sustainability

MDPI

Article

Exploring Artificial Intelligence in Smart Education: Real-Time Classroom Behavior Analysis with Embedded Devices

Liujun Li [1], Chao Ping Chen [2], Lijun Wang [3], Kai Liang [4] and Weiyue Bao [5,*]

[1] School of Media and Art Design, Wenzhou Business College, Wenzhou 325035, China
[2] Smart Display Lab, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[3] SenseTime Education Research Institute, SenseTime Group Inc., Shanghai 201900, China
[4] Mel Science (Shanghai) Co., Ltd., Shanghai 200040, China
[5] School of Fine Arts, Shanghai Institute of Visual Arts, Shanghai 201620, China
* Correspondence: baoweiyue@siva.edu.cn

Abstract: Modern education has undergone tremendous progress, and a large number of advanced devices and technologies have been introduced into the teaching process. We explore the application of artificial intelligence to education, using AI devices for classroom behavior analysis. Embedded systems are special-purpose computer systems tailored to an application. Embedded system hardware for wearable devices is often characterized by low computing power and small storage, and it cannot run complex models. We apply lightweight models to embedded devices to achieve real-time emotion recognition. When teachers teach in the classroom, embedded portable devices can collect images in real-time and identify and count students' emotions. Teachers can adjust teaching methods and obtain better teaching results through feedback on students' learning status. Our optimized lightweight model PIDM runs on low-computing embedded devices with fast response times and reliable accuracy, which can be effectively used in the classroom. Compared with traditional post-class analysis, our method is real-time and gives teachers timely feedback during teaching. The experiments in the control group showed that after using smart devices, the classroom teaching effect increased by 9.44%. Intelligent embedded devices can help teachers keep abreast of students' learning status and promote the improvement of classroom teaching quality.

Keywords: education science; emotion recognition; behavior analysis; embedded devices

Citation: Li, L.; Chen, C.P.; Wang, L.; Liang, K.; Bao, W. Exploring Artificial Intelligence in Smart Education: Real-Time Classroom Behavior Analysis with Embedded Devices. *Sustainability* **2023**, *15*, 7940. https://doi.org/10.3390/su15107940

Academic Editor: Hao-Chiang Koong Lin

Received: 1 March 2023
Revised: 8 May 2023
Accepted: 9 May 2023
Published: 12 May 2023

1. Introduction

1.1. Classroom Behavior Analysis

Classroom behavior analysis has received increasing attention from researchers in the fields of education and psychology. Classroom instruction, as a fundamental form of teaching and learning, has always been at the heart of education. Student behavior in the classroom is crucial to the educational process and cannot be ignored. Obtaining information about student behavior not only helps to grasp the learning, personality, and psychological characteristics of students but is also an effective way to evaluate the quality of education. In the classroom, student behaviors include body movements, language, and emotions. Classroom emotion analysis is a branch of classroom behavior analysis. With the development of camera equipment and AI technology, many scholars have begun to pay attention to classroom emotion analysis. H. Zeng proposed analyzing student engagement through classroom videos [1]. Putra, W.B. proposed understanding students' learning stress through emotional recognition [2].

Teachers teach mainly in classrooms, which is a one-to-many format, and it is difficult for teachers to evaluate each student's classroom performance in depth and comprehensively in the classroom. In the classroom, teachers obtain feedback by observing students' facial emotions or behaviors during class. However, in the classroom, students' behaviors

are complex and variable. Therefore, it is difficult for teachers to obtain a comprehensive understanding of students' performance in the classroom. This is because their main focus is on teaching. This is not conducive to the improvement of teaching quality and students' development.

The traditional classroom behavior analysis method focuses on analysis scales. Observers record teaching behaviors in the classroom based on the developed scales, which are then computed and analyzed. The observer records the verbal interaction behaviors of teachers and students to form an interaction analysis matrix. Using the analysis scale approach, teachers are able to conduct classroom behavior analyses and reflect on their teaching. The analysis scale method requires specialized observers to record classroom situations for analysis, which is time-consuming and costly. Moreover, in the analysis scale method, an analysis is conducted after the fact and does not provide support for teaching in real-time during class.

Modern teaching has undergone tremendous progress, and a large number of advanced devices and technologies have been introduced into the teaching process. Video and projection equipment is used in teaching, and some scholars have developed video analysis systems based on analysis scales [3]. With this video tool, classroom behavioral details can be recorded accurately. In addition, the scale scores analyzed by the system can be recorded without the influence of subjective observations and omissions, which is a great step forward in scale analysis. However, the video tool method requires the recording of classroom videos for analysis, which has a pitfall. The subjects for conducting classroom behavior analysis are students, mostly minors. Video analysis requires recording student video information, and there is a potential risk of student privacy leakage.

1.2. Requirement

Classroom behavior analysis is helpful in grasping students' learning, analyzing teachers' teaching effectiveness, and improving teaching quality. We hope to put new technology and new equipment into education to promote educational development. Taylor and other scholars [4,5] proposed the requirement of rapid behavior analysis for classroom teachers.

1. Privacy. Classroom behavior analysis should capture and analyze students' classroom behaviors but also protect students' personal privacy. Students' personal privacy includes two parts: biometric privacy and personal behavioral privacy. Biometric privacy is the student's image data, mainly facial images. Personal behavioral privacy is mainly students' special behaviors in the classroom, such as confrontational emotions, negative emotions, etc. The leakage of these data may have adverse effects on students' lives and psychology.

2. Real-time. The purpose of classroom behavior analysis is to help teachers improve the quality of teaching and learning. Post-event analysis methods can help teachers reflect on problems in teaching, but with a long interval of time, they may be omitted. It would be more helpful if the statistics of student behavior could be used to remind teachers to adjust their teaching style in real time. For example, if the classroom behavior analysis system counts that more than 40% of students are confused, teachers can see the statistics and know that many students are not listening and can adjust their teaching style by asking appropriate questions or repeating the lecture in a different way. This will help students master the teaching content and improve the teaching effect.

3. Efficient. Teachers have a heavy teaching load and cannot spend much time learning to use the equipment and operate the analysis system. The classroom behavior analysis system is a tool to help teachers improve the quality of their teaching, and it should be easy to operate and give the analysis results visually.

4. Data-based. Analysis of classroom behavior is a long-term endeavor, and the data gathered from classroom behavior only a few times may not be very useful. The data accumulated over a long period of time help teachers analyze their own teaching situation at various times and in different classes for comparative analysis. The comparative analysis of

classroom behavior analysis data from different teachers is also helpful in finding effective teaching approaches to deal with various classroom behaviors.

1.3. Challenges

In modern teaching, classroom behavior analysis is still an effective method to improve the quality of teaching and learning. Students from different regions at different times have their own characteristics. For example, Asian students are more restrained in class and less interactive with teachers. In contrast, European and American students are more active in class and like to interact with teachers. Classroom behavior analysis helps teachers grasp students' characteristics, adjust teaching methods in a targeted manner, and improve teaching effectiveness. Traditional classroom behavior analysis has challenges, mainly the following three points.

1. The inability to protect privacy. The use of camera equipment makes classroom behavior analysis convenient but also poses security risks, as students' biometric information and personal behavior information may be compromised.

2. Time-consuming. Traditionally, it is an after-the-fact analysis, which is time-consuming and inefficient. Due to limited educational resources, it is impossible to provide every teacher with an observer and an analyst, and classroom behavior analysis is not time-sensitive and is of limited help to teachers.

3. Difficulty in data statistics and analysis. The improvement of teachers' teaching quality through classroom behavior analysis is a gradual improvement after long-term implementation. Classroom behavior analysis data are desensitized statistics that need to be accumulated, compared, and analyzed over a long period of time in order to provide additional help to teachers. Ordinary teachers generally do not have such an ability and need help from information technology systems.

1.4. Research Content

We propose the solution of embedded devices plus AI to apply intelligent and portable embedded devices to analyze students' behavior in classroom teaching. It can facilitate visual analysis and evaluation of students' classroom behavior and learning motivation in the teaching process so that teachers can make real-time adjustments according to the actual situation in classroom teaching and improve the quality of teaching.

As the main focus of teachers in the classroom is teaching, the equipment and system for classroom behavior analysis must be easy to use. It must be convenient to view at any time. It must also not interfere with teachers' teaching behavior. We use a wearable MR (Mixed Reality) device, which is essentially a combination of a pico-projector, a camera, and an embedded system, and is a smart portable device. MR devices create an interactive feedback loop between the real world, the virtual world, and the user to enhance the user's experience with realism, real-time interaction, and conceptualization [6].

Embedded systems are special-purpose computer systems based on an application. Embedded devices can flexibly tailor software and hardware according to application scenarios to form a final system that meets requirements. As can be seen in Figure 1, an embedded MR glass is a miniaturized computer system. Embedded MR glasses are wearable devices, and application scenarios require a small size, light weight, and long battery life. Therefore, the design of MR glasses needs to tailor to the hardware and functions. The application scenarios of MR glasses determine that they have low CPU computing power, small storage capacity, and low energy consumption. Such an operating environment is difficult to run large-scale AI network models. Embedded MR glasses must use lightweight models capable of responding quickly in low-computing environments.

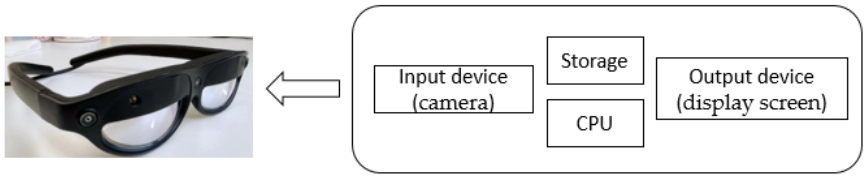

Figure 1. Schematic diagram of embedded MR device.

The use scenario of the MR device is shown in Figure 2, the camera of the MR device will shoot the real-time image and pass it to the embedded system, and the AI model in the embedded system of the MR device will judge the learning status of each student according to the recognition of facial emotion, and show the current classroom student learning status statistics so that teachers can understand the current student learning status, make a real-time adjustment and improve the teaching quality, for example, when students appear confused emotion, it should represent that the students are not understanding or grasping the teaching content, thus, they can ask appropriate questions and repeat the explanation; when some students appears tired, they can relax with light-hearted topics to soothe the atmosphere; when some students appears focused, it indicates concentration and a better teaching effect. Teachers' energy in the classroom is mainly focused on teaching, and it is impossible to observe students' dynamics at all times. Smart glasses with a classroom behavior analysis model can help teachers understand students' emotions and make corresponding adjustments in real-time to improve teaching quality.

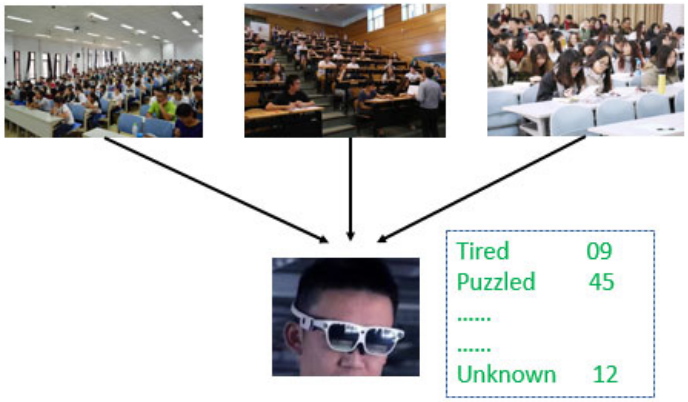

Figure 2. Scenarios of MR equipment used for classroom behavior analysis.

The process of using MR devices in classroom behavior analysis is shown in Figure 3.

Figure 3. Flow chart of classroom behavior analysis with MR device.

1. Image acquisition. The camera of the MR device acquires classroom image data.
2. Segmentation. Classroom behavior analysis is based on students' facial emotions. Therefore, the faces in the images need to be segmented into many separate images to prepare for AI model analysis.
3. Analysis. The analysis module is the core component of the system, which uses the AI model to identify the facial emotion of each segmented image. As a result of the principle of privacy protection, our AI model does not recognize physical features but only facial emotion features.
4. Data statistics. After the analysis of each segmented image is completed, the quantity is counted separately according to emotional features.

5. Display. MR's projector forms a virtual screen in front of a person's eyes to display statistical results, as shown in Figure 1. In order not to interfere with the normal sight of the teacher, the virtual screen was placed at the edge of the sight line of a moderate size.

From Figure 2, image acquisition and semantic segmentation are mature techniques and are not the focus of the study. The focus of the solution is the analysis module. Analysis modules need to meet the requirements of classroom behavior analysis in modern teaching. The modules should protect students' privacy and present analysis results quickly and efficiently. Therefore, the AI model of the analysis module does not only have the function of emotion recognition. We need to optimize the analysis module.

The workflow of the optimized analysis module is shown in Figure 4.

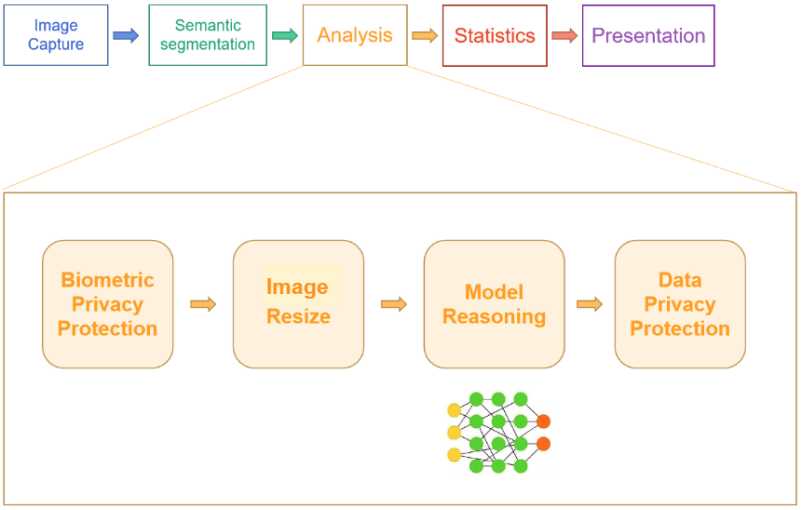

Figure 4. Optimization of analysis module.

1. Biometric privacy protection. The segmented facial images are protected for privacy. We call this biometric privacy protection because it protects the privacy of the facial features.

2. Image resize. The size of semantically segmented images varies. To ensure that the input size of the AI model is consistent, the images are resized uniformly here.

3. Model reasoning. The AI model needs to run in a low-power embedded device environment and be able to respond quickly.

4. Data privacy protection. The results of the recognition are privacy-protected, and the results of emotion recognition are not associated with individuals. In other words, the result of the AI model is not "Tom is paying attention to the lecture" but "there is a student who is paying attention to the lecture".

By optimizing the analysis module, the system can meet the requirements of classroom behavior analysis and run on embedded MR devices. This can protect students' privacy and present the analysis results quickly and efficiently.

2. Literature Review

2.1. AI Technology in Education

Since the late 1970s, AI has been used in education for rule-based expert systems, AI-based data processing systems since the mid-1980s, and machine learning-integrated AI since the mid-2000s. AI algorithms and systems are increasingly being used in education.

As seen in Figure 5, there has been an increase in the number of papers on the topic of AI applications in education since the mid-2010s. The applications of AI in education fall into two main categories.

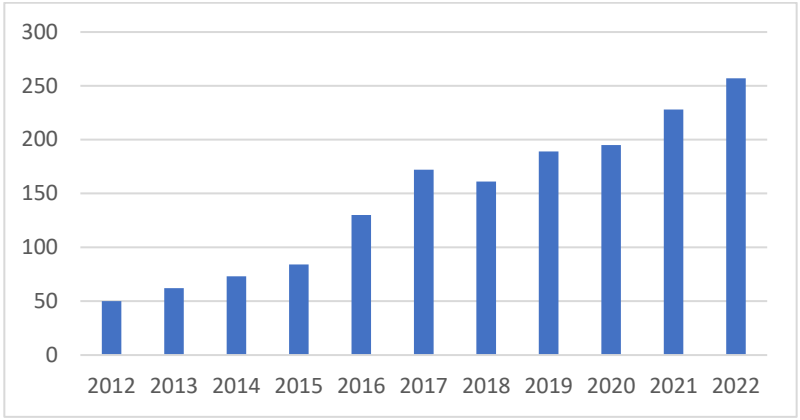

Figure 5. Papers on Google Scholar in the last ten years, with the keyword "AI" and "Education".

1. Improving learning effectiveness and the learning experience.

AI applications in this category mainly involve collecting and organizing learning materials, intelligent search, personalized recommendations, etc. Current developments are in the direction of intelligent teaching robots and virtual reality applications to demonstrate or present concepts and scenarios to students. These applications use 3D technology and highly interactive simulations as teaching tools, which help students better understand the concepts presented [7,8]. This category of AI applications is mainly aimed at learners to increase interest and deepen the learning experience.

2. Teaching management and assessment.

Artificial intelligence is used in education in a variety of forms and functions. As education evolves, researchers are trying to apply advanced AI techniques to deal with more complex issues, such as assisting in the performance of educational administration and managerial functions. It enables teachers to perform their administrative functions, such as grading and providing feedback to students, more effectively [9–11]. The application of AI can reduce teachers' paperwork and workload, especially in performing various administrative functions, thus allowing them to focus on their core task, teaching [12].

The application of AI to instructional administration is a critical component of current research. The classroom behavior analysis studied in this paper is one of the methods of teaching assessment that aims to assist teachers in improving the quality of their teaching.

2.2. Computer Vision Technology Applications

Computer vision is currently an area of intensive research activity in AI, with automated facial recognition (AFR) being a priority in AI applications in the context of education. Rafika [13] suggested using facial recognition to track student attendance. Roy [14] and Savchenko [15] propose the use of facial emotion recognition for identifying students' attention during online learning. This is a solution to the problem of decreasing student attention due to the lack of human supervision in online courses.

The above studies have explored the use of computer vision technology for educational applications but have overlooked two issues. The first is data leakage. The application of computer vision technology in educational scenarios mainly deals with underage students, and image data leakage may have a negative impact.

While facial recognition technology has brought great help to education, it has also brought many problems, and data privacy issues are receiving increasing attention. The way facial images are acquired and used may have negative effects due to data leakage [16]. Andrejevic [17] suggests that AFC technology can solve problems such as campus security, automatic registration, and student emotion detection, but data leakage is a cause for concern and worry.

The second is the running environment. Artificial intelligence vision models usually require high-performance servers to work, which is inconvenient in terms of use. If AI

models can run on mobile and embedded devices, it will greatly increase the ease of use. This would also allow for the wider use of AI in the education industry.

Many companies are working on lightweight models. The localized mode of operation of lightweight models reduces network consumption and maximizes the use of local computing resources. In addition, lightweight models no longer require high-performance servers and can be deployed on smart embedded devices. This makes the application of AI more promising and can be applied to more fields and scenarios [18,19]. However, there is less research on applying it to embedded devices in educational scenarios.

We propose a solution to apply AI to embedded devices. The solution optimizes a model that can adapt to the operating environment of embedded devices, avoiding the problem of data leakage and protecting student privacy.

3. Materials and Methods

3.1. Solution Structure

The embedded device plus AI solution applied to classroom behavior analysis requires models that are privacy-preserving, lightweight, and have a fast response. Based on the requirements of the application, we propose the following solution, as shown in Figure 6.

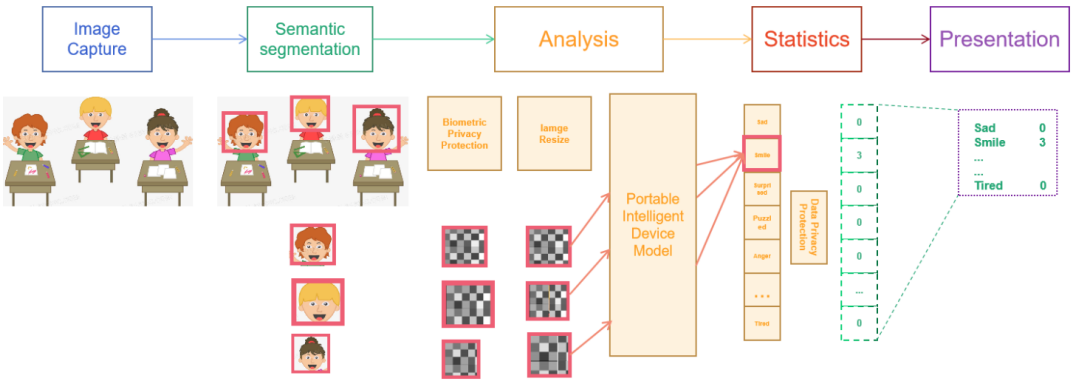

Figure 6. Solution structure.

1. MR device images capture classroom image data.

2. The classroom images contain multiple images of students, which cannot be used directly for analysis, and the faces in the images need to be segmented into many separate images to prepare for AI model analysis.

3. The analysis module is the core part of the model and contains four components

- Biological privacy protection. The Fourier transform is used to remove the low-frequency information from the images and process them into high-frequency images. The images thus processed are not recognizable to the naked eye but do not affect the training effect of the AI model.
- Uniform image size. The size of the images after semantic segmentation varies. To facilitate model training, we unify the size of the images to $224 \times 224 \times 3$.
- Visual model analysis. Vision models need to run on embedded devices, so they must be small in size, fast in response, and accurate. We propose the model PIDM (Portable Intelligent Device Model), which is only 76% of the size of MobileNet, with higher accuracy and a faster response rate.
- Behavior privacy protection. The results of the visual model analysis are not associated with individuals to protect the privacy of student behavior.

4. After the analysis of each image after segmentation is completed, the number of images is counted separately according to emotional features.

5. MR's projector forms a virtual screen in front of a person's eyes to display the statistical results.

The analysis module is the most important part of the whole system and is the main design and optimization part of this study, as shown in Figure 7.

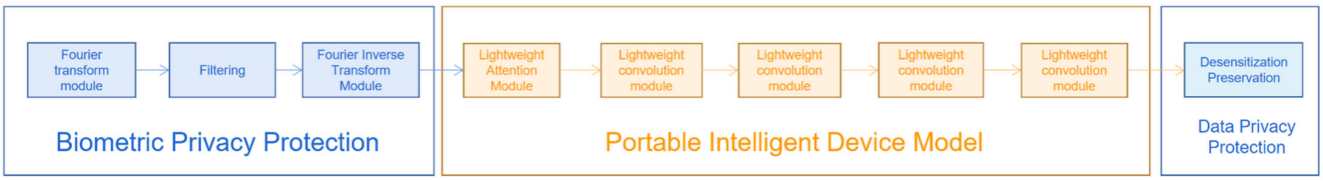

Figure 7. Structure of analysis module.

Our architecture starts with two privacy-preserving modules at the front and back to ensure the leakage of biometric information and data feature information.

The face images used for analysis are cut from a large image of the entire classroom. The quality of the images cannot be guaranteed. To improve the accuracy, the model needs to process the details better. Additionally, it is very subtle between multiple expressions. In some cases, crying and laughing are actually very similar. More attention to detail is needed to improve accuracy. Since the detailed analysis is at the front of the deep model, we put the attention module at the forefront of the whole model. The surface feature information is extracted to capture the details to improve accuracy.

3.2. Function Module Description

There are two function modules in the solution, privacy protection and the lightweight model.

3.2.1. Privacy Protection Module

Traditional privacy protection methods are data encryption and decryption, but they are not efficient for neural network applications. We adopted a privacy-preserving method based on Fourier high-channel filtering. In order to protect the biological features of the analyzed objects, we designed a picture feature processing module, as shown in Figure 8.

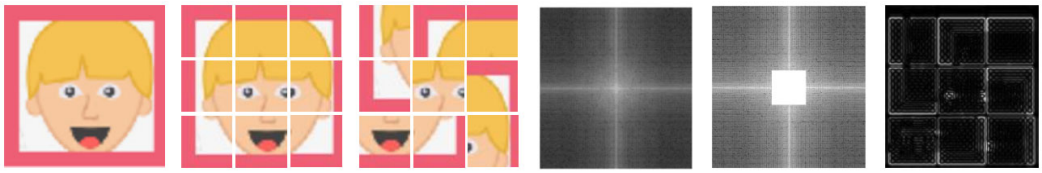

Figure 8. Privacy protection method based on Fourier high channel filtering.

In the first step, we split the image into small pieces and shuffle the order. This will increase the difficulty of naked-eye recognition but will not affect the recognition of AI models. The way the AI model recognizes the picture is through convolution operation, which is to divide the picture into many different feature pictures. Therefore, this operation is aimed at people and has no effect on the computer. In the second step, we use the Fourier transform technique to remove the low-frequency information of the image. For pictures, low-frequency information is generally picture texture and color, and high-frequency information is some edges and pixel-sharp areas. The shape is recognized by its high-frequency information, so the high-frequency information is retained, and the low-frequency information is clipped. The image with the low-frequency information removed is shown on the far right of Figure 8. They look like multiple little black squares. AI models can continue to use these images to identify the information in them.

The comparison in Figure 9 shows that the low-frequency pictures feel similar to the original pictures, while the high-frequency pictures are more different from the original pictures. This is because humans are more sensitive to low-frequency pictures. Therefore, the high-frequency pictures can ensure model training and increase the difficulty of naked

eye recognition. Through this processing, we can effectively protect image privacy and not affect the picture recognition effect.

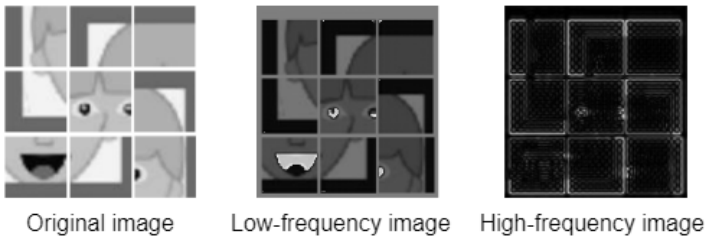

Figure 9. Image privacy protection processing.

3.2.2. Lightweight Model

This study proposes a high-performance vision model, PIDM, which needs to be applicable to the use scenario of smart portable devices with small computation and storage space that can respond quickly and has a certain degree of accuracy. The structure of the PIDM model is shown in Figure 10.

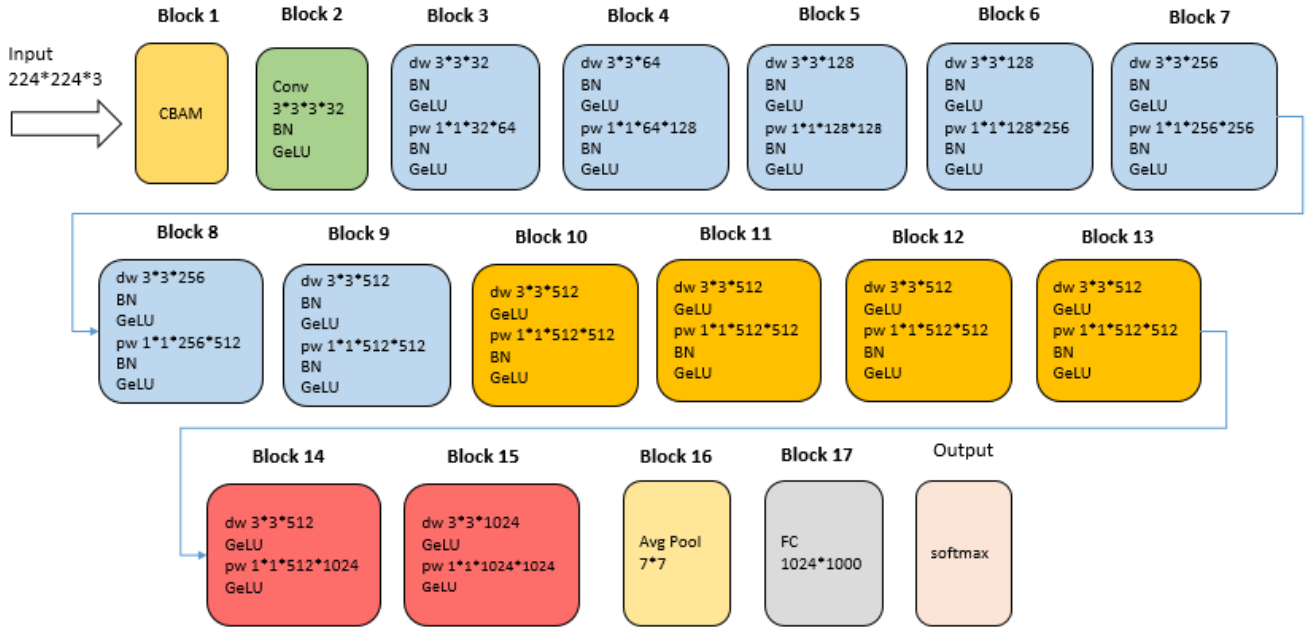

Figure 10. Structure of PIDM model.

1. Input

In the privacy protection module, the collected images are divided into 224 × 224 × 3 dimensions according to face. A commonly used public sentiment recognition dataset, such as Fer-2013, is of this size. Our input uses the 224 × 224 × 3 size to facilitate training with a public dataset.

2. Block 1

Block 1 is the CBAM (Convolutional Block Attention Module) module with the Attention Mechanism. The Attention Mechanism [20] is a data processing method in machine learning, which is widely used in different types of machine learning tasks, such as natural language processing, image recognition, and speech recognition. In layman's terms, the attention mechanism means that the network is expected to automatically learn what needs to be noticed in a sequence of images or text. For example, when the human eye looks at a picture, it does not allocate attention equally to all the pixels in the picture but instead allocates more attention to the areas that people pay attention to. From the

implementation point of view, the attention mechanism operates through a neural network to generate a mask; the value on the mask is a scoring that evaluates the current points that need attention.

Attention can be divided into three types, as follows: Channel Attention Mechanism, which generates a mask for the channel and scores it, is represented by SE (Squeeze-and-Excitation). The Spatial Attention Mechanism generates a mask for space and scores it. The CBAM [21] is a Hybrid Domain Attention Mechanism that evaluates both channel attention and spatial attention, as shown in Figure 11.

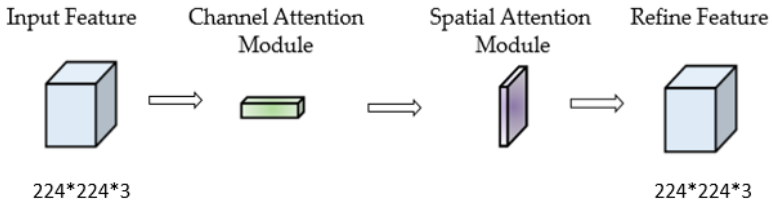

Figure 11. CBAM schematic diagram.

The CBAM is a lightweight, general-purpose module, and it can be seamlessly integrated into any neural network architecture with negligible additional overhead. The performance of the model is consistently improved after integrating CBAM into different models on different classification and detection datasets, demonstrating its wide applicability.

The amount of CBAM parameters is usually around 0.1 million, which can improve the accuracy of lightweight models without increasing the size of the model. CBAM does not change the input size and can be seamlessly integrated into the network model. In PIDM, CBAM is the first module that connects to the input layer to improve the model's accuracy.

3. Block 2

Block 2 is a regular convolution that extracts features from data.

In the PIDM model, we use the GeLU activation function (Gaussian Error Linear Units) instead of ReLU, which is commonly used in neural networks. The activation function was introduced to improve the neural network model's nonlinearity. The GeLU activation function introduces stochastic regularity in activation. This is a probabilistic description of neuronal input and is intuitively more in line with natural understanding, and the experimental results are better than ReLU. GeLU has been used in many of the most advanced models available today. The industry's top models, such as BERT and RoBERTa, use this activation function.

4. Block 3–Block 9

A total of 7 Blocks, from Block 3 to Block 9, use the standard DSC structure.

Sifre and MallatL [22] proposed the concept of Depthwise Separable Convolution (DSC). DSC is composed of two parts: Depthwise (DW) Convolution and Pointwise (PW) Convolution. One layer is used for filtering and the other for combination. This decomposition process can greatly reduce the amount of calculation and model size. Therefore, DSC is widely used in lightweight network models [23–25], and the famous MobileNet model also uses DSC.

The computational effort of DSC convolution is about 1/9 of that of ordinary convolution, and the computational effort is greatly reduced, but of course, this high efficiency of DSC comes at the expense of accuracy. At present, the design of high-efficiency and high-precision DSC variant modules is still one of the very hot research directions [26–28].

5. Block 10–Block 13

Among these 4 blocks, we use an optimized DSC structure named Block A. The BN layer after dw in DSC is removed in Block A. Block 10, Block 11, Block 12, and Block 13 are all in the form of Block A. This is the network pruning method, which aims to find redundant connections and eliminate them [29]. Recent studies have reduced the

structural redundancy of models in convolutional neural networks by reducing the number of channels and pruning BN layers [30]. Based on the study of BN, we believe that BN has the effect of improving learning speed and suppressing overfitting for deep neural network models, but the overuse of BN layers may also have an impact on deep neural network models.

Block A is applied in the middle part of the whole model, which can effectively reduce the model parameters without affecting the accuracy of the model.

6. Block 14 and Block 15

Block 14 and Block 15 use an optimized DSC structure named Block B. Block B removes the BN layer after dw and pw in the DSC. The BN layer has a considerable impact on the image classification problem when placed before the softmax output layer and destroys the role of the softmax output as a measure of uncertainty. Block 14 and Block 15 before the softmax layer of the model is both in the form of Block B, which prevents the BN layer from amplifying the fine feature weights and affecting the image classification effect.

7. Block 16

Block 16 is the average-pooling layer that reduces data dimensions. Compared with max-pooling, average-pooling emphasizes the downsampling of the overall feature information layer by layer, which reduces the dimension and is more conducive to passing the information to the next module for feature extraction.

8. Block 17

Block 17 is a fully connected layer. The fully connected layer converts the convolutional output two-dimensional feature map into a one-dimensional vector. The output of the fully connected layer is a highly refined feature that can be handed over to the final classifier or regression.

9. Output

The output layer uses the softmax function. In multi-classification problems, softmax is used as the activation function of the neural network output layer, which allows the model to output the probability value that the sample belongs to each category.

4. Experiment

4.1. Experimental Environment

4.1.1. Hardware and Software

In our experiments, we place the model training on the cloud server provided by Kaggle. The trained models are installed into the embedded devices for testing. We used two embedded devices running environment for testing: one is Amlogic T962 with Android 9.0, and the other is Raspberry Pi 3B with Windows NT 6.1; see Table 1 for details.

Table 1. Hardware and Software.

	Model Training Environment	**Amlogic T962 Embedded Environment**	**Raspberry Pi 3B Embedded Environment**
CPU	Kaggle kernels	Cortex-A53 1.5 GHz	Cortex-A53 1.2 GHz
RAM	13 G	2 G	1 G
GPU	p100 16 G	Mali 450 MP	Broadcom VideoCore IV
System	Kaggle platform	Android 9.0	Windows NT 6.1

4.1.2. Dataset

Three publicly available datasets for emotion recognition, FERPlus, ExpW, and AffectNet, were used in this experiment.

FERPlus is a re-labeling of the dataset FER-2013. The training set contains 28,709 samples, and the test set contains 3589 samples. Each sample is a grayscale image of a roughly centered

head with emotion annotation, which is suitable for facial emotion recognition model training. It is worth noting that the FER-2013 dataset has a large number of labeling errors, and in 2016, Microsoft re-labeled the FER-2013 dataset with three additional categories and named it the FERPlus dataset. The FERPlus dataset re-labeled by Microsoft is used in this experiment.

The Expression in-the-Wild (ExpW) dataset was used for facial expression recognition and contains 91,793 faces manually labeled with expressions. Each face image is annotated as belonging to one of seven basic expression categories.

AffectNet is a large facial expression dataset containing approximately 400,000 manually labeled images containing eight facial expressions. The AffectNet dataset is so large that we randomly selected a portion of it as the dataset.

The images commonly used in computer vision research are 224 × 224 pixels, and the input layer size of the pre-trained model provided by the Keras library is usually 224 × 224 × 3. For comparison with other models, the input layer size of our model is also defined as 224 × 224 × 3. This does not correspond to the image size provided by the dataset, e.g., the FERPlus dataset provides an image of 48 × 48 pixels. Therefore, we use the python imaging library (PIL) to pre-process and convert the size of the emotion recognition dataset to 224 × 224 × 3.

4.2. Purpose of the Experiment

The purpose of this study is to use MR devices for classroom behavior analysis, and experiments are used to test the efficiency and accuracy of AI models running on embedded devices in the following ways.

1. The PIDM model can run in a variety of embedded environments. Two embedded environments are prepared in the experiment: one is Amlogic T962 equipped with Android 9.0 system, and the other is Raspberry Pi 3B equipped with the Windows NT 6.1 system.

2. The effect of the PIDM model running on embedded devices is examined in terms of model size, accuracy, and response speed. We use the same dataset in our experiments for comparison with classical models such as VGG16, GoogLeNet InceptionV3, MobileNet, etc. We hope that the PIDM model has the smallest size, the highest accuracy, and the fastest response speed.

3. The effect of privacy-preserving treatment on accuracy. In the model, we subject the images to be segmented to privacy-preserving treatment in order to protect privacy. In the experiment, we want to compare the accuracy of the images without the privacy-preserving treatment.

5. Results

5.1. Running in an Embedded Device Environment

As shown in Table 2, the experiments demonstrate that PIDM can run well in both of the embedded environments shown above.

Table 2. Embedded device environment.

	Hardware: Amlogic T962 System: Android 9.0	**Hardware: Raspberry Pi 3B System: Windows NT 6.1**
PIDM	√	√

Note: √ indicates that it can run in this environment.

5.2. Better Performance

To compare the effectiveness of the PIDM model, we used the pre-trained classical model provided by the Keras library and tested it using the FERPlus, ExpW and AffectNet datasets. The test results are shown in Table 3.

Table 3. Comparison of multiple models.

Model	FERPlus Accuracy	ExpW Accuracy	AffectNet Accuracy	Million Parameters	Response time/k
VGG16	**83.76%**	**81.52%**	**78.25%**	138	3.5387 s
GoogLeNet InceptionV3	78.67%	76.75%	74.87%	23	3.4376 s
MobileNet	81.25%	80.46%	76.32%	4.2	2.5783 s
PIDM	82.43%	80.65%	76.24%	**3.2**	**1.8703 s**

Note: 1. Bold fonts are the best in this item. 2. VGG16 cannot be run in an embedded environment, and its data are measured on the Kaggle cloud server. The other three models are data obtained from testing on the embedded environment after being trained in the cloud and installed. The model response speed is the time for the model to recognize 1000 images.

VGG [28] is a convolutional neural network model proposed by Simonyan et al., and its name comes from the abbreviation of the Visual Geometry Group at the University of Oxford. VGG 16 model achieved excellent results in the 2014 ImageNet Image Classification and Localization Challenge. It ranked second in the classification task and first in the positioning task. GoogLeNet [31] won the 2014 ImageNet classification task by beating VGG16. Moreover, GoogLeNet is only 16.7% of the size of VGG16, which is one of the representatives of lightweight models. The Inception module proposed by GoogLeNet can reduce parameters while increasing the depth and width of the network.

VGG16 has the highest accuracy of all three datasets. However, the data contained within VGG16 are obtained from testing under high-performance cloud servers and are not comparable. The accuracy of the PIDM model in the three datasets is 82.43%, 80.65%, and 76.24%, respectively; only the AffectNet dataset is lower than MobileNet 0.08%, but the number of parameters of the PIDM model is only 76% of MobileNet. From a lightweight perspective, the PIDM model is more suitable for applications in smart portable devices. The accuracy of the PIDM model is second only to VGG16 running on a high-performance server, and the number of parameters is the smallest, and the response speed is the fastest. The PIDM model is suitable for embedded MR devices.

5.3. Impact of Privacy Protection on Model Accuracy

Before the emotion recognition task, we performed the privacy-preserving treatment on the pictures, and only the high-frequency information was retained. In the two embedded environments, we conducted three sets of comparison experiments with privacy-preserving processed pictures and unprocessed pictures, and the results are shown in Table 4.

Table 4. Comparison of accuracy of privacy protection.

	Hardware: Amlogic T962 System: Android 9.0		**Hardware: Raspberry Pi 3B System: Windows NT 6.1**	
	Protected	**Unprotected**	**Protected**	**Unprotected**
1	82.43%	**82.58%**	**81.25%**	81.16%
2	**82.35%**	82.26%	**81.65%**	80.87%
3	81.82%	**82.47%**	80.76%	**81.38%**

Note: Bold fonts indicates that it is the best effect in the group.

Experiments demonstrate that bioprivacy protection loses the low-frequency information of the images but has no effect on classification recognition. The difference in model accuracy is less than 1% in both embedded environments.

The experimental results show that the PIDM model meets the design requirements and runs in an embedded environment with reliable accuracy.

6. Discussion

6.1. Analysis of Technology

The models are optimized to reduce the computational power and storage requirements. Not only does the model run only on a PC or server, but it also runs well on embedded devices and achieves a good accuracy rate. The optimized model does not decrease in accuracy with reduced size and faster computational power. The images processed using Fourier transform did not affect the model's learning of image features while privacy was protected. The error of multiple experiments is within 1%.

6.1.1. Information Hiding Based on Fourier Transform

Humans can only perceive low-frequency components, while CNNs can perceive both low-frequency and high-frequency components. For any image, it should include both semantic information (texture information or low-frequency information) and high-frequency information. For any dataset, it should include both semantic information (texture information or low-frequency information) and high-frequency information, only that the proportion is variable, and for the same distribution dataset, its semantic distribution and high-frequency distribution should have their own distribution characteristics. It can be simply considered that for the same category labeled dataset, assuming that the dataset is collected from multiple scenes, the semantic distribution within each scene should be nearly the same (otherwise, it would not be labeled as the same category), but the high-frequency distribution is not necessarily related to a specific domain, and the high-frequency component may include specific information related to the category. Of course, it is also possible to include noise outside the distribution, and that noise is harmful to model training and can affect generalization ability. For humans, annotation relies solely on semantics because they cannot perceive the high-frequency components and ignore them.

The CNN will first use semantic information or low-frequency information for training in the early stage of training, and when the loss can no longer fall, additional high-frequency components will be introduced to further reduce the loss. Therefore, the direct use of high-frequency images does not have an impact on accuracy.

We use Fourier transform for privacy protection after image segmentation. Experiments demonstrate that the processed images are not recognizable to the naked eye and do not affect model training and recognition.

6.1.2. Ablation Experiments

Based on the study of BN, we believe that BN can improve the learning speed and inhibit the overfitting of the deep neural network model, but the excessive use of the BN layer may also affect the efficiency of the deep neural network model.

In the final stage of the deep neural network, the input data of the convolution layer gradually become regular. Although the data can be further normalized by continuing the normalization process, the performance improvement is less. From the perspective of cost, especially from the perspective of MR device application, we believe that it is low-cost performance [32,33]. In other words, when the BN layer is used in the early stages of a deep neural network, the performance improvement is obvious; when the BN layer is used in the later stages of a neural network model, the performance improvement is obtained by consuming the same computation and consumption is limited. Therefore, we propose a lightweight model.

1. Reduce the BN layer gradually in the later stages of the deep neural network model

The BN layer has low-cost performance in the later stages of the deep neural network model, and it may enlarge the weight of non-major features so as to reduce the robustness of the model. Our idea is to gradually reduce the BN layer from less to more in the second half of the network model.

2. Do not use the BN layer before the softmax layer

The BN layer before the softmax layer destroys the function of the softmax output as an uncertainty measure, and the performance gain can still be achieved after the BN layer is removed.

Through ablation experiments, we compared the size and accuracy of the model under various pruning methods.

Following different pruning ideas, we tested more than twenty model structures, and several feature representations are listed in Table 5.

Table 5. Ablation experiment using various pruning methods.

Name	Description
Model_1	Remove a DSC layer, Block 9.
Model_2	Modify the channel of pw convolution in Block12 from 512 to 256.
Model_3	Remove all BN layers.
Model_4	Remove all BN layers of Block3 and Block4.
Model_5	Remove all BN layers of Block14 and Block 15.
Model_6	Remove all BN layers of Block13, Block14 and Block 15.
PIDM	1. Remove all BN layers of Block14 and Block 15. 2. Remove BN layers after dw in Block10, Block11, Block12, and Block13.

Model_1 tries to prune off one DSC layer, MobileNet has five convolutional layers (Block8-Block12) with the same operation, and Block9 convolution is removed in Model_1.

Model_2 tries to reduce the channel of convolution.

Model_3 tries to remove all BN operations.

Model_4 tries to remove the BN operations of the first and second DSC layers.

Model_5 tries to remove the BN operations of the last two DSC layers.

Model_6 tries to remove the BN operations of the last three DSC layers.

In PIDM, BN operations are removed from the last two DSC layers and BN operations are removed in Block10, Block11, Block12, and Block13 after dw convolution.

The performance of the model still needs to be checked for accuracy, and we use the FERPlus dataset to train and verify the effect.

Comparing the above training curves, we can draw the following conclusions.

1. From Figure 12, we can see that after pruning one DSC layer, the model accuracy is less than 50%, which is not good.

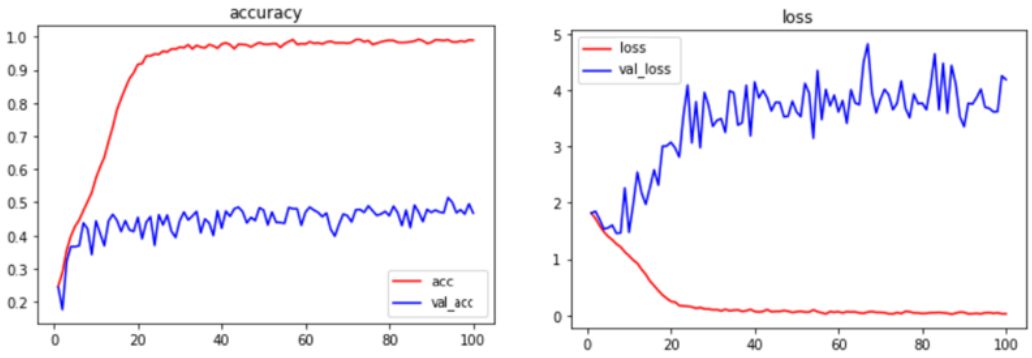

Figure 12. Training curve of Model_1.

2. From Figure 13, we can see that after pruning the number of channels of one DSC layer, the model accuracy is less than 50%, which is not good.

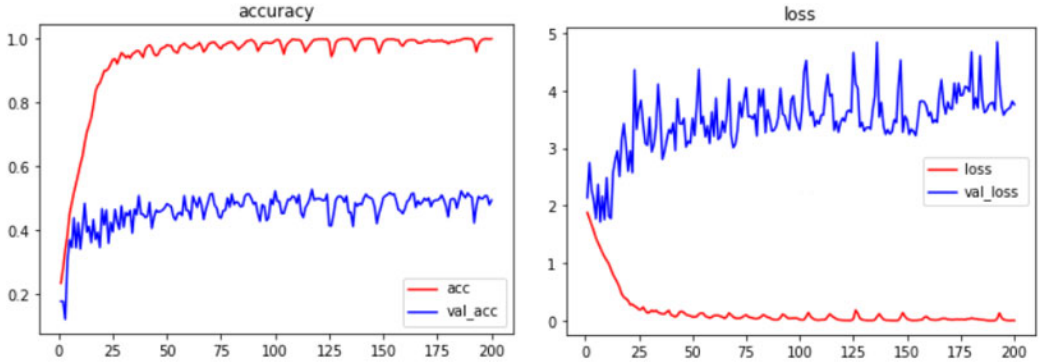

Figure 13. Training curve of Model_2.

3. From Figure 14, we can see that the model crashes after pruning all BN layers. This indicates that BN layers have a rather critical role in the model and cannot be removed completely.

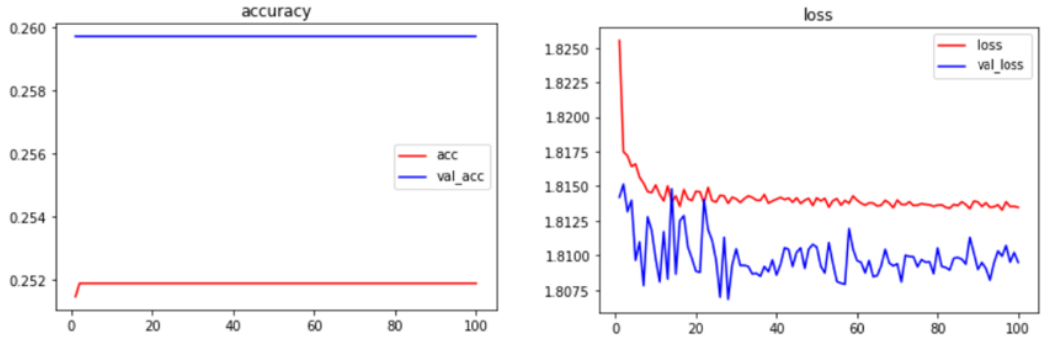

Figure 14. Training curve of Model_3.

4. Comparing Figures 15 and 16, model_4 is pruning BN at the beginning layers of the model, and model_5 is pruning BN at the final layers of the model. Model_5 has obviously produced a better result. It shows that the role of the BN layer at the beginning of the model is very obvious. However, it is not very important in the last few layers of the model. Therefore, the focus of pruning BN layers should be on the final part of the model.

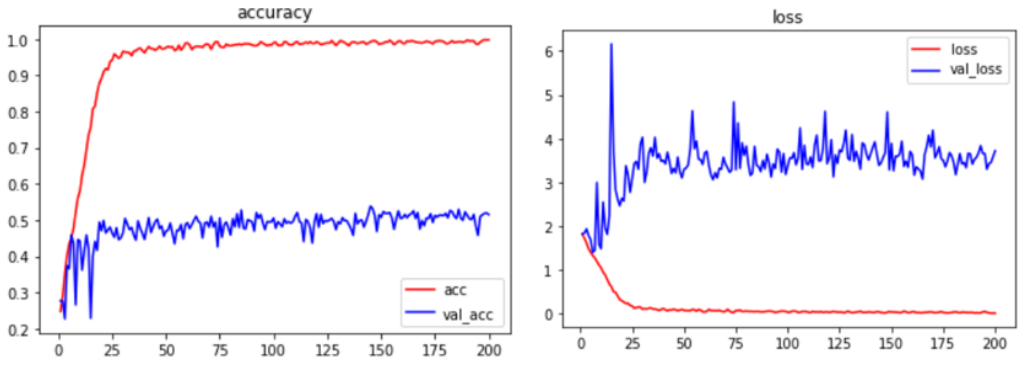

Figure 15. Training curve of Model_4.

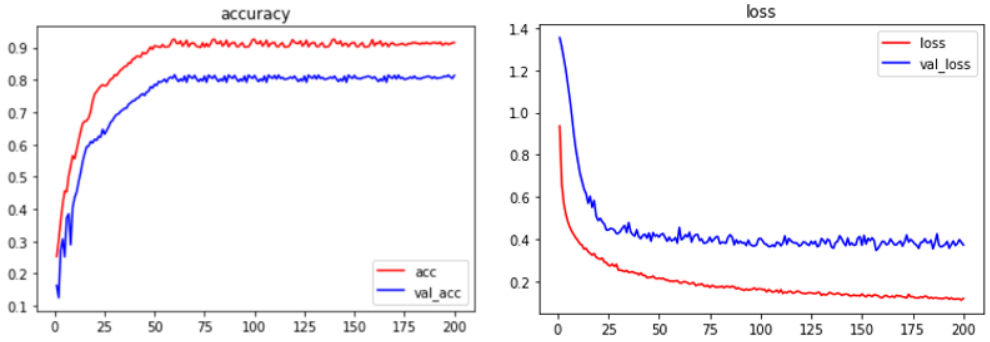

Figure 16. Training curve of Model_5.

5. Model_6 prunes all BNs in the last three DSC layers, and the accuracy rate decreases significantly, as shown in Figure 17. It can be seen that it is not feasible to prune BN excessively in the final part of the model.

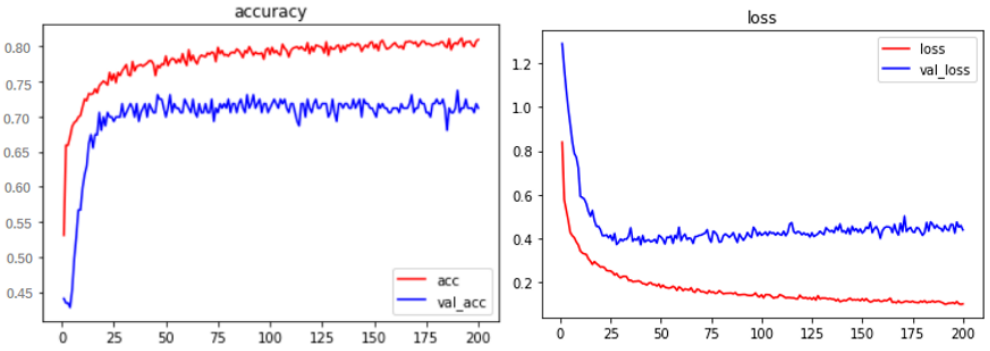

Figure 17. Training curve of Model_6.

6. PIDM is in line with our idea to step down the BN layers in the second half of the neural network model. In addition, it does not use the BN layers before the softmax layer. As shown in Figure 18, PIDM achieves acceptable results.

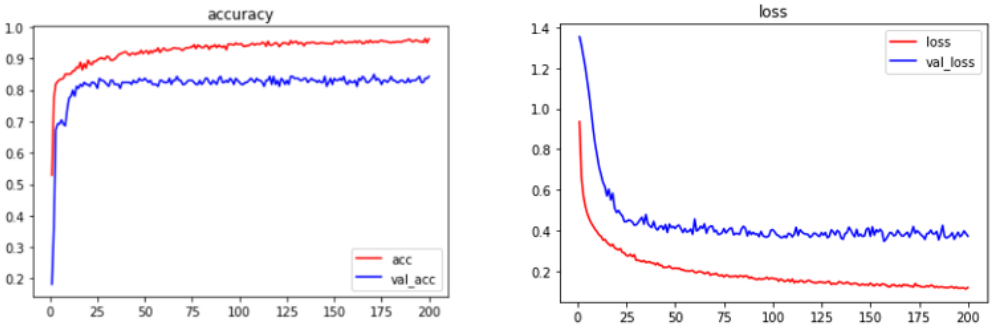

Figure 18. Training curve of PIDM.

Comparing the training results of the above models, as shown in Table 6, the accuracy of Model_1, Model_2, Model_3, and Model_4 are too low to be applied. The accuracy of Model_6 is 71.46%, which is also too low for practical application. The accuracies of Mod-el_5 and Model_7 are above 80%, and these are acceptable. However, the size of the parameters of Model_7 is only 3.22 M, which is 0.04 M smaller than that of Model_5, and the accuracy is 0.54% higher than that of Model_5.

Table 6. Comparison of training effects of multiple models.

Name	Parameter Size	FERPlus Accuracy
Model_1	2.96 M	46.32%
Model_2	2.98 M	43.21%
Model_3	3.02 M	26.00%
Model_4	3.26 M	50.68%
Model_5	3.26 M	81.92%
Model_6	3.24 M	71.46%
PIDM	3.22 M	82.46%

Through the use of ablation experiments, we believe that PIDM is the best model obtained after pruning, and it can be effectively used for MR devices. The effectiveness of our optimized design was demonstrated through ablation experiments.

6.2. Analysis of Teaching Applications

To understand the effect of embedded smart glasses in actual teaching, we invited 10 teachers and 166 students. Four classes participated in the experiment: two art classes with 82 students, and two science classes with 84 students. One part of the course was taught with a smart glasses device that allowed teachers to see students' emotions in real time. The other part of the course was regular. The experiment lasted one month. The teachers and students who participated in the experiment received examinations and questionnaires afterward.

6.2.1. Analysis of Teaching Effect

In order to understand the actual teaching effect of smart devices in the classroom, we arranged two control groups to conduct teaching experiments, in which smart devices were used when teaching Team A, and smart devices were not used when teaching Team B. Team A and Team B come from different classes and are randomly arranged after voluntary registration by students. To reduce the influence of teacher factors, we arranged for two teachers to teach designated courses for Team A and Team B, respectively. Teacher A and Teacher B teach different courses, each lasting 16 class hours. An exam is held in each of the three stages 2 class hours, 8 class hours, and 16 class hours. The test scores are calculated on a 100-point scale. The average score statistics of the three examinations are shown in Table 7.

Table 7. Average score of exams in the control group.

	Student	2 Class Hours	8 Class Hours	16 Class Hours
Teacher A	Team A	79.23	85.67	84.12
	Team B	76.41	73.53	77.49
Teacher B	Team A	82.19	80.67	83.27
	Team B	74.16	76.57	74.28

Teacher A teaches the same course for Team A and Team B. The difference is that for Team A lessons, smart portable devices are used to understand students' emotional feedback in real time, while Class B is taught without smart devices. As can be seen from Table 7, Team A's average test score is higher than Team B's. Teacher B is the same. In order to compare the teaching effectiveness of Team A and Team B, we calculate the average score and the overall average score for the three exams of the two courses, respectively, as shown in Table 8.

Table 8. Statistics of three exam scores.

	2 Class Hours	8 Class Hours	16 Class Hours	Overall Average
Team A	80.71	83.17	83.70	82.53
Team B	75.28	75.05	75.89	75.41

The average scores of Team A on the three exams were 80.71, 83.17, and 83.70, all higher than Team B. The overall average score of the three exams for Team A was 82.53, and the overall average score for Team B was 75.41. Team A's overall average score was 7.12 higher than Team B's. Experiments show that Team A using smart devices has a better teaching effect, which is 9.44% higher than Team B.

6.2.2. Analysis of Teacher Questionnaire

In the teachers' questionnaires, they rated the ease of use and usefulness of the smart glasses. Ratings were based on a 5-point scale, with 5 being the highest and 1 being the lowest. In the experiment, we divided the teachers into two groups: Group A, who had been teaching less than 2 years, and Group B, who had been teaching longer. The statistical results of the teachers' questionnaires are shown in Table 9.

Table 9. Statistics of teachers' questionnaire results.

	Convenience Rating (Average)	Teaching Effectiveness Rating (Average)
Group A	4.33	3.53
Group B	4.57	4.31
Total	4.45	3.92

In terms of convenience, there was a small difference between the ratings of the two groups of teachers. In terms of teaching effectiveness, the ratings of the two groups of teachers differed significantly. Rookie teachers rated teaching effectiveness at 3.53, while veteran teachers recognized smart devices as a teaching aid at 4.31.

From the teachers' questionnaire, the average rating of all teachers on the ease of use of the smart glasses devices was 4.45, which is a satisfactory result when viewed on a 5-point scale. This indicates that smart glasses devices are easy to use and that teachers can easily learn to use them. A large difference emerged in the ratings of the two groups of teachers in terms of teaching effectiveness. We originally thought that younger teachers would be more receptive to new devices and technologies in their teaching. The actual rating of smart glasses' effectiveness in teaching was only 3.53 for younger teachers with less teaching experience and 4.31 for veteran teachers. This indicates that veteran teachers found smart glasses more helpful. Our analysis suggests that rookie teachers with less teaching experience focus their main efforts in the classroom on the course content. They have no room to examine the information provided by smart MR glasses. To some extent, the information provided by the smart glasses may have distracted the rookie teachers. In contrast, veteran teachers are experienced and familiar with course content. They were more concerned about the students' reactions to the lecture when they taught. Even when they do not have smart glasses, veteran teachers will divert their energy to observe their students' progress. As a supplement to naked-eye observations, smart glasses provide statistical information about students' emotions, allowing teachers to accurately assess their students' emotions, adjust teaching according to their responses, and improve teaching effectiveness.

6.2.3. Analysis of Student Questionnaire

The student questionnaire was a rating of satisfaction with the effectiveness of teaching in all courses. Again, a score of 5 is the highest satisfaction level, and 1 is the lowest satisfaction level. The statistical results of the students' questionnaires are shown in Table 10.

Table 10. Statistics of students' questionnaire results.

	With Smart Devices	Without Smart Devices
Liberal arts students	4.03	4.14
Science students	3.62	4.18
Sample variance	0.83	0.75

Due to the large number of students involved in the scoring, we calculated the sample variance of the scores. This was to observe the scoring data volatility. The sample variance in the statistical results is less than 1. This indicates that students' rating data do not fluctuate much, and there are not many extreme ratings. It can be assumed that the mean of the ratings truly reflects the majority of students' perceptions.

As seen in Table 6, the influence of whether or not to use smart devices on satisfaction ratings with teaching effectiveness is not significant for literature students. In the results of science students' ratings, the average satisfaction rating of courses without smart devices is 3.62, while the average satisfaction rating of courses with smart devices is 4.18. Compared to courses without smart devices, there is a greater increase in satisfaction with teaching.

We found that there was only a 0.11 difference in the mean rating of satisfaction with smart devices among liberal arts students. This was not much different from the mean rating of teaching effectiveness satisfaction. Our analysis suggests that liberal arts courses have no strong logical connection. There is usually no situation where students cannot understand subsequent courses without understanding a certain knowledge point. Therefore, the effectiveness of teaching liberal arts courses depends more on teachers' teaching level, and the influence of smart glasses devices is smaller. In contrast, the satisfaction rating of science students who used smart glasses devices was higher than that of those who did not use smart glasses by 0.56. From the analysis of the structure of knowledge points, the knowledge points of science courses are very logically related, and the lecture explanation is a reasoning process with strong continuity and cause-and-effect relationships. With smart glasses, teachers can adjust the speed of lectures when many students are confused. They can also achieve better teaching results by asking questions or repeating lectures.

6.3. Contribution

In this paper, we propose an embedded device plus AI solution applied to classroom behavior analysis. This solution will adapt to the development of modern teaching, assist teachers in classroom teaching, and improve teaching quality. Our contribution has three main points.

1. We propose a lightweight model to analyze students' classroom behaviors through emotion recognition and statistics with the following features.

- Accuracy

Compared with the traditional manual scale recording by observers, the AI model avoids the subjective influence and omissions of manual recording. The recording is more complete and accurate, which is more effective for classroom behavior analysis.

- Efficiency

The traditional model of scale analysis requires the participation of experts, which is time-consuming and less efficient. Through MR devices, we can collect information and analyze and display it in real time to assist classroom instruction.

2. The vision models we use can run on embedded devices, which brings the following advantages.

- Ease of use

Embedded systems are widely used in wearable devices, and the MR glasses investigated in this study are one of them. MR glasses can be used in the classroom without

disrupting the original teaching, and students can easily see the statistics of classroom behavior analysis.

- Real-time display of analysis data

MR displays the analysis statistics in real-time and efficiently and can display statistics for multiple periods based on classroom time.

- Localization services

AI vision models can be run locally on embedded devices. This avoids data privacy leaks caused by image uploads to the network and also allows for a fast response, energy savings, and increased endurance.

6.4. Deficiency and Improvement Direction

6.4.1. Technology

In the model optimization, we only achieve the basic purpose of running on embedded devices. Many excellent optimization methods have not been used yet. In later studies, we will try to use techniques such as Network quantization and knowledge distillation for model optimization. Additionally, in our research, we used the SE (Squeeze-and-Excitation) module in mobilenetV3, but it did not work very well. We also need to study how to use the mature compression module in conjunction with our scenario in our later research.

Regarding the insufficient generalization ability of the model, the image's high-frequency components can be divided into two parts: the useful high-frequency components related to data distribution, A, and the noisy harmful high-frequency components unrelated to data, B. In the data training process, the CNN may use two kinds of high-frequency components AB for overfitting, and because the proportion of AB used cannot be determined, so there will be different generalization abilities of CNN models, and if more noise components are introduced, then the corresponding generalization ability decreases. We cannot distinguish component A and component B well now and can only obtain better results for specific datasets. Later, we improve the generalization ability of the model by discriminating between component A and component B.

6.4.2. Applications in Teaching

The core function of the PIDM model is emotion recognition, and the recognition results are presented to teachers with statistics categorized by emotion. Currently, the amount of information displayed is relatively small. In subsequent applications, the amount of information can be increased to provide teachers with more statistics. For example, gender classification for each emotion, statistics for a time period, etc., are displayed.

We can also combine recognition results with teaching suggestions. Smart devices provide some information for teachers, and it allows teachers to adjust their teaching style based on their experience to improve teaching effectiveness. It can be helpful for teachers who are experienced in teaching. However, for young teachers with less experience, this result is not very meaningful. Our subsequent research should ensure that the system can learn the relationship between recognition results and teaching effectiveness. Based on the characteristics of teachers, courses, and student classes, it can give teaching suggestions to improve teaching effectiveness.

Classroom behavior analysis is a long-term continuous educational management work. Real-time feedback data in the classroom can help, but even more valuable is the analysis of large amounts of accumulated data. Through this data analysis, we can understand the common characteristics of students, the common difficulties in the curriculum, and which teachers have achieved good teaching results in these difficulties. From this, teaching suggestions are extracted and provided to teachers in advance so that teachers can carry out targeted teaching work. These are not available in the current model and are the direction of our follow-up research.

7. Conclusions

In this paper, we propose an embedded device plus AI solution applied to classroom behavior analysis to adapt to the development of modern teaching, assist teachers in classroom teaching, and improve teaching quality. The solution can better help teachers, continuously improve the quality of teaching and learning, ensure the educational competence of different teachers, and guarantee the equity and sustainability of education.

We propose a lightweight model to analyze students' classroom behavior through emotion recognition and statistics with the following features: high accuracy and high efficiency. We collect data analysis through embedded MR devices and analyze and display them in real time to assist with classroom teaching.

The inevitable problem of computer vision models applied to the education industry is data leakage in image processing and the violation of personal privacy. Our vision model performs biological privacy protection and behavioral privacy protection before and after image recognition, respectively.

The PIDM models we use can run on embedded devices. It is easy to use and does not interfere with teachers' teaching work. It has a fast response time and a real-time display of the analysis data. Furthermore, it has local operation on embedded devices, which can avoid data privacy leakage caused by images uploading to the network, but also a fast response, saving energy and improving endurance.

AI in education needs further exploration, and we believe the following is the direction of future research.

Reduce equipment costs: We have made some progress using AI in embedded smart devices and teaching. As smart devices become more prevalent, the requirements for their use need to be further reduced so that they can be used on ordinary devices as well.

Reduce learning costs: At present, the teaching effect of smart devices still relies heavily on the experience of teachers. Future research should introduce expert systems so that rookie teachers can easily use and promote teacher training.

Promotion of AI in education: Because AI recognition involves personal privacy, some students and parents are not yet acceptable. Future research should adopt more privacy-preserving methods, which should be easy to understand for students and parents as well as secure. In this way, AI can be accepted by students and parents and promote its wide application in the field of education.

Author Contributions: Conceptualization, L.L. and W.B.; methodology, W.B.; software, L.W. and K.L.; validation, L.L., C.P.C. and W.B.; formal analysis, L.L.; investigation, C.P.C.; resources, C.P.C.; data curation, L.W. and K.L.; writing—original draft preparation, W.B.; writing—review and editing, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: Mixed Reality Holographic Teaching Application Project of Science and Technology Development Center, Ministry of Education (2018C01014); Development of Teaching Resources for Artificial Intelligence Major Course, Department of Higher Education, Ministry of Education Industry-University Cooperation Collaborative Education Project (202102476009); National Natural Science Foundation of China (61831015).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Shanghai Institute of Visual Arts (sv-20200816, approval date: August 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The FER-2013 dataset can be downloaded from https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data (accessed on 1 May 2023). The AffectNet dataset can be downloaded from http://mohammadmahoor.com/affectnet/ (accessed on 1 May 2023). The ExpW dataset can be downloaded from https://www.kaggle.com/datasets/mohammedaaltaha/expwds (accessed on 1 May 2023).

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

References

1. Zeng, H.; Shu, X.; Wang, Y.; Wang, Y.; Zhang, L.; Pong, T.-C.; Qu, H. *EmotionCues*: Emotion-Oriented Visual Summarization of Classroom Videos. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 3168–3181. [CrossRef] [PubMed]
2. Putra, W.B.; Arifin, F. Real-Time Emotion Recognition System to Monitor Student's Mood in a Classroom. *J. Phys. Conf. Ser.* **2019**, *1413*, 012021. [CrossRef]
3. Li, Y.Y.; Tang, Z.G. Design and implementation of the interactive analysis system software ET Toolbox FIAS 2011 based on Flanders. *China Educ. Technol. Equip.* **2011**, 102–104.
4. Taylor, S.S. Behavior basics: Quick behavior analysis and implementation of interventions for classroom teachers. *Clear. House A J. Educ. Strateg. Issues Ideas* **2011**, *84*, 197–203. [CrossRef]
5. Alberto, P.; Troutman, A.C. *Applied Behavior Analysis for Teachers*; Pearson: London, UK, 2013.
6. Chen, C.P.; Cui, Y.; Chen, Y.; Meng, S.; Sun, Y.; Mao, C.; Chu, Q. Near-eye display with a triple-channel waveguide for metaverse. *Opt. Express* **2022**, *30*, 31256–31266. [CrossRef] [PubMed]
7. Timms, M.J. Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 701–712. [CrossRef]
8. Mikropoulos, T.A.; Natsis, A. Educational virtual environments: A ten-year review of empirical research (1999–2009). *Comput. Educ.* **2011**, *56*, 769–780. [CrossRef]
9. Rus, V.; D'mello, S.; Hu, X.; Graesser, A. Recent Advances in Conversational Intelligent Tutoring Systems. *AI Mag.* **2013**, *34*, 42–54. [CrossRef]
10. Sharma, R.C.; Kawachi, P.; Bozkurt, A. The Landscape of Artificial Intelligence in Open, Online and Distance Education: Promises and concerns. *Asian J. Distance Educ.* **2019**, *14*, 1–2. [CrossRef]
11. Pokrivcakova, S. Preparing teachers for the application of AI-powered technologies in foreign language education. *J. Lang. Cult. Educ.* **2019**, *7*, 135–153. [CrossRef]
12. Chassignol, M.; Khoroshavin, A.; Klimova, A.; Bilyatdinova, A. Artificial Intelligence trends in education: A narrative overview. *Procedia Comput. Sci.* **2018**, *136*, 16–24. [CrossRef]
13. Rafika, A.S.; Sudaryono; Hardini, M.; Ardianto, A.Y.; Supriyanti, D. Face Recognition based Artificial Intelligence with AttendX Technology for Student Attendance. In Proceedings of the 2022 International Conference on Science and Technology (ICOSTECH), Batam City, Indonesia, 3–4 February 2022; pp. 1–7. [CrossRef]
14. Roy, M.L.; Malathi, D.; Jayaseeli, J.D.D. Facial Recognition Techniques and Their Applicability to Student Concentration Assessment: A Survey. In Proceedings of the International Conference on Deep Learning, Computing and Intelligence; Springer: Singapore, 2022; pp. 213–225. [CrossRef]
15. Savchenko, A.V.; Savchenko, L.V.; Makarov, I. Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network. *IEEE Trans. Affect. Comput.* **2022**, *13*, 2132–2143. [CrossRef]
16. Bu, Q. The global governance on automated facial recognition (AFR): Ethical and legal opportunities and privacy challenges. *Int. Cybersecur. Law Rev.* **2021**, *2*, 113–145. [CrossRef]
17. Andrejevic, M.; Selwyn, N. Facial recognition technology in schools: Critical questions and concerns. *Learn. Media Technol.* **2020**, *45*, 115–128. [CrossRef]
18. Kumalija, E.J.; Nakamoto, Y. MiniatureVQNet: A Light-Weight Deep Neural Network for Non-Intrusive Evaluation of VoIP Speech Quality. *Appl. Sci.* **2023**, *13*, 2455. [CrossRef]
19. Aloufi, B.O.; Alhakami, W. A Lightweight Authentication MAC Protocol for CR-WSNs. *Sensors* **2023**, *23*, 2015. [CrossRef] [PubMed]
20. Mnih, V.; Heess, N.; Graves, A. Recurrent models of visual attention. *Adv. Neural Inf. Process. Syst.* **2014**, 27. [CrossRef]
21. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018. [CrossRef]
22. Sifre, L.; Mallat, S. Rigid-Motion Scattering for Texture Classification. *Comput. Sci.* **2014**, *3559*, 501–515.
23. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
24. Gomez, A.N.; Kaiser, L.M.; Chollet, F. Depthwise Separable Convolutions for Neural Machine Translation. Available online: https://arxiv.org/abs/1706.03059 (accessed on 1 May 2023).
25. Prasetyo, E.; Purbaningtyas, R.; Adityo, R.D.; Suciati, N.; Fatichah, C. Combining MobileNetV1 and Depthwise Separable convolution bottleneck with Expansion for classifying the freshness of fish eyes. *Inf. Process. Agric.* **2022**, *9*, 485–496. [CrossRef]
26. Yoo, B.; Choi, Y.; Choi, H. Fast depthwise separable convolution for embedded systems. In Proceedings of the Neural Information Processing: 25th International Conference (ICONIP 2018), Siem Reap, Cambodia, 13–16 December 2018.
27. Hossain, S.M.M.; Aashiq Kamal, K.M.; Sen, A.; Deb, K. *Tomato Leaf Disease Recognition Using Depthwise Separable Convolution*; Springer International Publishing: Cham, Switzerland, 2022.
28. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Available online: https://arxiv.org/abs/1409.1556 (accessed on 1 May 2023).
29. Blalock, D.; Gonzalez Ortiz, J.J.; Frankle, J.; Guttag, J. What is the state of neural network pruning? *Comput. Sci.* **2022**, *2*, 129–146. [CrossRef]

30. Wang, Z.; Li, C.; Wang, X. Convolutional Neural Network Pruning with Structural Redundancy Reduction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 14908–14917. [CrossRef]
31. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *IEEE Comput. Soc.* **2014**, 1–9. [CrossRef]
32. Kim, G.H.; An, S.H.; Kang, K.I. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Build. Environ.* **2004**, *39*, 1235–1242. [CrossRef]
33. Zhang, Y.F.; Fuh, J.Y.; Chan, W.T. Feature-based cost estimation for packaging products using neural networks. *Comput. Ind.* **1996**, *32*, 95–113. [CrossRef]