electronics

MDPI

Article

Study of Process-Focused Assessment Using an Algorithm for Facial Expression Recognition Based on a Deep Neural Network Model

Ho-Jung Lee [1] and Deokwoo Lee [2,*]

[1] CTO Division, LG Electronics, Seoul 07336, Korea; ripia1313@naver.com
[2] Department of Computer Engineering, Keimyung University, Daegu 42601, Korea
* Correspondence: dwoolee@kmu.ac.kr; Tel.: +82-53-580-5268

Abstract: This study proposes an approach for process-focused assessment (PFA) utilizing the concept of deep neural networks with a sequence of facial images. Recently, process-based assessment has received significant attention compared to result-based assessment in the field of education. Continuously evaluating and quantifying student engagement, as well as understanding and interacting with teachers in study activities are considered important factors. However, to achieve PFA, from the technical and systematic perspectives, the real-time monitoring of the learning process of students is desired, which requires time consumption and extremely high attention to each student. This study proposes an approach to develop an efficient method for evaluating the process of learning and studying students in real time using facial images. We developed a method for PFA by learning facial expressions using a deep neural network model. The model learns and classifies facial expressions into three categories: easy, neutral, and difficult. Because the demand for online learning is increasing, PFA is required to achieve efficient, convenient, and confident assessment. This study chiefly considers a sequence of 2D image data of students solving some exam problems. The experimental results demonstrate that the proposed approach is feasible and can be applied to PFA in classrooms.

Keywords: expression recognition; process-focused assessment; face detection; deep neural network; machine learning

Citation: Lee, H.-J.; Lee, D. Study of Process-Focused Assessment Using an Algorithm for Facial Expression Recognition Based on a Deep Neural Network Model . *Electronics* **2021**, *10*, 54. https://doi.org/10.3390/electronics10010054

Received: 13 November 2020
Accepted: 15 December 2020
Published: 31 December 2020

1. Introduction

The recognition of objects (e.g., faces), expressions, and emotions using visual images is a challenging task in computer vision and pattern recognition. Because visual information is widely used in several fields, recognizing target objects is an important task, and recognition activities have significantly improved from the theoretical and practical perspectives [1–4]. Notably, face recognition is a popular research area, receiving significant interest, and is extensively studied. Furthermore, its application exists in several practical areas [5–7]. Based on the technology of face recognition with image analysis, image processing and classification, and expression and (or -based) emotion recognition, notwithstanding being challenging and occasionally appearing as an ill-posed problem, recognition has gained considerable attention in practical areas [8,9]. In the field of face recognition, information, such as using landmarks, 3D curves representing the geometric shape of faces, intensity, and eigenfaces, has provided successful recognition results. This area is fundamental and central in practical fields such as certification, surveillance, security, and finance. In our previous study on face recognition, the proposed approach mainly focused on the analysis of images that are predominantly defined in 2D space. Moreover, in the previously proposed approaches, the efficient representation of face data is of interest because images usually contain more than a million pixels, and current cameras provide billions of pixels, leading to the production of high-resolution images. A major challenge of face (or any object) recognition is to classify objects that are usually defined

in high-dimensional space and represent them in a lower dimensional space. Such an excessive number of pixels leads to the degradation of processing speed, hindering application to real practices. Therefore, the projection of the original images defined in a high-dimensional space to a lower dimensional space is a crucial task, which is called dimensionality reduction [10]. To achieve dimensionality reduction for efficient representation without information loss, several algorithms have been introduced and have shown successful results for classification, representation, and recognition. Principal component analysis (PCA) focuses on the representation of the original data in lower dimensional space, and the eigenface algorithm uses PCA for face recognition. Linear discriminant analysis (LDA) focuses on determining the criteria for data classification. PCA, LDA, independent component analysis (ICA), and Fisher discriminant analysis (FDA) employ linear feature extraction [11]. Although there are several algorithms that achieve accurate recognition and classification results and alleviate the limitations that usually arise from ambient light, varying lighting conditions, and varying poses, PCA is widely employed [12]. Because the original data are usually defined in high-dimensional space, in practice, feature extraction and representation in a non-linear manner are of interest. Random projection is a nonlinear-based method, which matches a feature point in high-dimensional space to a point in a lower dimensional space. Establishing a relationship between these two points is crucial for the nonlinear method. A popular method is to employ kernel-based analysis such as kernel-PCA and kernel-LDA. In nonlinear methods, the intrinsic characteristics or structures of feature points at a high dimension should be preserved, whereas the dimension of the points that are projected into a lower dimensional space should be reduced. Local linear embedding (LLE), self-organizing map (SOM), curvilinear component analysis (CCA), curvilinear distance analysis (CDA), and other approaches have been used for nonlinear dimensionality reduction in recognition and classification. The details and surveys on dimensionality reduction using linear or non-linear approaches can be found in [13–16]. Expression pose or light-invariant face recognition has also been proposed to achieve a robust system for recognition and classification. To achieve the aforementioned robust results, the 3D coordinates of a target face, that is the geometric information of a face, have been fully exploited; consequently, the accuracy has shown considerable improvement [17,18]. However, as reported by previous studies, the improvement of the results is not significant, although 3D coordinate information has been used. Recently, learning-based face recognition has received significant attention, particularly recognition using convolutional neural networks (CNNs), regression, etc. [19–21]. Extended from face recognition, emotion recognition from visual information is of interest in practical fields. Speech signals are used to recognize emotions in humans, and image-based recognition has received significant attention in practical fields. To recognize emotions using visual information, we can exploit facial expression, as it performs a significant role in interpersonal communication. Visual expression is more popularly employed than the voice (verbal) signal. Similar to face recognition, the conventional recognition of emotion is based on localizing landmark points on a face image [22]. That is, emotion recognition depends on accurate feature extraction and detection. Recently, learning-based methods, in particular deep learning-based methods, have been extensively investigated, and their accuracies have been significantly improved. The details of the review on emotion recognition can be found in [23,24].

This study considers two major contributions: the first is expression-based emotion recognition, and the second is process-focused assessment (PFA). Between the two contributions, expression-based emotion recognition is significantly considered in this study. PFA (also called process-focused evaluation) is one of the methodologies used to evaluate the learning and studying performances of students. PFA has provided new approaches in the field of student assessment. In the past few decades, result-based assessment has been a major method for monitoring the learning performance of students, for example during midterm and final exams in classrooms. However, recently, monitoring how students understand course materials has received significant attention in the field of education.

Although the application of expression recognition in education is scarce, it is necessary for the PFA of students' performance and concentration. In particular, remote education has been extensively used in distance education for part-time students who face difficulties in physically attending classes. Moreover, in the case of remote education, it is difficult to assess or evaluate how students feel about the level of difficulty of problems or assignments given to them in class. In addition to technical perspectives, PFA has become popular in educational fields because, recently, result-based evaluation has been considered as not effective [25]. Because there are differences between students' abilities in learning, understanding, problem solving, etc., it is necessary for teachers or advisors to carefully observe students in real time. The emotions of students in class are worth observing and significantly affect education (both offline and online learning). Some studies have proposed estimating the emotional state of students by observing or quantifying the movement of the eyes, head, entire face, and other parts that are related to learning activities [26,27]. In a recent study, an approach was proposed to evaluate students' engagement in the virtual environment of education [28]. We use facial images to recognize expression-based emotions of persons solving problems (in the experiments, the test problems are related to programming languages). If the emotion of each student can be recognized using facial expression (from visual images), it can be a significant clue to teachers or instructors, so that they can adjust the level of difficulty or rearrange course materials that are suitable for each student's ability. In the problem solving test, the video camera records their face in real time, and the sequence of facial images is stored. The expressions are categorized into three classes: easy, neutral, and difficult; however, several studies on expression recognition categorize them into seven classes: happy, angry, sad, neutral, surprise, disgust, and fear [29]. It is challenging to recognize emotions based on facial expressions when solving assigned problems because there are variations in the facial position. Therefore, the sequence of facial images is recorded until the students compare their answers to the true one. The expressions of the person vary with time, and a 2D image of a face with the expression is recorded and used as an input for the deep neural network (DNN) model. The output of the model is categorized into three classes. Based on the approach proposed in this paper, the results can be applied to PFA in the field of education as follows:

- The ability to understand materials can be observed in real time.
- Based on the learned expression data, teachers can determine the next level of difficulty of a problem (or studying materials).
- Teachers can prepare teaching materials in a more precise manner such that the materials reflect the learning ability of each student.

The rest of this paper is organized as follows. Section 2 describes the overall architecture of the proposed approach, and Section 3 details the detection and classification of facial expressions based on the sequence of 2D facial images. Section 4 verifies the proposed approach by presenting the experimental results, and Section 5 concludes this paper.

2. Overall Architecture

This section describes the overflow of the proposed algorithm for the recognition of facial expressions that is used to develop an evaluation framework to achieve and focus on PFA, instead of result-based evaluation. The proposed approach mainly employs a DNN model; the DNN is a popular method for recognition and classification. Rather than face recognition or expression recognition, this study focuses on developing a framework to continuously evaluate and monitor students' engagement and concentration in solving the problems. All of these problems are official exam problems provided by the national agency of South Korea. To achieve our purpose in this study, we use facial images that are defined in 2D space and stored as a sequence of 2D images using a video camera. As the paradigm of education changes from result-based assessment to process-based assessment, continuously observing students' faces is required to establish a technical framework for expression-based emotion recognition. This type of evaluation system is necessary for offline and online classes. The proposed approach provides image-based recognition of

facial expressions, and we categorize the expressions into three classes, each of which corresponds to the level of difficulty of the problems. The expressions are recorded in a real-time manner with a sequence of facial images. The overall flow of this study is shown in Figure 1.

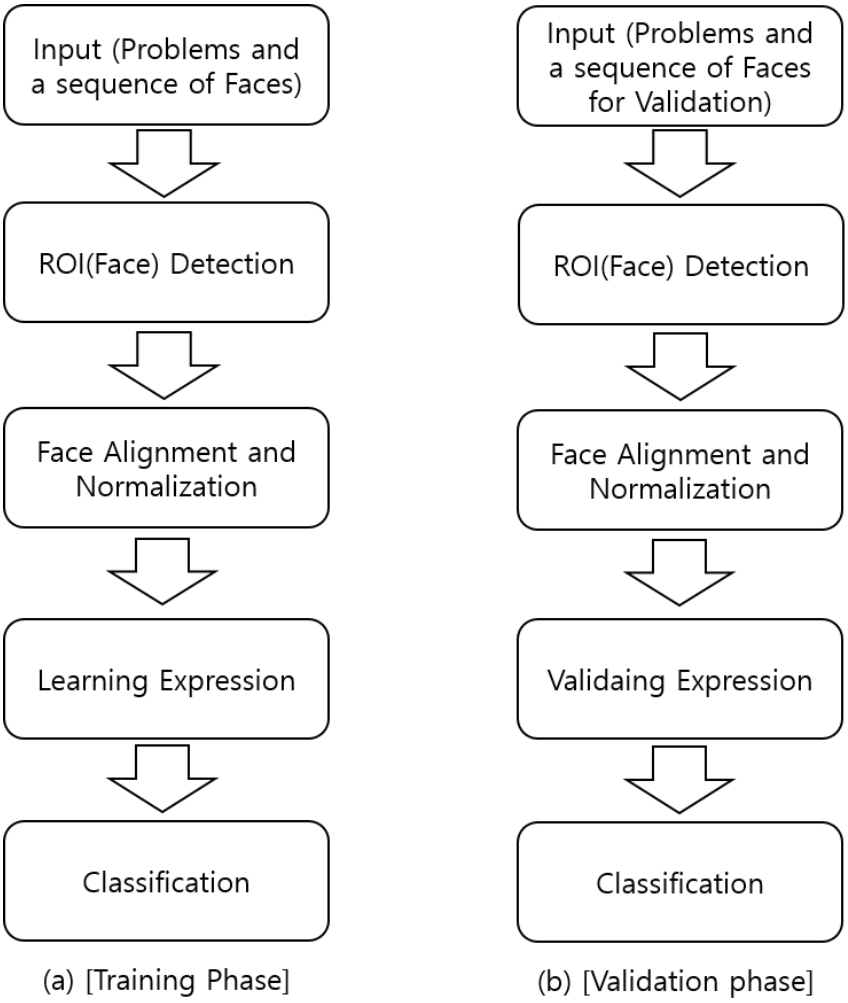

Figure 1. Overall flow of the proposed approach for process-focused assessment (PFA) using recorded face images. The input for this system is a sequence of face images and problems categorized into three classes: easy, neutral, and difficult. (**a**) Training phase to learn facial expressions using a sequence of 2D images that are recorded in real time. (**b**) Validation phase to test and apply the proposed approach on face images.

The participants of the experiments in this study are undergraduate and graduate students majoring in computer science engineering. Once the participants are given the problems (in the experiments, all of the problems are displayed on a computer monitor), they start solving them by watching the monitor, and they occasionally require a pencil and paper to write and summarize their ideas. We assume that the facial expressions of the students vary according to the difficulties of the problems. Because the proposed approach is based on facial images (in 2D space), we need to preprocess the input images (from the recorded images, a sequence of multiple facial images is stored and used as the input for the recognition system) such as face detection, smoothing, alignment (registration), and normalization. Preprocessing ensures that facial expressions with the same or different persons are properly compared. The former is an intra-class comparison, and the latter is an inter-class comparison. In the course of face detection, it employs AdaBoost and Haar feature-based detection algorithms to detect the regions of human faces by extracting the feature points

of a face using a cascade function [30]. In the learning phase, a DNN is applied to the face images, and the classification is performed by categorizing the expression into three classes: easy, neutral, and difficult. The accuracy of classification (learning and validation accuracy) is improved with the increase of the epochs (the number of iterations of learning via the DNN model). This study employs a CNN model to recognize and classify facial expressions. In the study conducted by Correa et al., they classified facial expressions into seven classes: angry, disgusted, fearful, happy, neutral, sad, and surprise [31]. However, in this study, we focus on categorizing facial expressions into three classes based on the assumption that the expression varies according to the level of difficulty of the problem. Once the learning procedures of the expressions and classification are complete, validation is performed using the test data. This approach can be applied to PFA during the real-time monitoring of students who are solving problems. Based on the observation of facial expression or its variation, the appropriate feedback or adjustment of the difficulty level of the remaining problems can be possibly performed from the educational perspective.

3. Detection and Classification of Facial Expression

This section presents the details of an important contribution in this study, i.e., the detection of the face and classification for expression recognition. Detecting the regions of interest is a crucial and fundamental stage in the fields of image processing, computer vision, and pattern recognition. In addition to those areas, to achieve classification using machine learning techniques, detecting the region of interest should be performed a priori. This study involves the detection of a region of a human face in real time from the captured images (i.e., sequence of recorded images), followed by the recognition and classification of expressions; therefore, the region of the human face should be detected from the video sequence. In the course of face detection, the recorded sequences of images are partitioned into frames, and face detection is performed in each frame. A rectangular shape is used to present the result of the face detection, and these bounded faces are stored as datasets that are to be used for the classification procedure. Once detection is performed, the result, which is described using a bounded rectangular box, is resized to sizes of 48×48 and 227×227, such that different image sizes can be used for the learning procedure, leading to a detection and classification system that is robust to variations in size or scale. In addition, we can investigate if the resolution of the input image affects the accuracy of the classification. Once the region of a face from a sequence of frames is successfully detected, the feature points are extracted; the extraction should be robust to the scale of the image. In this procedure, a DNN is employed. In this study, to extract feature points such as eyes, nose, and mouth, the adaptive boosting algorithm (AdaBoost) and Haar feature-based cascade are used [30,32]. The AdaBoost algorithm determines the common characteristics of the human face in the captured images by applying a weak classifier; then, numerous weak classifiers are linearly added to generate a strong classifier by optimizing the weighted parameters as follows.

$$\mathbf{f}(\overline{\mathbf{x}}) = \sum_{i=1}^{N} \boldsymbol{\alpha}_i^T \mathbf{h_i}(\overline{\mathbf{x}}), \tag{1}$$

where $\mathbf{f}(\overline{\mathbf{x}})$, $\boldsymbol{\alpha}_i$ ($\boldsymbol{\alpha}_i^T$ is the transpose of $\boldsymbol{\alpha}_i$) and $\mathbf{h_i}(\overline{\mathbf{x}})$ are the strong classifiers, weighted parameters, and weak classifiers, respectively. ($\overline{\mathbf{x}} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_M}\}$) represents the image data defined in a 2D space, that is $\overline{\mathbf{x}} \in \mathbb{R}^{M \times M}$ and $x_1 \subset \mathbb{R}^{M \times 1}$. The region of the face is detected and bounded by a rectangular box; thus, the size of the detected region of interest is $M \times M$. $\mathbf{h_i}(\overline{\mathbf{x}})$s are added with iteration, and at the n^{th} iteration ($i = n$), $\boldsymbol{\alpha}_{\boldsymbol{i=n}}$ is updated in an adaptive manner. Consequently, the final step of the classifier $\mathbf{f}(\overline{\mathbf{x}})$ performs well. In the case of face detection, our $\mathbf{f}(\overline{\mathbf{x}})$ shows improved feature extraction accuracy. Similar to linear classification (e.g., single layer perception or linear regression), $\mathbf{h_i}(\overline{\mathbf{x}})$ is applied to input data $\overline{\mathbf{x}}$ and the optimized current state of feature extraction (in general). This is a classification that classifies the eye(s), nose, mouth, etc., in the face image and determines

α_i. The AdaBoost algorithm is considered suitable for face detection with Haar-like feature extraction, which is an appropriate method for detecting feature points in a single target object. Because our target image is a single face in a captured image, the proposed approach uses AdaBoost and Haar-based feature extraction. Given a color image including a face, it is transformed into a grayscale image, and features such as eyes, nose, and mouth are detected. To classify facial expressions based on whether a participant is watching the problems to be solved, DNN-based classification is employed. In particular, the CNN model categorizes the facial expressions into three: easy, neutral, and difficult. In this study, we utilized the FER2013 database to perform learning procedures and compare the proposed method with those developed by studies on classification [33]. To validate the proposed approach, real faces are captured by a video camera, and the learning process is performed in three cases as follows.

- Case 1: A participant is confronting and solving an easy problem.
- Case 2: A participant is confronting and solving a neutral problem.
- Case 3: A participant is confronting and solving a difficult problem.

The hypothesis is that the facial expression varies with the level of difficulties, and our approach increases the size of the first layer of CNN when using images of higher resolution. To classify facial expressions after the face region is detected, the face that is bounded by a rectangular box is used as the input for the CNN-based model that classifies the input images into three classes. Figure 2 depicts the procedures for classifying facial expressions, and Figure 3 depicts the model that is based on the CNN, which is employed in this study [31]. The input data are the segmented region of a face (face detection is performed prior to entering the DNN model), whose size is 48×48 or 227×227. Moreover, 3×3 kernels and ReLU function ($f(x)$) (from Equation (2)) are used for the learning and activation functions, respectively.

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

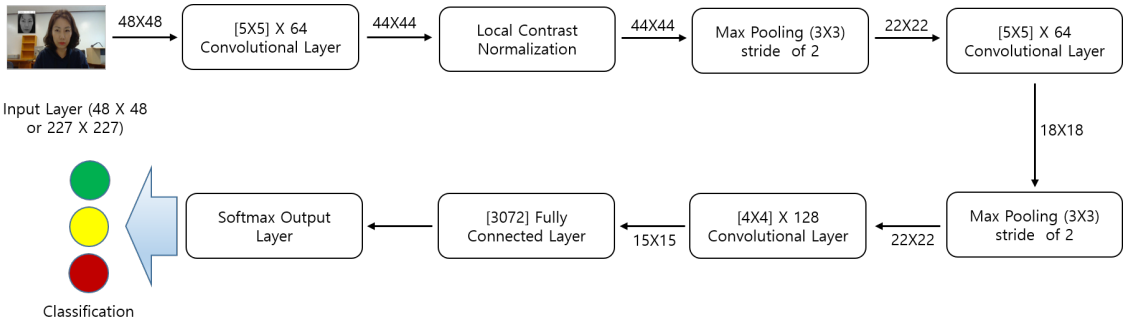

Figure 2. Overall procedure for expression classification using deep neural network (DNN) (convolutional neural networks (CNN) model).

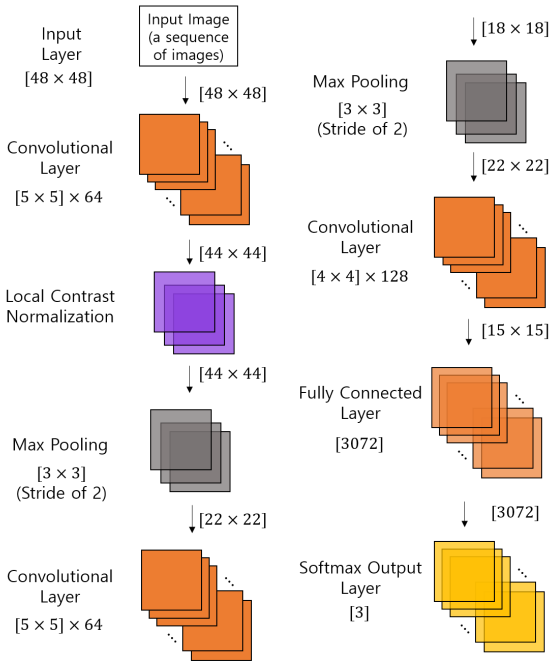

Figure 3. Basic model employed for expression recognition.

In the proposed model, analogous to other studies using CNN, it is important to design the number of hidden layers and neurons in each layer. It is also important to determine the activation function and parameters that affect the filters applied to the input of each layer (it can be the original input or the output of another layer). The parameters for this model are optimized via error back-propagation. As the participant solves the problems, the recorded sequence of images is classified into three expressions: easy, neutral, and difficult. The detected faces are aligned and normalized, and the result of detection is represented as a 2D face bounded by a rectangle. In the recorded face, the extent of concentration and immersion in studying materials can also be evaluated. This procedure enables us (in particular, teachers or instructors) to achieve PFA. Figure 4 shows the details of the process of training and validation for the classification of facial expressions.

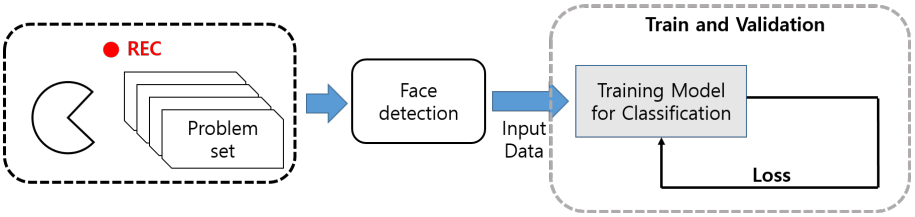

Figure 4. Overall flow of a system for classifying facial expressions using DNN.

To calculate the loss, a cross-entropy function is employed in this study. Because the expression is categorized into three classes, the output of the softmax function (y) can be written as follows:

$$y_i = \frac{e^{E_i}}{\sum_{j=1}^{3} e^{E_j}}, \tag{3}$$

where E_i is the i^{th} expression of a face; for example, $i = 1$, $i = 2$, and $i = 3$ correspond to easy, neutral, and difficult, respectively. The cross-entropy function is expressed as follows:

$$H_{\text{cross}} = -\sum_{j=1}^{3} t_i log(y_i), \tag{4}$$

where H_{cross} represents the cross-entropy and t_i and y_i represent the true and estimated values, respectively. The aforementioned procedures are iterated to optimize the parameters that are included in the network model. The adaptive moment (Adam) optimizer is used [34].

4. Experimental Results

To substantiate the proposed method, we present the experimental results in this section. The system environment is presented in Table 1.

Table 1. Experimental environment.

Category	Version
Operation system	Window 10
CPU	Intel(R) Core(TM) i5-8250U CPU @1.80 GHz 1.60 GHz
System type	64 bits
Memory (RAM)	8.0 GB
Simulation environment	Anaconda 4.7.5

Python 3.6.8, OpenCV-Python 4.1.0.25, Numpy 1.16.4, Keras 2.2.4, and TensorFlow 1.13.1 are used to fully exploit the deep learning library. We aim to accomplish a framework for PFA by employing the concept of the classification of facial expressions using a DNN model. First, to record and recognize facial expressions, an appropriate selection of the problems enables us to observe variations in facial expressions while solving the problems. The overall flow of the experiments is shown in Figure 5. The problems are selected for students majoring in computer science and engineering (one student's major is computer education). They used two personal computers; i.e., one for displaying problems to be solved and the other for recording. Once problem solving starts, a video camera (usually attached to a laptop PC) simultaneously starts recording the face. The recording ends if the participants finish solving the problems and comparing the true answer with the selected answer. When participants are solving problems and comparing their answers to the true ones, their expressions vary. This variation plays a key role in recognition and classification. In the training phase, the expressions are categorized into three classes, each of which is assigned according to the student's emotion and whether the selected answer is true or not. Once the training phase is complete, other images (faces) are used as validation datasets. In this experiment, we recorded 27,110 images with a 48×48 resolution and 18,338 images with a 227×227 resolution. Some images are used as the training set, and the rest are used as validation datasets (the ratio of the training and validation sets is approximately 4:1).

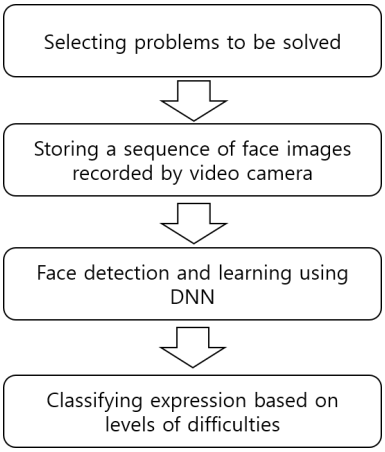

Figure 5. Overall flow of the experiments conducted in this research to establish a framework for process-focused assessment (PFA).

There are 45 problems, and the participant is given approximately 5 min to solve and check the true answer for each problem. The selected problems are from officially approved exams, and the level of difficulty is selected based on the percentages of correct answers that are also officially published (from the official organization in South Korea). An example is shown in Figure 6.

```
#include <stdio.h>

int main ()
{
     int a, b;
     a = b = 1;
     if (a = 2)
          b = a + 1;
     else if (a == 1)
          b = b + 1;
     else
          b = 10;
     printf("%d, %d\n", a, b)
}
```

Question : What are final output values of "a" and "b"?

① 2, 3

② 2, 2

③ 1, 2

④ 2, 10

Figure 6. Example of the problem assigned to a student in this experiment (from https://www.gosi.kr/). Question: Find an answer that shows the final output of this source code (since the question is in Korean, we have translated it to English).

To magnify the variation of facial expression for our study, the problems require different levels of the depth of thought. Once the problem is solved, the camera starts to record the face and finishes recording once the participant compares the true answer with his/her answer, such that all variation in the facial expression can be observed. Analysis and classification are performed after all the facial images are stored in memory. This is because real-time analysis and classification are difficult owing to the hardware performance of laptop PCs. Once the participants check the correct answer, the result of the answer's correctness affects their facial expressions; thus, the recording is conducted until correctness is confirmed. To record the expression, two cameras are used, each of which records the face from different perspectives. The resolution of the recorded image is 1280×720, and the frame rate is 30 frames per second. The region of a face is detected with a feature extraction algorithm, and the dataset is categorized into two groups: training and validation. The selection of problems is randomly performed such that the participant cannot manipulate the experimental conditions. The examples of detected faces using the Haar cascade are presented in Figure 7.

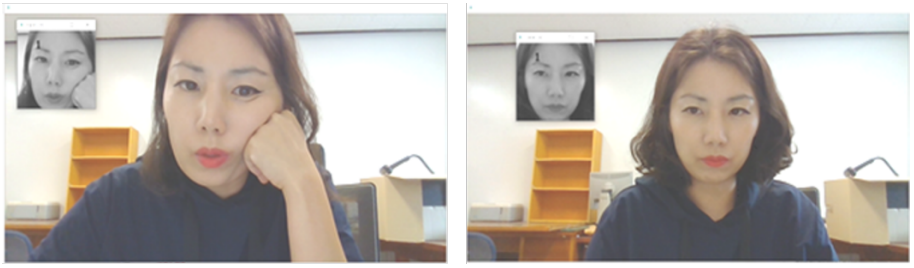

Figure 7. In face detection, the Haar cascade algorithm is employed, and the grayscale image is stored and used as the input for the learning and training processes.

The datasets for the experiments are summarized in Tables 2 and 3.

Table 2. Dataset (size 48 × 48) used for the experiments.

Classification	Training Set	Validation Set	Total
Easy	7001	1780	8781
Hard	9922	2520	12,442
Neutral	4322	1565	5887
Total	21,245	5865	27,110

Table 3. Dataset (size 227 × 227) used for the experiments.

Classification	Training Set	Validation Set	Total
Easy	5220	1308	6537
Hard	7229	1810	9039
Neutral	2217	554	2711
Total	14,666	3672	18,338

In setting up the hyperparameters of the network model, the training step (parameter `train_step`), size of the batch (`batch_size`), learning rate (`learning_rate`), and rate of dropping a number of neurons out of the total number of neurons (`dropout_keep_prob`) are 5000, 30, 10^{-4}, and 0.7, respectively. Once the training process is complete, validation is performed to evaluate the accuracy of the classification. Optimization is performed to avoid the overfitting problem. The size of the input image affects the optimization parameter, which prevents overfitting. In the experiments, to validate the effect of the number and resolution of the input images, three experiments are performed using different datasets, as shown in Table 4.

Table 4. Different experimental setups for training and validation.

Experiments Setup	Size (Resolution)	Training Set	Validation Set	Total
Setup I	48 × 48	2250	460	2710
Setup II	48 × 48	14,602	3488	18,090
Setup III	227 × 227	5871	1250	7121

To explain the experimental results, Table 5 presents the accuracy of training and validation.

Table 5. Loss and accuracy values of training and validation for classification.

Experiments Setup	Size (Resolution)	Number of Samples	Training Accuracy (%)	Validation Accuracy (%)
Goodfellow et al. [33]	48 × 48	35,887	–	64.24%
Setup I	48 × 48	2710	70	75
Setup II	48 × 48	18,090	52	65
Setup III	227 × 227	7121	83.9	82

In the first experiment (Setup I), the input image is 48 × 48, and two-thousand seven-hundred and ten images with 15 epochs are used. The results show that the optimal epoch is 10, and an epoch number larger than 15 leads to an overfitting problem. The loss and accuracy of the training and validation are shown in Figures 8–10. Furthermore, the number of datasets and the size of the input image affect the training and validation accuracy. We can conclude that, to achieve accurate results based on DNNs, the quality and quantity of input images are crucial for the overall performance. In the first case (Setup I), two-thousand seven-hundred and ten input images of 48 × 48 pixels are used. In Figure 8, once the epoch exceeds 10, overfitting is observed (loss or accuracy is sharply degraded).

Moreover, the accuracy and model loss are improved until the epoch reaches 10 in training and validation. This means that, as shown in Figure 9, the optimal number of iterations needs to be determined to achieve the optimal result. In Figure 9, the accuracy and model loss show better performance than in Figure 8, although the quantity of improvement is not significant. From the experiments in Figures 8 and 9, the number of input samples affects the performance, although the improvement is not significant. In Figure 10, the resolution of the input images increases from 48 × 48 to 227 × 227 (pixels × pixels). In Figure 10, from the observation of accuracy and model loss, the performance is significantly improved compared to the case of Setups I and II. Based on the experimental results, the classification of facial expressions is demonstrated to perform accurately with input images of high quality. Analogous to the findings from previous studies, particularly those on DNNs for classification, the quality of the input data plays a key role in the overall performance of the system.

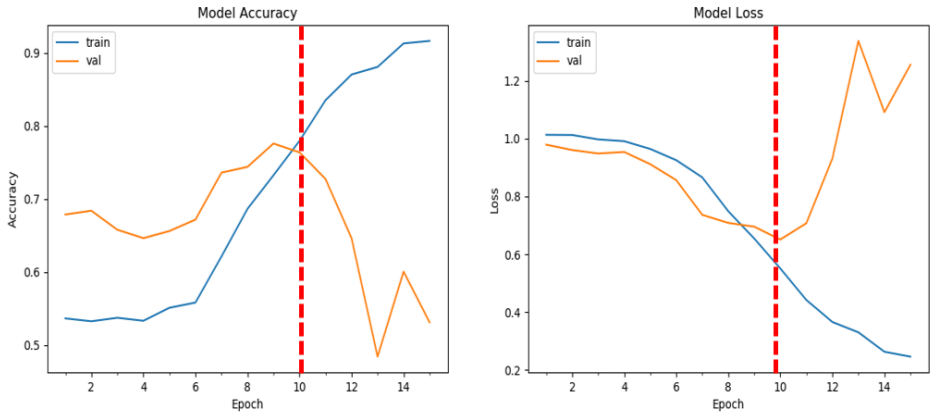

Figure 8. Accuracy and loss values of training and validation using 48 × 48 sized images (Setup I in Table 4).

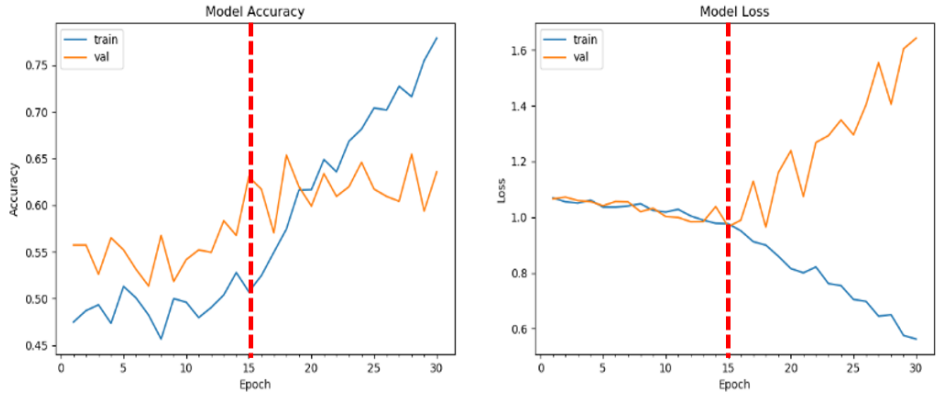

Figure 9. Accuracy and loss values of training and validation using 48 × 48 sized images (Setup II in Table 4).

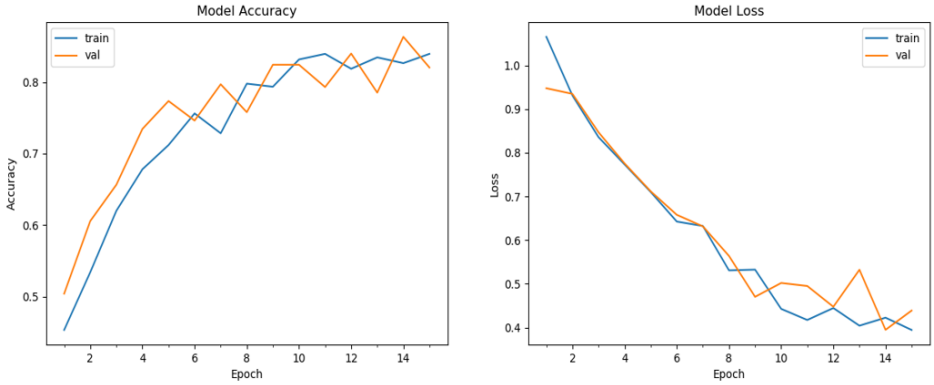

Figure 10. Accuracy and loss values of training and validation using 227 × 227 sized images (Setup III in Table 4).

From the experiment, the system to supervise and observe students' engagement in the classroom requires high-quality facial images. Once real-time classification is feasible, the instructor or teacher can arrange and adjust the environment for the study in a continuous real-time manner.

5. Conclusions

In this study, we propose an approach to classify facial expressions to achieve PFA. This study mainly considers the educational aspect by employing the concept of machine learning based on a DNN model. To achieve PFA, we record the facial expressions of participants while they solve problems. A recording was obtained using a video camera during the process of solving the problems, and it was stopped after the participant checked the answers. The model learned facial expressions based on a CNN for classifying and identifying the feelings of the participant on the difficulty level of the given problems. The experiments showed reasonable accuracy in training and validation. In this study, we have two important considerations. The quality of input images plays an important role in the overall accuracy for training and validation. Moreover, we cannot expect a significant improvement of overall accuracy only with an increase in the number of iterations (the accuracy increased slightly and not significantly). Notably, for recognition, this study employs a CNN; an open library, Keras, is used to implement the network. However, we focus on the recognition of facial expressions while students solve exam problems. The important aspect of this study is that the performance (i.e., accuracy) highly depends on the image quality. The number of datasets or iterations (epochs) highly affects the results, and this study shows that image quality is important. In future studies, we will attempt to increase the number of classes of expressions and face data for the testing phase, such that the teachers or supervisors can analyze the participants precisely, and PFA can be performed in a practical environment.

Author Contributions: Conceptualization, H.-J.L. and D.L.; methodology, H.-J.L.; validation, H.-J.L.; formal analysis, H.-J.L.; investigation, H.-J.L. and D.L.; data curation, H.-J.L.; writing—original draft preparation, H.-J.L.; writing—review and editing, D.L.; visualization, H.-J.L. and D.L.; supervision, D.L.; project administration, D.L.; funding acquisition, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partly supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2019R1G1A110017212).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to privacy. The data presented in this study are available on request from the corresponding author.

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

References

1. Loncomilla, P.; Ruiz-del-Solar, J.; Martinez, L. Object recognition using local invariant features for robotic applications: A survey. *Pattern Recognit.* **2016**, *60*, 499–514. [CrossRef]
2. Serban, A.; Poll, E.; Visser, J. Adversarial examples on object recognition: A comprehensive survey. *ACM Comput. Surv.* **2020**, *53*, 1–38. [CrossRef]
3. Bucak, S.; Jin, R.; Jain, A. Multiple kernel learning for visual object recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1354–1369. [PubMed]
4. Jain, A.; Duin, R.P.W.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37. [CrossRef]
5. Zhao, W.; Chellappa, R.; Phillips, P.J.; Rosenfeld, A. Face recognition: A literature survey. *ACM Comput. Surv.* **2003**, *35*, 1–61. [CrossRef]
6. Jafri, R.; Arbnia, H. A survey of face recognition techniques. *J. Inf. Process. Syst.* **2009**, *5*, 41–68. [CrossRef]
7. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188. [CrossRef]
8. Pantic, M.; Patras, I. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Syst. Man Cybern. Part B* **2006**, *36*, 433–449. [CrossRef]
9. Fasel, B.; Luettin, J. Automatic facial expression analysis: A survey. *Pattern Recognit.* **2003**, *36*, 259–275. [CrossRef]
10. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality Reduction by Learning an Invariant Mapping. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; pp. 1735–1742.
11. Delac, K.; Grgic, M.; Grgic, S. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *Int. J. Imaging Syst. Technol.* **2006**, *15*, 252–260. [CrossRef]
12. Seow, M.-J.; Tompkins, R.C.; Asari, V.K. A New Nonlinear Dimensionality Reduction Technique for Pose and Lighting Invariant Face Recognition. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)—Workshops, San Diego, CA, USA, 21–23 September 2005.
13. Huang, W.; Yin, H. On nonlinear dimensionality reduction for face recognition. *Image Vis. Comput.* **2012**, *30*, 355–366. [CrossRef]
14. Wismüller, A.; Verleysen, M.; Aupetit, M.; Lee, J.A. Recent Advances in Nonlinear Dimensionality Reduction, Manifold and Topological Learning. In Proceedings of the ESANN 2010 Proceedings, European Symposium on Artificial Neural Networks—Computational Intelligence and Machine Learning, Bruges, Belgium, 28–30 April 2010; pp. 71–80.
15. Cao, L.J.; Chua, K.S.; Chong, W.K.; Lee, H.P.; Gu, Q.M. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing* **2003**, *55*, 321–336. [CrossRef]
16. Geng, X.; Zhan, D.-C.; Zhuo, Z.-H. Supervised nonlinear dimensionality reduction for visualization and classification. *IEEE Trans. Syst. Man Cybern. Part B* **2005**, *35*, 1098–1107. [CrossRef]
17. Lee, D.; Krim, H. 3D face recognition in the Fourier domain using deformed circular curves. *Multidimens. Sys. Signal Process.* **2017**, *28*, 105–127. [CrossRef]
18. Drira, H.; Amor, B.; Srivastava, A.; Daoudi, M.; Slama, R. 3D face recognition under expressions, occlusions, and pose variations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2270–2283. [CrossRef] [PubMed]
19. Kuanar, S.; Athitsos, V.; Pradhan, N.; Mishra, A.; Rao, K.R. Cognitive Analysis of Working Memory Load from EEG, by a Deep Recurrent Neural Network. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2576–2580.
20. Wu, D.; Zheng, S.-J.; Zhang, X.-P.; Yuan, C.-A.; Cheng, F.; Zhao, Y.; Lin, Y.-J.; Zhao, Z.-Q.; Jiang, Y.-L.; Huang, D.-S. Deep learning-based methods for person re-identification: A comprehensive review. *Neurocomputing* **2019**, *337*, 354–371. [CrossRef]
21. Kuanar, S.; Rao, K.R.; Blias, M.; Bredow, J. Adaptive CU mode selection in HEVC intra prediction: A deep learning approach. *Circuits Syst. Signal Process.* **2019**, *38*, 5081–5102. [CrossRef]
22. Wu, Y.; Ji, Q. Facial landmark detection: A literature survey. *Int. J. Comput. Vis.* **2019**, *127*, 115–142. [CrossRef]
23. Ko, B.C. A brief review of facial emotion recognition based on visual information *Sensors* **2018**, *18*, 401. [CrossRef]
24. Kumari, J.; Rajesh, R.; Pooja, K.M. Facial expression recognition: A survey. *Procedia Comput. Sci. Second Int. Symp. Comput. Vis. Internet* **2015**, *58*, 486–491. [CrossRef]
25. Lee, K.-H.; Kang, H.; Ko, E.-S.; Lee, D.-H.; Shin, B.; Lee, H.; Kim, S. Exploration of the direction for the practice of process-focused assessment. *J. Educ. Res. Math.* **2016**, *26*, 819–834.
26. Krithika, L.B.; Lakshmi, P.G.G. Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. *Procedia Comput. Sci.* **2016**, *85*, 767–776. [CrossRef]
27. Rao, K.; Chandra, M.; Rao, S. Assessment of students' comprehension using multimodal emotion recognition in e-learning environments. *J. Adv. Res. Dyn. Control Syst.* **2018**, *10*, 767–773.

28. Olivetti, E.C.; Violante, M.G.; Vezzetti, E.; Marcolin, F.; Eynard, B. Engagement evaluation in a virtual learning environment via facial expression recognition and self-reports: A preliminary approach. *Appl. Sci.* **2020**, *10*, 314. [CrossRef]
29. Tarnnowski, P.; Kolodziej, M.; Majkowski, A.; Rak, R. Emotion recognition using facial expressions. *Procedia Comput. Sci.* **2017**, *108*, 1175–1184. [CrossRef]
30. Ma, S.; Bai, L. A Face Detection Algorithm Based on AdaBoost and New Haar-Like Feature. In Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016; pp. 651–654.
31. Correa, E.; Jonker, A.; Ozo, M.; Stolk, R. *Emotion Recognition using Deep Convolutional Neural Network*; Tech. Report IN4015; TU Delft: Delft, The Netherlands, 2016; pp. 1–12.
32. Kwak, J.-H.; Woen, I.-Y.; Lee, C.-H. Learning algorithm for multiple distribution data using Haar-like features and decision tree. *KIPS Trans. Softw. Data Eng.* **2013**, *2*, 43–48. [CrossRef]
33. Goodfellow, J., II; Erhan, D.; Carrier, P.L.; Courville, A.; Mirza, M.; Hamner, B.; Cukierski, W.; Tang, Y.; Thaler, D.; Lee, D.-H.; et al. Challenges in representation learning: A report on three machine learning contests. *Neural Netw.* **2015**, *64*, 59–63. [CrossRef]
34. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.