Hindawi
Security and Communication Networks
Volume 2023, Article ID 9865069, 1 page
https://doi.org/10.1155/2023/9865069

Retraction

Retracted: Computational English Online Teaching Monitoring System Based on Deep Learning Algorithm

Security and Communication Networks

Received 16 November 2022; Accepted 16 November 2022; Published 24 January 2023

Security and Communication Networks has retracted the article titled "Computational English Online Teaching Monitoring System Based on Deep Learning Algorithm" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] J. Feng and W. Michalak, "Computational English Online Teaching Monitoring System Based on Deep Learning Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 7145129, 10 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7145129, 10 pages
https://doi.org/10.1155/2022/7145129

Research Article

Computational English Online Teaching Monitoring System Based on Deep Learning Algorithm

Julin Feng[1] **and Wazid Michalak**[2]

[1]*School of Foreign Languages, Yulin University, Yulin, Shaanxi 719000, China*
[2]*School of Computer Science, International Ataturk Alatoo University, Bishkek, Kyrgyzstan*

Correspondence should be addressed to Wazid Michalak; prof.michalak@mail.cu.edu.kg

Received 12 March 2022; Accepted 18 April 2022; Published 7 May 2022

Academic Editor: Muhammad Arif

In order to improve the monitoring effect of English online teaching, this paper combines the deep learning algorithm to construct the English online teaching monitoring system and conduct real-time supervision of the English online teaching process. Furthermore, by altering the original DCGAN, this research seeks to apply the approach of constructing a deep convolutional adversarial network to tackle the issue of small-sample target recognition in a given scene and develop an appropriate learning model. According to the results of the experimental research, the English online teaching monitoring system proposed in this paper, which is based on the deep learning algorithm, can play an important role in online English teaching monitoring and effectively improve the efficiency of online English management.

1. Introduction

With the development of artificial intelligence, machine vision technology is more and more widely used in people's lives. Mechanical vision technology is a technology that relies on photoelectric equipment to collect images and relies on computers to analyze and judge the images to identify objects. With the continuous development of education in our country, various advanced technologies are used in classrooms, but machine vision technology has not been widely and efficiently used in classrooms. At present, machine vision has been widely used in various fields and plays an extremely important role in China's production and life. Moreover, the efficient and accurate characteristics of machine vision make it a modern identification method, which is valued by more and more people. At present, machine vision is mainly used in China's industrial production and medical science and technology fields. Because of its high precision and high efficiency, it can help enterprises reduce labor costs and improve labor productivity, which is of great significance for promoting industrial development and economic development. Because of the enormous market and severe rivalry, product quality and look, in addition to performance, have become key assessment elements, and surface quality has become an essential comparative feature, for which machine vision was substantially created. As a result, machine vision is a technology with several benefits. It has no touch and causes no harm, and it may be used to achieve equipment automation, intelligence, and accuracy. It also offers a high level of safety, a broad spectrum response range, high manufacturing efficiency, and great flexibility. In our nation, schools are the front lines of education, and high-efficiency classrooms are sought after and monitored by the whole public.

This work uses a deep learning algorithm to build a monitoring system for English online teaching, as well as to oversee the process in real time and to enhance the quality of English online teaching.

2. Related Work

Reference [1] believes that an important model of deep learning is convolutional neural network. Its unique network structure can make a certain degree of translation, scaling,

and distortion highly invariant, and the performance of image recognition is good. Reference [2] uses face recognition technology to collect real-time face images of students for face recognition and completes the analysis of individual students' concentration. Reference [3] uses the convolutional neural network VGG pretraining network model to transfer learning, extracts the characteristics of abnormal behavior of students in the classroom, and realizes the detection and analysis of abnormal behaviors such as playing with mobile phones and sleeping, but does not involve "doing other coursework, mind wandering."

Literature [4] proposes a multimodal emotional feature fusion method based on the genetic algorithm, which uses genetic algorithm to select, cross, and recombine the emotional features of a single modality. Literature [5] proposes a method based on electrical skin signals and text. Literature [6] proposed a bimodal emotion recognition method based on bilateral sparse partial least squares method for expression and gesture. Reference [7] pointed out that emotion plays an important role in human cognitive process, so a new emotion computing module is proposed, which adopts biological, physical (heart rate, skin galvanic, and blood volume pressure), and facial expression methods. To extract the emotional state of learners; literature [8] proposed an emotion recognition method based on physiological signals. To convert the original physiological signals into spectral images, we use bidirectional long short-term memory recurrent neural network (LSTM-RNNS) to learn features and use deep neural network (DNN) for prediction.

Negative classroom habits include being late, leaving early, truancy, playing with cell phones, talking and chattering, and other behaviours that interfere with the regular development of classroom instructional activities. Negative conduct among students in the classroom may be classified into "explicit negative behaviour" and "implicit negative behaviour." According to the literature [9], students' explicit and negative classroom behaviours include coming late, leaving early, truancy, conversing, and talking, while unobservable activities include playing mobile phones, napping, mind wandering, and completing other schoolwork in class. Head-down behaviour (secretly playing with mobile phones and doing other coursework), head-turning behaviour (focusing on objects outside the blackboard range), and pseudo-listening behaviour (mind wandering and dozing off) are the three categories of implicit negative classroom behaviours identified by literature [10]. Bowing is mostly exhibited in bowing for a length of time, whether it is discreetly playing with a cell phone or performing other coursework. Within a certain amount of time, the head-turning behaviour manifests itself as head-turning and a particular range of body-turning. For a length of time, mind wandering in pseudo-lecture behaviour, also known as daydreaming, displays as eyes open but no blinking and no head movement activity. Closed eyelids and lack of control with head motions are the most common signs of drowsiness in pseudo-listening habits.

Reference [11] pointed out that emotions can be distinguished by facial expressions, and facial expressions and emotional labels may be connected, but this connection is variable in different cultural contexts. Literature [12] argues that determining a specific emotional state cannot ignore context, body, and culture, infer happiness from smiles, anger infer anger, frowns infer sadness, and current technologies try to build on these misunderstood scientific facts.

Reference [13] found through experiments that the template matching method is better than the feature-based method, and its advantage lies in the invariance of illumination, but its algorithm cannot exclude the influence of facial expression changes. The face recognition method proposed in [14] first uses principal component analysis (PCA) to reduce the dimensionality of the image's apparent features and then calculates the Euclidean distance from the target feature according to the dimensionality reduction feature to identify the identity. Another elastic graph matching technique extracts facial features to obtain the attribute map of the input image. However, these methods are more sensitive to changes in conditions such as light, age, and expression, and when certain conditions change, the recognition effect is not ideal. Deep learning has made great achievements in face feature extraction, weakening the influence of external factors, improving the reliability of face recognition, and promoting the practical application of face recognition technology. Aiming at the application of face recognition in classroom roll call, literature [15] proposed a classroom face recognition system based on a mobile platform. The Haar face detection method and the VGG face feature extraction network method were used to analyze the faces of students collected by mobile phone camera identification. However, due to the limited shooting area of the system, it did not play the role of classroom roll call. Reference [16] proposes a classroom face recognition system that combines AdaBoost's face detection algorithm and principal component analysis PCA algorithm, but the PCA algorithm is more sensitive to conditions such as light, age, and expression and cannot guarantee the extracted face features. The information is consistent, but the recognition effect is poor.

3. Image Processing Algorithm of English Online Teaching Based on Deep Learning

The first produced picture is not adequate for direct detection due to the involvement of several environmental elements in the image capture process. It entails removing the effect of irrelevant data and replacing it with valuable genuine data. A good picture preprocessing may make the next detection a lot easier, enhance the algorithm's feature extraction capabilities, and increase its matching and recognition reliability. In the object detection procedure, image preprocessing is crucial.

The majority of photographs taken in natural environments are colour photos based on RGB (red, green, and blue) three-channel data. The majority of colours in natural sceneries may be created by varying the proportions of the three main hues. Each colour channel component's size is

split into 256 values ranging from 0 to 255. The colour is whiter when the value is higher, and vice versa. A picture with $R = G = B$ is referred to as a grayscale image in the colour model, and the value of $R = G = B$ is referred to as a grayscale value. The grayscale value may be calculated in three different methods. The mean technique, the greatest value method, and the weighted approach are the three options. Formula (1) shows the solution using the mean value technique. The needed grey value may be calculated by adding and totalling the three channels' values and then averaging the results.

$$\mathrm{GRAY}(R, G, B) = \frac{(R + G + B)}{3}. \tag{1}$$

The maximum value method is to take the maximum component of the three channels in the RGB image as the grey value of the current image:

$$\mathrm{GRAY}(R, G, B) = \mathrm{Max}(R, G, B). \tag{2}$$

The weighted method is currently the most widely used grayscale conversion method. It uses artificially specified three-channel weights as reference parameters for weighted summation. Common normalization parameters have values of 0.3, 0.59, and 0.11:

$$\mathrm{GRAY}(R, G, B) = 0.3R + 0.59G + 0.11B. \tag{3}$$

Geometric feature transformation of an image generally refers to the transformation of the image in space, and its transformation methods include operations such as rotation, scaling, and translation. By using this kind of transformation in space, the error caused by the object in the process of motion or the system error caused by the instrument itself can be eliminated to a certain extent. After the image is geometrically transformed, an interpolation algorithm is usually used for spatial value mapping to map the transformed coordinates to an integer coordinate system. At present, common interpolation methods include bilinear interpolation and nearest neighbor interpolation, as shown in Figures 1 and 2.

3.1. Bilinear Interpolation. The value of a point x on the line between two points $(x_0, y_0), (x_1, y_1)$ on the line can be expressed in equation (4). It is known that its monolinear interpolation can be obtained by formula (5) [17].

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0}, \tag{4}$$

$$y = y_0 + \frac{(x - x_0)y_1 - (x - x_0)y_0}{x_1 - x_0}. \tag{5}$$

Bilinear interpolation is an extension of linear interpolation, which is linear interpolation in two directions by an interpolation function with two different variables. For a known point $N_{11} = (x_1, y_1), N_{12} = (x_1, y_2), N_{21} = (x_2, y_1), N_{22} = (x_2, y_2)$, the interpolation in the x direction can be expressed as (6) and (7), and the interpolation in the y direction is shown in formula (8) [18].

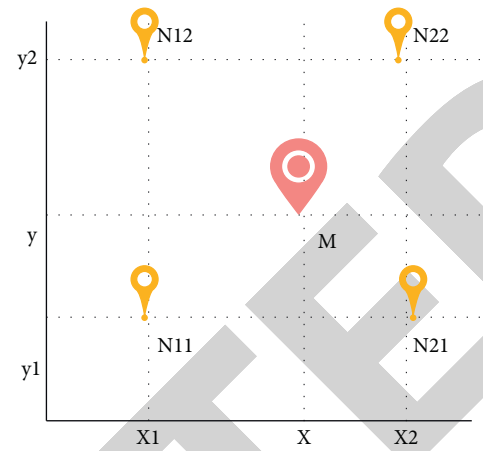

Figure 1: Schematic diagram of M-point bilinear interpolation.

$$f(x, y_1) \approx \frac{x_2 - x}{x_2 - x_1} f(N_{11}) + \frac{x - x_1}{x_2 - x_1} f(N_{12}), \tag{6}$$

$$f(x, y_2) \approx \frac{x_2 - x}{x_2 - x_1} f(N_{12}) + \frac{x - x_1}{x_2 - x_1} f(N_{22}), \tag{7}$$

$$f(M) \approx \frac{y_2 - y}{y - y_1} f(x, y_1) + \frac{x - x_1}{x_2 - x_1} f(x, y_2). \tag{8}$$

By combining the two, the final bilinear interpolation result is obtained.

$$\begin{aligned} f(x, y) \approx{}& \frac{f(N_{11})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y_2 - y) \\ &+ \frac{f(N_{12})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y_2 - y) \\ &+ \frac{f(N_{12})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y - y_1) \\ &+ \frac{f(N_{22})}{(x_2 - x_1)(y_2 - y_1)} (x - x_1)(y - y_1). \end{aligned} \tag{9}$$

3.2. Nearest Neighbor Interpolation. The nearest neighbor interpolation is the simplest interpolation method. The pixel value of a certain point after the image is scaled is the value directly calculated according to the scaling ratio of the original image. Occasionally, calculations result in decimals, and nearest neighbor interpolation takes its rounded integer pixel value directly. Using this interpolation method to calculate an enlarged image will produce a mosaic effect and calculate a reduced image will produce severe distortion, so the application scenarios are rare.

Data enhancement may increase the image's visual performance while also enhancing the expression of particular attributes in the sample. It expands the contrast between the characteristics of distinct objects or changes the representation information of the immediate region to enhance the image's information. Denoising techniques

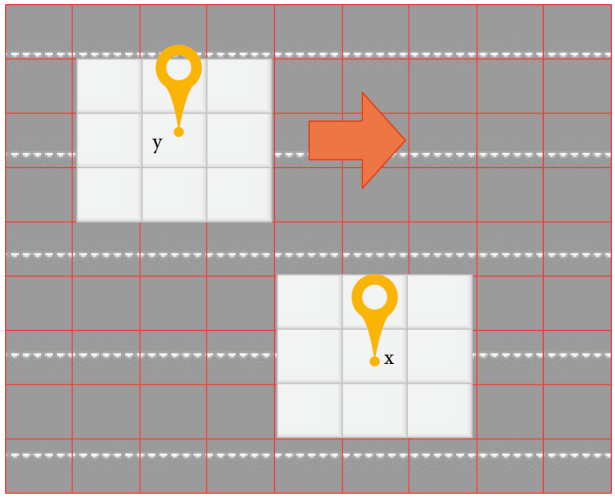

Figure 2: Schematic diagram of pixel-centered neighborhood block traversal.

based on spatial domain approaches, smoothing algorithms, and sharpening algorithms are all common ways. The NL-Means algorithm is a nonlocal averaging method. To assess the similarity between pixels, the NL-means method scans all of the given pixel blocks on the picture and assigns a weight w to the neighbourhood pixel blocks of y.

The grayscale calculation at x in the image u obtained by denoising the source image v is shown in formula (10). The similarity weight w is determined by the distance between the x and y pixel blocks, which can be calculated by formula (11).

$$u(x) = w(x, y) * v(y), \tag{10}$$

$$w = \frac{1}{Z(x)} \exp \frac{\|V(x) - V(y)\|^2}{h^2}. \tag{11}$$

Among them, $Z(x)$ is the image normalization coefficient, which can be calculated by formulas (12) and (13). The parameter h controls the smoothness of the image. Increasing h can improve the degree of image denoising but will blur the image; decreasing h can retain more edge features but leaves too much noise. The specific value depends on the situation.

$$\|V(x) - V(y)\|^2 = \frac{1}{d^2} \sum \|v(x + z) - v(y + z)\|^2, \tag{12}$$

$$Z(x) = \sum_y \exp\left(-\frac{\|V(x) - V(y)\|^2}{h^2}\right). \tag{13}$$

Mean filtering: the filtering algorithm is a common algorithm in image smoothing processing. In a square area composed of 9 pixels in the image, the average value of the pixels in the area is used as the pin point pixel value, that is,

$$g(x, y) = \frac{1}{M} \sum_{f \in s} f(x, y). \tag{14}$$

The mean filter algorithm is simple to calculate and has high efficiency, but the way it uses the mean will blur the edges of the image and lose part of the feature information. The mean filter usually has a good performance in smoothing Gaussian noise.

When dealing with picture smoothing challenges, median filtering employs a slightly different processing approach than mean filtering. The median filter technique utilises the median of 9 pixel values instead of the centre pixel value in the same 33 rectangular pixel block. This has the benefit of avoiding visual blurring and distortion induced by the average. The median filter has a clear benefit in smoothing the pulse signal since it uses a nonlinear filtering approach and can keep the image's edge properties throughout the smoothing process. As a result, it performs well in dealing with salt and pepper noise; nevertheless, the median filter is unsuitable for applications involving more Gaussian noise.

Laplacian: the Laplacian operator is a common method of image sharpening algorithms, which is used to enhance the content contained in the image. This operator is a second-order differential operator contained in the n-dimensional space and is usually used to represent the concept of gradient and divergence. The Laplacian operator template can be written in the form Reference [19].

$$\begin{aligned} \nabla^2 f(x, y) = {} & f(x - 1, y - 1) + f(x - 1, y) \\ & + f(x - 1, y + 1) + f(x, y + 1) \\ & + f(x + 1, y + 1) + f(x + 1, y) \\ & + f(x + 1, y - 1) + f(x, y + 1) - 8f(x, y). \end{aligned} \tag{15}$$

The Laplacian operator-enhanced image obtained from the template can be expressed by

$$f_E(x, y) = f(x, y) - \nabla^2 f(x, y). \tag{16}$$

Sample augmentation is very important for target detection, especially the deep learning method requires a large number of samples as a training set. Traditional sample augmentation includes rotation, translation, scaling, mirroring, Gaussian noise, blurring, and more.

Image rotation is a common sample augmentation strategy. Multiangle target samples can be obtained by specifying different angle parameters. This augmentation method can simulate the multiangle characteristics of the target to a certain extent and has a certain augmentation effect, as shown in Figure 3.

The principle of mirror symmetry is to obtain a new image by horizontally flipping the original image, which can effectively improve the generalization ability of the neural network while maintaining the physical characteristics of each pixel in the original image (Figure 4).

The term "random cropping" refers to cropping a portion of a picture containing the target and deleting the rest, such that the target appears in any place on the cropped image at random. The purpose of effective data expansion may be performed by such processes. Some research have

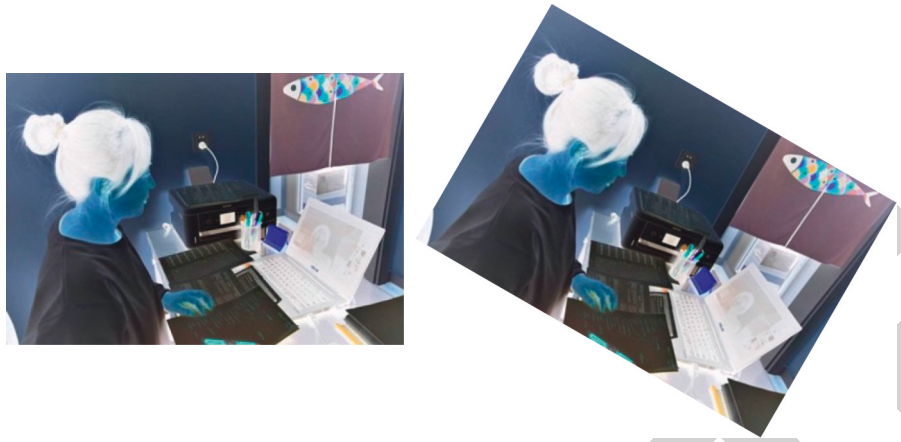

Figure 3: An example of the angle transformation of the original image.

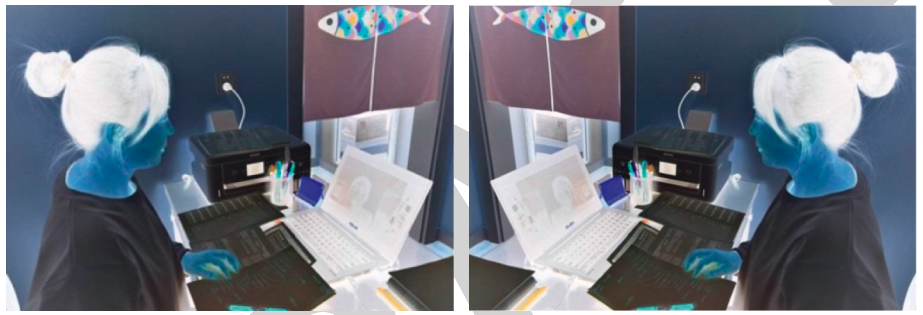

Figure 4: Example of mirror symmetry.

shown that, by utilising random cropping to increase the data set, the model's accuracy can be successfully enhanced, and the model's generalization ability can be improved to a degree (Figure 5).

The translation operation refers to moving an image relative to the original image by a distance defined by the number of pixels. Through image translation, some objects in the image will be cropped because they appear outside the image area. In this way, the effect of partial occlusion of the target can be achieved to a certain extent, thereby improving the generalization ability of the classifier in extracting local features of the target (Figure 6).

Image blur is also a strategy for sample augmentation. Common image blur algorithms include motion blur and Gaussian blur, which have slightly different effects on target features. Gaussian blur is also called Gaussian smoothing, which is not only a kind of smoothing algorithm but also a common strategy in image augmentation. It is mostly used in the image preprocessing stage in the field of target detection. The essence of Gaussian blur is a filter that uses normal distribution to calculate each pixel in the image. For an image, the normal distribution in the N-dimensional space can be expressed as [20]

$$G(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-r^2/(2\sigma^2)}. \tag{17}$$

Among them, r is the blur radius of the Gaussian blur, usually the radius of the spatial convolution kernel it uses, and σ is the standard deviation in the normal distribution equation. In the process of actual image preprocessing, using the above formula will generate a circular structure based on normal distribution. The original picture is used to compute the convolution kernel, which consists of nonzero pixel values in the image. Each pixel value after computation is the average of several pixels next to the pixel. The bigger the pixel value of the originally set pixel, the larger the weight value, and therefore, the larger the reserved value after the computation; conversely, the lower the weight, the smaller the reserved value. After computation, the interference regions that appear as tiny pixels will have their brightness greatly lowered and blended into the background to eliminate interference, but the bigger feature areas with clear brightness characteristics in the picture will be kept (Figure 7).

Motion blur-based sample augmentation has a good effect in some specific application scenarios. It is necessary to add motion blurred images to the training samples because the device may shake during the shooting process, and sometimes, the movement of the target object will also cause the target to be blurred in the image. By using Gaussian blur and motion blur, the classifier can learn the target features better and improve the classification accuracy.

Image scaling is usually used to simulate the multiscale features of the target, and it can be used when the scale difference between the images in the training dataset and the images to be tested is too large. The neural network is

Figure 5: Example of random cropping of the original image.

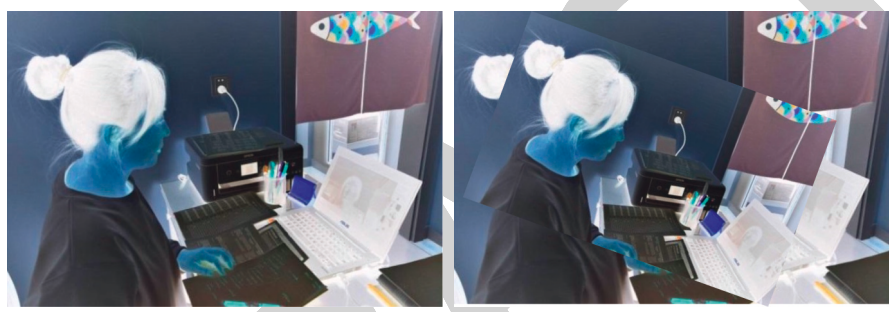

Figure 6: Example of original image translation.

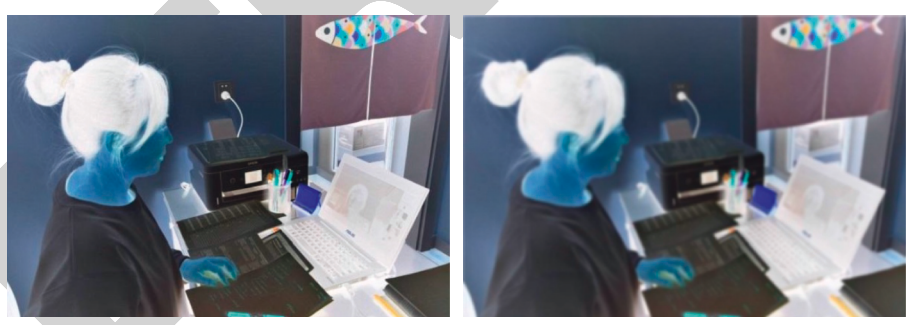

Figure 7: Example of Gaussian blur of the original image.

sensitive to the size of the target sample, and reducing or enlarging the target image can significantly improve the feature extraction ability of the network at multiple scales (Figure 8).

Salt and pepper noise and Gaussian noise are two examples of common noises that are used to improve data. Different filtering techniques emphasise the image's local characteristics and noise may improve the classifier's generalization capacity to cope with the instrument's faults. Gaussian noise is defined as noise with a noise density function that follows a normal distribution equation. Instrument noise is formed in real application situations when image collection equipment is used in an unsuitable working environment (e.g., high temperature and excessive exposure). This is owing to the equipment's difficulties. As a result, including such images in the training set improves the classifier's robustness under harsh situations to some degree (Figure 9).

For the data distribution $P_{\text{data}}(x)$ of all images in the high-dimensional space, the learning model needs to learn the data distribution law of the images through the provided data set. For the prerandomized initial parameter distribution $P_G(x;\theta)$, deep learning is used to make it reach the maximum likelihood state:

$$\theta^* = \text{argmax}_\theta L = \prod_{i=1}^{m} P_G\left(x^i;\theta\right),; \text{Set}\left\{x^1, x^2, \ldots, x^m\right\}. \quad (18)$$

For the problem of deep learning, its essence is a process of optimizing the target divergence, and this divergence

optimization network is a generative network. For a known initialization vector z (obeying the standard normal distribution), it can be transformed into a sample of unknown distribution after calculation, and the distribution of this sample may be completely different from vector z. The solution process of the generator is to minimize the divergence distance between the two by calculation and finally obtain the approximate solution of the real data distribution. A picture generated by a network G that generates pictures is denoted as $G(z)$. The input of the generator is artificially defined noise, and simulated samples are generated through a series of operations. The mathematical expression of the generator can be described as

$$G = \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right). \tag{19}$$

Among them, m is the number of input samples and z is the number of input random noise; what the generator needs to do is to maximize the $D(G(z(i)))$ term to fool the discriminator.

For a discriminator network D, the probability judgment of the input image is denoted as $D(x)$, and for the discriminator D, its mathematical expression can be denoted as

$$D = \nabla_{\theta_{gd}} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)\right]. \tag{20}$$

The purpose of the discriminator is to maximize the $\log D(x^{(i)})$ term, and the maximum value of this term indicates that the discriminator classifies the real sample as true. Similarly, minimizing the $\log(1 - D(G(z^{(i)})))$ term means that the discriminator classifies the simulated sample as false. The loss function of the discriminator can be expressed as

$$L_D = -\int_x^n \left[P_{\text{data}}(x) \log D(x) + P_G(x) \log(1 - D(x))\right] \mathrm{d}x. \tag{21}$$

When $\partial L_D / \partial D = -[P_{\text{data}}/D(X) - P_G(X)/1 - D(X)] \longrightarrow 0$, the optimal discriminator can be obtained:

$$\widehat{D}(x) = \frac{P_{\text{data}}}{P_G + P_{\text{data}}}. \tag{22}$$

For any probability distribution, the index data to describe its difference usually refers to the KL divergence (Kullback–Leibler divergence), and its definition is

$$KL\left(P_r \| P_f\right) = E_{x \sim P_r}\left[\log \frac{P_r(x)}{P_f(x)}\right] = \int P_r(x) \log \frac{P_r(x)}{P_f(x)} \mathrm{d}x. \tag{23}$$

Therefore, the working process of the discriminator is actually to maximize the divergence between the real samples and the samples generated by the generator. The structure of the adversarial network can be summarized as

$$\begin{aligned} \min_G \max_D V(D,G) = {} & E_{x \sim P_{\text{data}}}[\log D(x)] \\ & + E_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \end{aligned} \tag{24}$$

Compared with the previous unsupervised learning methods, generative adversarial networks can generate more realistic and effective samples. Different from traditional methods such as Boltzmann machine and GSNS, GAN does not introduce any deterministic bias in the training process. If the discriminator is trained well, the model can learn a perfect sample distribution.

The role of the bottleneck restriction mechanism of mutual information for deep learning was first pointed out by Alemi in 2016, and its optimization model can be expressed as formula (25). The initial optimization model is very prone to overfitting, so the concept of regularization needs to be introduced to improve the generalization ability of the model.

$$\min_q E_{x,y \sim P(x,y)}[-\log(y|x)]. \tag{25}$$

The idea of VDB is that the drum model only focuses on the most easily distinguishable feature information to limit the model. The objective function for the mutual information restriction between the data x and the result z of the encoder is

$$\begin{aligned} J(D,E) = {} & \min_{D,E} E_{x \sim p^*(x)}\left[E_{z \sim e(z|x)}[-\log(D(z))]\right] \\ & + E_{x \sim G(x)}\left[E_{z \sim e(z|x)}[-\log(1 - D(z))]\right], \end{aligned} \tag{26}$$

$$\text{s.t.} \quad E_{x \sim \widetilde{p}(x)}[KL[E(z|x) \| r(z)]] \le I_c. \tag{27}$$

Among them, $\widetilde{p} = 1/2p^* + 1/2G$ represents the constraint condition of the variational discriminator bottleneck, p^* is the real data distribution, and the generated fake target sample is $G(x)$. The Lagrangian coefficient β is introduced to optimize the objective function to obtain the final mutual information restriction result. The encoder E and the discriminator D can finally be expressed by formulas (28) and (29).

$$\begin{aligned} J(D,E) = {} & \min_{D,E} \max_{\beta \ge 0} E_{x \sim p^*(x)}\left[E_{z \sim e(z|x)}[-\log(D(z))]\right] \\ & + E_{x \sim G(x)}\left[E_{z \sim e(z|x)}[-\log(1 - D(z))]\right], \end{aligned} \tag{28}$$

$$\text{s.t.} \quad \beta\left(E_{x \sim \widetilde{p}(x)}[KL[E(z|x) \| r(z)]] - I_c\right). \tag{29}$$

After introducing the variational discriminator bottleneck mechanism, the original unstable adversarial network structure becomes robust because the discriminator is constrained. In the process of training the discriminator to the optimal one, the gradient of the generator rarely disappears, from the evaluation results of the adversarial network. The V-DCGAN that introduced the VDB mechanism has a good performance in terms of network robustness and the authenticity of the generated samples.

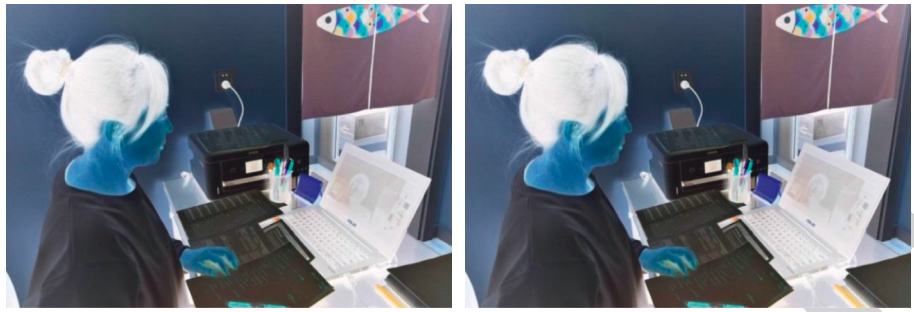

Figure 8: Example of original image scaling.

Figure 9: Example of adding Gaussian noise to the original image.

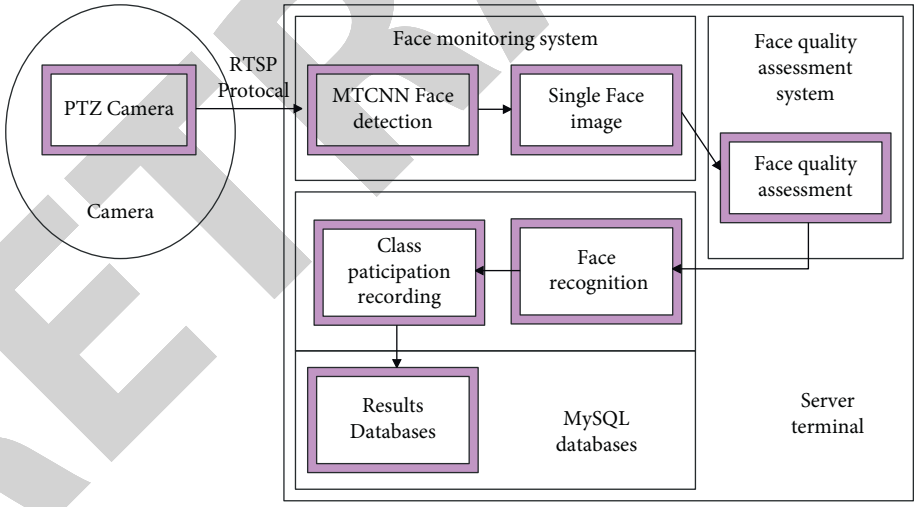

Figure 10: Face recognition online teaching monitoring system for English based on deep learning.

4. English Online Teaching Monitoring System Based on Deep Learning Algorithm

This paper is a deep learning-based face recognition English online teaching monitoring system. The overall design block diagram of the system is shown in Figure 10, which is mainly composed of two parts: the camera and the server.

Following the collection of data from diverse sources, a complete assessment strategy must be established. It is required to utilise mathematical and statistical approaches to generate scientific statistics, as well as research data to monitor the status of each student's class attendance statistically. Each student's visual and aural data are collected and analysed to find group issues and trends. At the same time, as shown in Figure 11, identifying individual distinctions, as well as variances in each person's learning and class stages, may be more effectively used to monitoring students' class status and adjusting instructors' teaching, therefore strengthening the teaching reform.

After constructing the English online teaching monitoring system based on the deep learning algorithm, the system is evaluated and analysed, and the monitoring effect

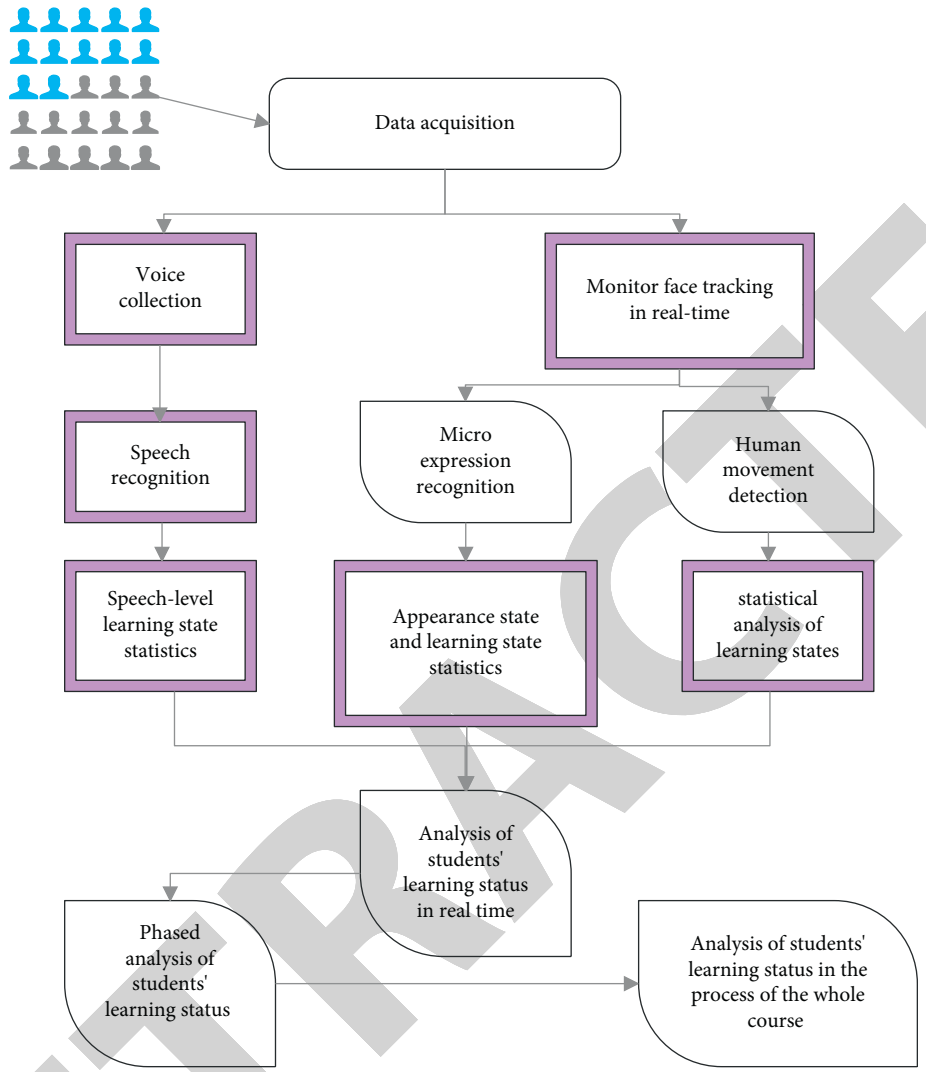

Figure 11: The process of teaching evaluation.

Table 1: Statistical table of monitoring effect of English online teaching monitoring system based on the deep learning algorithm.

Num	Monitoring effect	Num	Monitoring effect	Num	Monitoring effect	Num	Monitoring effect
1	88.34	21	88.38	41	88.41	61	85.02
2	84.84	22	86.15	42	83.84	62	85.84
3	84.52	23	83.22	43	80.55	63	87.29
4	88.87	24	83.63	44	88.58	64	87.60
5	85.84	25	87.18	45	85.12	65	81.97
6	80.76	26	80.59	46	88.45	66	84.37
7	84.10	27	84.16	47	80.32	67	87.04
8	87.00	28	86.86	48	87.67	68	88.80
9	83.51	29	80.12	49	81.22	69	85.89
10	80.56	30	80.13	50	87.45	70	87.17
11	82.84	31	83.39	51	82.33	71	84.98
12	83.75	32	81.60	52	82.35	72	87.08
13	81.39	33	88.15	53	88.09	73	83.20
14	87.80	34	87.88	54	87.01	74	85.11
15	83.52	35	83.48	55	86.68	75	83.40
16	85.49	36	87.39	56	83.42	76	86.60
17	81.39	37	80.59	57	84.61	77	80.87
18	82.67	38	80.68	58	81.41	78	86.44
19	81.57	39	84.48	59	87.97	79	83.68
20	85.63	40	80.08	60	83.75	80	88.95

of the English online teaching monitoring system based on the deep learning algorithm on the online teaching is counted, and the results shown in Table 1 are obtained.

From the above research, it can be seen that the English online teaching monitoring system based on the deep learning algorithm proposed in this paper can play an important role in the online monitoring of English teaching and effectively improve the efficiency of online English management.

5. Conclusion

In order to urge students to learn efficiently in English online teaching classes, teachers should often criticize and remind students in class. However, teachers have limited energy and time, and cannot remind students all the time. Schools can strengthen control through monitoring and teacher patrols, but doing so is time-consuming, labor-intensive, inefficient, and will also cause misjudgments and missed judgments. However, machine vision-based detection methods can solve this problem to a large extent. Through this method, the students' listening status can be judged by identifying the students' facial expressions, which saves manpower and improves the classroom efficiency. In this paper, a deep learning algorithm is combined to construct a monitoring system for English online teaching, to supervise the process of online English teaching in real time, and to improve the quality of online English teaching. The experimental research results show that the English online teaching monitoring system based on the deep learning algorithm proposed in this paper can play an important role in the online monitoring of English teaching and effectively improve the efficiency of online English management.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. Bogusevschi, C. Muntean, and G. M. Muntean, "Teaching and learning physics using 3D virtual learning environment: a case study of combined virtual reality and virtual laboratory in secondary school," *Journal of Computers in Mathematics and Science Teaching*, vol. 39, no. 1, pp. 5–18, 2020.

[2] S. F. M. Alfalah, "Perceptions toward adopting virtual reality as a teaching aid in information technology," *Education and Information Technologies*, vol. 23, no. 6, pp. 2633–2653, 2018.

[3] G. Cooper, H. Park, Z. Nasr, L. P. Thong, and R. Johnson, "Using virtual reality in the classroom: preservice teachers' perceptions of its use as a teaching and learning tool," *Educational Media International*, vol. 56, no. 1, pp. 1–13, 2019.

[4] J. Zhao, X. Xu, H. Jiang, and Y. Ding, "The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies," *BMC Medical Education*, vol. 20, no. 1, pp. 1–10, 2020.

[5] S. J. Bennie, K. E. Ranaghan, H. Deeks et al., "Teaching enzyme catalysis using interactive molecular dynamics in virtual reality," *Journal of Chemical Education*, vol. 96, no. 11, pp. 2488–2496, 2019.

[6] S. F. M. Alfalah, J. F. M. Falah, T. Alfalah, M. Elfalah, N. Muhaidat, and O. Falah, "A comparative study between a virtual reality heart anatomy system and traditional medical teaching modalities," *Virtual Reality*, vol. 23, no. 3, pp. 229–234, 2019.

[7] M. Reymus, A. Liebermann, and C. Diegritz, "Virtual reality: an effective tool for teaching root canal anatomy to undergraduate dental students - a preliminary study," *International Endodontic Journal*, vol. 53, no. 11, pp. 1581–1587, 2020.

[8] V. L. Dayarathna, S. Karam, R. Jaradat et al., "Assessment of the efficacy and effectiveness of virtual reality teaching module: a gender-based comparison," *International Journal of Engineering Education*, vol. 36, no. 6, pp. 1938–1955, 2020.

[9] O. Hernandez-Pozas and H. Carreon-Flores, "Teaching international business using virtual reality," *Journal of Teaching in International Business*, vol. 30, no. 2, pp. 196–212, 2019.

[10] V. Andrunyk, T. Shestakevych, and V. Pasichnyk, "The technology of augmented and virtual reality in teaching children with ASD," *Econtechmod: Scientific Journal*, vol. 7, no. 4, pp. 59–64, 2018.

[11] R. Mayne and H. Green, "Virtual reality for teaching and learning in crime scene investigation," *Science & Justice*, vol. 60, no. 5, pp. 466–472, 2020.

[12] M. Taubert, L. Webber, T. Hamilton, M. Carr, and M. Harvey, "Virtual reality videos used in undergraduate palliative and oncology medical teaching: results of a pilot study," *BMJ Supportive & Palliative Care*, vol. 9, no. 3, pp. 281–285, 2019.

[13] K. E. McCool, S. A. Bissett, T. L. Hill, L. A. Degernes, and E. C. Hawkins, "Evaluation of a human virtual-reality endoscopy trainer for teaching early endoscopy skills to veterinarians," *Journal of Veterinary Medical Education*, vol. 47, no. 1, pp. 106–116, 2020.

[14] X. Xu, P. Guo, J. Zhai, and X. Zeng, "Robotic kinematics teaching system with virtual reality, remote control and an on-site laboratory," *International Journal of Mechanical Engineering Education*, vol. 48, no. 3, pp. 197–220, 2020.

[15] P. W. Chang, B. C. Chen, C. E. Jones, K. Bunting, C. Chakraborti, and M. J. Kahn, "Virtual reality supplemental teaching at low-cost (VRSTL) as a medical education adjunct for increasing early patient exposure," *Medical Science Educator*, vol. 28, no. 1, pp. 3-4, 2018.

[16] J. Zhang and Y. Zhou, "Study on interactive teaching laboratory based on virtual reality," *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 30, no. 3, pp. 313–326, 2020.

[17] R. Ramlogan, A. U. Niazi, R. Jin, J. Johnson, V. W. Chan, and A. Perlas, "A virtual reality simulation model of spinal ultrasound," *Regional Anesthesia and Pain Medicine*, vol. 42, no. 2, pp. 217–222, 2017.

[18] Y. C. Hsu, "Exploring the learning motivation and effectiveness of applying virtual reality to high school mathematics," *Universal Journal of Educational Research*, vol. 8, no. 2, pp. 438–444, 2020.

[19] J. D. Anacona, E. E. Millán, and C. A. Gómez, "Aplicación de los metaversos y la realidad virtual en la enseñanza," *Entre ciencia e ingeniería*, vol. 13, no. 25, pp. 59–67, 2019.

[20] P. Calvert, "Virtual reality as a tool for teaching library design," *Education for Information*, vol. 35, no. 4, pp. 439–450, 2019.