Expression Recognition Based on Gabor Filter

WeiChen Wang

Abstract:
We propose a comprehensive approach that integrates traditional methods with deep learning techniques to address the challenges of insufficient feature extraction, limited feature discriminability, and dimensionality catastrophe in facial expression recognition. This method begins by artificially extracting image features using Gabor filters to effectively capture relevant information within the images. Subsequently, Principal Component Analysis (PCA) is employed to reduce feature dimensions, controlling feature space while eliminating redundant information introduced by Gabor feature extraction, resulting in a novel feature representation. The reduced-dimensional features are fed into a Convolutional Neural Network (CNN), where multi-layered convolution and pooling operations extract abstract features and enable classification training. Experimental results demonstrate that the proposed method achieves impressive performance on the CK+ dataset, with an accuracy of 98.98%, signifying a substantial improvement over traditional approaches.

Keywords: Facial expression recognition, CK+ dataset, Gabor filters, PCA (Principal Component Analysis), Convolutional Neural Network,

1. Introduction

1.1 Background

Facial expression analysis[1] is a significant research area in computer vision and pattern recognition. It aims to understand and recognize human emotional states by analyzing and interpreting facial expressions. With the rapid advancements in computer vision and artificial intelligence technologies, computer-based facial expression recognition has become a reality[2]. Facial expression analysis finds wide-ranging applications in emotion recognition, user experience assessment, intelligent human-computer interaction[3], psychological health monitoring, safe driving[4], and various other fields.

Human facial expressions[5] convey emotions and information, making facial expression analysis crucial for understanding human emotional cognition and social behavior. Psychologist Mehrabian[6] demonstrated through extensive experiments that in daily communication, approximately 55% of emotional communication is conveyed through facial expressions.

Traditionally, facial expression analysis heavily relied on manually designed features and machine learning algorithms. However, these methods had limitations when dealing with complex facial images and diverse expressions.

In recent years, the rise of deep learning technologies has introduced new opportunities for facial expression analysis. Deep learning models, particularly Convolutional Neural Networks (CNNs), can automatically learn feature representations from raw image data, enabling them to learn facial expression features and achieve accurate classification. This deep learning-based approach exhibits improved generalization and robustness, leading to significant advancements in facial expression analysis.

1.2 Proposed Approach to Address the Shortcomings

In this study, we aim to explore deep learning methods for facial expression analysis and propose a framework that combines Gabor filter feature extraction with a CNN classifier. Gabor filters are widely used for texture analysis and edge detection, allowing us to capture fine details in facial expressions by extracting texture features and edge information from facial images. By integrating Gabor filters with deep learning models, our goal is to enhance the accuracy and robustness of facial expression analysis.

This research aims to develop an efficient and accurate facial expression analysis system that can automatically recognize and classify different expression categories. We will conduct experiments and evaluations using the publicly available CK+ dataset, employing common evaluation metrics such as accuracy, precision, recall, and F1 score to assess the performance of our method.

Through this research, we aim to provide valuable insights into facial expression analysis and offer useful references and guidance for researchers and developers in related domains.

The key innovations in this paper can be summarized as follows:

(1)A new facial expression analysis method has been proposed by combining traditional machine learning with deep learning. Gabor filters are used to extract local features from images. PCA is applied to reduce the dimensionality of the extracted feature images, highlighting feature details in the data input to the CNN neural network while minimizing irrelevant image information that could mislead the model.

(2)On the CK+ dataset, our method achieves an accuracy of 98.98% on the validation set, demonstrating excellent performance in facial expression classification.

2. Related Work

In the following section, we will introduce the latest research developments in facial expression recognition from two perspectives: traditional machine learning and deep learning.

2.1 Traditional Facial Expression Recognition

Traditional machine learning methods primarily rely on digital data extraction from facial expression images for feature extraction and processing. Feature extraction methods based on machine learning are generally categorized into two approaches: global extraction and local extraction.

Global extraction methods focus on analyzing the overall facial expression features, such as extracting holistic expression information from a facial photo or video frame. Some classic algorithms for global extraction include Linear Discriminant Analysis (LDA)[7], Principal Component Analysis (PCA)[8], and Independent Component Analysis (ICA)[9]. Global extraction methods are suitable for capturing global expression information but may overlook the finer details of facial expressions, necessitating local feature extraction methods.

Local feature extraction methods analyze specific facial regions, such as eyes, eyebrows, mouth, wrinkles, etc. A more detailed understanding of different expressions can be achieved by examining subtle variations in these specific regions. Classic local extraction methods include Gabor wavelets[10] and Local Binary Pattern (LBP)[11]. Local feature methods can capture subtle differences in expressions but may require more computational resources and complex algorithms to handle multiple local features.

After feature extraction, a classifier is needed to associate the extracted features from facial images with predefined expression categories. During training, the classifier learns to map features to the correct expression category. Common classifiers include Support Vector Machines (SVM), decision trees, random forests, and naive Bayes.

2.2 Deep Learning-Based Expression Recognition

In recent years, deep learning methods have emerged as a powerful approach to facial expression analysis, leading to significant breakthroughs in the field. The resurgence of deep learning began with a paper[12] by Hinton and Ruslan in 2006 addressing the problem of vanishing gradients in deep network training. Deep learning models, particularly Convolutional Neural Networks (CNNs)[13], can automatically learn feature representations from raw image data without requiring manual feature design. This advantage has resulted in outstanding performance of deep learning methods in facial expression classification.

2.3 Chapter Summary

In conclusion, facial expression analysis is a diverse and challenging research field. The continuous development of feature extraction and classification methods has significantly improved the accuracy and robustness of facial expression classification. In particular, deep learning methods have shown remarkable performance in enhancing classification accuracy. In this study, we aimed to further enhance facial expression classification by combining the Gabor filter and PCA feature extraction with deep learning CNN. Satisfactory results were achieved. This field holds immense potential, and with ongoing technological advancements, we believe that facial expression analysis will continue to positively impact our society and daily lives.

3. Methods

3.1 Structural Diagram

Based on Gabor filters and deep learning, the facial expression recognition network proposed in this paper is structured as shown in the diagram. The workflow is divided into four main parts: data preprocessing, feature extraction, data dimensionality reduction, and model training. Data preprocessing involves data splitting for training and validation, image normalization and other operations. Feature extraction includes local feature extraction using Gabor filters and dimensionality reduction using Principal Component Analysis (PCA) to extract facial image features effectively, thereby enhancing facial expression recognition performance.

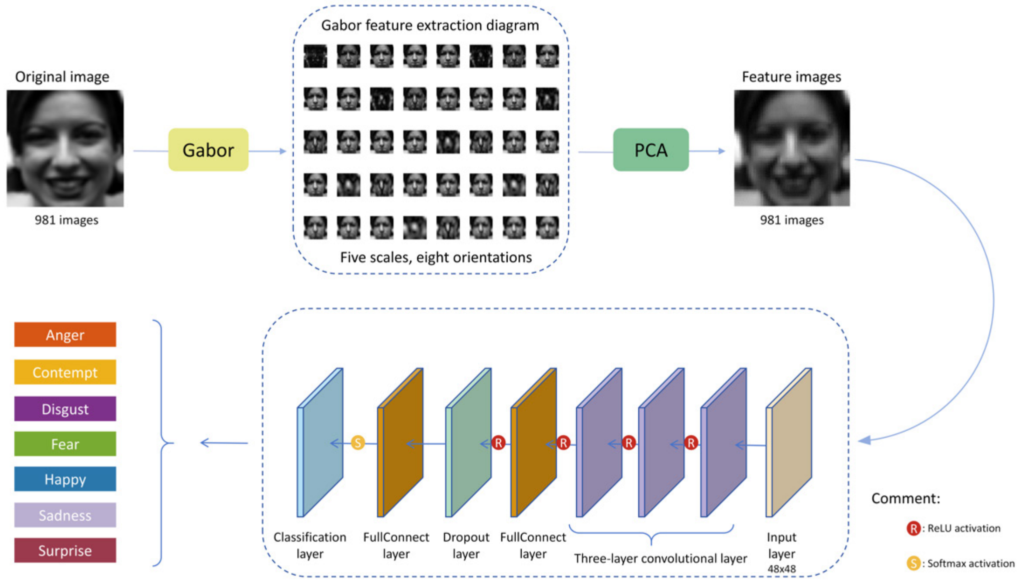


Table1. The facial expression recognition network that combines Gabor+PCA+CNN

3.2Introduction of the Experimental Process Around the Structural Diagram

3.2.1 Data Preprocessing

For data preprocessing, we performed the following steps:

(1) Data Loading: Firstly, we defined the folder path of the CK+ dataset and category labels. Images for each facial expression category are stored in separate folders. By specifying the folder path and category labels, we obtained all images' paths and corresponding labels.

(2) Image Reading: We used MATLAB's `read` function to read each image. We obtained the image paths by iterating through image files in each folder and then used the `read` function to read the images as grayscale.

(3) Storing Images and Labels: The read images were stored in a four-dimensional array (48x48x1xN), where N is the number of images. Additionally, we created a classification label array to store the corresponding facial expression category labels for each image.

(4)Dataset Splitting: For ease of training and validation, we split the dataset into training and validation sets. We used the common split ratio of 80% for training and 20% for validation. Using MATLAB's `crosswind` function, we randomly partitioned the data and obtained the respective indices for the training and validation sets.

(5) Preprocessing Steps: Before feature extraction, we applied preprocessing steps, such as normalization and transformation, to the image data. Specifically, we normalized the image data to scale pixel values between 0 and 1. This preprocessing step helps the model handle image data more effectively and enhances training stability.

Through the dataset above loading and preprocessing steps, we obtained image data for the training and validation sets and their corresponding labels, preparing them for subsequent feature extraction and model training.

3.2.2 Feature Extraction and Model Training

This research employs methods such as Gabor filter feature extraction and a Convolutional Neural Network (CNN) classifier to achieve facial expression analysis.

Gabor Filter Feature Extraction

Gabor filters are widely used to extract texture information from images. We utilize Gabor filters to capture texture and edge features in facial images, which are crucial for facial expression analysis. Specifically, we create a set of Gabor filters, including five different orientations and eight scales. These filters are applied to images in the training and validation sets to obtain Gabor filter response feature maps. The mathematical representation of a Gabor function is as follows:

$$G(x,y;\lambda,\theta,\psi,\sigma,\gamma)=\exp\left(-\frac{2\sigma^2 x'^2+\gamma^2 y'^2}{2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\psi\right)$$

Here, x and y are pixel coordinates, $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$ are the coordinates rotated along the filter's direction, λ is the wavelength, θ is the orientation angle, ψ is the phase offset, σ is the standard deviation, and γ is the ellipticity parameter.

Gabor filter response images contain information related to image texture and edges. By applying Gabor filters to the images, we transform them into feature maps, which will serve as inputs to the deep learning model.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a commonly used data analysis technique for dimensionality reduction of high-dimensional data. This study applies PCA to fuse features from 40 Gabor filter response images generated for each image. This reduces data dimensionality and extracts primary features, making inputting the data into the CNN easier. The algorithmic steps for implementing PCA are as follows:

Given m data samples, each with n dimensions:

(1) Combine the original data into a matrix X with n rows and m columns.

(2) Normalize each row of X to have zero mean.

(3) Calculate the covariance matrix $C = \frac{XX^T}{m}$.

(4) Obtain the eigenvalues and corresponding eigenvectors of C.

(5)Sort the eigenvectors in descending order according to their corresponding eigenvalues and form a matrix P with the top k rows.

(6) Transform the data as Y = PX to obtain a k-dimensional representation.

Convolutional Neural Network (CNN) Classifier

We constructed a CNN classifier specifically designed for facial expression analysis to achieve end-to-end image classification. CNN is a deep learning model tailored for image classification tasks, as it can automatically learn feature representations from data without manual feature engineering.

Our CNN model consists of convolutional layers, ReLU activation layers, pooling layers, and fully connected layers. By stacking multiple convolutional layers, the model captures abstract features from images. Fully connected layers map these features to different facial expression categories.

We use the cross-entropy loss function as the classifier's loss function and employ the stochastic gradient descent (SGD) optimization algorithm to train the model. We added a Dropout layer to the model to prevent overfitting, randomly dropping some neurons during training, enhancing the model's generalization.

3.2.3 Training and Validation

In the training phase, we use the training set's image data and corresponding labels to train the CNN classifier. We update the model's weights and biases using the backpropagation algorithm to minimize the loss function. We monitor the model's performance using the validation set and adjust hyperparameters such as learning rate and regularization based on validation performance.

4. Experiment

4.1. Dataset and Experimental Parameters

4.1.1 Dataset Introduction

To conduct facial expression analysis, we utilized publicly available datasets to train and evaluate our model. This study used the CK+ (Cohn-Kanade) dataset, a classic facial expression database containing images from various expression categories. The CK+ dataset comprises posed samples from individuals, each with corresponding expression labels.

We divided the CK+ dataset into seven facial expression categories: anger, contempt, disgust, fear, happiness, sadness, and surprise. Each category contains approximately 100 facial images, totaling around 981 images. These images are provided in grayscale format and have a size of 48x48 pixels.

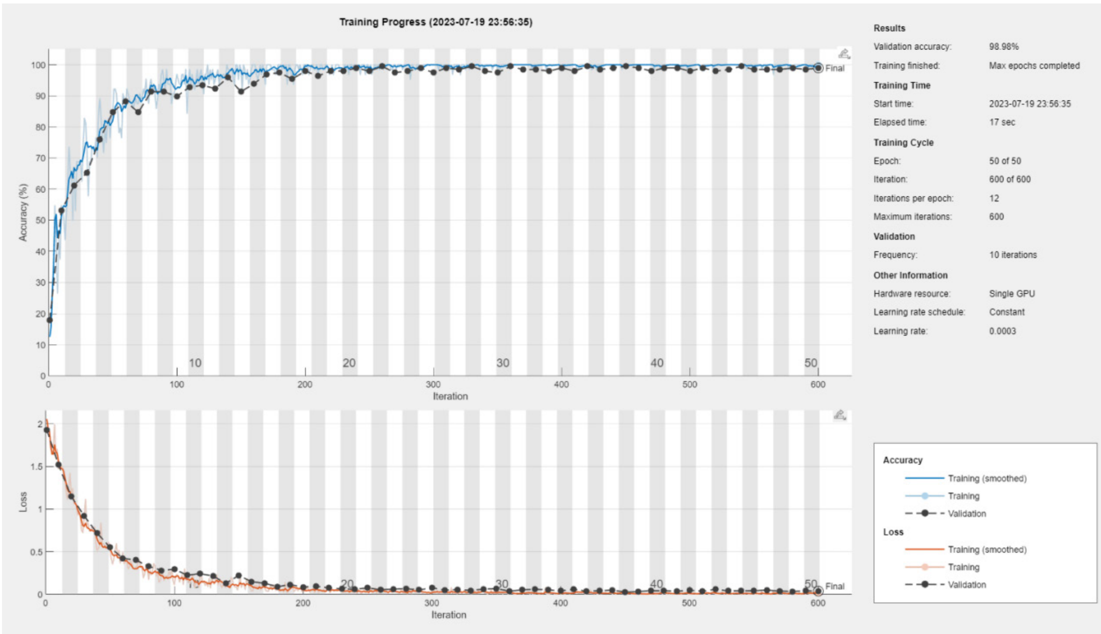
4.1.2 Testing Protocol

Based on our experimental results, our approach achieved an accuracy of 98.98%

on the validation set, indicating good performance in facial expression classification.

The training process graph is shown below:

Table2. Training Process



Accuracy is a crucial evaluation metric in facial expression analysis. Our model achieved high accuracy, demonstrating the effectiveness of our approach in recognizing and classifying different facial expression categories. The combination of Gabor filter feature extraction and deep learning CNN classification provides a strong foundation for automatic learning and extracting facial expression features from images, leading to accurate classification.

Additionally, we calculated precision, recall, and F1 scores to perform a more detailed analysis of the model's classification performance.

Table3. Precision, recall, and F1 scores

Facial expression tag	Class-wise accuracy	Class-wise recall	Class-wise F1 scores
Anger	100.00%	93.10%	0.9643
Contempt	81.82%	100.00%	1.0000
Disgust	100.00%	100.00%	1.0000
Fear	100.00%	100.00%	1.0000
Happy	100.00%	100.00%	1.0000
Sadness	100.00%	100.00%	1.0000
Surprise	100.00%	100.00%	1.0000

4.1.3 Experimental Parameters

The experiments were conducted on a system running Windows 11 with an NVIDIA GeForce GTX 4070 GPU, using MATLAB to execute the program. The sample image size was 48x48 pixels. The training learning rate was set to 0.0003, the maximum number of training iterations was 50, the batch size was 64, and L2 regularization was 0.0001.

4.2 State-of-the-Art Comparison

To investigate the facial expression recognition performance of the proposed Gabor+PCA+CNN method, we conducted experiments on the CK+ dataset and compared it with some existing models. The accuracy comparison is shown in the following figure:

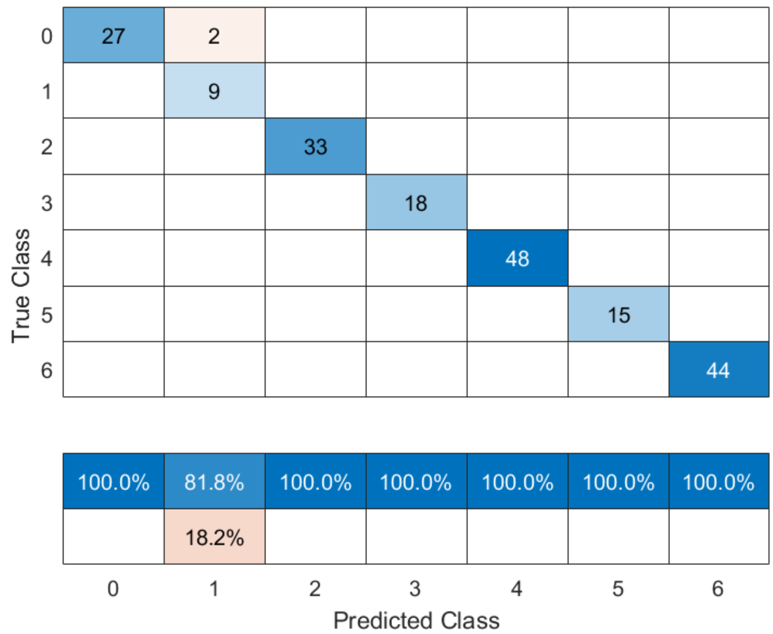
Table4. Accuracy comparison of different models

Method	Accuracy rate
Gabor+PCA+CNN	98.98%
PHOG+LBP[14]	94.6%
SEPL[15]	94.7%
DTAGN[16]	96.7%
Ali-Net[17]	93.2%

The proposed method achieved an accuracy of 98.98% on the CK+ dataset, outperforming the other four methods. Among the several advanced algorithms we compared, it achieved the highest accuracy.

We also visualized the model's classification results using a confusion matrix:

Table5. Confusion Matrix



The confusion matrix reveals that the model struggled to distinguish between "contempt" and "anger," leading to difficulty and confusion. However, the model performed well in recognizing "anger," "disgust," "fear," "happiness," "sadness," and "surprise," achieving an accuracy of 100%.

5. Ablation Experiments

Comparative experiments were set up to validate the importance of Gabor filters and PCA in improving model performance. Group A represents the model without Gabor filters, Group B represents the model without Gabor filters and PCA, and Group C represents the original model with Gabor filters and PCA. Here are the results of five experiments conducted using the CK+ dataset:

Table6. Comparative trial

Round	A	B	C
1	96.09%	97.01%	98.89%
2	97.01%	97.87%	98.55%
3	96.61%	98.10%	99.31%
4	95.78%	98.03%	98.70%
5	96.00%	97.66%	98.99%
Average Value	96.30%	97.74%	98.89%

The experimental data demonstrates that Group B achieved the highest accuracy, indicating a significant improvement due to the feature extraction capabilities of Gabor filters and PCA. Group A's lower accuracy than Group B might be because using only PCA for feature extraction resulted in the loss of some local features in the images, thus reducing the model's recognition performance. These experiments show that utilizing Gabor filters and PCA for facial expression features leads to greater accuracy and improves model performance.

6. Conclusion

This research explored a deep learning approach for facial expression analysis by combining Gabor filter feature extraction with a CNN classifier framework. Through experiments and evaluations using the publicly available CK+ dataset, we draw the following conclusions:

6.1 Innovation and Advantages:

(1)The effectiveness of deep learning: Leveraging deep learning models, especially Convolutional Neural Networks (CNNs), allowed us to learn facial expression features from images automatically. Experimental results demonstrated the excellent performance of our deep learning method in facial expression classification, achieving high accuracy and recall.
(2)The advantage of Gabor filter feature extraction: Integrating Gabor filters with deep learning models enabled more efficient capture of fine-grained facial expression details and texture features. Introducing Gabor filters significantly improved the model's expressive recognition performance, particularly in certain expression categories.

6.2 Challenges and Future Directions:

Possibility for Further Optimization: While our method performs exceptionally well in most cases, there is still room for improvement. Future work can optimize the model structure, integrate more sample data, and explore different feature extraction methods to further enhance facial expression analysis performance.

6.3 Future Research Directions:

(1)Multi-Modal Emotion Recognition: Combining facial expression analysis with other emotion recognition modalities such as audio, speech, and posture information can lead to a more comprehensive and accurate multi-modal emotion recognition system.
(2)Real-Time Emotion Analysis: Optimizing the model architecture and algorithms for real-time emotion analysis will enable the system to perform facial expression recognition and emotion analysis on video streams or real-time camera data.
(3)Long-Term Emotion Modeling: Research on how to model emotions over long-time sequences will enable the system to capture and understand the evolution and changes of emotions over time.

6.4 Chapter Summary

Facial expression analysis is an important research area in emotion recognition and human-computer interaction. It holds promising applications and extensive research opportunities. We believe that with ongoing technological advancements, facial expression analysis will continue to positively impact our society and daily lives.

[1]Masi I, Wu Y, Hassner T, et al. Deep face recognition: A survey[C]//2018 31st SIBGRAPI conference on graphics, patterns, and images (SIBGRAPI). IEEE, 2018: 471-478.
[2]Bosch N, D'Mello S K, Baker R S, et al. Detecting student emotions in computer-enabled classrooms[C]//IJCAI. 2016: 4125-4129.
[3]Samara A, Galway L, Bond R, et al. Affective state detection via facial expression analysis within a human–computer interaction context[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(6): 2175-2184.
[4]Zhang Dianye, Cheng Jing, Zhang Yi.Study on the influence of emotion on driving behavior [J].Chinese Journal of Safety Science, 2018, 28 (10): 19-24.
[5]Sariyanidi E, Gunes H, Cavallaro A. Automatic analysis of facial affect: A survey of registration, representation, and recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 37(6): 1113-1133.
[6] Mehrabian A, Ferris S R. Inference of attitudes from nonverbal communication in two channels[J]. Journal of Consulting Psychology, 1967, 31(3): 248.
[7]Zong Y, Zheng W, Huang X, et al. Transductive transfer lda with riesz-based volume lbp for emotion recognition in the wild[C]//Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. 2015: 491-496.
[8]Uddin M Z, Lee J J, Kim T S. An enhanced independent component-based human facial expression recognition from video[J]. IEEE Transactions on Consumer Electronics, 2009, 55(4): 2216-2224.
[9]Abdulrahman M, Gwadabe T R, Abdu F J, et al. Gabor wavelet transform-based facial expression recognition using PCA and LBP[C]//2014 22nd Signal Processing and Communications Applications Conference (SIU). IEEE, 2014: 2265-2268.
[10]Zheng W, Zhou X, Zou C, et al. Facial expression recognition using kernel canonical correlation analysis (KCCA)[J]. IEEE transactions on neural networks, 2006, 17(1): 233-238.
[11]Zhu Xiaoming, Yao Minghai. Facial expression recognition based on local binary pattern [J].Computer system application, 2011,20 (06): 151-154.
[12]Hinton G E, Salakhutdinov R R. Reducing data dimensionality with neural networks[J]. science, 2006, 313(5786): 504-507.
[13]Li W, Li M, Su Z, et al. A deep-learning approach to facial

expression recognition with candid images[C]//2015 14th IAPR International Conference on Machine Vision Applications (MVA). IEEE, 2015: 279-282.
[14]Happy S L, Routray A. Robust facial expression classification using shape and appearance features[C]//2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR). IEEE, 2015: 1-5.
[15]Happy S L, Routray A. Automatic facial expression recognition using features of salient facial patches[J]. IEEE Transactions on Affective Computing, 2014, 6(1): 1-12.
[16]Jung H, Lee S, Yim J, et al. Joint fine-tuning in deep neural networks for facial expression recognition[C]//Proceedings of the IEEE international conference on computer vision. 2015: 2983-2991.
[17]Mollahosseini A, Chan D, Mahoor M H. Going deeper in facial expression recognition using deep neural networks[C]//2016 IEEE Winter conference on applications of computer vision (WACV). IEEE, 2016: 1-10.