes called optimization experiment design.

This paper proposes an algorithm based on active learning, which reconstructs the facial action unit
2. Related Work

2.1 Main Framework
Based on the facial expression system needs to extract the relevant action unit, and the action unit maps to the corresponding expression. This paper proposes a facial expression analysis algorithm based on active learning and using support vector machine assisted. The flow chart of this algorithm is shown below figure 1.

![Figure 1. The flowchart of Facial expression analysis algorithm](image)

2.2 Measure Method
In traditional image classification, manual labeling of related data is time-consuming and expensive. In recent years, some scholars have adopted the learning methods in machine learning, such as supervised learning, semi-supervised learning, and active learning, and carries out relevant sample training to form the necessary models to classify images and reduce related costs. This article mainly utilizes active learning [5]. It actively submits some facial action unit annotation requests, delivers a part of the filtered data to the experts for annotation, constantly corrects the required model, and reduces the consumption of data annotation.

2.2.1 Basic Principle of Active Learning. Active learning [5], through the process of simulating human learning, adopts existing training models to obtain new knowledge, and accumulates information and corrects existing models in order to gain more accurate models. Different from passive learning, passively acquiring relevant information, active learning acquires knowledge through selective methods. The active learning algorithm model consists of the following five parts:

\[
A = (C, L, S, Q, U)
\]

Where C is a classifier or set of classifiers; L is a set of labeled training sample sets; S is a supervisor, and the unlabeled sample in U can be correctly labeled; Q is a query function for unsigned samples Pool U queries the large sample of information; U is the entire unlabeled sample set.

The active learning algorithm is mainly divided into two stages:

1. The first stage is the initialization stage. We randomly select a small part from the unlabeled sample and the supervisor marks it to establish the initial classifier model as the training set.
2. The second stage is the cyclic query stage. In the unlabeled sample set U, according to a certain query criterion Q, a certain unlabeled sample is selected for labeling, added to the training sample set U, and the classifier is retrained until training is achieved. Stop the standard so far.

The relevant steps are shown in the following figure 2:
The active learning algorithm is an iterative process. The classifier is trained by adopting iterative feedback samples that continuously improves the efficiency.

2.3 Support Vector Machine (SVM)

2.3.1 Support Vector Machine Principle. Support vector machine (SVM) [5] has many advantages in solving small samples, nonlinear and high-dimensional modes, and can be applied to other problems such as function fitting.

Tong and Chang proposed using the SVM classifier to provide relevant feedback results for video retrieval. Similarly, our method is based on the SVM classifier output used for the tagging of FACS [6] sequences. The SVM algorithm process is described in detail from the following aspects.

Given a specific training set:

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

Solving quadratic programming problems:

$$\min_{\alpha} \frac{1}{2} \sum_i \sum_i \alpha_i \alpha_j y_i (x_i \cdot x_j) - \sum_i \alpha_i$$

$$s.t. \sum_i \alpha_i y_i = 0, \alpha_i \geq 0$$

$$\alpha^* = (\alpha_1^*, \ldots, \alpha_n^*)^T$$

Calculation parameters w, select one of the positive components \(\alpha_i^*\), calculation b

$$w^* = \sum_i \alpha_i^* y_i x_i$$

$$b^* = y_j - \sum_i \alpha_i^* y_i (x_j \cdot x_i)$$

Structure decision boundary \(g(x) = (w^* \cdot x) + b^*\), solution function:

$$f(x) = \text{sgn}(g(x))$$

The important difference from Tong and Chang is that this article arranges image segments based on a combination of facial actions, as facial action units can appear in different combinations, which adds diversity to positive and negative training sets.

In order to obtain a large number of positive samples for each motion unit, this paper adopts active learning to reconstruct the model. The following figure 3 describes the process of active learning of AU02 (outside eyebrow lift) and AU04 (eyebrow droop) in detail. The AU02 and AU04 classifiers are initialized, and the RBF kernel SVMs are used to generate a predictive result for each existing
behavior.

Figure 3. Active learning strategy generates a positive sample flow for an action unit
(a) Fragmentation of AU fragments and marking is preferred.
(b) The FACS training coder marks the fragments.
(c) Use the new sample to train the AU classifier.

The predicted results of each existing facial behavior are generated with the initialized AU02 and AU04 classifiers and RBF core SVMs. We initialize the smile with 8392 pictures, AU02 and AU04, and use 0-100 to determine the classifier's level. The sigmoid function is used to calibrate the output of the classifier.

The training set used in this article is CAS-PEAL [7]. CAS-PEAL was completed by the Institute of Computing Technology of the Chinese Academy of Sciences in 2003, with a total of 99,450 face images, including 1040 volunteers. This article uses the CAS-PEAL face database expression subset (including neutral expressions and several common expressions)

Figure 4. CAS-PEAL [7] face library expression subset

Without active learning, the amount of data containing AU02 and AU04 will be reduced by 2% and the amount of data for smiling will be reduced by 20%. On the contrary, with active learning, 30% of the data volume will be used by the AU02 and AU04 classifiers. In order to obtain more positive training sets, at the same time, more negative expressions are obtained.

About RBF kernel functions or feature vectors $x_i$ and $x_j$, $k_{RBF}(x_i, x_j) \leq \phi(x_i), \phi(x_j) \geq \exp(-\gamma(x_i - x_j)^T(x_i - x_j))$ where $\gamma$ is a hyper function, we find the approximate value of the mapping function.

Since the samples used for training will be projected onto (RN) three-dimensional space, at the same time, the RBF kernel SVM method is used, causing scale problems. For this purpose, a method of selecting a subset of training samples at random is adopted. This process is called the Nystrom method. Specifically, consider randomly selecting $N_s$ samples from the training data set $(x_i)_{i=1}^{N_s}$, then the mapping of any sample $X$ is $\tilde{\phi}$:

$$\left(\tilde{\phi}(X)\right)_i = \exp\left(-\gamma(x - x_i)^T(x - x_i) / \sqrt{s}_{ii}\right) \quad \text{(For } i=1\ldots N_s\text{)} \quad (9)$$

Where $s$ is the eigenvalue matrix of the $N_s$ sample kernel. The normalization process is as follows,
This mapping applies to all data in this article, then linear SVM is learned in RNs three-dimensional space. In this case, the time-consuming classification mainly depends on the calculation of the feature vector $\phi$, which is directly proportional to $N$. We accelerate the system corresponding by reducing the $N$s or increasing $N$s to obtain an approximation closer to the RBF kernel. By training each action unit, the action unit detection module required by this article is formed.

To verify the effectiveness of the proposed algorithm, the obtained action units are mapped to the corresponding emotions. We improve the EMFACS facial expression recognition method proposed by Friesen and Ekman, and reconstruct the mapping of facial expressions to emotions.

### 2.4 Emotional Facial Action Coding System (EMFACS)

Emotional Facial Action Coding System (EMFACS) [8] was proposed by Friesen, W. Ekman in 1983. The main advantage is that anyone can use FACS to select the appropriate coding standard to write their own EMFACS. However, the disadvantages are also obvious. With the loss of FACS prediction data, the robustness of EMFACS becomes worse. The table 1 is shown as follows.

**Table 1.** Relationship between emotion and action units in EMFACS [8]

| Emotion | Action units |
|---------|--------------|
| Joy     | AU01+AU04+AU015 |
| Sadness | AU01+AU02+AU05+AU26 |
| Surprise| AU01+AU02+AU04+AU20+AU25, AU26 or AU27 |
| Fear    | AU04+AU05+AU07+AU17+AU23 |
| Anger   | AU09+AU17 |
| Hate    | RAU12/LAU12 or RAU14/LAU14 |

Note: The letters R/L indicate the side of the face where muscle movements are occurring. R is the right side and L is the left side. This article presents the relationship between emotion and movement units, as shown in the table below.

The relationship between emotion and unit of motion is presented in the following table 2.

**Table 2.** Relationship between emotion and action units in this article

| Emotion | Bonus movement | Subtraction action | Action unit |
|---------|----------------|--------------------|-------------|
| Joy     | Smile          | Brow Raise         | AU06+AU12   |
|         | Brow Furrow    |                    |             |
|         | Lip Suck       | Brow Raise         | AU01+AU04+  |
|         | Eye Widen      | Smile              | AU01+AU04+  |
|         |                | Lip Press           | AU15        |
|         |                | Mouth Open          |             |
| Surprise| Inner Brow Raise| Smile              | AU01+AU02+  |
|         | Jaw Drop       | Brow Furrow         | AU05B+AU26  |
|         | Eye Widen      |                    |             |
| Fear    | Inner Brow Raise| Brow Furrow        | AU01+AU02+  |
|         | Brow Raise     |                    | AU04+AU05+  |
|         | Eye Widen      | Brow Furrow         | AU07+AU20+  |
|         | Lip Stretch    |                    | AU26        |
| Anger   | Brow Furrow    | Inner Brow Raise    | AU04+AU05+  |
|         |                |                    |             |
EYE WIDEN  
CHIN RAISE 

BROW RAISE
SMILE

AU07+AU23

HATE
INNER BROW RAISE
BROW FURROW
LIP CORNER DEPRESSOR

AU09+AU15+
AU16

DISDAIN
BROW FURROW
SMIRK

RAU12A+RAU14A

Note: The letters A/B indicate the direction of the muscle movement, with A pointing upwards and B pointing downwards.

3. Experimental results and analysis

In the experiment, the pictures were tested using a face expression database, CK[9], established by the Robot Research Center and the Department of Psychology at Carnegie Mellon University in the United States. Each volunteer's expression sequence was collected. We select 583 pictures for experimentation. The following figure 4 shows the six basic facial expressions and neutral facial expressions of a volunteer. From left to right, the order is neutral, fear, surprise, sadness, anger, hate, joy. The table 3, table 4 and figure 5 show the results of male and female facial expression recognition rate, and the overall results of the facial expression recognition rate.

Figure 5. neutral expression and 6 basic expressions

Table 3. CK [9] female expression data set test

| Number | Emotion | Sample number | Correct number | Recognition rate | Average recognition rate |
|--------|---------|---------------|----------------|------------------|--------------------------|
| 1      | Joy     | 80            | 80             | 100%             |                          |
| 2      | Sadness | 36            | 33             | 91.67%           |                          |
| 3      | Surprise| 80            | 76             | 95%              |                          |
| 4      | Fear    | 36            | 34             | 94.44%           | 94.07%                   |
| 5      | Anger   | 54            | 50             | 92.59%           |                          |
| 6      | Hate    | 72            | 65             | 90.28%           |                          |
| 7      | Neutral | 30            | 27             | 90%              |                          |

Table 4. CK [9] male expression data set test

| Number | Emotion | Sample number | Correct number | Recognition rate | Average recognition rate |
|--------|---------|---------------|----------------|------------------|--------------------------|
| 1      | Joy     | 38            | 38             | 100%             |                          |
| 2      | Sadness | 22            | 20             | 90.90%           |                          |
| 3      | Surprise| 54            | 51             | 94.44%           |                          |
| 4      | Fear    | 14            | 12             | 85.71%           | 90.77%                   |
| 5      | Anger   | 32            | 29             | 87.88%           |                          |
| 6      | Hate    | 29            | 22             | 82.76%           |                          |
| 7      | Neutral | 6             | 5              | 83.33%           |                          |
Figure 6. Female, male facial expression recognition rate and overall facial expression recognition rate

From the above figure 6, it can be seen that women's facial expression recognition rate is better than male's facial expression recognition rate, and the accuracy rate is much higher than that of men, indicating that women have richer expression changes than men.

Compared with the document PCA [10], this paper achieves better results. The table 5 lists the recognition rates of individuals using the PCA [10] method. The third column of the table shows the correct rate from human self-recognition expressions. From the data, it can be seen that the study of expression recognition methods is difficult for computer recognition, and it is also very difficult for humans to distinguish between them.

Table 5. Comparison of our algorithm and PCA [10] and human recognition rate

| Emotion   | PCA recognition rate | Human recognition rate | Our recognition rate | Effectiveness |
|-----------|----------------------|------------------------|----------------------|--------------|
| Joy       | 98%                  | 98%                    | 100%                 |              |
| Sadness   | 72%                  | 74%                    | 91.38%               |              |
| Surprise  | 98%                  | 98%                    | 94.78%               |              |
| Fear      | 97%                  | 76%                    | 92.00%               |              |
| Anger     | 86%                  | 73%                    | 95.18%               |              |
| Hate      | 72%                  | 79%                    | 86.14%               |              |
| Neutral   | 69%                  | 88%                    | 89.92%               |              |

Note: red arrow represents increase, green arrow represents decrease.
Figure 7. Comparison of our algorithm, PCA [10] and human recognition rate

As can be seen from the above figure 7, the accuracy of this algorithm has been greatly improved in some aspects, but in some facial expressions, the recognition rate is still at a low level, and the false recognition rate is high, indicating that the facial expression recognition still has great development potential.

4. Conclusion
This paper is based on the active learning algorithm for the study of facial expression recognition method, extracting relevant action units, and reconstructing the relationship between the action unit and emotion. The experiment shows that the extracted related action units are better mapped to various expressions, which effectively improves the recognition accuracy of each expression. However, the algorithm still lacks certain recognition accuracy for facial expressions, such as facial expression fear. In the next step, the robustness of our algorithm will be further improved, and the actual video expression sequence will be combined to enhance the practical effect of the research on expression recognition methods.

References
[1] P Ekman 1978. A technique for the measurement of facial action. Palo alto [J]. Facial Action Coding System.
[2] Gautama T, Van Hulle MA 2002, 13(5). A phase-based approach to the estimation of the optical flow field using spatial filtering. Neural Networks IEEE Transactions on, 1127-36.
[3] S Polikovsky, Y Kameda, Y Ohta 2013, 96(1). Facial micro-expression detection in hi-speed video based on facial action coding system (FACS) [J]. Ieice Transactions on Information & Systems, 81-92.
[4] Wang S J, Chen H L, Yan W J, et al. (2014). Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. Neural Processing Letters, 39(1), 25-43.
[5] Abdul-Majjed I O 2017. Emotion recognition system based on facial expressions using svm.
[6] P Ekman, WV Friesen 1978, 47(2). Facial action coding system (FACS): a technique for the measurement of facial actions. Rivista Di Psichiatria, 126-38.
[7] X Zhang, S Shan, B Cao, W Gao 2005, 17(1). Cas-peal: a large-scale chinese face database and some primary evaluations. Journal of Computer Aided Design & Computer Graphics, 9-17.
[8] Friesen, W.; Ekman, P. (1983). EMFACS-7: Emotional Facial Action Coding System. Unpublished manual, University of California, California.

[9] P Lucey, JF Cohn, T Kanade, J Saragih 2010, 36(1). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. Computer Vision and Pattern Recognition Workshops IEEE, 94-101.

[10] Chen J, Jenkins W K 2017. Facial recognition with PCA and machine learning methods. IEEE, International Midwest Symposium on Circuits and Systems, 973-976.