Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9806796, 1 page
https://doi.org/10.1155/2023/9806796

Retraction

Retracted: Real Time Facial Expression Recognition for Online Lecture

Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/ participant consent to participate, and/or agreement to publish patient/ participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Wu, "Real Time Facial Expression Recognition for Online Lecture," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 9684264, 11 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 9684264, 11 pages
https://doi.org/10.1155/2022/9684264

Research Article

Real Time Facial Expression Recognition for Online Lecture

Haobang Wu

University of Sydney Faculty of Engineering, Australia

Correspondence should be addressed to Haobang Wu; hawu3115@uni.sydney.edu.au

Received 25 January 2022; Revised 8 March 2022; Accepted 22 March 2022; Published 14 April 2022

Academic Editor: Vinayakumar Ravi

In order to better improve online teaching during the epidemic, teachers can adjust the teaching according to the students' understanding. During the epidemic period, online teaching has become a basic way for teachers to teach. One big problem with online teaching is that the lecturers find it hard to see the students' facial expressions. According to the study, the students' facial expressions are essential for the instructor to understand the students' understanding of the course material. There are some relevant studies on the recognition of human facial expressions. However, most of them did not consider some expressions such as "Enlightened," "Confused," or "Bored." This study helps teachers to improve the quality of teaching by designing a program to provide them with students' facial expressions. The study, implemented on students' computers, can capture their faces from a camera, identify some common and useful facial expressions, and send them to lecturers regularly during online lectures. The main research methods used in this paper are as follows: (1) Use the face recognition algorithm provided by OpenCV to capture faces. (2) Train a convolutional neural network model to recognize the expression "Happy," "Surprised," "Neutral," "Enlightened," "Confusion," and "Boredom." (3) Use the message protocol EMQX to transmit the expression information. This study can successfully capture faces, and about 80% accuracy can identify expressions and successfully transmit expression information. This study contains some expressions rarely studied by others and is innovative in the field of facial expression recognition. Datasets, deep learning models, and results are available for other research teams.

1. Introduction

Online teaching has become an essential method for the university to continue holding semesters for students, researchers, and professors during the pandemic. The online conference platforms are playing a significant role in delivering the online lecturer. These platforms can help restore the traditional lecture model by allowing the lecturers to show the course materials through screen sharing or to write notes on the virtual whiteboard. However, what online teaching can not restore is efficient interaction between the lecturers and their students. Although the online conference platforms can allow attendees to see and talk to each other through the webcams, it is not realistic for the lecture to interact with many students due to network bandwidth and video processing limitations. From the lecturers' perspective, students' videos are showed in windows which occupied a part of the lecturer's screen. There will be more windows showing on the screen if more students are attending the lecture. When the number of windows reaches the capacity of a screen, the windows will shrink to smaller sizes if more students are coming in the lecture room. For courses with a large number of students, lecturers will have trouble seeing the faces of students through the small and packed windows on the screen, not to mention reading their facial expressions. According to research, the students' facial expressions are the most used nonverbal communication mode in a lecture and are related to their emotions, which can help the lecturer recognize their comprehension of the lecture. Suppose there is a tool that can provide students' facial expressions to the lecturers periodically during an online lecture. In that case, it can restore the reading facial expression scenarios, and the lecturer can get useful interaction feedbacks from their students and can adjust their ongoing lectures accordingly.

This study designed and constructed a program on identifying common facial expressions of students in class. The program can automatically capture students' facial expressions from the camera and identify some common and useful facial expressions, including Boring, Confused, Cheerful, Happy, Surprised, and Neutral, and expression messages can be sent to the lecturer during online lectures.

2. Literature Review

Studies in computer vision show that machine learning algorithms are effective in human facial expression recognition. Among all the subfields of machine learning, deep learning is one of the most popular algorithms used by researchers.

In this paper, Shengjun designed a face recognition device with fill light performance [1]. Wu and others designed a privacy security scheme for face recognition based on Facenet and state secret algorithm [2]. Li et al. designed the face recognition system from the aspects of MTCNN and Facenet [3]. The paper written by Gory et al. summarizes the performance of different machine learning models in recognizing facial expressions. The research team applied and tested several machine learning algorithms, including AdaBoost, logistic regression, and two deep learning models: dense neural network (DNN) and convolutional neural network (CNN), which are used to identify the seven most common expressions in the paper. Gory et al. tested the performance of different machine learning models and proved that the CNN may be the most suitable [4].

Jiankui and others used the AdaBoost algorithm to study the application of face recognition technology in smart construction site [5]. Yanxiu and others studied the hotel management system and used facial recognition technology to build an intelligent hotel management system [6]. Zhenqian and Weizeng applied biometric technology to the design of online authentication module [7].

As one of the most remarkable developments in the computer vision field, the convolutional neural network (CNN) has a significant ability to recognize objects from an image [8]. The CNN is a type of deep learning model that contains convolutional layers and dense layers. For each convolutional layer, several convolutional kernels are used to detect a certain feature of the image. The different convolutional layers can detect different features of an image. With all the convolutional layers stacked together, the image's features will be collected and delivered to the following fully connected layers. The fully connected layers can learn the nonlinear combinations of the high-level features extracted by the convolution layers and do the classifications.

To investigate how the CNN model can be used to do the recognition, I read the paper of Dr. Hussain and Dr. Balushi. The authors propose using a deep CNN model to do a real-time face emotion classification and recognition. The paper's core approach is to build a CNN model based on a well-known model called VGG16. VGG16 is a CNN model with a structure containing five different convolutional layers, each with different numbers and sizes of kernels and three fully connected layers. The final accuracy of the trained model described in the paper is 0.88.

The convolutional neural network has an excellent performance on human facial expression recognition. Chenhao used a complex convolutional neural network to study and verify face recognition in mixed scenes [9]. Zhichao and others designed a driver fatigue dangerous driving detection system based on face recognition [10]. Li et al. studied the multispectral face recognition system [11]. Kezheng et al. used two-dimensional linear discriminant analysis and collaborative representation for face recognition [12]. From privacy concerns to improving recognition accuracy, Zijian is concerned about the dynamic application of face recognition technology in business application Governance [13].

Liu et al. designed a CNN model with an accuracy of 65% [14]. Fathallah et al.'s CNN models can have an accuracy of 90% when testing on datasets CK+, RaFD, and MUG [15]. The data that can be potentially used for my thesis is not sufficient. Most of the available collected datasets from other researchers do not contain emotions like boredom, confusion, and enlightened. It is hard to build a large dataset in a short time. Therefore, a technique that can do expression recognition on a small dataset is needed. Heidari and Fouladi-Ghaleh introduced and applied such a technique: transfer learning in their paper. Transfer learning can help to utilize the knowledge of a pretrained model onto a new model so that only a portion of the parameters in the new model needs to be trained. Therefore, the new model can be trained with a smaller dataset and within a shorter time. The research team froze the trained parameters of the first four convolutional layers (these parameters are untrained) and only trained the last convolutional layer and the fully connected dense layers. According to the paper, the result accuracy is 95.62%, which is higher than some other common methods.

Although the goal of the paper is different from mine, the idea of using transfer learning with the help of pretrained models will be potentially helpful in the situation where data is not enough. The key to transfer learning is to choose a proper pretrained model. Caroppo et al. compared them by evaluating three widely used pretrained models of facial expression recognition (VGG16, AlexNet, and GoogLeNet) on four different benchmark datasets (FACES, Lifespan, CIFE, and FER2013) [16]. Full experiments and results provide the reader with good performance of transfer learning in different modalities and help me to make decisions in the paper.

3. Methods

Let us first recap the goal of the thesis: to design and build a program that can automatically capture students' faces from cameras; can recognize the most common and useful facial expressions, including boredom, confused, enlightened, happy, surprised, and neutral; and can send the expression information to the lecturer periodically during an online lecture.

There are three steps to accomplish the goal: (1) Capture students' faces with their cameras. (2) Design and build a deep learning model that can accurately recognize the target facial expressions. (3) Build an output pipeline that can return the information to the lecturer. Different methods are used for each step.

The purpose of the model is to recognize the facial expressions after training using a dataset with limited data. Inspired by the research of Caroppo et al. and the paper of

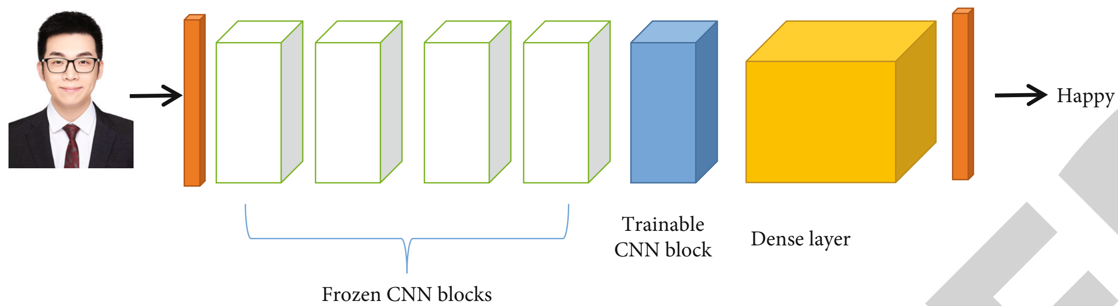

Figure 1: The complete model structure: including one input layer, 4 frozen CNN blocks, 1 trainable CNN block, dense layers, and one output layer.

Heidari and Fouladi-Ghaleh [17], I built a convolutional neural network by applying the transfer learning method. The pretrained model I chose is VGG16, provided by a python library called Keras. As I mentioned in Literature Review, VGG16 contains five convolutional blocks and three fully connected layers. My CNN model had the same structure but with some add-ons and modifications. The complete model structured is shown in Figure 1. The input layer is modified to accept an input image of size $160 * 160 * 3$. There are two convolutional layers and one 2D max-pooling layer for each convolutional block (Figure 2). The five convolutional blocks contain 64, 128, 256, 512, and 512 convolutional kernels of a size of $3 * 3$ for each convolutional layer. Three fully connected layers with a number of 512 neural follow the convolutional blocks. Each layer is followed by a "RELU" activation function. Besides, two dropout layers of a rate of 50% are added after the second and the third fully connected layer. The final output layer contains six output neural followed by a "Softmax" activation function. The last layer's output is a vector consisting of six values, and each value represents the prediction probability of each category. The prediction of the model is the expression with the highest prediction probabilities. I chose to freeze the first four convolution layers so that the parameters are not trained. The last convolutional block plus the following linear layers is trainable. Although the training data is not sufficient, the fact that fewer parameters are trainable can guarantee an acceptable result of the model performance. The two dropout layers are helping to fight against potential overfitting problems caused by a lack of data.

To build the dataset, I collected 1073 images: 200 images each for expression boredom, confusion, happy, surprised, and neutral and 73 images for expression enlightened. The images are collected from three sources. Most of the data are the self-collected. I asked volunteers to take selfies about the six required expressions. This part of data is the most validated because the labels of the pictures are provided directly by the person who makes the expressions. Plus, the volunteers are all students and the students are exactly the objects for the recognition task. Another big portion of the data comes from open sources, including the dataset provided by Sara Zhalehou and Bahcesehir University ECE Department, "Bahcesehir University Multimodel Face Database of Spontaneous Affective and Mental States" [18]. The data are collected during research experiments with the

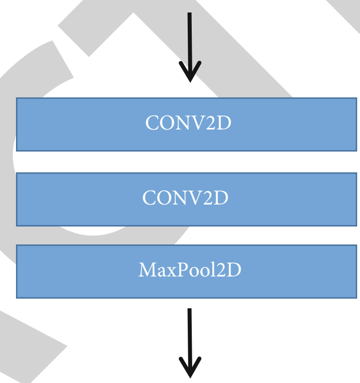

Figure 2: The structure of each convolutional block.

volunteered students and are labeled by the researchers. The data are carefully examined and organized before providing to the public. I used the pictures labeled with the six same facial expressions to expand my dataset. The final part of the data is collected from online sources like image webpages and social media. I searched the pictures using the facial expression terms as the keywords and collected after manually selection by a selection team. The team consists of my supervisor, my classmates, and me. The pictures are chosen and labeled if all members agreed that they correspond to the facial expression keywords.

With the collected raw data, a preprocessing step is needed before the deep learning model train. The pictures are converted into greyscale and are resized into a resolution of $160 * 160$ pixels. A greyscale face image can reduce the unexpected problems caused by different skin colors, and uniform image sizes are required for the model training. The label of the data is represented using one hot vector. In this case, the one hot vector is a vector with six values of 1 or 0. There is one value of "1" representing the corresponding category of the data, and the rest are "0."

For model training, the dataset is divided into a training dataset and a validation dataset with a ratio of 4 : 1. The input of the CNN model is the $160 \times 160 \times 3$ cropped face images from the training dataset. The output of the model is a list of prediction probabilities for the six expressions. The model is trained with a learning rate of 0.001 and is trained for 100 epochs. For each epoch, the loss is calculated with a cross-entropy loss function (Equation (1)),

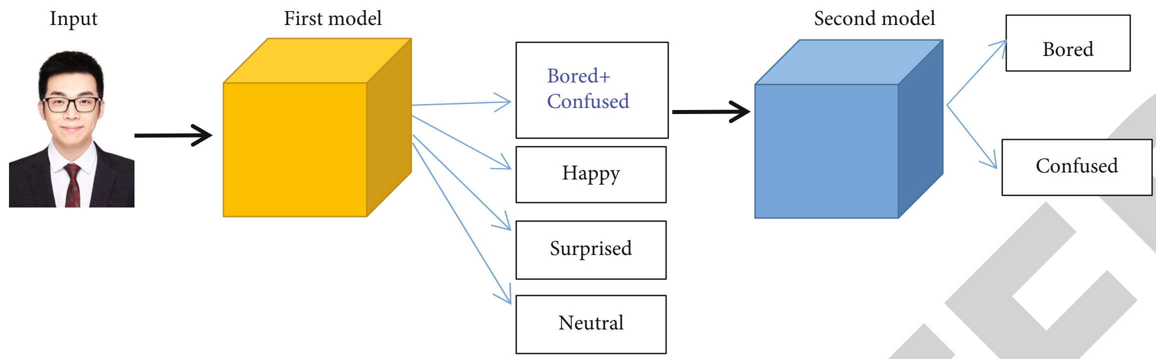

Figure 3: The structure of the third implementation.

and the model is optimized with stochastic gradient descent (Equation (2)). The accuracy of the output equals the portion of the number of correct predictions among all predictions.

$$\text{Cross entropy loss} = -\Sigma t_i * \log(p(x)), \tag{1}$$

$$W = W - \eta \Sigma \frac{\partial \text{Loss}}{\partial \text{w}_{\text{j}}}. \tag{2}$$

For testing the performance of the trained model, three-fold cross-validation testing is used. The model's performance is evaluated by watching the pattern of the training and validation loss/accuracy graph (against the number of epochs) and the accuracy of predictions for each expression. The whole coding for training the model is written using Python. The Python library Keras is used to create, train, and test the model.

The method of how to build the CNN model for training is clearly stated in the previous section. In my thesis work, the method is implemented in different ways to get different aspects of the results. The first implementation uses the CNN model to train with the data that contains all six facial expressions. Because the number of images about Enlightened is only 73, which is much less than the number of other expressions (200), I have to choose a dataset that contains 73 images for each expression for the training/testing procedure. The insufficient data may lead to a result that can be badly overfitting on training data. This is the trade-off I have to make due to the limited size of my dataset.

The second implementation uses a slightly different CNN model from the first one. Because the result from the first implementation implies that the model's performance in recognizing the expression "Enlightened" is very poor, I chose to evaluate the model without the "Enlightened." This time, the numbers of data for training and testing are larger: 200 images for each expression. With more training data involved and less output category, the result of the second implementation should have a better performance than the first one.

The third implementation is a combination of two CNN models. (Figure 3) After combining the two models, the implementation can output a classification for the five facial expressions. The method is innovative and sounds reasonable, and you can expect a better result than the second implementation.

4. Results and Discussion

4.1. Results and Discussion about First Implementation. The first implementation is to use the CNN model to recognize all six facial expressions (Boredom, Confusion, Neutral, Enlightened, Happy, and Surprised). The CNN model is the modified model from pretrained VGG16 and has an output layer with six output neural. The dataset used for this implementation contains 438 images in total: 73 images for each emotion. The data is divided into a training/validation dataset with a ratio of 4 : 1. Before the training, the data is enlarged using the default data augmentation method provided by Keras. The model is trained with cross-entropy loss, SGD optimizer, and a learning rate of 0.001. The performance of the model is evaluated using three-fold cross-validation. There are three different sets of the training/validation dataset, and the performances of the model on all the sets are counted.

Figures 4 and 5 are the plots for the training/validation loss against the number of epochs and the training/validation accuracy against the number of epochs for the model performed on one of the sets. From the plot of loss vs. epochs, it is clear that the validation loss (orange curves) stopped decreasing after around 30 epochs and finally floated around a loss of 0.8 to 1.0 while the training loss (blue curves) kept decreasing and converged to zero in the end. A similar pattern was also showed in the plot of accuracy vs. epochs. The validation accuracy increased very slowly and floated around 72%. Table 1 shows the average accuracy of the cross-validation data. The specific accuracy on recognizing each of the six expressions is shown in the table. The model has accuracies around 80% for Happy, Surprised, and Neutral. Accuracies for Boredom and Confusion are lower and are around 70%. Accuracy for recognizing Enlightened faces is the worst and is only 56%. The total accuracy for the model is 72%. The confusion matrix of predictions and labels is shown in Figure 6.

The overall result of the first implementation is poor. The trained model is not an acceptable one for the design. The performance of the expression "Enlightened" is the worst because the training data about "Enlightened" is

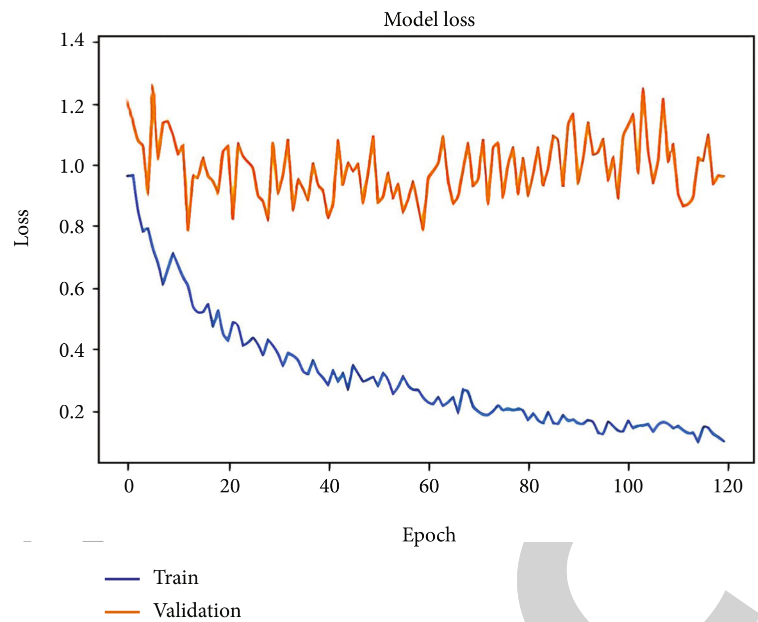

Figure 4: First implementation: loss.

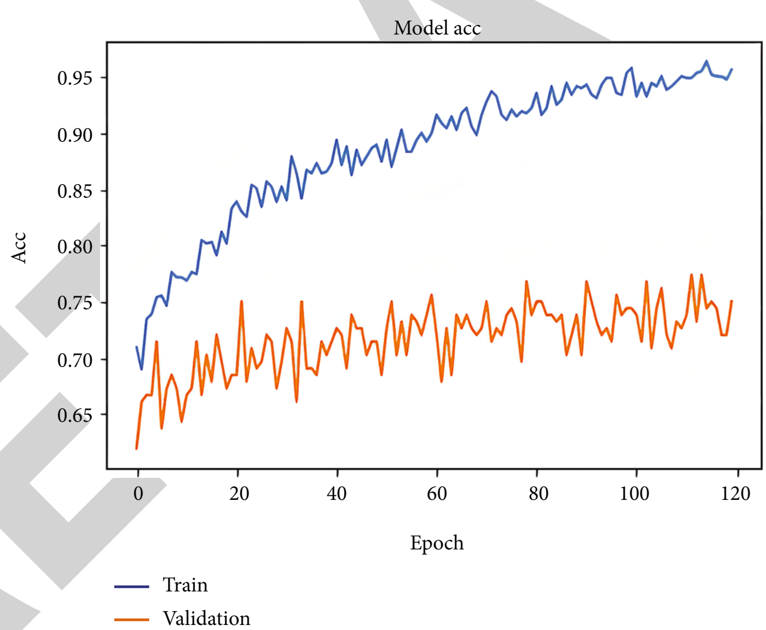

Figure 5: First implementation: accuracy.

Table 1: Experiment results for the first implementation.

	Total	Happy	Surprised	Neutral	Boredom	Confusion	Enlighten
Accuracy	0.72	0.86	0.81	0.78	0.67	0.69	0.56

collected and labeled only based on my own opinion. Since there is no convincing and valid dataset about this facial expression, I decided to exclude the expression "Enlightened" from my thesis goal and focus on classifying the other five facial expressions, which is my second implementation.

4.2. Results and Discussion about Second Implementation. Second implementation uses the CNN model to recognize the five facial expressions, excluding "Enlightened" (Boredom, Confusion, Neutral, Happy, and Surprised). The CNN model is similar to the one used in the first

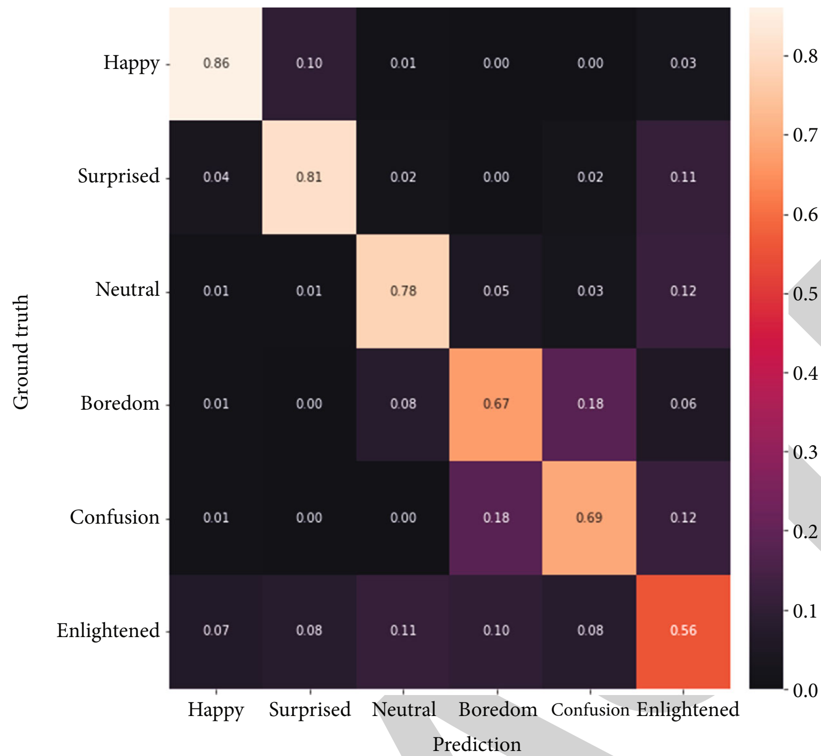

Figure 6: First implementation: confusion matrix.

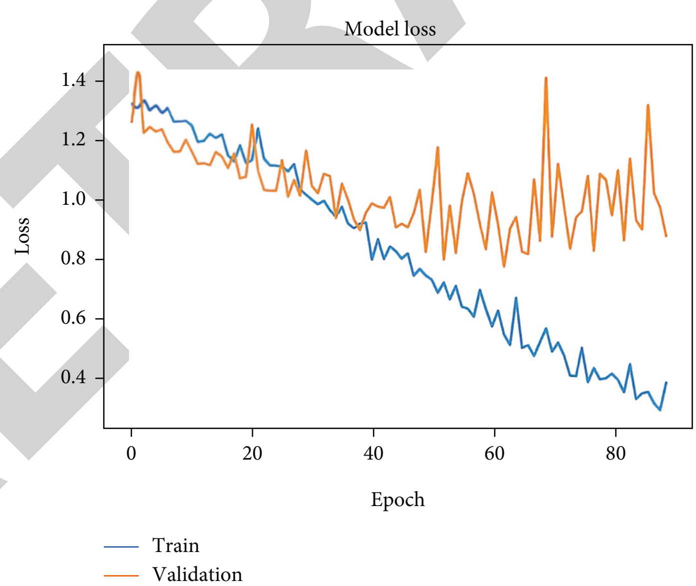

Figure 7: Second implementation: loss.

implementation. Again, it is a modified model from pre-trained VGG16 but has an output layer with five output neural this time. The dataset now used for this implementation contains 1000 images in total: 200 images for each emotion. The data is still divided into a training/validation dataset with a ratio of 4 : 1 and is enlarged with Keras data augmentation function. The model is trained with cross-entropy loss, SGD optimizer, and a learning rate of 0.001, and the model's performance is evaluated using three-fold cross-validation.

Figures 7 and 8 are the plots for the training/validation loss against the number of epochs and the training/validation accuracy against the number of epochs for the model performed on one of the sets. From the loss vs. epochs plot, the validation loss stopped decreasing after around 50 epochs and finally floated around a loss of 1.0 while the

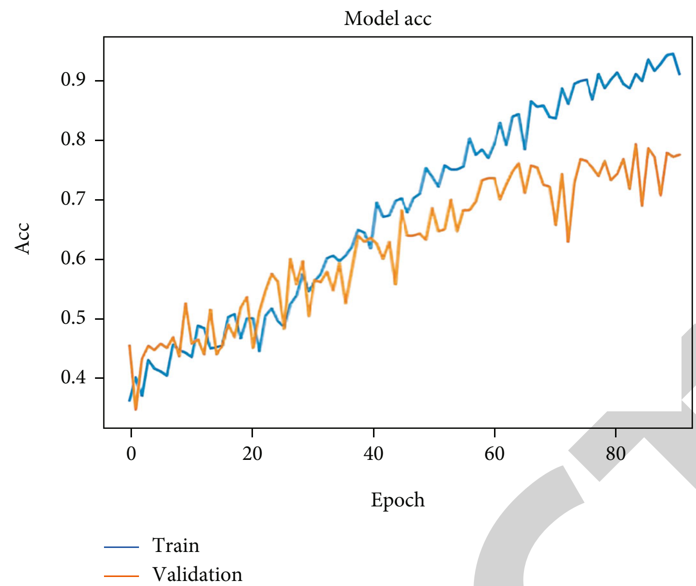

Figure 8: Second implementation: accuracy.

Table 2: Experiment results for the second implementation.

	Total	Happy	Surprised	Neutral	Boredom	Confusion
Accuracy	0.75	0.85	0.81	0.80	0.65	0.63

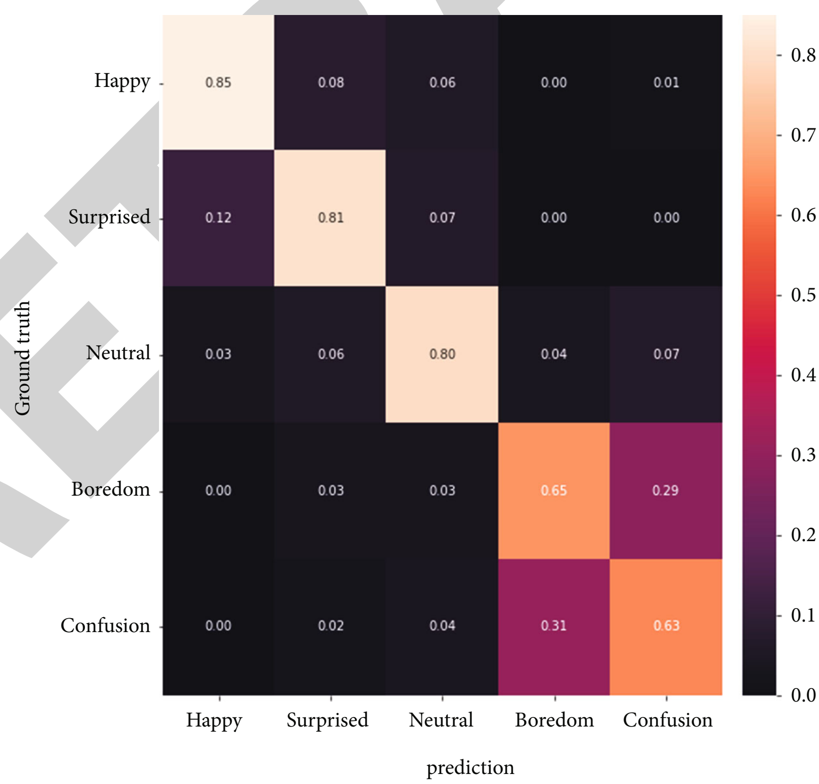

Figure 9: Second implementation: confusion matrix.

training loss kept decreasing. In the accuracy vs. epochs plot, the validation accuracy increased along with the training accuracy and finally floated around 75%. Table 2 shows the model's average accuracy in predicting the facial expressions of all three sets of data. The model has accuracies around 85% for Happy and 80% for Surprised and Neutral.

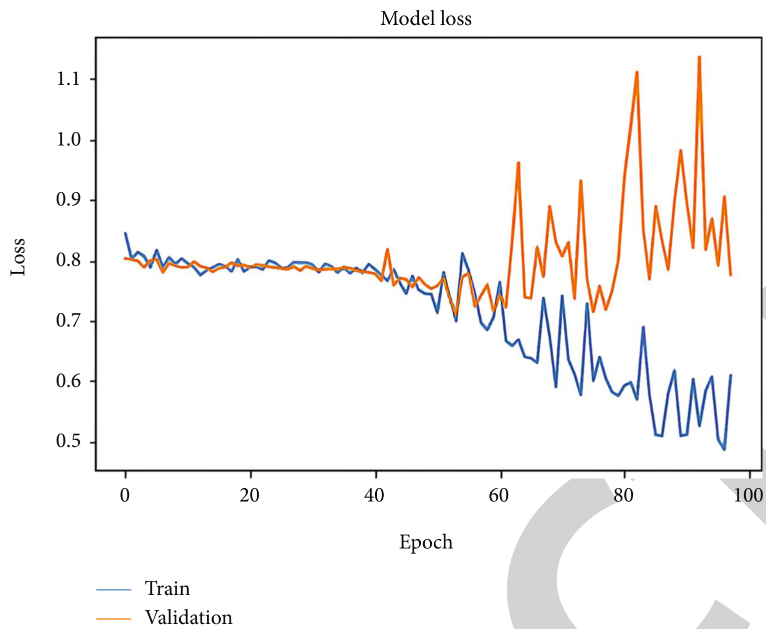

FIGURE 10: Third implementation, B+C model: loss.

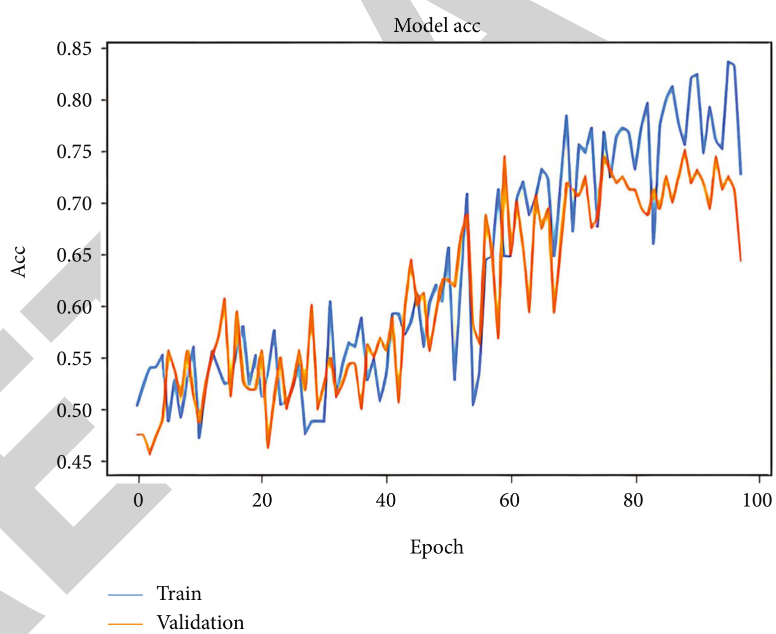

FIGURE 11: Third implementation, B+C model: accuracy.

Accuracies for Boredom and Confusion are still lower: they are below 70%, just like the first implementation. The model's total accuracy is 75% because the worst accuracy of Enlightened is excluded and can no longer drag down the average. The confusion matrix of predictions and labels is shown in Figure 9.

From the graphs of the "loss/accuracy vs. epochs" (Figures 10 and 11), it is clear that the model is still overfitting on training data. The validation loss finally converged to around 0.9, and the validation accuracy finally converges to around 75%. The results from the two plots show the same problem as stated in the first implementation. The model that tends to fit on training data is not good at generalizing for unseen data. From the table of accuracy, the accuracies for "Happy," "Surprised," and "Neutral" are still around 80%, which is similar to the last result, whereas accuracy

TABLE 3: Experiment results for B+C model.

	Total	Boredom	Confusion
Accuracy	0.75	0.76	0.74

Table 4: Experiment results for the combination model.

	Total	Happy	Surprised	Neutral	Boredom	Confusion
Accuracy	0.77	0.83	0.79	0.81	0.71	0.69

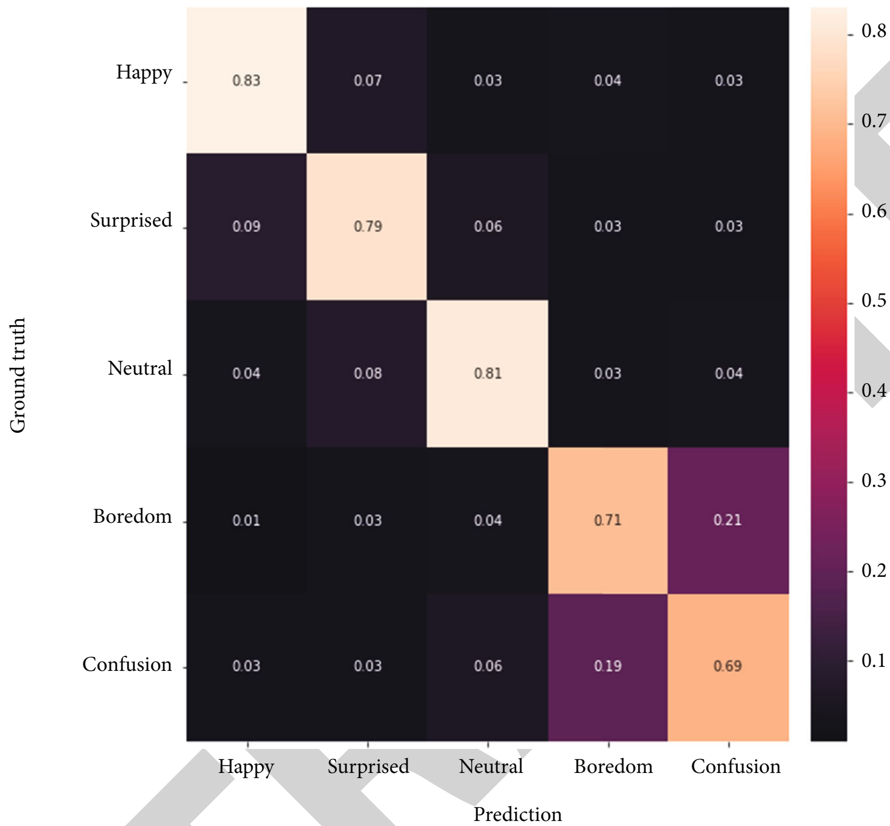

Figure 12: Combination implementation: confusion matrix.

for "Boredom" and "Confusion" decreased slightly compared to the first implementation. The model's total accuracy is increased because the low accuracy category "Enlightened" is not involved this time.

It is out of my expectation that the model did not improve a lot after getting rid of "Enlightened" and training with more data. The performance on recognizing "Happy," "Surprised," and "Neutral" is acceptable when compared to the result of other related researches. However, the model had a poor performance on recognizing "Boredom" and "Confusion." To find out why, I check every single prediction of the validation data for these two expressions. Surprisingly, I found out that the images of these two expressions are often classified to be the other one incorrectly.

4.3. Results and Discussion about Third Implementation (Combination Model). The third implementation combines two CNN models to recognize the five facial expressions (Boredom, Confusion, Neutral, Happy, and Surprised). Both two CNN models are similar to those used in the previous implementations but with different output layers. The first model regards "Boredom" and "Confusion" to be one category. Thus, it contains an output layer with four neurons. The dataset contains 800 images in total: 200 images for each category are used to train the model. If the output of the first model is the "B+C" category for an image, the image will go through the second model to be further classified as either "Boredom" or "Confusion." The second model is trained to classify these two expressions, so it has a binary output layer. The dataset to train this model contains 200 images for each expression. The datasets are still divided into training/validation dataset with a ratio of 4 : 1 and are enlarged using Keras default data augmentation method. The models are trained with cross-entropy loss, SGD optimizer, and a learning rate of 0.001. The results of the two models will be discussed separately, and the performance for the combination of the models will be evaluated as well.

From the result, we can conclude that the model still has an acceptable performance on recognizing "Happy," "Surprised," and "Neutral," and this time, it can clearly classify expressions "Boredom" and "Confusion" with the other three expressions. If my second binary classification model can also have a good performance, the accuracy for recognizing these two will be higher than before.

The combination model contains the first model after training for 100 epochs and the second model after training for 60 epochs. To evaluate the performance of the third implementation, I used the combination model to predict the validation dataset and check the accuracy. The data first went through the first model and then recognized as one of the four categories (Happy, Surprised, Neutral, or B+C). If the data is one of the first three expressions, it will be output and compared to the ground truth. If the data is B+C, it was further classified by the second model and was compared to the ground truth. The accuracy of the further B+C model is showed in Table 3. The accuracy of the combination model is showed in Table 4. The confusion matrix of predictions and labels is shown in Figure 12.

When compared to the result of the model in the second implementation (Table 2), the accuracy for Boredom increases from 65% to 71%, and the accuracy for Confusion increases from 63% to 69%. The total accuracy, therefore, increased from 75% to 77%. Although the third implementation indeed improved from the second one, it still did not reach what I have expected: an average accuracy of 80%.

5. Conclusion

The main goal of this study is to improve the interaction between lecturers and students during the online lectures. The program designed and constructed in this paper achieves this goal by informing the teacher of students' facial expressions. This study contains some expressions rarely studied by others and is innovative in the field of facial expression recognition. Datasets, deep learning models, and results are available for other research teams. The dataset insufficiency is a big limitation that prevents me from building more accurate programs that makes the final program not perfect. The dataset insufficiency is a big limitation that prevents me from building more accurate programs that makes the final program not perfect. This can be an experiment in a university to let students take pictures of the required facial expressions. The second step is to test message transmission performance between two computers. Tests can include delayed testing and message loss rate testing. The final step in the future is to better display the facial expression information on the teacher's screen. The idea of this study was to use emojis to represent each student in a small window, consistent with the received facial expression, which is a direct expression for the teacher to see more intuitively.

Data Availability

The data underlying the results presented in the study are available within the manuscript.

Conflicts of Interest

There is no potential conflict of interest in our paper, and all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

References

[1] W. Shengjun, "A face recognition device design with light-filling performance," *Electronic Technology*, vol. 51, no. 2, pp. 172-173, 2022.

[2] J. Wu, P. Changgen, T. Weijie, and W. Zhenqiang, "Face Enc Auth: face recognition privacy and security scheme based on Face Net and national security algorithm," *Computer Engineering and Application*, vol. 18, no. 3, pp. 1–7, 2022.

[3] L. Zhihua, Z. Jianyu, and W. Zhongcheng, "Design of the face recognition system based on MTCNN and Facenet," *Modern Electronic Technology*, vol. 45, no. 4, pp. 139–143, 2022.

[4] S. Gory, M. Al-Khassaweneh, and P. Szczurek, "Machine learning approach for facial expression recognition," in *2020 IEEE International Conference on Electro Information Technology (EIT)*, pp. 32–39, Chicago, IL, USA, 2020.

[5] L. Jiankui, C. Yang, H. Xiaoxing, and L. Hui, "Application of facial recognition technology of Adaboost algorithm in intelligent construction site," *Single-chip microcomputer and embedded system application*, vol. 21, no. 4, 2021.

[6] Y. Yanxiu, Y. Rui, and Y. Feijie, "An intelligent hotel management system based on facial recognition," *Innovations*, vol. 8, no. 1, 2021.

[7] J. Zhenqian and X. Weizeng, "Design of an online authentication module based on biometric identification technology," *Information system Engineering*, vol. 10, 2020.

[8] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 5, pp. 699–714, 2005.

[9] Y. Chenhao, "Research and implementation of mixed scene face recognition based on complex convolutional neural network," *Electronic World*, vol. 2, 2022.

[10] H. Zhichao, Z. Hongmei, E. Chan, and T. Yaosen, "Driver fatigue dangerous driving detection system based on facial recognition," *Electromechanical Engineering Technology*, vol. 50, no. 12, pp. 143–146, 2021.

[11] C. Li, L. Zhihong, L. Yanping, R. Xiaoguang, and B. Changchun, "Review of multispectral facial recognition systems studies," *Technology Wind*, vol. 35, 2021.

[12] L. Kezheng, X. Deng, and Z. Yulun, "Two-dimensional linear discriminative analysis and collaborative representation of facial recognition methods," *Small Microcomputer system*, vol. 42, no. 8, pp. 1688–1693, 2021.

[13] Z. Zijian, "Face recognition technology business application governance dynamics: from privacy concerns to improved recognition accuracy concerns," *Competition Policy Research*, vol. 3, pp. 81–91, 2021.

[14] K. Liu, M. Zhang, and Z. Pan, "Facial expression recognition with CNN ensemble," in *2016 International Conference on Cyberworlds (CW)*, pp. 163–166, Chongqing, China, 2016.

[15] A. Fathallah, L. Abdi, and A. Douik, "Facial expression recognition via deep learning," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, pp. 745–750, Hammamet, Tunisia, 2017.

[16] A. Caroppo, A. Leone, and P. Siciliano, "Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults," *Journal of Computer Science and Technology*, vol. 35, no. 5, pp. 1127–1146, 2020.

[17] M. Heidari and K. Fouladi-Ghaleh, "Using Siamese networks with transfer learning for face recognition on small-samples datasets," in *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pp. 1–4, Iran, 2020.

[18] S. Zhalehpour, O. Onder, Z. Akhtar, and C. E. Erdem, "BAUM-1: a spontaneous audio-visual face database of affective and mental states," *IEEE Transactions on Affective Computing*, vol. 8, no. 3, pp. 300–313, 2017.