applied sciences

MDPI

Article

Application of Noise Detection Using Confidence Learning in Lightweight Expression Recognition System

Yu Zhao [1,2,3], Aiguo Song [1,2,3,*] and Chaolong Qin [1,2,3]

[1] The State Key Laboratory of Bioelectronics, Southeast University, Nanjing 210096, China; 220193286@seu.edu.cn (Y.Z.); 230198306@seu.edu.cn (C.Q.)
[2] Jiangsu Key Lab of Remote Measurement and Control, Southeast University, Nanjing 210096, China
[3] School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
* Correspondence: a.g.song@seu.edu.cn

Abstract: Facial expression is an important carrier to reflect psychological emotion, and the lightweight expression recognition system with small-scale and high transportability is the basis of emotional interaction technology of intelligent robots. With the rapid development of deep learning, fine-grained expression classification based on the convolutional neural network has strong data-driven properties, and the quality of data has an important impact on the performance of the model. To solve the problem that the model has a strong dependence on the training dataset and weak generalization performance in real environments in a lightweight expression recognition system, an application method of confidence learning is proposed. The method modifies self-confidence and introduces two hyper-parameters to adjust the noise of the facial expression datasets. A lightweight model structure combining a deep separation convolution network and attention mechanism is adopted for noise detection and expression recognition. The effectiveness of dynamic noise detection is verified on datasets with different noise ratios. Optimization and model training is carried out on four public expression datasets, and the accuracy is improved by 4.41% on average in multiple test sample sets. A lightweight expression recognition system is developed, and the accuracy is significantly improved, which verifies the effectiveness of the application method.

Keywords: confidence learning; expression recognition system; dynamic noise adjustment; datasets optimization

check for updates

Citation: Zhao, Y.; Song, A.; Qin, C. Application of Noise Detection Using Confidence Learning in Lightweight Expression Recognition System. *Appl. Sci.* **2022**, *12*, 4808. https://doi.org/10.3390/app12104808

Academic Editor: Yosoon Choi

Received: 28 March 2022
Accepted: 5 May 2022
Published: 10 May 2022

1. Introduction

The expectation of a convenient daily life, an aging society, child care in two-job families, and the shortage of nursing staff and other issues have created a large demand for intelligent service robots. Due to the need to directly provide corresponding services for users, the human–computer interaction of intelligent service robots needs to be comprehensive and natural. Service robots need to recognize users' emotions and respond in many application scenarios which are defined as emotional interaction.

Facial expression is an important way for humans to express emotions and accounts for 55% of the information in communication [1]. Facial expression recognition is widely used in medical [2], business, teaching [3], criminal investigation, and other fields to analyze people's emotional states by capturing facial expressions.Therefore, the emotional interaction intelligent robot based on expression recognition has become a research hotspot.

Expression recognition has been gradually studied by many subjects such as computer science, psychology, cognitive science, neural computing, and so on since the 1990s. Two core steps of expression recognition are feature extraction and expression classification. Feature recognition is particularly important in traditional methods. Global feature extraction methods and local feature extraction methods represented by geometric and texture feature extraction have been proposed. Machine learning is first applied to the expression classification after traditional expression feature extraction when developed, including SVM [4], KNN [5], and so on. However, the usage of complex algorithms and artificial feature extraction has the disadvantages of incomplete feature definition, inaccurate feature extraction, and being too cumbersome and time-consuming.

Many recognition tasks have made revolutionary leaps with the development of deep learning. The approach to feature extraction becomes extremely different with a deep artificial neural network. Instead of using complex algorithms or even manual feature extraction for pattern recognition, a large amount of data is “fed” into the black box of the convolutional neural network, and feature extraction and expression classification are carried out at the same time. The extracted features include color, texture, edge, and other features that human beings can understand, as well as a large number of features that human beings cannot even understand and explain directly. The weights of extracted features are updated iteratively through the back-propagation algorithm and optimizer, and an excellent effect is achieved.

A type of deep artificial neural network named convolutional neural network (CNN) originated from Hubel and Wiesel [6], who put forward the concept of “receptive field” in the study of cat visual structure and functional mechanism. CNN has been proved to have great advantages in image recognition and classification and has gradually become the mainstream method of expression recognition. In the study, most researchers improve the emotion recognition accuracy by increasing the size and depth when designing convolution network such as AlexNet [7], VGG [8], GoogLeNet [9], ResNet [10], and DenseNet [11]. However, this strategy of increasing the model complexity greatly increases the amount of calculation and puts forward higher requirements for hardware in the application so that the algorithms have poor portability for platform equipment with limited computing power, which is not conducive to real-time emotional interaction. Some current methods to make the neural network lightweight, including quantization, pruning, low-rank decomposition, teacher–student network, and lightweight network, design are used by researchers to solve the above problems. Some lightweight models for the domain of image recognition have been proposed through the study of efficient networks and lightweight methods at present [12]. A lightweight facial expression recognition system is characterized by small memory consumption and strong portability. It is generally constructed by a small-scale convolutional neural network model generated from the simplified deep CNN using lightweight methods. However, the lightweight model structure will lose part of the learning ability. The recognition model trained on specific datasets has a gap in performance from that on the verification dataset in the face of complex and diverse real environment samples, thus the generalization performance is difficult to guarantee. Such systems have poor performance in real interactive applications.

The seven recognized categories of expressions include happiness, surprise, sadness, fear, disgust, and anger [13]. Facial expression recognition has strong data-driven properties as a kind of fine-grained classification and the dataset samples at the input of the network also have an important influence on the performance, especially the generalization performance of the model in addition to the network structure and training algorithm as a consequence. The approach to improving the generalization performance is to increase the sample size, but the cost is high. In addition, it is liable to produce many incorrectly labeled samples that are difficult to identify and evaluate, namely, noise samples in the process of labeling. In this regard, a Self-Cure Network (SCN) is proposed to suppress the uncertainty in expression datasets because of the belief that noise samples may lead to insufficient learning of effective features [14], whereas SCN is aimed at large-scale expression recognition. On the one hand, it increases the weight module fully connected with the samples, changes the training process, and greatly increases the amount of calculation. On the other hand, it depends on the self-learning of the model. To optimize through the noise adjustment of datasets is a feasible solution, and there is less relevant research at present. A Confidence Learning (CL) algorithm [15] for the problem of noise samples is proposed, and certain noise detection effects on datasets such as ImageNet are achieved. A Confidence

Learning (CL) algorithm for the problem of noise samples is proposed in Reference [15], which achieved certain noise detection effects on datasets such as ImageNet. Therefore, the confidence learning algorithm for noise sample recognition and cleaning of datasets can be feasible and effective.

This paper builds a lightweight facial expression recognition system based on a network structure combining deep separation convolution and attention mechanism, and a confidence learning application method is proposed to detect and adjust the noise sample of the training dataset, in which appropriate noise samples are reserved and the robustness of the model is improved. The purpose of this approach is to generalize the suitable model chosen on specific training datasets to the real environment and improve the performance of robot emotional interaction in engineering applications. Specifically, this paper's main contributions lie in the following:

- In the expression recognition of human–computer natural emotional interactions, we find that the actual lightweight recognition system has poor recognition effect on the real environment, insufficient recognition rate, and interaction stability no matter how excellent the network recognizing accuracy is on a certain dataset in the theoretical simulation. We use a confidence learning algorithm to adjust the dataset to improve the generalization ability in the recognition environment in reality.
- In the confidence learning algorithm, we improve the reliability of noise detection by modifying the self-confidence.
- In the application of confidence learning, we propose to transfer the concept of hyper-parameters in machine learning to confidence learning and set the hyper-parameters according to the related problems affecting the noise adjustment effect. By manually adjusting the learning effect, the portability and flexibility of this method are improved.
- A lightweight expression recognition system suitable for human–computer emotional interaction is established, and the recognition effect is significantly improved through application optimization.
- This study provides an optimization idea for other lightweight human–computer interaction systems based on deep neural networks.

This paper introduces this strategy in five parts. Section 2 reviews the works related to the research background, algorithms, and engineering applications of this paper. Section 3 introduces the algorithm, application ideas, and methods of confidence learning, and briefly describes the model construction and training of the lightweight expression recognition system. Section 4 verifies the constructed system and the proposed scheme by three designed experiments. The experiments are carried out, and the results are analyzed and discussed. Section 5 succinctly summarizes the work of this paper and analyzes the shortcomings and further research improvement. The proposed method in this paper can play an important role in improving the emotional interaction experience of intelligent robots.

2. Related Works

2.1. Human–Computer Emotion Interaction Based on Facial Expression

Human–computer emotion interaction based on facial expression is applied in many scenes. For example, drivers' emotions can be recognized by facial expressions to monitor their emotional state to react promptly so that some traffic accidents can be avoided [16], and the robot recognizes the expression and interacts with the user at the same time so that the experience of the human–computer interaction will become better [17]. The corresponding emotional interaction development can be carried out based on some famous robot platforms such as NAO and Pepper. Using a lightweight expression recognition system is a feasible method to improve portability in different emotion-based interactions.

2.2. Lightweight CNN for Recognition Tasks

Several typical lightweight CNNs at present include Xception [18], MobileNet [19], ShuffleNet [20], and SqueezeNet [21]. Module Inception [9,22–24] is an important lightweight idea for models. In the inception module, the spatial correlation between the channels of the convolution layer can be separated and mapped, respectively. The main idea is to use the 1 × 1 convolution kernel to map each channel of the characteristic graph to a new small space, learn the correlation between channels in this process, and then use the conventional 3 × 3 or 5 × 5 convolution kernel or Concat operation to correlate the correlation between space and channels.

SqueezeNet is the earliest lightweight model. It draws on the idea of Inception that 1 × 1 convolution is used to reduce the number of channels input to the subsequent convolution kernel, and the dimension of the characteristic graph is increased in the subsequent module. In addition, the feature map resolution is maintained by delayed downsampling to obtain higher classification accuracy.

ShuffleNet and MobileNet are both built on depthwise separable convolutions. This network reduces the amount of computation and retains more information through depthwise convolution and pointwise convolution.

Xception is the abbreviation of Extreme Inception with the core idea of 2D depthwise separable convolution to completely separate the correlation of channel space. It first carries out the spatial convolution of each channel, then the channels are correlated, and simplifies the nonlinear activation function after the module. It is stacked in the form of a residual structure, which makes the model framework flexible to optimize.

2.3. Application of Confidence Learning in Engineering

Confidence learning has been applied in some engineering practices. CL is used by Li Wenna et al. [25] to detect errors in the knowledge base. In this application, the dataset has positive and negative binary classification samples, and the recognition model used is a multi-layer perceptron. In the application of confidence learning, samples that are discriminated wrong by the threshold are directly dropped. Better results may be achieved in a specific application, but in this case, the sample retention will decrease and there will be an overfitting risk in deep network learning.

Confidence learning is used by Zhang Minghua et al. [26] to evaluate air traffic complexity with noise. Serial machine learning algorithms are used to jointly study and judge the threshold. The threshold obtained can be more objective compared to judging according to one model. However, this method is time-consuming and computationally expensive if it is used in a deep network. Samples also are retained by dropping noise at a certain rate but the rate still seems to be high and not flexible.

3. Materials and Methods

3.1. Sample Noise Detection Based on Confidence Learning

3.1.1. Confidence Learning

A confidence learning algorithm for noise sample recognition and cleaning of datasets [15] is proposed. The algorithm trains the dataset through the noise recognition model, and the trained model reversely test the samples in the original dataset. Its core idea is to take the newly learned knowledge as the criterion to re-examine the source of knowledge. The steps are as follows:

1. Select noise detection model according to application requirements;
2. The threshold value of the sample being wrongly labeled as a certain type of label, which is defined as the self-confidence is determined by the model's prediction probability of the samples;
3. Compute the confusion matrix.

In this paper, the structure of the noise detection model is consistent with that of the recognition model. The set of seven categories of label values in the expression dataset is denoted as **Y** that is defined by the following Formula (1):

$$\mathbf{Y} = \{y|0 \leq y \leq 6, y \in \mathbb{Z}\} \tag{1}$$

The set of the dataset samples is denoted as $\mathbf{X}$ that is defined by the following Formula (2):

$$\mathbf{X} = \left\{ (\mathbf{x}, \tilde{y})^n | \mathbf{x} \in \mathbb{R}^d, \tilde{y} \in \mathbf{Y} \right\} \tag{2}$$

where $\mathbf{x}$ is $1 \times d$ sample pixel data, and d is the product of the length and width pixels of the sample image. $\tilde{y}$ is the corresponding label of the sample, and n is the sample size of the dataset. The sample count matrix of labeled label $\tilde{y}$ and potential real label y^* is denoted by $\mathbf{C}$, $\mathbf{C} \in \mathbb{N}^{7\times7}$, and the elements in $\mathbf{C}$ are evaluated by the following Formula (3):

$$\mathbf{C}_{i,j} = \left| \mathbf{X}_{i,j} \right| \tag{3}$$

In the formula above $\mathbf{X}_{i,j}$ is computed as the following Equation (4):

$$\mathbf{X}_{i,j} = \left\{ x | x \in \mathbf{X}_{\tilde{y}=i}, p(j;x) \geq t_j \right\} \tag{4}$$

where $X_{\tilde{y}=i}$ is the subset of all samples labeled with i in the sample set. $p(j;x)$ represents the predicted probability of sample $\mathbf{x}$ for label j. t_j is the self-confidence of label j and is computed as Equation (5):

$$t_j = \frac{1}{\left| \mathbf{X}_{\tilde{y}=j} \right|} \sum_{x \in X_{\tilde{y}=j}} p(j;x) \tag{5}$$

3.1.2. Application Method

The framework of the lightweight expression recognition system is shown in Figure 1.

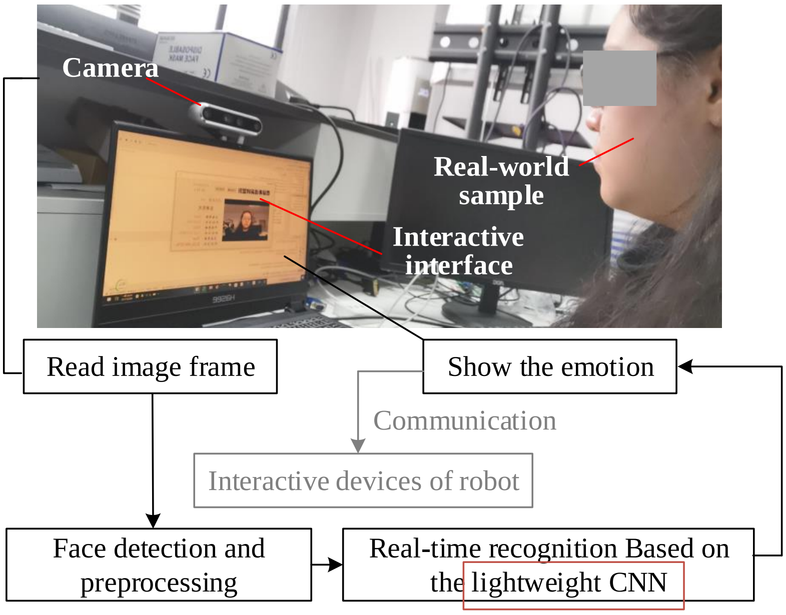

Camera
Real-world sample
Interactive interface
Read image frame
Show the emotion
Communication
Interactive devices of robot
Face detection and preprocessing
Real-time recognition Based on the lightweight CNN

Figure 1. Block diagram of the system.

The application idea of confidence learning sample noise detection in the system is shown in Figure 2. By adjusting the noise to optimize the training dataset, the dependence of the lightweight model on the specific training set can be reduced and the recognition ability in the actual emotional interaction can be enhanced.

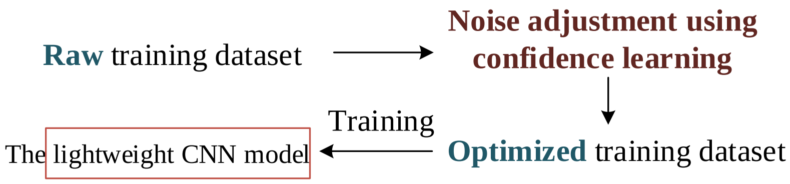

Raw training dataset
Noise adjustment using confidence learning
Optimized training dataset
Training
The lightweight CNN model

Figure 2. The Idea of the application.

In the optimization of expression recognition dataset, there are the following main problems:

1. The existing noise-cleaning mechanism cannot meet the demand. If the noise samples are cleared at the scale of 10^{-1} [15,26], it is easy to cause the loss of a large number of high-quality facial expression image samples in large-scale datasets. If the sample noise is too low, the model is over-fitting;
2. The identification of noise samples is inadequate. There are also some noise samples in the sample set which are not judged as noise by the confidence learning mechanism.

This paper introduces hyper-parameters aiming at the above problems. Hyper-parameters are parameters defined and valued before model learning in machine learning, such as the learning rate, the number of network layers, the number of hidden units, and the kernel size of layers in CNN [27] that are used to manually adjust the learning effect. The concept can be applied to the confidence learning process to more flexibly adjust the noise detection range according to the application requirements without increasing the complexity of the model, so it is suitable for lightweight systems.

Specifically, to solve problem 1 , sample labels were revised to reduce sample loss [14]. However, the revision discriminant threshold of noise detection is strongly dependent on the noise detection model. Therefore, modified self-confidence is proposed in this paper, which mainly includes two aspects: First, the self-confidence of a certain type of label is the probability mean. In order to improve the reliability of the mean, the low outliers are eliminated by the triple standard deviation method to reduce the impact of extreme data. Then, $p(j;x)$ in Equation (5) needs to meet the following Formula (6):

$$p(j;x) > t_j - 3 \cdot \sqrt{\sum_{x \in X_{\tilde{y}=i}} (p(j;x) - t_j)^2} \tag{6}$$

Secondly, the sample revision threshold is increased according to a certain proportion to reduce the influence of model recognition unreliability and noise detection range. In this paper, a set of distrust coefficients (DC) is introduced which is denoted as **DC**, then there is $\mathbf{DC} \in \mathbb{R}^7$. The seven values in the coefficient vector are multiplied by the self-confidence of the seven expression categories in the detection model, respectively. Each coefficient has a lower limit of 1 and an upper limit of $\max_j$, and the following inequality 7 holds:

$$\max_j = \frac{\max\limits_{x} (p(j;x))}{t_j} \tag{7}$$

To solve problem 2 , potential noise samples were mined and cleared of the sample set that had been recognized incorrectly by the detection model but not recognized as noise in a certain proportion. In this paper, a Random Noise Reduction Coefficient (RC) denoted as RC is introduced and $0 \leq RC \leq 1$. In order to ensure small sample loss and test objectivity, the magnitude is generally no more than 10^{-2}.

The application process of the above method is shown in Figure 3.

Three steps are mainly included in the process:

- Raw processing. The lightweight CNN model selected is trained on the initial sample set. All predictions including the probability of each facial expression for each sample can be calculated by the trained original model. Two groups from the sample set can be divided according to the comparison between the predicted label and the origin label. The self-confidence of an expression category can be simultaneously calculated from predicted probabilities of the certain facial expression for all samples.

- New dataset generation. The two groups of the sample set and self-confidence obtained in the first step are processed. The eliminated outliers are checked and removed for the calculation of self-confidence. Self-confidences are then adjusted by the set of distrust coefficients. The incorrectly predicted group is divided into two parts according to the modified self-confidences, the relabeled one and the random noise reduced one. The final adjusted dataset is combined with the two parts and the correctly predicted group.
- Optimized data training. The final adjusted model is the same model structure trained on the adjusted training set.

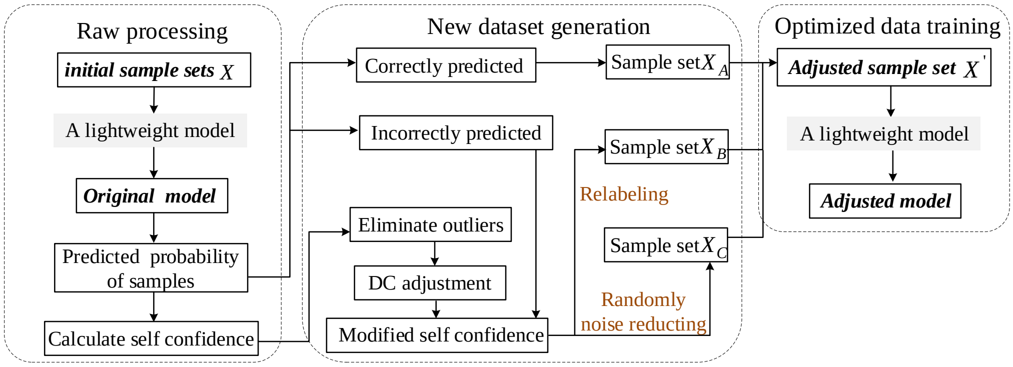


Figure 3. Application method.

3.1.3. Evaluation of Noise Detection Effect

Accuracy, Precision, Recall, F1 Score [25], Relabel Accuracy, and Retention Rate are statistically determined in a noise detection. The formulas for the above estimates are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \tag{11}$$

$$Relabel\ Accuracy = \frac{TN_{rl}}{N_{rl}} \tag{12}$$

$$Retention\ Rate = \frac{N_R}{N_O} \tag{13}$$

where TP is the number of true noise samples recognized, FP is the number of non-noise samples incorrectly recognized as noise, FN is the number of unrecognized true noise samples, and TN is the number of non-noise samples correctly recognized. N_{rl} is the sample number to be relabeled, TN_{rl} is the sample number relabeled correctly; N_O is the sample number of the original dataset, and N_R is the sample number that has been retained after noise detection. It can be perceived from the formulas that Accuracy reflects the accuracy of overall prediction, Precision reflects the reliability of prediction results of noise samples, and Recall reflects the accuracy of noise samples in the original samples. F1 score reflects the comprehensive evaluation of the above three values on the noise detection effect and the higher the value, the better the noise detection effect and the larger the detection range. The Relabel Accuracy is the proportion of the correct revised samples in all the revised samples, reflecting the positive effect of relabeling. The higher the revision accuracy is, the more effective the relabeling is. The Retention Rate is the proportion of the total amount of optimized samples to the total amount of original samples. The larger the sample retention rate is, the fewer samples that are lost and the better the data retention.

3.2. Lightweight Expression Recognition System

3.2.1. Image Preprocessing

Image preprocessing includes image graying, face detection, data normalization, resizing and dimension adjustment adapting to the requirements of the model, and data augmentation:

- Image graying. Image graying is necessary for recognition because of the feature complexity in RGB color images. In fact, the facial expression features in the gray image have met the recognition requirements, and it can greatly reduce the amount of calculation compared with color images. The average method in Formula (14) is used to convert RGB images into gray images:

$$\mathrm{Gray}(x,y) = [R(x,y) + G(x,y) + B(x,y)]/3 \tag{14}$$

- Face detection. Face detection refers to recognizing the region and location of the face using corresponding algorithms. The MTCNN [28] algorithm, which is accepted to do nice work, is adopted. The MTCNN algorithm is a face detection and face alignment method based on deep learning. It can complete the tasks of face detection and face alignment at the same time. Compared with traditional algorithms, it has better performance and faster detection speed. Image pyramid, P-Net, R-Net, and O-Net are the main steps included.
- Data normalization. We normalize each value of the pixel data of the input image to the [−1,1] interval through Formula (15) in order to better use the standardized data for training:

$$\boldsymbol{N} = (\boldsymbol{I}/255.0 - 0.5) \times 2 \tag{15}$$

  N is the result of normalization, and I is the input pixel matrix in Formula (15). After this processing, the illumination and other factors of the image are weakened, the extraction of useful features is easier, and the amount of calculation is reduced, so as to accelerate the convergence.
- Data augmentation. Data augmentation is a technique to expand the dataset and reduce overfitting based on some transformations of the data samples of the existing dataset. The main operations include cutting pictures, stretching, adjusting the brightness and angle, etc.

3.2.2. Selection of CNN Model Structure

A depthwise separable convolutional network [17] is adopted for facial expression recognition and noise detection in order to improve the operating efficiency of the recognition system. The network decouples the channel correlation and spatial correlation of the convolutional neural network, simplifies the network, and removes parameters with the idea of the residual module, which greatly reduces the model size. Its core processes, including depthwise convolution and pointwise convolution, are shown in Figure 4.

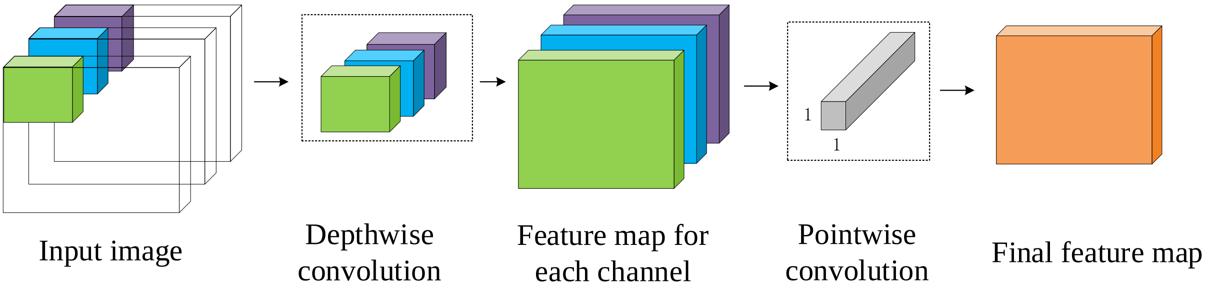


Figure 4. Depthwise separable convolution.

In order to focus on important feature weights in model learning, this paper adds the attention mechanism module CBAM [29] after the depth separation convolution layer of the first residual module of the CNN model [17]. The CBAM module has two sequential sub-modules: channel attention module and spatial attention module. For the channel attention module, the process can be summarized as:

$$\begin{aligned} \mathbf{F}' &= \mathbf{M_c}(\mathbf{F}) \otimes \mathbf{F} \\ \mathbf{F}'' &= \mathbf{M_s}\left(\mathbf{F}'\right) \otimes \mathbf{F}' \end{aligned} \tag{16}$$

where $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$ denotes an intermediate input feature map, $\mathbf{M_c} \in \mathbb{R}^{C\times 1\times 1}$ denotes a 1D channel attention map, $\mathbf{M_s} \in \mathbb{R}^{1\times H\times W}$ denotes a 2D spatial attention map, and $\otimes$ denotes element-wise multiplication. The channel attention module is computed as:

$$\begin{aligned} \mathbf{M_c}(\mathbf{F}) &= \sigma(MLP(\mathrm{Avg\,Pool}(\mathbf{F})) + MLP(\mathrm{Max\,Pool}(\mathbf{F}))) \\ &= \sigma\left(\mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{F^c_{avg}}\right)\right) + \mathbf{W_1}(\mathbf{W_0}(\mathbf{F^c_{max}}))\right) \end{aligned} \tag{17}$$

where $\mathbf{F^c_{avg}}$ and $\mathbf{F^m_{max}}$ denote average-pooled features and max-pooled features, respectively. $\mathbf{W_0} \in \mathbb{R}^{C/r\times C}$ and $\mathbf{W_1} \in \mathbb{R}^{C\times C/r}$, and they are the weights of the multi-layer perceptron (MLP). σ denotes the sigmoid function. The sigmoid function is calculated as:

$$y = \frac{1}{1+e^{-x}} \tag{18}$$

As the specific model structural parameters have been described in the reference [17], they will not be described here. A simple schematic diagram of the model structure is shown in Figure 5.

BN refers to Batch Normalization, which can solve the problem that the data distribution changes in the training process, so as to prevent the gradient from disappearing or exploding and speed up the training speed. It is calculated as:

$$y^{(i)} \leftarrow \gamma \odot \hat{x}^{(i)} + \boldsymbol{\beta} \tag{19}$$

where γ denotes the Stretch parameter, and β denotes the Offset parameter. The Relu function is calculated as:

$$f(x) = \max(0, x) \tag{20}$$

The scale test comparison of some lightweight models is shown in Table 1.

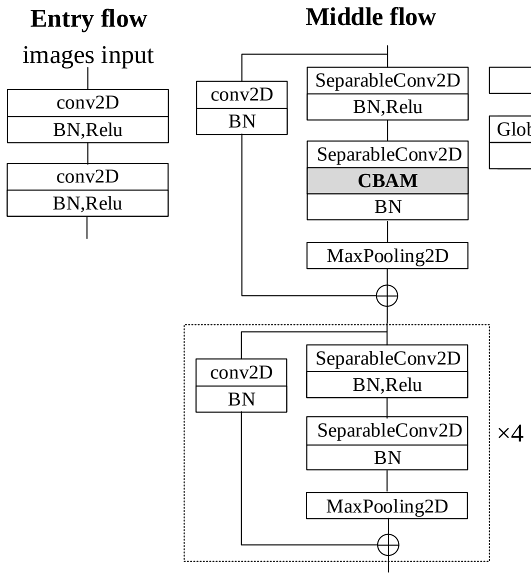


Figure 5. Diagram of model structure.

Table 1. Scale tests of some lightweight network.

Network	Number of Parameters (Thousand)	Size of Model (M)
Xception [18]	20,902.5	239
MobileNet [19]	3235.5	37.1
ShuffleNet [20]	1947.1	23.8
Deep separable CNN [17]	203.4	2.53
CNN [17] + CBAM	203.7	2.55

It can be found that the network of reference [17] is very tiny, which is suitable for building a lightweight facial expression recognition system. However, the proposed CNN [17]+CBAM can improve the accuracy of the reference [17] network by about 1 percentage point in training convergence on the FER2013 [30] dataset, but the model size is almost unchanged, so it is feasible.

3.2.3. Training of the Model

At the input layer, the model structure is matched to normalize the image pixel data and data augmentation is used in the training process to improve the robustness of the model. The softmax function is adopted at the end of the network, and the output result of expression category i of a sample x is recorded as $S_i(x)$, so the calculation Formula (21) is as follows:

$$S_i(x) = \frac{e^{p_i(x)}}{\sum_j e^{p_j(x)}} \tag{21}$$

where $p_i(x)$ is the probability score of sample x to the expression category i. The cross-entropy loss function is used to calculate the error, and the Formula (22) is as follows:

$$L = -\frac{1}{N}\sum_i \sum_{c=0}^{6} q_c(x)\log(S_i(x)) \tag{22}$$

where N is the number of samples, and $q_c(x)$ is the symbolic function representing whether the real label category of sample x is c.

4. Design and Results of the Experiments

4.1. Content and Purpose of the Experiments

Three experiments are carried out to verify the validity of the proposed method.

- Experiment 1: In order to prove the effect of confidence learning on the noise detection effect of facial expression dataset and the effect of the proposed hyper-parameters, a group of hyperparameters were used to evaluate datasets with different noise ratios;
- Experiment 2: In order to prove the effectiveness of the proposed method in facial expression recognition using the model above, four public facial expression datasets are used as training sets for optimization and comparison experiments;
- Experiment 3: In order to prove the function of the proposed method in the practical lightweight recognition system, the system is built and a real-time recognition experiment of real samples is carried out. The overall accuracy rate, the accuracy rate of various expressions, and specific frames are compared intuitively.

4.2. Datasets and Environment of the Experiments

The datasets are preprocessed before the experiment, including image graying, face detection and extraction, and resizing into 48 × 48 images. The datasets involved include FER2013 [30], FERPlus [31], CK+ [32], RAF-DB [33], KDEF [34], and JAFFE [35]. FERPlus [31] is a multi-label dataset formed by relabeling based on FER2013 [30]. In order to ensure the effectiveness of the noise detection effect test, in this paper, a total of 7942 samples which are in the original seven emoticon categories and with 10 voters voting for are selected from FERPlus as the correct label sample dataset. The correct sample labels of 10%, 20%, and 30% are randomly modified as noises and recorded as datasets FERplUS-10, FERPLUS-20, and FERPLUS-30, respectively.

The operating system for the experiment is win10, the CPU is Intel i7, and the running memory is 16G. The GPU is NVIDIA GeForce RTX 2070 with 8G video memory.

4.3. Results of the Experiments

4.3.1. Experiment 1

Six groups of hyper-parameters are tested on FERPLUS-10, FERPLUS-20, and FERPLUS-30. The first five groups are randomly selected within the range, and the sixth group does not exert influence. The proposed lightweight model is used for dynamic noise detection and adjustment. The experimental parameters are shown in Table 2.

Table 2. Parameters.

Number	DC	RC
1	[1.03, 1.03, 1.08, 1, 1.02, 1, 1.08]	0.01
2	[1.05, 1.065, 1.023, 1.01, 1.02, 1.01, 1.08]	0.002
3	[1.066, 1.08, 1.04, 1, 1.033, 1.02, 1.024]	0.008
4	[1.2, 1.3, 1.1, 1.1, 1.1, 1.2, 1.08]	0.02
5	[1.05, 1.3, 1.05, 1.06, 1.3, 1, 1.01]	0.03
6	[1, 1, 1, 1, 1, 1, 1]	0

The noise detection evaluation values are counted according to Section 3.1.3, as shown in Figure 6.

From the change in the F1 Score, it can be judged that the defined hyper-parameters can dynamically adjust the detection range of noise samples with high precision and accuracy and minimal sample loss in the process. Moreover, the proposed parameters show a certain positive correlation with the detection effect in different noise ratio datasets.

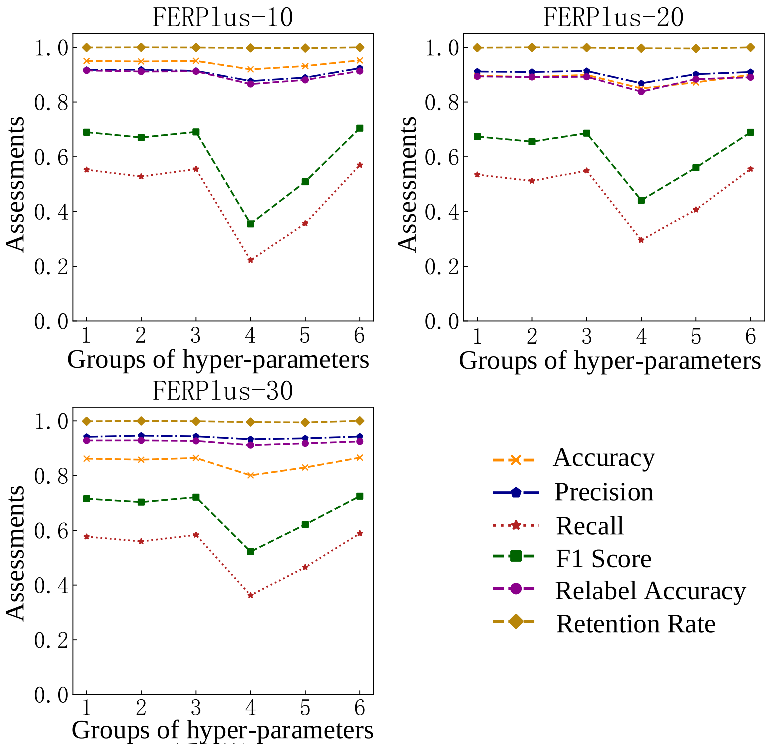


Figure 6. Dynamic noise detection effect of different noise proportions.

4.3.2. Experiment 2

The models are trained on RAF-DB, CK+, KDEF, and FER2013 before and after optimization and are verified by sampling in a certain proportion on original RAF-DB, CK+, JAFFE, KDEF, and FER2013. Among them, four datasets other than the training set are taken as the corresponding actual environment samples. In order to ensure the effectiveness of the experiment, the single validation sample set of the model obtained before and after the optimization of each specific training set is the same. In order to reduce contingency, five random samples are selected for each validation set, and the average recognition accuracy is taken as the result. The experimental results of the generalization performance are shown in Table 3.

Table 3. Model recognition performance before and after noise adjustment in the training set.

Training Dataset	Noise Adjustment [1]	RAF-DB	CK+	JAFFE	KDEF	FER2013
RAF-DB	×	**81.40%**	71.00%	39.50%	53.20%	51.10%
	✓	80.80%	**72.60%**	**47.30%**	**58.70%**	**51.60%**
CK+	×	30.70%	**99.40%**	25.60%	**40.00%**	29.90%
	✓	**35.70%**	94.80%	**30.20%**	36.10%	**30.20%**
KDEF	×	19.90%	29.70%	20.90%	**77.50%**	21.40%
	✓	**26.50%**	**48.70%**	**30.20%**	76.90%	**25.70%**
FER2013	×	60.80%	64.80%	41.90%	50.50%	**71.30%**
	✓	**61.40%**	**68.70%**	**46.50%**	**51.30%**	71.00%

[1] In the column 'Noise Adjustment', to the training set, × means the noise adjustment hasn't been made and ✓ means the noise adjustment has been made.

It can be perceived that the accuracy of the model optimized by the training set decreases slightly on the original training set. This is because the optimized model is trained based on the noise-adjusted training set, and the test sets are all original datasets. Therefore, the accuracy of the optimized model will inevitably decline in the original training set but will rise in other test sets that are not for training, which indicates that the dependence on specific training sets has decreased. From Table 3, the verification accuracy on most other datasets has indeed improved. We can analyze the performance of each training set in its corresponding real sample environment before and after noise adjustment from Table 3. The assessment is shown in Table 4.

It can be deduced from Table 4 that the generalization performance of the adjusted model has been improved to a certain extent. The Accuracy in other datasets is inferior for the training set KDEF. The reason is that KDEF has 4900 samples in total but includes only 70 subjects. The samples are photographed in the laboratory environment, and the image size, illumination, clarity, and posture are strictly controlled. The same expression of the same subject has multiple samples from different angles, but there is no difference in their facial expression characteristics. Therefore, its effective sample size is too small for deep neural network learning, resulting in overfitting. From the data, the relative increase for the model trained by KDEF before and after noise adjustment is observed to be much higher than the other training sets. It can be inferred that the more serious the overfitting, the better the optimization effect.

The virtue of the purpose of this study is to build a lightweight expression recognition system and adopt confidence learning noise detection to optimize the practical application effect. Only the network structure constructed in Section 3.2 is tested, and other networks will not be further discussed in this article.

Table 4. Improvements in accuracy relative to real sample environment.

Training Set	Adjustment [5]	Accuracy [1]	Absolute Increase [2]	Relative Increase [3]
RAF-DB	×	53.70%	3.85%	7.17%
	✓	**57.55%**		
CK+	×	31.55%	1.50%	4.75%
	✓	**33.05%**		
KDEF	×	22.98%	9.80%	42.66%
	✓	**32.78%**		
FER2013	×	54.50%	2.48%	4.54%
	✓	**56.98%**		
Total average [4]	-	-	4.41%	14.78%

[1] The accuracy here refers to the recognition effect of the model trained on a training set on the dataset other than the original training dataset. For example, the first row of data in Table 3 represents the model effect trained before the adjustment of RAF-DB, so the accuracy of the first row in Table 4 is calculated as the mean value of the accuracy of the first row in Table 3 on CK +, JAFFE, KDEF, and FER2013, that is, (71.00% + 39.5% + 53.2% + 51.1%) ÷ 4 = 53.7%. [2] The column Absolute increase means the increase in the data in column Accuracy before and after noise adjustment for the relative training set. For example, 3.85% is the result of 57.55% minus 53.70%. [3] The column Relative increase means the percentage of the absolute increase to the accuracy before noise adjustment for the relative training set. For example, 7.17% is the result of 3.85% divided by 53.70%. [4] Total average means the mean value of the column.[5] In the column 'Adjustment', to the training set, × means the noise adjustment hasn't been made and ✓ means the noise adjustment has been made.

4.3.3. Experiment 3

The Intel realistic depth camera D435 deployed to the interactive terminal of the intelligent robot is used to obtain facial images. The data acquisition range of d435 can reach 10m and the frame rate is 30 fps, which can better meet the requirements of the shooting environment and real-time performance. In the actual expression recognition, the depth image is not used. The purpose of using the depth camera is to cooperate with the robot terminal to facilitate the integration of the system with other depth image recognition interactive methods such as gesture and posture. The back end of the recognition system uses the deep learning framework Keras to train and load the recognition model and preprocesses the acquired image through OpenCV and NumPy. The front end adopts the pyqt5 framework for interactive interface development for displaying the recognition image, expression category, and total recognition time in real-time and for tracking the emotion polarity classification [36] information of the displayed expression. The system recognition frame rate is 16.9 fps.

Through further statistics on the data in Table 3, the accuracy of the training model is weighted based on the sample size of each test set before and after optimization of the four training datasets, and the standard deviation of the accuracy on all test sets are obtained, as shown in Table 5.

Table 5. Effect of models in Experiment 2.

Training Dataset	Noise Adjustment [1]	Accuracy	Standard Deviation
RAF-DB	×	59.00%	0.17
	✓	59.50%	0.14
CK+	×	32.60%	0.31
	✓	33.70%	0.28
KDEF	×	24.00%	0.25
	✓	29.10%	0.22
FER2013	×	**67.50%**	**0.12**
	✓	**67.60%**	**0.11**

[1] In the column 'Noise Adjustment', to the training set, × means the noise adjustment hasn't been made and ✓ means the noise adjustment has been made.

It can be seen from Table 5 that the model trained by FER2013 in Table 3 has the highest weighted average accuracy and the lowest standard deviation, meaning a high recognition accuracy and a stable performance. Therefore, this group of models is selected for the experiment at the back end of the system. Three subjects are randomly tested in real-time, and they make seven expressions to different degrees. The total number of frames and accuracy of the test is shown in Table 6.

Table 6. Tests of recognition system.

Subject	Total Frames	Adjustment [1]	Correctly Recognized	Accuracy	Absolute Increase	Relative Increase
A	2647	×	1722	65.05%	18.55%	28.52%
		✓	2213	**83.60%**		
B	2650	×	1901	71.74%	10.37%	14.46%
		✓	2176	**82.11%**		
C	2289	×	1575	68.81%	14.11%	20.51%
		✓	1898	**82.92%**		
Average	-	-	-	-	14.34%	21.16%

[1] In the column 'Adjustment', to the training set, × means the noise adjustment hasn't been made and ✓means the noise adjustment has been made.

The average accuracy rate increased by 14.34%, which is significantly higher than the test results of the public dataset. The analysis reason is that the expression change speed of the subjects is lower than the frame rate during the real-time test. Similar frames with correct recognition after optimization and wrong recognition before optimization would be continuously recorded, resulting in a larger gap. The comparison of test frame number and accuracy rate of various expressions are shown in Table 7 and Figure 7, respectively.

Table 7. Numbers of test frames for various expressions.

Emotion	Subject A	Subject B	Subject C
Angry	230	466	583
Disgust	292	221	415
Scared	383	459	444
Happy	387	361	210
Sad	540	394	286
Surprised	512	550	231
Neutral	303	199	120

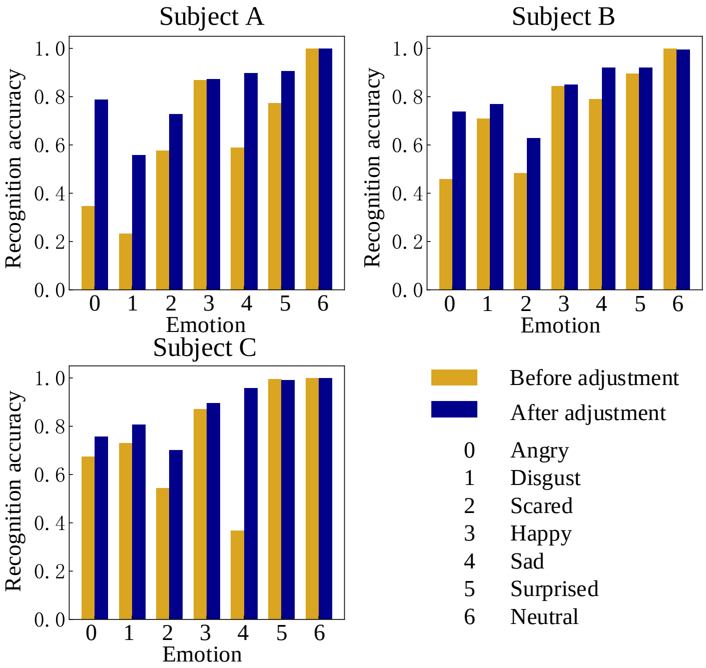
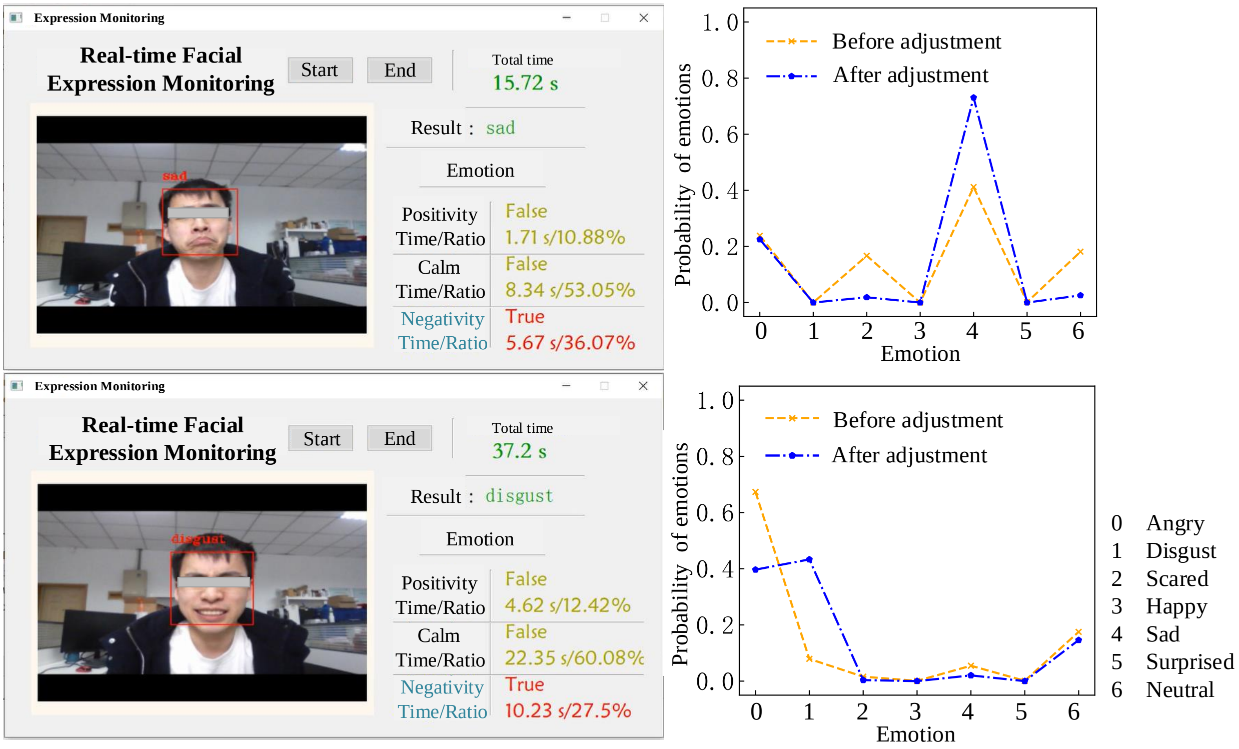


Figure 7. Comparison of various expression recognition tests.

It can be judged from Figure 7 that the accuracies of neutral and happy have high recognition using the model before optimization and were not significantly improved after optimization. However, the model greatly improved in the categories of anger, disgust, fear, and sadness. The system interface displays the optimized model recognition results in real-time. Some recognition frames and the corresponding expression probability outputs of subject A are shown in Figure 8.



Real-time Facial Expression Monitoring

Total time 15.72 s

Result : sad

Emotion

Positivity Time/Ratio: False 1.71 s/10.88%

Calm Time/Ratio: False 8.34 s/53.05%

Negativity Time/Ratio: True 5.67 s/36.07%

Real-time Facial Expression Monitoring

Total time 37.2 s

Result : disgust

Emotion

Positivity Time/Ratio: False 4.62 s/12.42%

Calm Time/Ratio: False 22.35 s/60.08%

Negativity Time/Ratio: True 10.23 s/27.5%

Figure 8. Expressions of sadness and disgust of Subject A.

It can be seen that the adjusted model also has a higher probability output of correct expression when recognizing the correct label at the same time.

5. Conclusions

This paper puts forward the problems of low recognition accuracy and poor generalization performance in the actual interaction and improves it from the perspective of a training dataset aiming at a lightweight expression recognition system in the emotional interaction scene of an intelligent robot. In order to reduce the dependence of the lightweight model on the specific training set, the application of a confidence learning algorithm is proposed to optimize the training set by noise sample detection. The relabeling method is applied at the same time. The hyper-parameters are innovatively introduced to manually adjust the noise detection range to match the application requirements and obtain better system robustness.

On the theoretical side, we add uncertainty to the confidence learning algorithm so that its learning effect can be adjusted by some preset hyper-parameters as the mode of machine learning, which increases flexibility and creates more possibilities for its better and wider application. However, the algorithm has not been accurately verified for other machine learning algorithms or deep learning networks and will be further studied in the future. On the practical side, a lightweight expression recognition system is established. We analyzed and found the insufficient recognition generalization performance caused by dataset dependence, and introduced confidence learning in an innovative way for noise detection and adjustment so that we realized a great improvement in the recognition effect in the actual system recognition, which is of great significance in the engineering application of human–computer emotional interaction. However, the generalizability of the application effect, such as the effect on special groups such as the elderly, children, and patients with facial diseases, needs to be further studied. In addition, frame-by-frame static expression recognition is proven to be effective after optimization in real-time recognition, but the recognition stability of the system is insufficient. More dynamic recognition methods will be further discussed in the future.

This paper further makes full use of the specific training set to optimize the model parameters without increasing the complexity of the model, changing the training method, and expanding the new dataset in other publications. It achieves good results in an actual lightweight expression recognition system. In the follow-up, the improvement of the lightweight model structure and the selection and optimization method of hyper-parameters will be further studied, and based on this, a robot emotional interaction system with a better experience will be built.

6. Patents

Expression recognition and emotion tracking method based on artificial and compound optimization dataset (CN202111173985.4).

Author Contributions: Conceptualization, Y.Z.; validation, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z. and A.S.; investigation and analysis, A.S. and C.Q.; resources, A.S.; project administration, A.S. and C.Q. All authors have read and agreed to the published version of the manuscript.

Funding: Special project of Jiangsu basic research plan (cutting-edge leading technology):BK20192004.

Institutional Review Board Statement: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed and written consent was obtained from all individual participants included in the study.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CK+ database is available at https://sites.pitt.edu/~emotion/ck-spread.htm accessed on 19 April 2022. The RAF-DB database is available at http://www.whdeng.cn/raf/model1.html accessed on 15 November 2021. The JAFFE database is available at https://zenodo.org/record/3451524 accessed on 26 March 2022. The KDEF dataset is available at https://kdef.se/download-2/register.html accessed on 26 March 2022. The FER2013 database is available at https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data accessed on 17 February 2020. The FERPlus database is available at https://github.com/Microsoft/FERPlus accessed on 20 February 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mehrabian, A. Communication without words. *Psychol. Today* **1968**, *2*, 53–55.
2. Girard, J.M.; Cohn, J.F.; Mahoor, M.H.; Mavadati, S.M.; Hammal, Z.; Rosenwald, D.P. Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image Vis. Comput.* **2014**, *32*, 641–647. [CrossRef] [PubMed]
3. Xu, L.L.; Zhang, S.M.; Zhao J.L. Summary of facial expression recognition methods based on image. *Comput. Appl.* **2017**, *37*, 3509. [CrossRef]
4. Tang, Y. Deep Learning Using Linear Support Vector Machines. *arXiv* **2015**, arXiv:1306.0239.
5. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theor.* **1967**, *13*, 21–27. [CrossRef]
6. Hubel, D.H.; Wiesel, T.N. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *J. Physiol.* **1962**, *160*, 106–154.
7. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
8. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arxiv:1409.1556.
9. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–9. [CrossRef]
10. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 6 June–1 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778. [CrossRef]
11. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2261–2269. [CrossRef]
12. Yang, H.T. Facial Expression Recognition Method Based on Lightweight Convolutional Neural Network. Master's Thesis, Beijing University of Civil Engineering and Architecture, Beijing, China, June 2020. [CrossRef]
13. Eckman, P. Universal and cultural differences in facial expression of emotion. *J. Pers. Soc. Psychol.* **1987**, *53*, 712–717. [CrossRef]
14. Wang, K.; Peng, X.; Yang, J.; Lu, S.; Qiao, Y. Suppressing Uncertainties for Large-Scale Facial Expression Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Seattle, WA, USA, 16–18 June 2020. [CrossRef]
15. Northcutt, C.; Jiang, L.; Chuang, I. Confident learning: Estimating uncertainty in dataset labels. *J. Artif. Intell. Res.* **2021**, *70*, 1373–1411. [CrossRef]
16. Xiao, H.; Li, W.; Zeng, G.; Wu, Y.; Xue, J.; Zhang, J.; Li, C.; Guo, G. On-Road Driver Emotion Recognition Using Facial Expression. *Appl. Sci.* **2022**, *12*, 807. [CrossRef]
17. Xu, G.Z.; Zhao, Y.; Guo M.M.; Jin M. Research on real-time interaction for the emotion recognition robot based on deepwise separable convolution. *Chin. J. Sci. Instrum.* **2019**, *40*, 8. [CrossRef]
18. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19._1. [CrossRef]
19. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
20. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856. [CrossRef]
21. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5 MB Model Size. *arXiv* **2016**, arXiv:1602.07360.
22. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the Machine Learning Research, Lille, France, 7–9 July 2015; Volume 37, pp. 448–456.
23. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2818–2826. [CrossRef]

24. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 4 February 2017; pp. 4278–4284.
25. Li, W.N.; Zhang, Z.X. Research on Knowledge Base Error Detection Method Based on Confidence Learning. *Data Anal. Knowl.* **2021**, *5*, 1–9. [CrossRef]
26. Zhang, M.H.; Xie, H.; Zhang D.F.; Ge, J.M.; Cheng, H.Y. Handling label noise in Air Traffic Complexity Evaluation Based on Confidence Learning and XGBoost. *Trans. Nanjing Univ. Aeronaut.* **2021**, *37*, 936–946. [CrossRef]
27. Deng, S. Hyper-parameter optimization of CNN based on improved Bayesian optimization algorithm. *Appl. Res. Comput.* **2019**, *36*, 1984–1987. [CrossRef]
28. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
29. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Honolulu, HI, USA, 21–26 July 2017; pp.1800–1807. [CrossRef]
30. Goodfellow, I.J.; Erhan, D.; Carrier, P.L.; Courville, A.; Mirza, M.; Hamner, B.; Bengio, Y. Challenges in representation learning: A report on three machine learning contests. In Proceedings of the International Conference on Neural Information Processing, Daegu, Korea, 3–7 November 2013; pp. 117–124. [CrossRef]
31. Barsoum, E.; Zhang, C.; Ferrer, C.C.; Zhang, Z. Training deep networks for facial expression recognition with crowd-sourced label distribution. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016; pp. 279–283. [CrossRef]
32. Lucey, P.; Cohn, J.F., Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101. [CrossRef]
33. Li, S.; Deng, W.; Du, J. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild. In Proceedings of the IEEE Conference On computer Vision and Pattern Recognition(CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2584–2593. [CrossRef]
34. Lundqvist, D.; Flykt, A.; Öhman, A. Karolinska directed emotional faces. *Cogn. Emot.* **1998**. [CrossRef]
35. Lyons, M.; Akamatsu, S.; Kamachi, M.; Gyoba, J. Coding facial expressions with Gabor wavelets. In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998; pp. 200–205. [CrossRef]
36. Miao, M.M.; Xu, B.G.; Hu W.J.; Wang, A.M.; Song, A.G. Emotion EEG recognition based on the adaptive optimized spatial-frequency differential entropy. *Chin. J. Sci. Instrum.* **2021**, *3*, 221–230. [CrossRef]