

Article

Not peer-reviewed version

Emotion Detection Based on Consecutive Facial by Combining CNN and LSTM Implemented on FPGA Chip

[Shing-Tai Pan](#)* and Han-Jui Wu

Posted Date: 22 April 2024

doi: 10.20944/preprints202404.1307.v1

Keywords: Consecutive facial images; Emotion recognition; convolution neural networks (CNNs); Long short-term memory (LSTM); Field-programmable gate array (FPGA)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Emotion Detection Based on Consecutive Facial by Combining CNN and LSTM Implemented on FPGA Chip

Shing-Tai Pan * and Han-Jui Wu

Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan, R.O.C.; m1105505@mail.nuk.edu.tw

* Correspondence: span@nuk.edu.tw

Abstract: This paper proposed emotion recognition methods for consecutive facial images, and implements the inference of neural network model on a field-programmable gate array (FPGA). The proposed emotion recognition methods are based on a neural network model architecture that combines convolution neural networks (CNNs), long short-term memory (LSTM), and fully connected neural networks (FCNNs), called CLDNN or ConvLSTM-FCN. This type of neural network model can analyze the local feature sequences obtained through convolution of data, making it suitable for processing time-series data such as consecutive facial images. In this paper, sequences of facial images are sampled from videos corresponding to various emotional state of subjects. The sampled images are then pre-processed with the processes includes facial detection, greyscale conversion, resize, and data augmentation if necessary. The 2-D CNN in ConvLSTM-FCN is used for feature extraction for these pre-processed facial images. These sequences of facial image features are time sequences with dependent properties between the elements within them. The LSTM is then used to model these time sequences followed by fully connected neural networks (FCNNs) for classification. The proposed consecutive facial emotion recognition method achieves an average recognition rate of 99.51% on RAVDESS dataset, 87.80% on BAUM-1s dataset and 96.82% on eNTERFACE'05 data set, using 10-fold cross-validation on the PC. Some comparisons of recognition accuracies between the proposed method and the other existing related works are conducted in this paper. According to the comparisons, the proposed emotion recognition methods outperform the existing related researches. The proposed emotion recognition methods are then implemented on an FPGA chip using the neural network model inference algorithms in this paper, and the accuracies of the experiments conducted on the FPGA chip are identical to those obtained on the PC. This verifies that the proposed neural network model implemented on the FPGA chip performs well.

Keywords: consecutive facial images; emotion recognition; convolution neural networks (CNNs); long short-term memory (LSTM); field-programmable gate array (FPGA)

1. Introduction

In recent years, with the continuous development of artificial intelligence (AI) technology, image and speech emotion recognition technologies have become one of the important applications. The AI technologies have significant application in many fields, such as driver monitor [1], fraud detection [2], medical care [3] and education [4]. In many application scenarios, embedded devices play an important role, so it has become particularly important to achieve neural network model inference on embedded devices such as FPGA chips.

For data with temporal dependencies, such as consecutive images, speech, and natural language, using temporal model can enhance the performance of the model. Therefore, the model architecture used in this paper first applies CNNs to extract local features from the data. To enhance the model's

learning performance on time-series level, LSTM neural networks are then used to model the feature sequences. Finally, FCNN is employed for results classification.

Emotion recognition is a technology that utilizes various signals, such as facial images and speech, to analyze and identify emotional states. In the case of using facial images for emotion recognition, the current main approach is to extract feature values from the image using 2-D convolutional neural networks, followed by prediction and classification, as demonstrated in papers [5,6]. Emotion recognition technologies are crucial for understanding and analyzing human emotions, as it can assist us in better understanding the emotional state of the user and provide more accurate and human-like solutions for related application areas [7]. Unlike speech recognition, where data preprocessing is crucial and involves steps such as audio cropping, noise reduction, and feature extraction, image recognition requires less preprocessing since raw images are typically used with minimal manipulation. Speech recognition heavily depends on high-quality preprocessing to handle variability, while image recognition can often learn features directly from raw images. Many studies have shown that the quality of data preprocessing and feature extraction significantly affects the performance of speech emotion recognition models. Experimental results in paper [8] show that the accuracy of machine learning models is affected by using different speech features for training. Consequently, in order to obtain accurate and stable performance of artificial intelligent models on emotion recognition, this paper focuses on the emotion recognition based on consecutive facial images.

However, the accuracies of facial recognition mainly depend on the quality of the input signals, such as the contrast, brightness, and focus of images for image signals. Consequently, it may significantly reduce the reliability of emotion recognition if any of the elements in the image fails to meet the standard requirements, such as overexposure or blurring. For the above reasons, this paper proposes consecutive facial pre-processing and recognition methods, to complete more suitable emotion recognition methods based on the user's environment.

1.1. Field-Programmable Gate Array

The implementation of intelligent systems such as automatic emotion recognition technology in embedded systems faces many challenges, e.g., real-time requirements, resource constraints, and low power consumption requirements. Therefore, when implementing facial emotion recognition technology, the choice of hardware platform is crucial to the system's efficiency and performance. Consequently, the FPGA (Field-Programmable Gate Arrays) have become one of the ideal platforms for implementing emotion recognition due to their highly customizable features, parallel processing capabilities, and low power consumption.

The FPGA is a reconfigurable embedded device commonly used in digital logic and digital signal processing applications. FPGA's high flexibility and programmability enable their wide application in various fields, including IC testing [9], embedded systems [10–12], and the IoT (Internet of Things) [13].

The features of FPGA include:

- A. **Reconfigurability:** FPGAs are reconfigurable [14] and can define their digital logic circuits through programming, allowing developers to redesign the FPGA's functions according to application requirements repeatedly.
- B. **High parallel processing capability:** FPGAs have multiple independent logic circuits and data paths that can run in parallel, enabling them to efficiently perform parallel processing for multiple tasks and hence provide high-performance computing power.
- C. **Low latency and high-frequency operation:** Due to the fact that FPGA's logic circuits composed of gate arrays and have high optimization capabilities, it can achieve low latency and high-frequency operation. This makes it ideal for applications requiring high-speed processing.
- D. **Customizability:** FPGAs are highly flexible in customization and can be designed and optimized according to application requirements. This includes design of logic circuits, data paths, memory, and interfaces.
- E. **Software and hardware co-design:** FPGAs provide the ability to co-design software and hardware on a single chip [15]. This provides higher system integration and performance.

- F. Suitable for rapid development and testing: FPGAs have a rapid development cycle. Developers can quickly develop and test them within a shorter period [16].

1.2. Experimental Protocol

This paper utilizes two deep learning frameworks, TensorFlow and Keras, to train emotion recognition models for consecutive facial images signals on a PC. The parameters of the models are transferred to an FPGA chip. The neural network model inference algorithms are used to simulate computation of model inference in the deep learning frameworks and then obtain the final classification results. For consecutive facial emotion recognition, this paper dynamically captures 30 frames of images from a video as the consecutive image data. The facial images are extracted by using the open-source face detection model from OpenCV. The CLDNN (Convolutional Long Short-Term Memory Fully Connected Deep Neural Networks) model architecture proposed in paper [17] is used to build and train the ML model. The trained model is then deployed on the FPGA chip for model inference.

2. Related Works

In this section, the CLDNN model [17] is introduced. Moreover, some related researches of consecutive facial emotion recognition, and the implementation of model inference on FPGA are also explored in this section.

2.1. CLDNN Model Architecture

The paper [17] proposes a neural network model architecture called Convolutional Long Short-Term Memory Fully Connected Deep Neural Networks, referred to as CLDNN or ConvLSTM-FCN, for processing time series data such as audio signals, consecutive images, and natural languages. The CLDNN model is composed of convolutional neural networks (CNNs), LSTM neural networks, and fully connected neural networks of DNNs. The CNNs and LSTM neural networks analyze and calculate local features and sequential patterns in the data, respectively. The fully connected layers are then classifying the prediction results. Since consecutive images have time-series characteristics, they are suitable for training with recurrent neural networks (RNNs) such as LSTM. According to [17], the CLDNN model outperforms those using only CNN or LSTM models in terms of the accuracies of the models.

Therefore, this paper applies and modifies the CLDNN model architecture for emotion recognition, combining CNNs, LSTM neural networks, and DNNs to build the ML model. The Figure 1 shows the architecture of the CLDNN model.

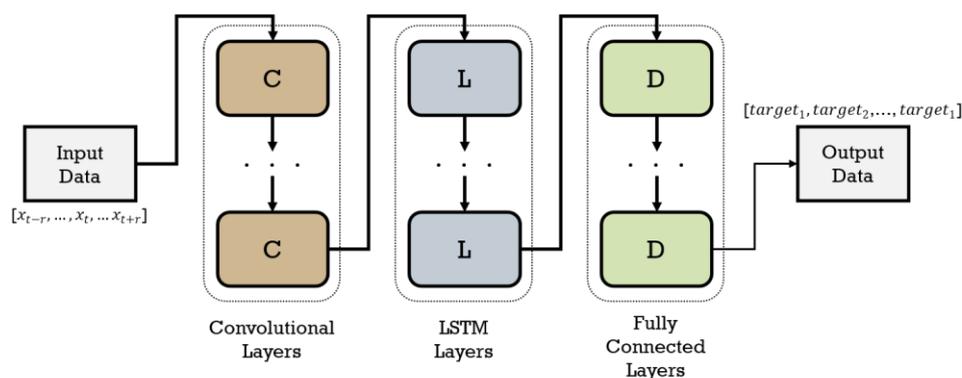


Figure 1. The CLDNN model architecture, composed of multiple CNNs, LSTM neural networks, and DNNs.

2.2. Consecutive Facial Emotion Recognition

Paper [18] proposes a method for consecutive facial emotion recognition. Firstly, the frame interval is calculated based on the average duration of all video files in the database. Frames are then extracted based on this interval to obtain a segment of consecutive image data. Next, facial landmark detection (68 points) [19] is used to identify the positions of the facial features in the images, and the Euclidean distances between these points are used as the feature value. The image feature values of the same image of consecutive image data are concatenated into a feature sequences. Finally, these feature sequences are used as inputs to train an LSTM model. This paper achieves an average accuracy of 98.9%.

A multimodal emotion recognition model that combines video data and audio data is proposed in [20]. For the emotion recognition of consecutive facial image, the authors randomly extract 30 images from video data. Then, using the MTCNN (multi-task cascaded convolutional network) proposed in paper [21], they extract the facial parts from all images in each image. The images are then reshaped uniformly to (160, 160) color images as data features. Hence, the input of the facial emotion recognition model is a data sequence of shape (N, 30, 160, 160, 3), where N is the number of video data. In the speech emotion recognition model, the log Mel-spectrogram is extracted from every segment of the speech signals and used as input data of ML model. In order to obtain the log Mel-spectrogram, the authors utilize 94 Mel-filter banks for each speech signal, with a 40ms Hanning window and 10ms overlapping. This result in a representation with dimensions of 94. After performing cross-validation with the ResNet-50 model, the paper find that the consecutive facial emotion recognition model achieves an accuracy of 95.49% and the speech emotion recognition model achieves an accuracy of 75.61%. The fusion model combining consecutive facial image and audio achieves an accuracy of 97.57%.

In paper [22], the authors propose a method that uses facial landmarks [19] to detect facial regions in images. The detected facial regions are then converted into grayscale images. The authors extract 32 features from the resulting images using Gabor filters. The features are then combined with the 68 positions of the facial landmarks. After processing all frames in the video, the authors obtain 2176 (32×68) features for each facial image. The proposed method in [22] achieves an accuracy of 96.53% on testing. Paper [23] proposes a multimodal emotion recognition model that combines consecutive facial image and speech. Regarding the consecutive facial emotion recognition model, the authors use the InceptionV3 model to extract feature values from single images, which are then sent to LSTM for time-series learning. This results in an accuracy of 94% on RAVDESS database. In terms of speech emotion recognition, the authors extract MFCCs from the speech as feature values and train them using the CLDNN model. The experimental results show that the speech emotion recognition model proposed in [23] achieves an accuracy of 82%. The fusion model combining consecutive facial image and audio achieves an accuracy of 96%.

3. Facial Emotion Recognition Methods and Parameter Setting

This section will introduce the training and testing process of the proposed consecutive facial emotion recognition method in this paper. Some experiments and the experimental environment will be also introduced in this section.

3.1. CLDNN Model

When conducting experiments using the RAVDESS database, this paper tests different model performances by employing both LSTM and GRU as temporal neural networks. The experimental results demonstrate that in the consecutive facial emotion recognition model, using LSTM to process temporal data achieves an accuracy of 99.51%, showing a 1.65% improvement compared to using GRU. As LSTM neural networks yield higher accuracy in consecutive facial emotion recognition models, and using an LSTM model for model inference on an FPGA chip does not show a significant increase in inference time compared to using a GRU model, this paper opts to use LSTM as the modeling approach.

Table 1. The difference in model accuracy and execution time on an FPGA chip between models built using LSTM and GRU.

CNN + LSTM + DNN		CNN + GRU + DNN	
Execution Time on FPGA	Accuracy	Execution Time on FPGA	Accuracy
11.70 sec	99.51%	11.67 sec	97.86%

3.2. Experimental Environment for Model Training on PC

This paper performs data preprocessing on the CPU of a PC, and trains a CLDNN model for consecutive facial emotion recognition using GPU. The trained parameters for CLDNN model are transferred to an FPGA chip for inference of the model. Table 2 shows the hardware specifications and software environment used for training the model on the PC.

Table 2. The experimental environment of the proposed emotion recognition methods.

Experimental environment	
CPU	Intel® Core™ i7-10700 CPU 2.90GHz Manufacturer: Intel Corporation, Santa Clara, CA, USA
GPU	NVIDIA GeForce RTX 3090 32GB Manufacturer: NVIDIA Corporation, Santa Clara, CA, USA
IDE	Jupyter notebook (Python 3.7.6)
Deep learning frameworks	TensorFlow 2.9.1, Keras 2.9.0

The experiments conducted in this paper uses 10-fold cross-validation method for training and testing. In order to ensure the reliability of machine learning model, the Scikit-learn package is conducted to randomly select 10% of the dataset as the data for testing, and the remaining 90% data are randomly divided into train data and validation data at a ratio of 9:1. Figure 2 illustrates the data proportions of the training set, validation set, and testing set.

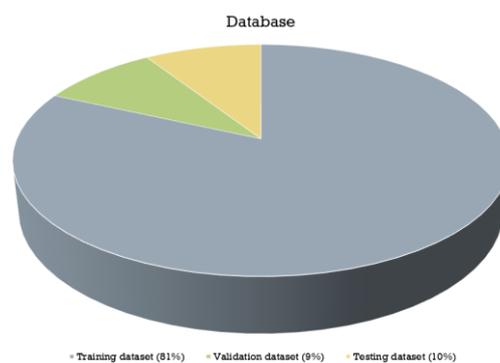


Figure 2. The database is split into training dataset, validation dataset and testing dataset.

3.3. Consecutive Facial Emotion Recognition

In this subsection, a detailed description of the databases, data preprocessing methods, neural network model architecture, and parameter settings for proposed consecutive facial emotion recognition method will be provided. Then, the experimental results of the proposed consecutive facial emotion recognition method will be compared with other related literature.

3.3.1. Databases

This paper uses the RAVDESS [24], BAUM-1s and eNTERFACE'05 databases for training and testing the proposed consecutive facial emotion recognition model.

3.3.1.1. RAVDESS

RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) is a database of consecutive emotional image and speech created by Ryerson University in Canada. It includes emotional expressions in both visual and auditory forms from different actors and can be used for research in the fields of emotion recognition, emotion analysis, audio processing, etc.

This database contains videos and voice data from 24 actors (12 males and 12 females) from Canada. Each actor records videos and audio in 8 different emotional states (Angry, calm, disgust, fear, happy, neutral, sad, and surprised), by speaking and singing. For each emotional state, the actor performs multiple times to provide data with different emotional intensities and expressions. The RAVDESS database has a total of 2,452 consecutive image and audio data, with the distribution and the amount of data for each emotional state shown in the Table 3.

Table 3. The quantity and proportion of data for each emotion in the RAVDESS database.

Label	Number of Data	Proportion
Angry	376	15.33%
Calm	376	15.33%
Disgust	192	7.83%
Fear	376	15.33%
Happy	376	15.33%
Neutral	188	7.39%
Sad	376	15.33%
Surprised	192	7.83%
Total	2,452	100%

3.3.1.2. BAUM-1s

The BAUM-1s database was created by the Department of Electrical and Electronics Engineering at Bahcesehir University in Turkey. This database consists of samples of videos and voice recordings, recorded by 31 professional actors. In this study, we utilize six emotional categories: angry, disgust, fear, happy, sad, and surprised, to conduct experiments of consecutive facial and speech emotion recognition. The data number and proportion of each emotion label in the BAUM-1s database is shown in Table 4.

Table 4. The quantity and proportion of data for each emotion in the BAUM-1s database.

Label	Number of Data	Proportion
Angry	59	10.85%
Disgust	86	15.81%
Fear	38	6.99%
Happy	179	32.90%
Sad	139	25.55%
Surprised	43	7.90%
Total	544	100%

3.3.1.3. eINTERFACE'05

The eINTERFACE'05 database was recorded by 44 professional actors and consists of six different emotion categories, with different sentences expressing each category. This database includes both videos and speech samples, making it commonly used in various multimodal emotion recognition research. The data number and proportion of each emotion label in the eINTERFACE'05 database is shown in Table 5.

Table 5. The quantity and proportion of data for each emotion in the eINTERFACE'05 database.

Label	Number of Data	Proportion
Angry	211	16.71%
Disgust	211	16.71%
Fear	211	16.71%
Happy	208	16.47%
Sad	211	16.71%
Surprised	211	16.71%
Total	1,263	100%

3.3.2. Pre-Processing

This paper refers to [18] and adopts a dynamic sampling method to capture 30 frames of video file to represent a consecutive image data. The frame intervals are determined by dividing the total number of frames in each video data by 30. Then the frame intervals are used as a spacing to capture 30 frames in each video. The Figure 3 shows the process of the dynamic capture of the consecutive images.

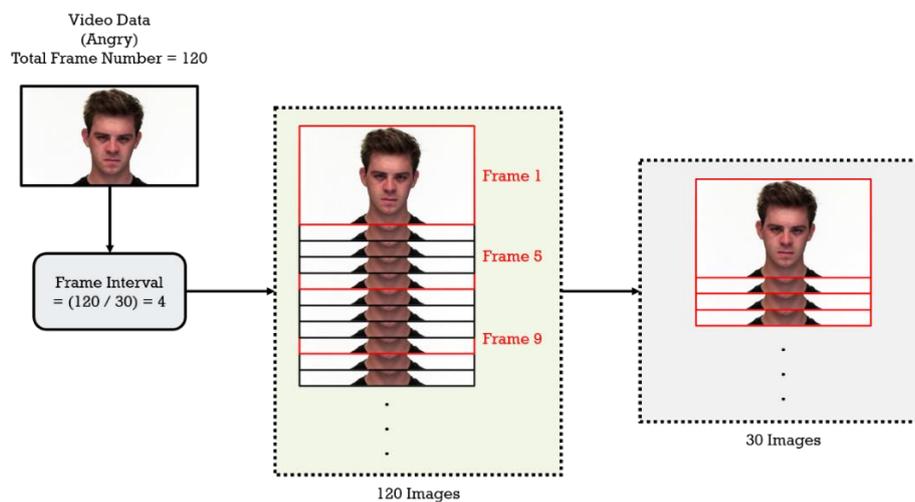


Figure 3. Calculates the frame interval based on the total number of frames in a video file, and captures 30 frames of video using the frame interval to represent a consecutive image data.

The OpenCV facial detection model is then used to locate the coordinates of the faces in the images, and any unnecessary data outside the coordinates is discarded. In order to reduce the computational load during model inference on the FPGA chip, the color images are converted into grayscale images. Moreover, all facial images are resized to a uniform size of 100×100 pixels. The Figure 4 shows the pre-processing flowchart for the facial emotion recognition method proposed in this paper.

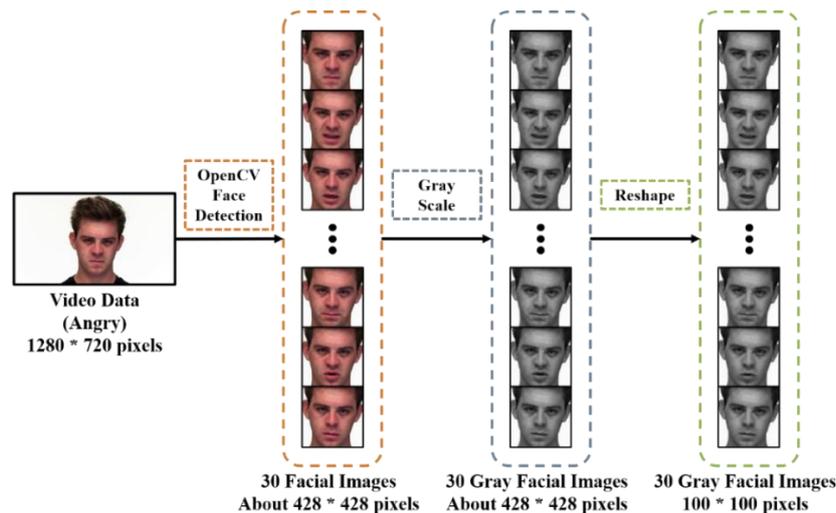


Figure 4. Pre-processing of the videos in proposed facial emotion recognition method.

3.3.2.1. Facial Detection

Facial detection is necessary before training facial emotion recognition models. Nowadays, there are many open-source facial detection models available, such as facial landmark detection and the MTCNN model proposed in [20]. In this paper, Haar cascade frontal face detection model provided by OpenCV is used, which is also used for facial detection in many related researches, such as the papers [25,26]. This facial detection model can locate the starting coordinates (x, y) of the facial region in the images, and the corresponding w (width) and h (height). The facial image within the region is then extracted, while unnecessary information outside of the region are discarded. The process of the facial image extraction is shown in the Figure 5.

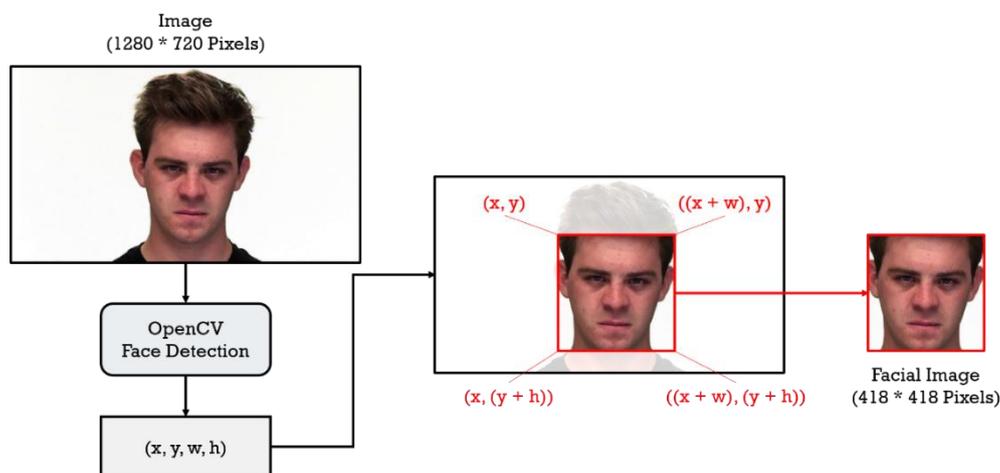


Figure 5. Uses the OpenCV face detection model to capture the facial part in the images.

3.3.2.2. Grayscale Conversion

Paper [27] compares the performance of color and grayscale images in facial recognition. It finds that in some cases, using grayscale images performs facial recognition better than using color images. Moreover, using color images increases the computations and parameters during recognition. Overall, there are advantages and disadvantages to using color or grayscale images for facial recognition. Considering the restriction in computational requirements of performing neural network model inference on embedded devices, this paper converts facial images to grayscale as the inputs to the neural network model.

3.3.2.3. Resize

As the OpenCV face detection model is used to extract facial images, the sizes of facial images are different. Therefore, all facial images need to be resized into a unified size of 100×100 pixels before being fed into the neural network model for training and testing.

3.3.3. Experiments

This paper proposes a consecutive facial emotion recognition method based on the CLDNN model architecture. The LFLBs (Local Feature Learning Blocks) with a CNN in the neural networks is used as the main body to extract local features of input facial images. Every 30 of image features are concatenated into consecutive image feature sequences. Then, the LSTM layer is applied to strengthen the learning of model on time-series. Finally, a fully connected layer is used for classification to output the recognition result.

3.3.3.1. Local Feature Learning Block

The LFLB is a type of neural network module used in deep learning, which is primarily used to extract local feature values from input data. LFLB consists of many sub-modules, such as convolutional layers, pooling layers, and fully connected layers, etc. It can extract features from different regions of the data and share weights to reduce the number of model parameters and improve the model's generalization ability. Some researches of facial emotion recognition also use LFLBs to extract local features from image data, such as the papers [28,29].

In this paper, a 2-D convolutional layer, a batch normalization layer, and a max pooling layer are used as the LFLB for extracting local features from the facial image data. Due to the local perception characteristics of the CNNs, it is suitable for calculating the local features of the image data and generating feature maps. The batch normalization layer can normalize the input data, scales the values to a specific range 0 ~ 1. It helps the model train the data faster and reduces the possibility of overfitting. The max pooling layer is mainly used to limit the sizes of the feature map. It helps preserve the most influential feature values (maximum values). This paper uses the zero-padding method in the design of sub-modules of the LFLB to ensure that the size of the feature map is the same as that of the input data, and to avoid data loss at the edges. Moreover, ReLU (rectified linear units) is used as the non-linear activation function of the LFLB. The Figure 6 shows the schematic diagram of the LFLB applied in this paper.

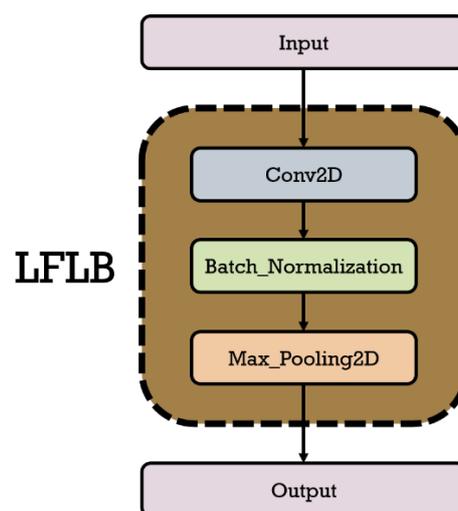


Figure 6. The LFLB used in this paper, including a 2-D convolutional layer, a batch normalization layer, and a max pooling layer.

3.3.3.2. Training Process and Parameters

When conducting experiments using the RAVDESS database, setting the number of local feature learning blocks to 6 results in 64 local features per image and achieves the highest accuracy of 99.51%. However, from Table 6, it can be seen that when increasing the number of local feature learning blocks to 7, the output of local features per image decrease to 16, leading to a sharp decline in accuracy, reaching only 32.68%. Therefore, this study sets the number of local feature learning blocks in the consecutive facial emotion recognition model to 6.

Table 6. The impact of using different numbers of LFLBs on model accuracy.

Number of LFLBs	Number of Local Features	Accuracy
3	2704	28.89%
4	784	52.96%
5	256	88.58%
6	64	99.51%
7	16	32.68%

The proposed strategy in this paper for consecutive facial emotion recognition is then described as follows. Firstly, local features will be extracted from single facial images using LFLB, and then concatenates the resulting 30 feature maps into a sequence. The sequence represents the features sequences of consecutive facial images. Then, the sequences are flattened and reshaped into the shape of [time step, data length], before being fed into the LSTM layer. Next, the output data are normalized after learning on time-series level in the LSTM layer. Finally, softmax method is performed when using a fully connected layer to extract the emotion label corresponding to the maximum output value for the sequence as the predicted result. The training process of the proposed consecutive facial emotion recognition method is shown in Figure 7.

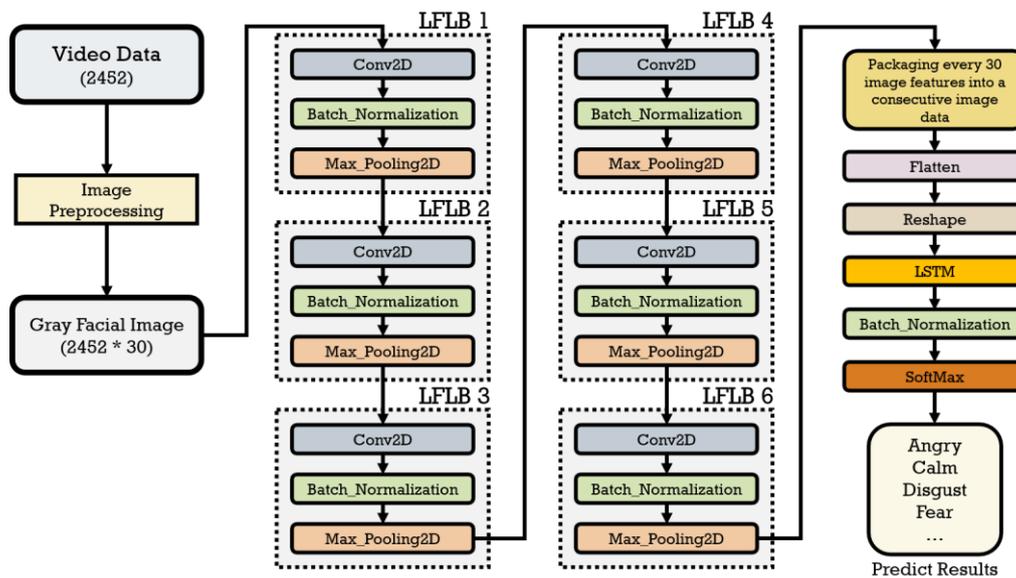


Figure 7. The flowchart of the proposed method for consecutive facial emotion recognition.

The number of memory units in LSTM layers can affect the training performance of machine learning models and needs to be adjusted according to the type of data and the length of the data sequences. In this paper, for the consecutive facial recognition CLDNN model, cross-validation is used to test the effect of different numbers of memory units in the LSTM layer for finding the best model's accuracy. The number of memory units is set up from 5 to 50, with an increment of 5. For

applying 10-fold cross-validation method, the experimental results show that the highest accuracy of 99.51% is achieved when the number of memory units in the LSTM layer is 20. Therefore, this paper sets the memory units of the LSTM layer in the CLDNN model for consecutive facial emotion recognition to be 20. Figure 8 shows the experimental results for the relationship between the number of memory units in the LSTM layer and the accuracy of the proposed consecutive facial emotion recognition model. The proposed model's parameters are listed in the Table 7.

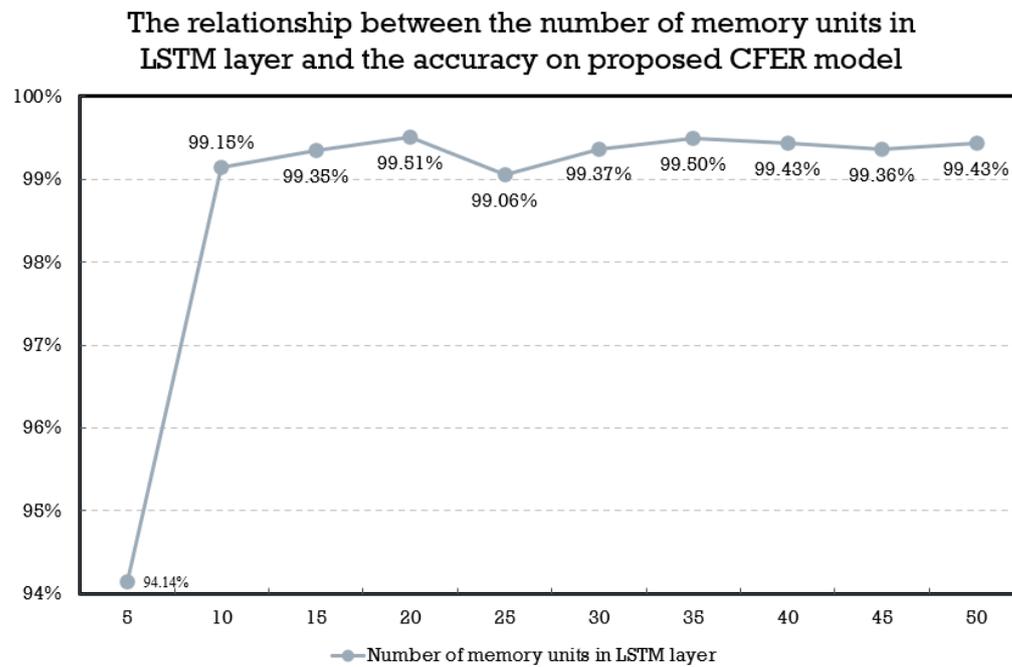


Figure 8. The relationship between the number of memory units in the LSTM layer and the accuracy of the proposed consecutive facial emotion recognition model.

Table 7. The parameters of the proposed CLDNN model for consecutive facial emotion recognition.

Model Architecture		Information
LFLB 1	Conv2d (Input) Batch_normalization Max_pooling2d	Filters = 16, Kernel_size = 5, Strides =1 Pool_size = 5, Strides = 2
LFLB 2	Conv2d Batch_normalization Max_pooling2d	Filters = 16, Kernel_size = 5, Strides =1 Pool_size = 5, Strides = 2
LFLB 3	Conv2d Batch_normalization Max_pooling2d	Filters = 16, Kernel_size = 5, Strides =1 Pool_size = 5, Strides = 2
LFLB 4	Conv2d Batch_normalization Max_pooling2d	Filters = 16, Kernel_size = 5, Strides =1 Pool_size = 5, Strides = 2
LFLB 5	Conv2d Batch_normalization Max_pooling2d	Filters = 16, Kernel_size = 3, Strides =1 Pool_size = 3, Strides = 2
LFLB 6	Conv2d Batch_normalization Max_pooling2d	Filters = 16, Kernel_size = 3, Strides =1 Pool_size = 3, Strides = 2
Concatenation		Packages every 30 image features into a consecutive facial image feature sequence

Flatten	
Reshape	
LSTM	Unit = 20
Batch_normalization	
Dense (Output)	Unit = 8, Activation = "softmax"

4. Experimental Results

In this section, the experimental results on RADESS, BAUM-1s and IEMOCAP database will be shown and discussed. Then, some comparison between the results of consecutive facial emotion recognition method proposed in this paper with those of other related researches will also be made in this section.

4.1. Experiments on RAVDESS Database

For RAVDESS database, the pre-processing and emotion recognition methods in Section 3 are applied for this experiments. According to the experiments, the proposed consecutive facial emotion recognition method achieves an average accuracy of 99.51% on testing through 10-fold cross-validation experiments. The experimental results for the loss and accuracy of each fold of training, validation, and testing are shown in the Table 8.

Table 8. The loss and accuracy of the proposed consecutive facial emotion recognition method during training, validation, and testing.

	Training		Validation		Testing	
	Loss	Acc	Loss	Acc	Loss	Acc
Fold 1	0.0308	1.0000	0.0749	1.0000	0.4998	0.9919
Fold 2	0.0366	1.0000	0.0745	1.0000	0.4517	1.0000
Fold 3	0.0192	1.0000	0.0415	1.0000	0.1363	1.0000
Fold 4	0.0206	1.0000	0.0428	1.0000	0.2667	0.9959
Fold 5	0.0369	1.0000	0.0593	1.0000	0.2978	0.9919
Fold 6	0.0310	1.0000	0.0703	1.0000	0.4527	0.9959
Fold 7	0.0179	1.0000	0.0382	1.0000	0.1913	0.9959
Fold 8	0.0118	1.0000	0.0206	1.0000	0.0459	0.9959
Fold 9	0.0225	1.0000	0.0378	1.0000	0.1757	0.9919
Fold 10	0.0348	1.0000	0.0769	1.0000	0.2238	0.9919
Average	0.0262	1.0000	0.0536	1.0000	0.2741	0.9951

Besides, the normalized confusion matrix obtained from cross-validation is shown in Figure 9. The accuracy, precision, recall, and F1-score for every emotion label can be calculated using equations from (1) to (4) with the confusion matrix in Figure 9, and the results are presented in Table 9.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} \quad (4)$$

where TP , FP , TN and FN is the number of *True Positive*, *False Positive*, *True Negative* and *False Negative*, respectively.

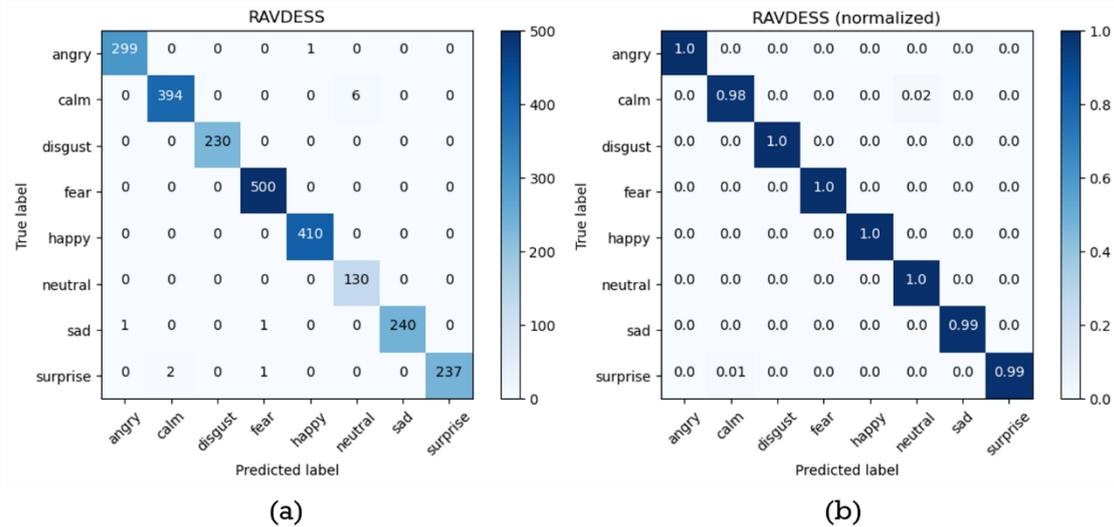


Figure 9. The confusion matrix before and after normalized, (a) and (b) respectively, obtained from 10-fold cross-validation of the consecutive facial emotion recognition method proposed in this paper.

Table 9. The accuracy, precision, recall, and F1-score of each emotion calculated by the confusion matrix of the proposed consecutive facial emotion recognition method.

Label	Accuracy	Precision	Recall	F1-score
Angry	0.9992	0.9967	0.9967	0.9967
Calm	0.9967	0.9949	0.9850	0.9899
Disgust	1.0000	1.0000	1.0000	1.0000
Fear	0.9992	0.9960	1.0000	0.9980
Happy	0.9996	0.9976	1.0000	0.9988
Neutral	0.9976	0.9559	1.0000	0.9774
Sad	0.9992	1.0000	0.9917	0.9959
Surprised	0.9988	1.0000	0.9875	0.9937
Average	0.9988	0.9926	0.9951	0.9938

This paper proposes a consecutive facial emotion recognition method and conducts experiments on the RAVDESS database. The experimental results in Table 9 show that this method performs well in terms of accuracy, precision, recall and F1-score, only performs a little poorly in predicting the emotion label of “neutral”. This is because that the facial features of the emotion label “calm” are similar to those of the emotion label of “neutral”. The degree of expression variation between these two emotions is small. This makes it more likely to be predicted incorrectly. However, our proposed method achieves an overall performance of average value over 99% in the accuracy, precision, recall and F1-score on the RAVDESS database. This verify the proposed model is precise and reliable. Based on these results, the method proposed in this paper is a feasible emotion recognition method and will have a stable performance in practical applications.

Table 10 presents a comparison of the experimental results of the proposed consecutive facial emotion recognition method with other related researches on RAVDESS database. It can be seen that the proposed method achieves an accuracy rate of 99.51%, which is higher than the methods proposed in other related researches.

Table 10. Comparison of the experimental results of cross-validation for the proposed consecutive facial emotion recognition method with other related researches on the RAVDESS database.

Method	Classes	Accuracy
E. Ryumina, et al. [18]	8	98.90%
F. Ma, et al. [20]	6	95.49%
A. Jaratrotkamjorn, et al. [22]	8	96.53%
Z. Q. Chen, et al. [23]	7	94%
Proposed model	8	99.51%

4.2. Experiments on BAUM-1s Database

Due to the data imbalance in the BAUM-1s database in which the emotion label “fear” accounts for only 6.99% of the data, and the emotion label “surprised” accounts for only 7.90%, the lower amount of learning data for these emotion labels during model training will affect the overall accuracy of the model. Therefore, this paper utilizes up-sampling to balance the data within the training dataset by replicating data instances from the minority classes, thus ensuring an equal number of data samples for all emotion categories. Figure 10 shows the distribution of the BAUM-1s database before and after up-sampling.

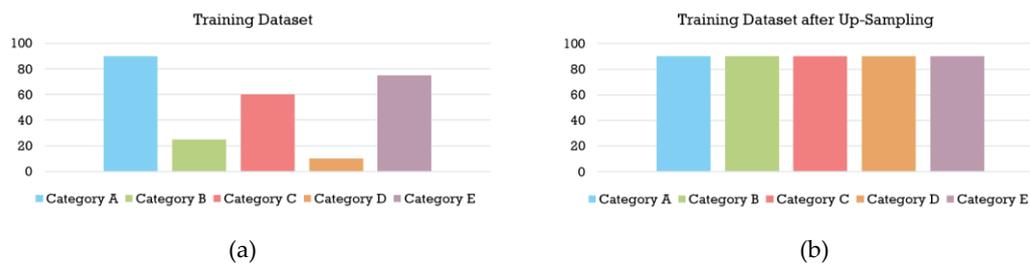


Figure 10. Up-sampling for BAUM-1s database. (a) before up-sampling (b) after up-sampling.

Applying the pre-processing and emotion recognition methods in Section 3 for this experiments, the proposed consecutive facial emotion recognition method in Figure 7 achieves an average accuracy of 87.80% on BAUM-1s database through 10-fold cross-validation experiments. The experimental results for the loss and accuracy of each fold of training, validation, and testing are shown in the Table 11.

Table 11. The loss and accuracy of the proposed consecutive facial emotion recognition method during training, validation, and testing on BAUM-1s database.

	Training		Validation		Testing	
	Loss	Acc	Loss	Acc	Loss	Acc
Fold 1	0.0505	0.9137	0.9556	0.9327	0.5868	0.8600
Fold 2	0.0623	0.9951	0.2346	0.9405	0.8079	0.8600
Fold 3	0.0852	0.9764	0.5582	0.9428	0.6573	0.8800

Fold 4	0.0705	0.9553	0.4763	0.9053	0.4489	0.8600
Fold 5	0.0792	0.9202	0.8127	0.9492	0.5791	0.8400
Fold 6	0.0801	0.9015	0.6274	0.9266	0.4527	0.8600
Fold 7	0.0928	0.9589	0.3468	0.9134	0.4802	0.9200
Fold 8	0.0893	0.9668	0.7501	0.9431	0.6054	0.9000
Fold 9	0.0934	0.9015	0.4307	0.9519	0.4205	0.9400
Fold 10	0.0683	0.9907	0.6919	0.9203	0.6596	0.8600
Average	0.0772	0.9480	0.5884	0.9326	0.5698	0.8780

Besides, the original and normalized confusion matrix obtained from cross-validation experiment on BAUM-1s database is shown in Figure 11 and the accuracy, precision, recall, and F1-score for each emotion are as shown in Table 12. Moreover, Table 13 shows the comparison of the accuracy of the proposed method for consecutive facial emotion recognition and those of other related researches for BAUM-1s database. Again, according to the comparison in Table 13, the proposed strategy for consecutive facial emotion recognition outperforms the methods in the other papers.

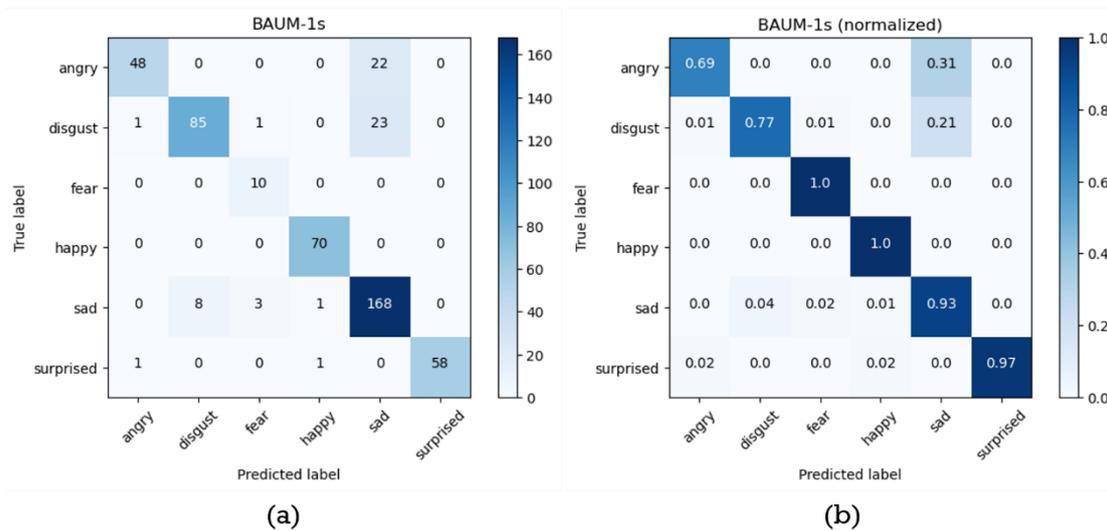


Figure 11. The confusion matrix before and after normalized, (a) and (b) respectively, obtained from 10-fold cross-validation of the consecutive facial emotion recognition method proposed in this paper on BAUM-1s database.

Table 12. The accuracy, precision, recall, and F1-score of each emotion of the proposed consecutive facial emotion recognition method on BAUM-1s database.

Label	Accuracy	Precision	Recall	F1-score
Angry	0.9520	0.9600	0.6857	0.8000
Disgust	0.9340	0.9140	0.7727	0.8374
Fear	0.9920	0.7143	1.0000	0.8333
Happy	0.9960	0.9722	1.0000	0.9859
Sad	0.8860	0.7887	0.9333	0.8550

Surprised	0.9960	1.0000	0.9667	0.9831
Average	0.9593	0.8915	0.8931	0.8825

Table 13. Comparison of the results of cross-validation for the proposed consecutive facial emotion recognition method with other related researches on the BAUM-1s database.

Paper	Classes	Accuracy
F. Ma, et al. [20]	6	64.05%
B. Pan, et al. [30]	6	55.38%
P. Tiwari [31]	8	77.95%
Proposed model	6	87.80%

4.3. Experiments on eINTERFACE'05 Database

Similar to the experiments conducted in Section 4.1 and Section 4.2 for the databases BAUM-1s database and RAVDESS database, respectively, the proposed consecutive facial emotion recognition method in Figure 7 achieves an average accuracy of 96.82% for the experiments on eINTERFACE'05 database through 10-fold cross-validation method. The experimental results for loss and accuracy of each fold of training, validation, and testing are shown in the Table 14.

Table 14. The loss and accuracy of the proposed consecutive facial emotion recognition method during training, validation, and testing on eINTERFACE'05 database.

	Training		Validation		Testing	
	Loss	Acc	Loss	Acc	Loss	Acc
Fold 1	0.0437	0.9752	0.2978	0.9563	0.4727	0.9603
Fold 2	0.0389	0.9747	0.1925	0.9632	0.2063	0.9683
Fold 3	0.0471	0.9968	0.3194	0.9491	0.2167	0.9762
Fold 4	0.0423	0.9604	0.3751	0.9578	0.4335	0.9603
Fold 5	0.0312	0.9823	0.3530	0.9684	0.5808	0.9603
Fold 6	0.0430	0.9521	0.2496	0.9467	0.3345	0.9524
Fold 7	0.0488	0.9816	0.1693	0.9546	0.4768	0.9683
Fold 8	0.0495	0.9873	0.3847	0.9619	0.3880	0.9762
Fold 9	0.0456	0.9768	0.2319	0.9443	0.2920	0.9841
Fold 10	0.0345	0.9765	0.3890	0.9691	0.4589	0.9762
Average	0.0424	0.9763	0.2962	0.9571	0.3860	0.9682

Besides, the original and normalized confusion matrix obtained from cross-validation experiment on eINTERFACE'05 database is shown in Figure 12 and the accuracy, precision, recall, and F1-score for each emotion are as shown in Table 15. Table 16 shows the comparison of the accuracy of the proposed method for consecutive facial emotion recognition and those of other related researches for eINTERFACE'05 database. According to Table 16, it is obvious that the proposed methods achieves a much higher recognition rate than those by using the methods in the other papers. This again verifies the performance of the proposed method.

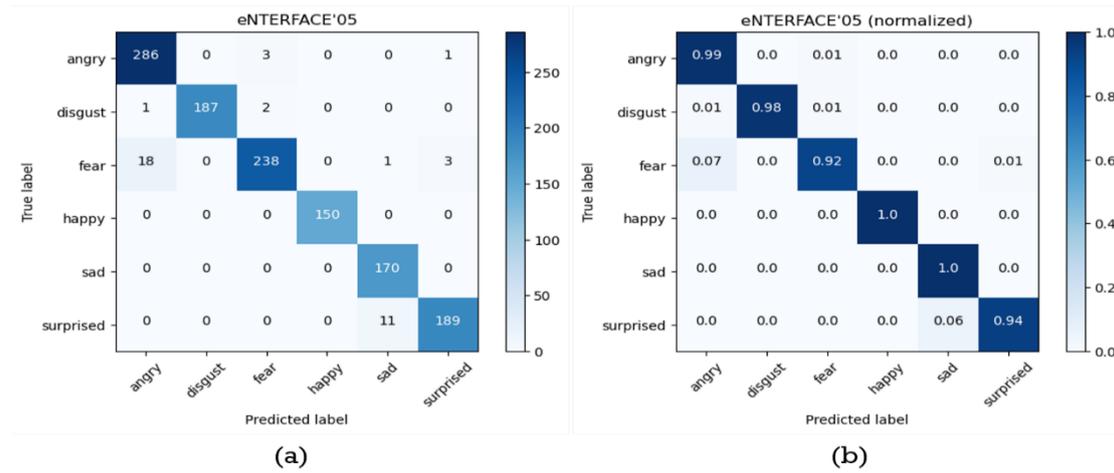


Figure 12. The confusion matrix before and after normalized, (a) and (b) respectively, obtained from 10-fold cross-validation of the consecutive facial emotion recognition method proposed in this paper on eINTERFACE'05 database.

Table 15. The accuracy, precision, recall, and F1-score of each emotion of the proposed consecutive facial emotion recognition method on eINTERFACE'05 database.

Label	Accuracy	Precision	Recall	F1-score
Angry	0.9817	0.9377	0.9862	0.9613
Disgust	0.9976	1.0000	0.9842	0.9920
Fear	0.9786	0.9794	0.9154	0.9463
Happy	1.0000	1.0000	1.0000	1.0000
Sad	0.9905	0.9341	1.0000	0.9659
Surprised	0.9881	0.9793	0.9450	0.9618
Average	0.9894	0.9718	0.9718	0.9712

Table 16. Comparison of the results of cross-validation for the proposed consecutive facial emotion recognition method with other related researches on the eINTERFACE'05 database.

Paper	Classes	Accuracy
F. Ma, et al. [20]	6	80.52%
B. Pan, et al. [30]	6	86.65%
P. Tiwari [31]	7	61.58%
Proposed model	6	96.82%

5. Experiments for FPGA Implementation

This paper implemented the proposed consecutive facial emotion recognition pre-processing methods and deep learning model introduced in Section 3 and Section 4 on FPGA chip and then tested the performance of the chip using 10-fold cross-validation. Each testing dataset consisted of 246 consecutive image data. There are 30 facial images in each consecutive image data. The FPGA in DE-10 Standard development board is equipped with a dual-core ARM Cortex-A9 processor. In order to accelerate computation of the model on FPGA, the test dataset is divided into two sub-datasets, and parallel computing was used to implement the neural network model inference. This reduces the

execution time by half. The accuracy of each fold in the test and the average test time for a single consecutive image are shown in Table 17.

Table 17. Accuracies and execution time of testing the proposed consecutive facial emotion recognition model on FPGA.

	Accuracy (%)	Execution time (sec)
Fold 1	99.19	11.19
Fold 2	100.00	12.04
Fold 3	100.00	12.15
Fold 4	99.59	11.24
Fold 5	99.19	11.20
Fold 6	99.59	12.17
Fold 7	99.59	12.06
Fold 8	99.59	11.45
Fold 9	99.19	11.92
Fold 10	99.19	11.59
Average	99.51	11.70

The experimental results show that the testing accuracies on the FPGA development board using the proposed neural network model inference algorithms are the same as those tested on PC using deep learning frameworks such as TensorFlow and Keras for implementing the proposed model. This verifies that the implementation of the proposed neural network model on FPGA chip using inference algorithms in this paper is efficient and performs well according to Table 17. Moreover, Table 18 shows the average execution time and proportion of each convolutional layer, batch normalization layer, max pooling layer, LSTM layer, and fully connected layer when testing a single consecutive image data for more detailed analysis. It can be seen that due to the first convolutional layer expanding the dimensions of the image data to 32, the input data for the second convolutional layer is much larger than that of the first layer. As a result, the execution time of the second convolutional layer is the longest at 6.9281 seconds (59.21%). Furthermore, since the proposed consecutive facial emotion recognition model incorporates max pooling layers, which can reduce the dimension of the feature maps, the execution time of each convolutional layer decreases sequentially. However, the average execution time for running the proposed deep learning model CLDNN on recognize the emotion from a video costs only 11.7 seconds. This verifies that the implemented AI (Artificial Intelligent) chip based on FPGA is feasible and is suitable for the AI edge computing application.

Table 18. Execution time and proportions of each layer in proposed consecutive facial emotion recognition model on FPGA.

Layer	Execution time (sec)	Proportion (%)
Conv2D_1	1.3649	11.66
Batch_Normalization_1	0.0008	Less than 0.01
Max_Pooling2D_1	0.8733	7.46
Conv2D_2	6.9281	59.21
Batch_Normalization_2	0.0010	Less than 0.01
Max_Pooling2D_2	0.2260	1.93
Conv2D_3	1.5648	13.37
Batch_Normalization_3	0.0009	Less than 0.01
Max_Pooling2D_3	0.0755	0.64

Conv2D_4	0.4721	4.03
Batch_Normalization_4	0.0006	Less than 0.01
Max_Pooling2D_4	0.0386	0.32
Conv2D_5	0.0468	0.40
Batch_Normalization_5	0.0006	Less than 0.01
Max_Pooling2D_5	0.0248	0.21
Conv2D_6	0.0322	0.27
Batch_Normalization_6	0.0005	Less than 0.01
Max_Pooling2D_6	0.0221	0.18
LSTM	0.0071	0.06
Batch_Normalization_7	0.0000	Less than 0.01
Dense (Softmax)	0.0002	Less than 0.01
Total	11.70	100%

6. Conclusions

This paper proposed the methods based on deep learning for consecutive facial emotion recognition. The proposed model was implemented on an embedded system with FPGA chip without the need for a deep learning framework during the model inference process. For consecutive facial emotion recognition, this paper captured 30 frames of an image sequence to represent a consecutive image segment. The Haar cascade frontal face detection model from OpenCV was utilized to extract the facial regions from the images, followed by grayscale conversion and resizing to reduce computational burden on the embedded device. The preprocessed images were then fed into local feature learning blocks to extract local features from individual frames. These features were then packaged into a feature sequence representing a consecutive image segment. The feature sequence was then passed through an LSTM layer for temporal sequence learning. Finally, a fully connected layer was used for classification.

Next, the parameters of the deep learning models for consecutive facial emotion recognition, as well as the test dataset, were fed into FPGA's memory for model inference. This research implemented the neural network model inference algorithms in Python. Then, through high-level synthesis, the algorithms were automatically transformed from the high-level language into circuit functionality. This allowed us to realize model inference on the embedded device without the need of deep learning frameworks. For the model inference of consecutive facial emotion recognition, the proposed method achieved the same test accuracy as that tested on a PC using deep learning frameworks. This indicated that the neural network model inference algorithms proposed in this paper can achieve the same performance as using the deep learning frameworks. The average testing time for a single consecutive image data was 11.70 seconds, with an average testing time of 0.39 seconds per single image of size 100×100 pixels. The implemented hardware had an FPS of 2.56. The experimental results for the designed FPGA chip verify that the implemented AI (Artificial Intelligent) chip based on FPGA is feasible and is suitable for the AI edge computing application.

Finally, according to the experimental results in Section 4, the proposed deep learning model applied on the three databases RADESS, BAUM-1s and IEMOCAP databases achieves much higher recognition rates than those in the other papers. This demonstrate that the proposed methods outperform the methods in the other literatures.

Author Contributions: Conceptualization, methodology, formal analysis, writing—review and editing, Shing-Tai Pan; software, validation, data curation, writing—original draft preparation, Han-Jui Wu. All authors have read and agreed to the published version of the manuscript.

Funding: This research work was supported by the Ministry of Science and Technology of the Republic of China under contract NSTC 112-2221-E-390-016.

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

References

- Izquierdo-Reyes, J.; Ramirez-Mendoza, R. A.; BustamanteBello, M. R.; Navarro-Tuch, S.; Avila-Vazquez, R. Advanced Driver Monitoring for Assistance System (ADMAS) Based on Emotions. *IEEE International Journal of Interactive Design and Manufacturing* **2018**, vol. 12(1), pp. 187-197.
- Prasad, N.; Unnikrishnan, K.; Jayakrishnan, R. "Fraud Detection by Facial Expression Analysis Using Intel RealSense and Augmented Reality," International Conference on Intelligent Computing and Control Systems, pp. 919-923, Madurai, India, Mar., 2018.
- Nijse, B.; Spikman, J. M.; Visser-Meily, J. M.; de Kort, P. L.; C. M. van Heugten, Social Cognition Impairments in the Long-Term Post Stroke. *Archives of Physical Medicine and Rehabilitation* **2019**, vol. 100(7), pp. 1300-1307.
- Ninaus, M.; Greipl, S.; Kiili, K.; Lindstedt, A.; Huber, S.; Klein, E.; Moeller, K. Increased Emotional Engagement in Game-based Learning-A Machine Learning Approach on Facial Emotion Detection Data. *Computers & Education* **2019**, vol. 142, no. 103641.
- Matsugu, M.; Mori, K.; Mitari, Y.; Kaneda, Y. Subject Independent Facial Expression Recognition with Robust Face Detection using a Convolutional Neural Network. *Neural Networks* **2003**, vol. 16(5), pp. 555-559.
- Pramerdorfer C.; Kampel, M. Facial Expression Recognition using Convolutional Neural Networks: State of the Art. arXiv preprint arXiv:1612.02903, Dec., 2016.
- Ayadi, M. E.; Kamel, M. S.; Karray, F. Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases. *Pattern Recognition* **2011**, vol. 44(3), pp. 572-587.
- Khalil, R. A.; Jones, E.; Babar, M. I.; Jan, T.; Zafar, M. H.; Alhussain, T. Speech Emotion Recognition Using Deep Learning Techniques: A Review. *IEEE Access* **2019**, vol. 7, pp. 117327-117345.
- Devika, K. N.; Bhakthavatchalu, R. Design of Reconfigurable LFSR for VLSI IC Testing in ASIC and FPGA. International Conference on Communication and Signal Processing, Chennai, India, Feb., 2017.
- Deschamps, J. P.; Bioul, G. J. A. *Synthesis of Arithmetic Circuits: FPGA, ASIC and Embedded Systems*. New York, USA: Wiley-Interscience, ISBN: 9780471687832, Feb., 2006.
- Fetcher, B. H. FPGA Embedded Processors: Revealing True System Performance. Embedded Systems Conference, pp. 1-18, San Francisco, USA, 2005.
- Bazil Raj, A. A. *FPGA-based Embedded System Developer's Guide*. Boca Raton, FL, USA: CRC Press, ISBN: 9781315156200, Oct., 2018.
- Rupani, A.; Sujediya, G. A Review of FPGA Implementation of Internet of Things. *International Journal of Innovative Research in Computer and Communication Engineering* **2016**, vol. 4, no. 9, pp. 16203-16207.
- Hauck S.; DeHon, A. *Reconfigurable Computing: The Theory and Practice of FPGA-Based Computation*. San Francisco, CA, USA: Morgan Kaufmann, ISBN: 9780080556017, Nov., 2007.
- Pellerin D.; Thibault, S. *Practical FPGA Programming in C*. Upper Saddle River, NJ, USA: Prentice Hall Press, ISBN: 9780131543188, Apr., 2005.
- Kilts, S. *Advanced FPGA Design: Architecture, Implementation, and Optimization*. New York, USA: Wiley-IEEE Press, ISBN: 9780470054376, Jun., 2007.
- Sainath, T. N.; Vinyals, O.; Senior A.; Sak, H. Convolutional Long Short-Term Memory Fully Connected Deep Neural Networks. IEEE International Conference on Acoustics Speech Signal Processing, South Brisbane, QLD, Australia, Apr., 2015.
- Ryumina, E.; Karpov, A. Facial Expression Recognition using Distance Importance Scores Between Facial Landmarks." International Conference on Computer Graphics and Machine Vision, pp. 1-10, Dec., 2020.
- Sagonas, C.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge. IEEE Workshops International Conference on Computer Vision, pp. 397-403, Sydney, NSW, Australia, Dec., 2013.
- Ma, F.; Zhang, W.; Li, Y.; Huang, S. L.; Zhang, L. Learning Better Representations for Audio-Visual Emotion Recognition with Common Information. *Applied Sciences* **2020**, vol. 10(20), pp. 7239.
- Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters* **2016**, vol. 23(10), pp. 1499-1503.
- Jaratrotkamjorn A.; Choksuriwong, A. Bimodal Emotion Recognition using Deep Belief Network. International Computer Science and Engineering Conference, pp. 103-109, Phuket, Thailand, Nov., 2019.
- Chen, Z. Q.; Pan, S. T. *Integration of Speech and Consecutive Facial Image for Emotion Recognition Based on Deep Learning*. Master's Thesis, National University of Kaohsiung, Kaohsiung, Taiwan, 2021.
- Livingstone S. R.; Russo, F. A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVD ESS): a Dynamic, Multimodal Set of Facial and Vocal Expressions in North American English. *PLoS One* **2018**, vol. 13(5), no. e0196391.
- Adeshina, S. O.; Ibrahim, H.; Teoh, S. S.; Hoo, S. C. Custom Face Classification Model for Classroom using Haar-like and LBP Features with Their Performance Comparisons. *Electronics* **2021**, vol. 10(2), no. 102.

26. Wu, H.; Cao, Y.; Wei, H.; Tian, Z. Face Recognition based on Haar Like and Euclidean Distance. *Journal of Physics: Conference Series* **2021**, vol. 1813, no. 012036.
27. Gutter, S.; Hung, J.; Liu, C.; Wechsler, H. *Comparative Performance Evaluation of Gray-Scale and Color Information for Face Recognition Tasks*. Heidelberg, Berlin, Germany: Springer, ISBN: 9783540453444, Aug., 2001.
28. Bhattacharya, S.; Kyal, C.; Routray, A. Simplified Face Quality Assessment (SFQA). *Pattern Recognition Letters* **2021**, vol. 147(4), pp. 108-114.
29. Khandelwal, A.; Ramya, R. S.; Ayushi, S.; Bhumika, R.; Adhoksh, P.; Jhavar, K.; Shah, A.; Venugopal K. R. Tropical Cyclone Tracking and Forecasting Using BiGRU [TCTFB]. *Research Square* preprint **2022**, PPR553621, DOI: <https://doi.org/10.21203/rs.3.rs-1955121/v2>.
30. Pan, B.; Hirota, K.; Jia, Z.; Zhao, L.; Jin, X.; Dai, Y. Multimodal Emotion Recognition Based on Feature Selection and Extreme Learning Machine in Video Clips. *Journal of Ambient Intelligence and Humanized Computing* **2023**, vol. 14, pp. 1903-1917.
31. Tiwari, P.; Rathod, H.; Thakkar, S.; Darji, A. Multimodal Emotion Recognition Using SDA-LDA Algorithm in Video Clips. *Journal of Ambient Intelligence and Humanized Computing* **2021**, vol. 14(5), pp. 1-18.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.