A Method of Describing Uncertain Emotions of Facial Expressions

Junhuan Lin\(^1\)\(^a\), Yuefen Chen\(^2\)\(^b\)

\(^1\)Faculty of Mechanical and Electrical Engineering, Taizhou Vocational and Technical College, Xueyuan Road 788#, Jiaojiang, Taizhou, Zhejiang, China, 318000

\(^2\)School of Electronics & Information Engineering, Taizhou University, Shifu Road 1139#, Jiaojiang, Taizhou, Zhejiang, China, 318000

\(^a\) linjunh@qq.com, +86-13456663933

\(^b\) chen-yuefen@163.com, +86-15988986294

ABSTRACT In order to describe uncertain and fuzzy emotions in recognizing facial expressions, a PAD regressive model based on deep convolutional neural network model is built to quantize the emotions, by which the facial expressions can be mapped to the emotional space of PAD. Then the emotional membership function is proposed to describe the uncertainty and fuzziness of the emotions in the space of PAD. Finally the testing results demonstrate that the regressive model performs well in recognizing facial expressions and the results from describing the emotions quantitatively agree with the human cognition to facial expressions.

1. INTRODUCTION

There are many advanced approaches to recognize facial expression. But mostly, the facial expressions are classified into some finite and discrete classes. In fact, human emotions are continuous and complex. Moreover, our emotional cognition is characterized by uncertainty and fuzziness. It is a challenge for researchers to describe emotions quantitatively in formalized presentation.

PAD emotional model is proposed by Mehrabian \(^1\), which can quantitatively describe emotion in three dimensions of Pleasure, Arousal and Dominance. Three dimensions respectively represent the positive and negative emotional state, the activation degree of individual neural state and controlling degree of individual to the object, which can realize emotional quantifiable calculation. Based on PAD emotion theory, Cao Jie et al. \(^2\) proposed EBM model, combined with Gabor features and SVM algorithm, to identify atypical emotion, and achieved good effect. Yang Ning \(^3\) proposed an improved PAD emotion model based on the basic emotional theory, and used the combination of active shape model and gray level co-occurrence matrix to extract features, finally realizing the fine classification of facial expressions. Sun Ying et al. \(^4\) established the relationship model between the features of speech signals and PAD model by using gray-scale correlation analysis, realized the prediction of PAD value and effectively recognized the voice emotion, solved the problem of recognizing continuous emotional state. The above researches use PAD emotion model to realize the quantitative description of emotion, but without involving the description of uncertainty and fuzziness of emotion. In this paper, regressive model of PAD is built based on deep convolutional neural network, which can map the facial expressions to the emotional space of PAD. Then the fuzzy membership function is used to quantitatively describe facial expression with uncertainty and fuzziness.
The rest of the paper is organized as follows. Section 2 describes deep convolutional neural network. Section 3 describes the PAD emotional regressive model. Section 4 describes the emotional membership based on PAD. Section 5 describes the procedure of experiments and discusses the results. Finally, some conclusions are made in section 6.

2. DEEP CONVOLUTIONAL NEURAL NETWORK

The deep convolution neural network used in this paper is based on the model of BVLC. Through two transfer learning processes shown in Figure 1, the model of BVLC is suitable for facial expression recognition. BVLC model has been trained on the large dataset of Imagenet, and has a good performance [5]. The first transfer learning is to transfer the lower convolutional layers of BVLC model to a new targeting DCNN. In the second transfer learning, at first the new targeting DCNN model is trained on the face attributes dataset named CelebA with 200000 samples, and then the medium convolutional layers are transferred to the final targeting DCNN. After the two stages of transfer learning, the final targeting DCNN is fine tuned on the facial expression dataset of CK+ and is used to recognize eight classes of facial expressions, i.e. neutrality, happiness, anger, disgust, surprise, contempt, fear and sadness. During each transfer learning, the feature visualization method proposed by Zeiler et al. [6] is used to check the feature extracting efficacy of the convolution layers and determine whether they can be transferred.

3. THE EMOTIONAL REGRESSIVE MODEL OF PAD

The emotional regressive model of PAD is inherited from DCNN proposed above, which is composed of the five convolutional layers and the Fc6 and Fc7 full connection layers that are both transferred from the proposed DCNN and the new Fc8 and Fc9 full connection layers, shown in Figure 2. The emotional regressive model is further trained by the CK+ dataset in which every sample is annotated quantitatively in space of PAD to establish the mapping relationship between facial expressions and PAD values.

![Figure 1. The pipeline of the dual transfer learning](image)

![Figure 2. The structure of the emotional regressive model of PAD](image)
4. THE EMOTIONAL MEMBERSHIP OF PAD

The emotional model of PAD is composed of three dimensions, namely, pleasure, arousal and dominance. The range of each dimension is [-1, 1]. The PAD values of seven basic emotions are shown in Table 1.

| Basic emotions | Pleasure | Arousal | Dominance |
|----------------|----------|---------|-----------|
| Happiness      | 0.63     | 0.40    | 0.29      |
| Anger          | -0.59    | 0.08    | 0.47      |
| Disgust        | -0.59    | -0.01   | 0.4       |
| Fear           | -0.08    | 0.18    | -0.39     |
| Surprise       | 0.41     | 0.55    | 0.19      |
| Contempt       | -0.4     | -0.2    | 0.45      |
| Sadness        | -0.28    | -0.12   | -0.37     |

In real life, human emotions are characterized by fuzziness and uncertainty, which can't be simply classified into distinct classes. The membership function in fuzzy mathematics can describe the fuzzy and uncertain objects very well. Therefore, we adopt the concept of fuzzy membership degree as the definition of emotional membership to describe the fuzziness and uncertainty of emotions.

Definition 1: Letting $e$ denote an emotion in universe of $E$ and existing an $A(e)$ corresponding to $e$, then $A$ is called an emotional set. $A(e)$ is the membership which represents the degree of $e$ belonging to $A$.

Lemma 1: According to Definition 1, in the emotional space of PAD, any emotion $e$ is expressed as $e = [p \ a \ d]$, and there are emotional sets: $A_H = \{\text{Happiness}\}$, $A_A = \{\text{Anger}\}$, $A_D = \{\text{Disgust}\}$, $A_F = \{\text{Fear}\}$, $A_S = \{\text{Surprise}\}$, $A_C = \{\text{Contempt}\}$, $A_Sa = \{\text{Sadness}\}$, then $A_H(e) = \exp(-\frac{(e - e_H)^2}{2\sigma_H^2})$, $A_A(e) = \exp(-\frac{(e - e_A)^2}{2\sigma_A^2})$, $A_F(e) = \exp(-\frac{(e - e_F)^2}{2\sigma_F^2})$, $A_S(e) = \exp(-\frac{(e - e_S)^2}{2\sigma_S^2})$, $A_C(e) = \exp(-\frac{(e - e_C)^2}{2\sigma_C^2})$, $A_Sa(e) = \exp(-\frac{(e - e_Sa)^2}{2\sigma_Sa^2})$, respectively indicates the membership degree of every emotion $e$ belonging to emotional sets of happiness, anger, disgust, fear, surprise, contempt and sadness, where $\sigma_H$, $\sigma_A$, $\sigma_D$, $\sigma_F$, $\sigma_S$, $\sigma_C$, $\sigma_Sa$ respectively represents the clustering radius of every emotion in emotional space of PAD, $e_H = [0.63 0.40 0.29]$, $e_A = [-0.59 0.08 0.47]$, $e_D = [-0.59 -0.01 0.4]$, $e_F = [-0.08 0.18 -0.39]$, $e_S = [-0.41 0.55 0.19]$, $e_C = [-0.4 -0.2 0.45]$, $e_Sa = [-0.28 -0.12 -0.37]$.

5. EXPERIMENTS

5.1 Annotation of the Dataset in Emotional Space of PAD

CK+ contains eight basic emotions having 39731 labeled samples, all of which are facial expression images selected from video sequences. Generally, the subjects expressed a kind of emotion from the beginning of neutral expression to the complete emotional expression that were recorded by video camera. In this paper, the samples are labeled one by one manually. For example, given an emotion labeled with surprise, checking Table 1, we can see that its PAD is $[0.41, 0.55, 0.19]$. There are 14 images in the sequence, and the first to the fifth are neutral, the PAD is $[0, 0, 0]$ and the seventh to the fourteenth are almost equal, the PAD is $[0.41, 0.55, 0.19]$ and the sixth is obviously different from the last one, so the PAD is $[0.4, 0.54, 0.18]$. 


5.2 Train of Emotional Regressive Model of PAD
Caffe framework is used to build the structure model of Figure 2 and CK+ dataset with PAD values is used for training the model. During the training process, the parameters of all convolutional layers keep invariable, and the full connection layers of Fc6 to Fc10 are trained. The objective function adopts Euclidean distance function and norm, as shown in formula (1).

\[
J = \frac{1}{2N} \sum_{n} \left\| Y_n - \hat{Y}_n \right\|_2^2 + \frac{1}{2} \|w\|_2^2
\]  

(1)

Where N is the number of dataset in each batch, \(Y_n\) is the labeled value of the nth sample in each batch, and \(\hat{Y}_n\) is the prediction value of the nth sample in each batch, \(w\) are parameters of model.

In Caffe environment, the related hyper parameters are set as shown in Table 2.

| Name          | base_lr | Min_batch | lr_policy |
|---------------|---------|-----------|-----------|
| value         | 0.0001  | 100       | step      |

Table 2. Hyper parameters of model

After the training, the model is tested on testing dataset, and the results are shown in Figure 3. This sample is selected from CK+. The emotion is labeled as ‘surprise’. The ‘True’ column of each image corresponds to the value of PAD, and the ‘Pre’ column corresponds to prediction value by the regressive model. It can be observed from the results that the gap between the emotion prediction value and the labeled value are small. The Euclidean distance function of formula (2) is used to measure the distance between values of PAD for samples and values of PAD for basic emotion shown in Table 1, and the results are shown in Table 3. According to the principle of Euclidean distance minimum, those samples can also be determined as ‘surprise’.

\[
L_n = \left\| Y_n - \hat{Y}_n \right\|_2
\]  

(2)

Where \(Y_n = [P_n, A_n, D_n]\) is the prediction value PAD for the sample n, \(Y_e = [P_e, A_e, D_e]\) is the value of PAD for basic emotion.

![Figure 3](image)

Figure 3. The prediction values of PAD for the samples labeled with disgust

| Emotion | 6th sample | 7th sample |
|---------|------------|------------|
| Happiness | 0.406     | 0.282     |
| Anger    | 0.987     | 1.136     |
| Disgust  | 0.991     | 1.159     |
| Fear     | 0.623     | 0.822     |

Table 3. Euclidean distance between prediction values and real values of PAD for the samples labeled with surprise
Figure 4 shows the results from testing of the regressive model on the samples without emotional labels in the CK+ dataset. Also, Euclidean distance is used to measure the distance between values of PAD for each sample and values of PAD for basic emotions, which is shown in Table 4. According to the principle of Euclidean distance minimum, those samples can also be determined as ‘disgust’, but it can’t be directly verified since those samples are unlabeled in CK+. However, it can be verified by comparing with the results from testing on the samples labeled with ‘disgust’. Figure 5 shows testing results of three samples labeled with ‘disgust’. The Euclidean distance measurement results are shown in Table 5. According to the Euclidean distance principle, the sample can also be determined as ‘disgust’. Comparing the two batches of samples in Figure 4 and Figure 5, we can find that the face has the same contour features on the nose and mouth parts, which present the disgusting features. It also demonstrates that emotions can be effectively recognized using the proposed regressive model.

| Emotion | fist sample | second sample | third sample |
|---------|-------------|---------------|--------------|
| Happiness | 1.275 | 1.321 | 1.333 |
| Anger | 0.098 | 0.128 | 0.137 |
| **Disgust** | **0.053** | **0.053** | **0.059** |
| Fear | 0.934 | 0.953 | 0.961 |
| Surprise | 1.138 | 1.185 | 1.198 |
| Contempt | 0.314 | 0.321 | 0.322 |
| Sadness | 0.827 | 0.828 | 0.831 |
| **Prediction** | **Disgust** | **Disgust** | **Disgust** |

Figure 4. The prediction values of PAD for the unlabeled samples

Figure 5. The prediction values of PAD for the samples labeled with disgust
Table 5. Euclidean distance between prediction values and real values of PAD for samples labeled with disgust

| Emotion   | fist sample | second sample | third sample |
|-----------|-------------|---------------|--------------|
| Happiness | 1.319       | 1.328         | 1.337        |
| Anger     | 0.124       | 0.126         | 0.136        |
| Disgust   | **0.031**   | **0.041**     | **0.053**    |
| Fear      | 0.967       | 0.973         | 0.970        |
| Surprise  | 1.189       | 1.198         | 1.205        |
| Contempt  | 0.296       | 0.304         | 0.314        |
| Sadness   | 0.839       | 0.843         | 0.839        |
| Prediction | **Disgust** | **Disgust**   | **Disgust**  |

5.3 Description of facial expression quantitatively

The trained regressive model can map any facial expression to a point in emotional space of PAD, which can be described by the emotional membership function proposed in this paper. Taking the first sample of Figure 4 as an example, the point in the emotional space of PAD is $e = [0.62, -0.01, 0.39]$. According to the basic emotions in Table 1, the clustering radius of each basic emotion can be determined as $\sigma_h = 0.1422, \sigma_a = 0.0570, \sigma_d = 0.0570, \sigma_f = 0.1805, \sigma_s = 0.1422, \sigma_c = 0.1366, \sigma_{sa} = 0.1805$. Then, according to lemma 1, it can be computed with $A_h(e) = 0.0025, A_a(e) = 0.8985, A_d(e) = 0.9878, A_f(e) = 0.0763, A_s(e) = 0.0075, A_c(e) = 0.7035$. According to the results, the degree of facial expression of the first image in Figure 4 belonging to ‘disgust’ is close to 1, and the degree of it belonging to ‘anger’ is up to 0.8985. We invite 10 subjects to fill a questionnaire by judging emotion for the first image in Figure 4. The results show 100% of the subjects judge this image as ‘disgust’, and 80% is ‘angry’, 30% is ‘contempt’, 10% is ‘surprise’. According to the results, we can confirm the proposed method of the quantitative description is in line with the human cognition.

6. CONCLUSIONS

Affective computing has been one of the research hotspots in the field of artificial intelligence. Facial expression recognition needs to solve the problem of the uncertainty and fuzziness in recognizing emotion. In this paper, a deep convolution neural network based on dual transfer learning is used to establish an emotion PAD regressive model, which can recognize the basic emotions and map emotions in the emotional space of PAD. Meanwhile, emotional membership is used to quantitatively describe emotions in emotional space of PAD, which solves the problem that emotions are difficult to describe under uncertainty and fuzziness. In the next step, we will continue to study the human-computer emotional interaction, where the proposed method of this paper will make the human-computer model more humanoid.

ACKNOWLEDGMENTS

This work was partially co-sponsored by the Foundation of Zhejiang Educational Commission (Grant number: Y201738147), the Taizhou Science Technology Plan projects (Grant number: 1501KY61), the project of Taizhou University (Grant number: 2019PY013).

REFERENCES

[1] Mehrabian, A. 1996. Pleasure-arousal-dominance: A general framework for Describing and measure individual differences in temperament. *Current Psychology: Developmental, Learning, Personality, Social*, 4(Apr.1996), 261-292.

[2] CAO Jie, PENG Hao, WANG Hong, HU Po. 2009. PAD Based Facial Emotion Recognition. *Journal of Image and Graphics*, 5(May 2009), 759-763.
[3] Ning Yang. 2012. The Study of Facial Expression Recognition Based on the Improved PAD Emotion Model. *Southwest University*, 2012.

[4] Ying SUN, Yan-xiang HU, Xue-ying ZHANG, Shu-fei DUAN. 2019. Prediction of emotional dimensions PAD for emotional speech recognition. *Journal of Zhejiang University (Engineering Science)*, 10(Oct.2019), 2041-2048.

[5] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. In *proceedings of 2014 ACM international conference on Multimedia*(Orlando, Florida, November 03-07,2014). MM ’14.

[6] M.D. Zeiler, R. Fergus. 2014. Visualizing and Understanding Convolutional Networks. In *proceedings of 2014 European Conference on Computer Vision*(Zurich, Switzerland, September 6-12, 2014). ECCV ’14.