Sensors and Materials, Vol. 35, No. 12 (2023) 4307–4324
MYU Tokyo

S & M 3482

Design and Experimentation of Face Recognition Technology Applied to Online Live Class

Mingyang Liu[1] and Haifeng Pang[2*]

[1]College of Arts and Information Engineering, Dalian Polytechnic University,
No. 117 Xuefu Street, Zhuanghe City, Dalian 116499, China
[2]Hi-Think Technology, Corp., No. 977 Huangpu Road, Hi-Tech Zone, Dalian 116085, China

(Received April 23, 2023; accepted November 10, 2023)

Keywords: face recognition, big data, expression analysis, online learning

In this study, we assess facial recognition technology, including detection, comparison, and attribute analysis, for addressing online education challenges. The system tracks facial expressions and records attendance and emotions, thereby improving educators' understanding of student performance and aiding in targeted interventions. Future enhancements will include statistical data analysis and seamless data transmission. In summary, facial recognition technology holds promise for enhancing online education, enabling real-time teaching strategy adjustments. In this work, we focus on precise student progress tracking, timely lecture analysis, and adaptive teaching in live web classes. An experimental platform with student, server, and teacher components, utilizing PyramidBox-based face recognition, is developed. Real classroom experiments, including roll call, knowledge review, content explanation, and classroom interaction, reveal face recognition's positive effect on real-time student monitoring and classroom assessment. This empowers educators to promptly adapt teaching strategies, enhancing lecture quality. The study also highlights emotional states' impact on learning.

1. Introduction

The rapid advancement of internet technology has propelled online live courses into prominence as a highly coveted mode of instruction within the contemporary educational landscape. These courses present a multitude of advantages, including heightened convenience, enhanced accessibility, and increased diversity, rendering them adaptable to a spectrum of educational requisites and specialized contexts. Nonetheless, akin to other pedagogical methods, online live courses confront a series of challenges, with one of the most conspicuous among them being the incapacity of educators to contemporaneously observe the facial expressions and emotional responses of students during virtual instructional sessions, thereby impeding their ability to gauge students' reactions and comprehension of the educational content.[1]

In contrast to the traditional in-person classroom milieu, where educators can glean insights into students' cognitive states and educational necessities through the direct observation of their

*Corresponding author: e-mail: 214634166@qq.com
https://doi.org/10.18494/SAM4436

ISSN 0914-4935

facial expressions and interpersonal interactions, online live courses are characterized by geographical separation between educators and students, constraining the feasibility of such face-to-face interactions and consequently compromising educators' acumen regarding students' affective responses and comprehension. In this milieu, facial recognition technology has emerged as a potent remedial tool. Harnessing the forefront of advancements in deep learning and computer vision, facial recognition technology is equipped to autonomously capture students' facial expressions and emotional fluctuations, while concurrently evaluating their levels of engagement and attentiveness.(2) Through the meticulous analysis of this dataset, educators can attain a more precise understanding of students' educational states and requisites, thereby facilitating the prompt modification of pedagogical strategies and the provision of personalized guidance and support.

Illustratively, should facial recognition technology discern indications of perplexity or distraction in a student, educators can undertake immediate interventions, such as supplementary elucidations or interactive engagement, to facilitate an enhanced assimilation of course content. Moreover, facial recognition technology affords the means to assess the prevailing level of tension or relaxation within the virtual classroom environment, offering educators valuable insights for the fine-tuning of instructional approaches and the cultivation of a more conducive ambiance for learning.

Hence, the integration of facial recognition technology within online live courses assumes paramount significance. It harbors the potential to ameliorate the limitations associated with traditional face-to-face interactions, elevate the quality of education, and augment students' scholastic attainments.

In this study, we seek to delve comprehensively into the promise of this technology and its core functionalities, encompassing facial detection, facial comparison, and facial attribute analysis, as avenues to address the aforementioned pedagogical challenges, thereby advancing the domain of online education.

2. Materials and Methods

2.1 Limitations of existing approaches

In a traditional offline classroom setting, educators are able to assess students' learning progress, comprehension, and classroom atmosphere in real time through direct observation and face-to-face communication. This interactive and observational approach empowers educators to promptly adjust teaching strategies to cater to individual student needs and enhance their learning experience.

However, in the context of online live-streamed teaching, educators often find it challenging to accurately observe and track each student's participation during class. This challenge primarily arises from the limitations of online education, which prevent educators from directly observing students' reactions and performance, potentially leading to communication barriers as educators may struggle to precisely comprehend each student's needs and confusion.

Furthermore, owing to the nature of online education, educators typically lack sufficient data for comprehensive post-classroom review and analysis. Although existing mainstream online live-streaming educational platforms, such as Massive Open OnlineCourse, Tencent Meeting, and DingDing Meeting,[2,3] offer some basic statistical functionalities like student enrollment and online duration, these functions often focus on rudimentary data and do not delve into the in-depth assessment of various critical factors during live-streamed classes, such as student engagement, classroom ambiance, and teaching quality.

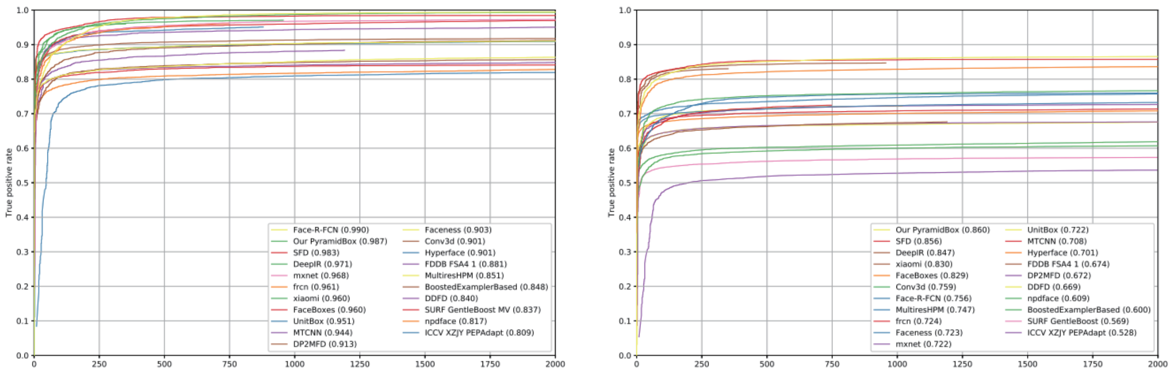
The absence of more detailed data and in-depth analytical features means that these platforms primarily prioritize ensuring the smooth delivery of courses but do not effectively address issues related to student learning performance and classroom atmosphere. Therefore, to enhance the quality and effectiveness of online education, there is an urgent need for a new solution to address these challenges.

2.2 Selection of face recognition algorithms

Among the numerous facial recognition algorithms, selecting a simple, stable, efficient, and reliable algorithm is of paramount importance.[4,5] We conducted a comparison of several commonly used algorithms on two of the most popular facial detection benchmarks: FDDB and WIDER FACE.

The FDDB dataset comprises 2845 images collected from the Yahoo News website, containing 5171 individual facial instances. Our findings indicate that PyramidBox achieved top-tier performance on this benchmark, as illustrated in Fig. 1. Note that our model was exclusively trained on the WIDER FACE training dataset, underscoring PyramidBox's robust ability to detect unconstrained faces.

The WIDER FACE dataset consists of 32203 images, with 393703 annotated faces characterized by variations in scale, pose, and occlusion. The dataset is divided into three subsets: a training set (40%), a validation set (10%), and a test set (50%). Both the validation and test sets are further categorized into " Easy", " Medium", and "Hard" subsets based on detection



(a) (b)

Fig. 1. (Color online) Evaluation on FDDB dataset. (a) Discontinous ROC curves and (b) continous ROC curves.

difficulty. We trained PyramidBox on the training set, evaluated its performance on the validation and test sets, and compared it to other leading detectors. Figure 2 illustrates the precision–recall curves and mean average precision values, demonstrating that PyramidBox outperformed other detectors across all three difficulty levels. Additionally, PyramidBox previously secured the top position in all three evaluation subsets, "Easy", "Medium", and "Hard" in the world's most prestigious public evaluation benchmark for facial detection, the WIDER FACE dataset.

It is currently open-sourced and maintained by Baidu Inc., and has gained a strong reputation through widespread adoption in Baidu's related products. Therefore, we have adopted this algorithm for our study.

The experimental platform is designed with the PyramidBox algorithm(5) as its core support for face detection. PyramidBox is a recently developed single-pass algorithm that utilizes contextual information to deal with face detection in complex scenes. It improves upon existing methods in three main aspects. First, it obtains information beyond the face, such as the background, head, and torso, by using a semisupervised approach to generate approximate label points and anchor them. Second, it uses a low-level feature pyramid network to combine high-level contextual semantic features with low-level facial features, allowing it to predict faces of all scales in a single shot. Third, it introduces a context-sensitive structure to enhance the prediction network's capacity and accuracy of the final output. Refer to Fig. 3 for the architecture of PyramidBox.

During the training process, PyramidBox utilizes datasets such as FDDB and WIDER FACE and increases the number of samples for various face sizes and poses at different scales through

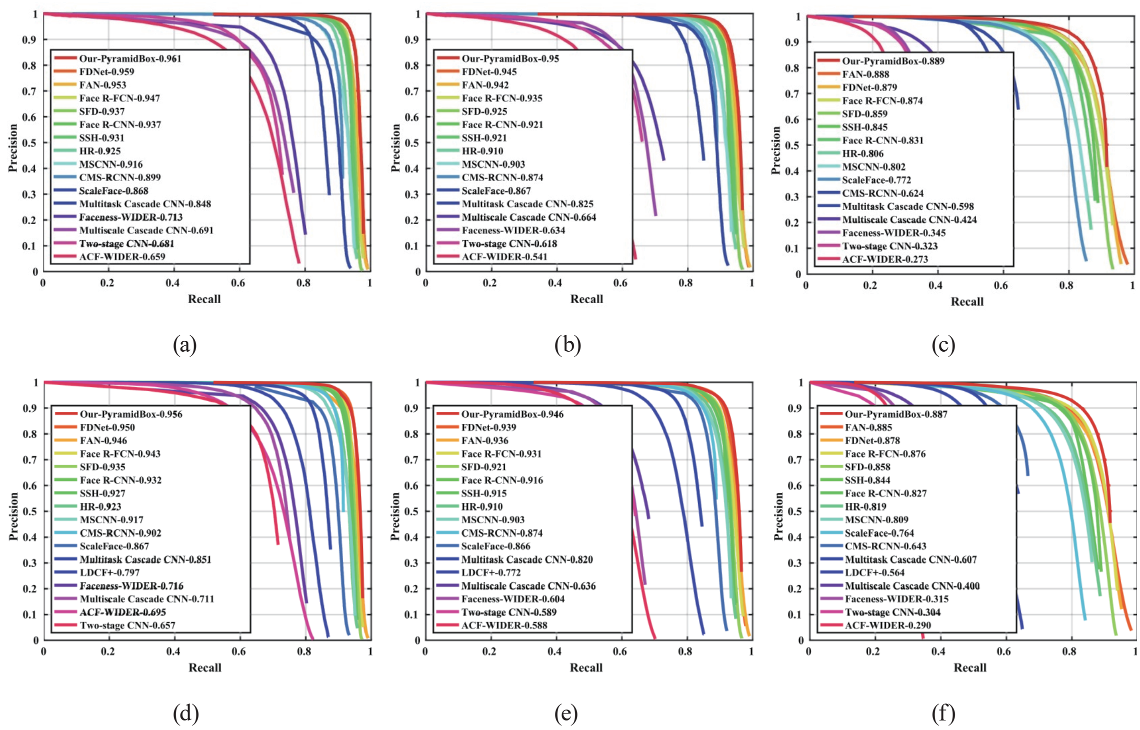


Fig. 2. (Color online) Precision–recall curves on WIDER FACE validation and test sets. (a) Val: easy, (b) val: medium, (c) val: hard, (d) test: easy, (e) test: medium, and (f) test: hard.

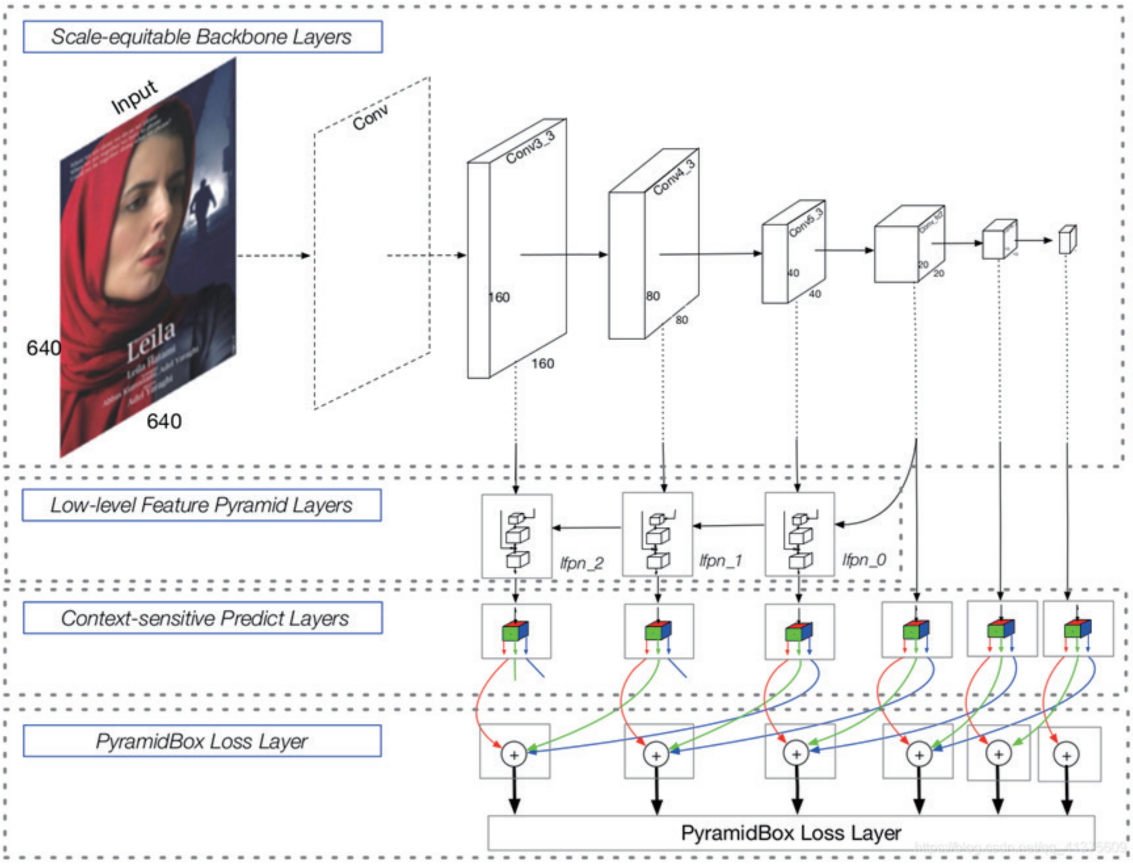


Fig. 3. (Color online) Architecture of PyramidBox.

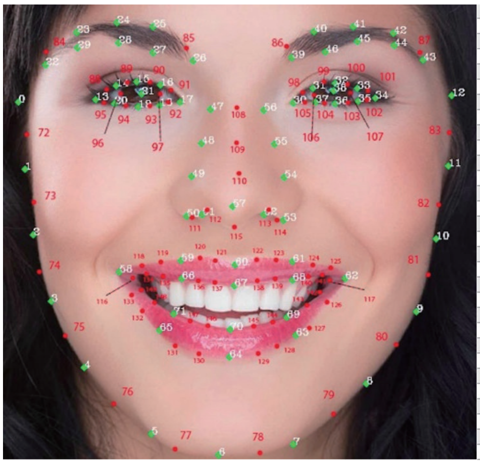
data anchor sampling, resulting in superior performance and one of the most accurate algorithms available.

To accurately identify facial information in acquired images, we parse their background and contextual content, isolate the face range, and extract 150 key points for further analysis. Refer to Fig. 4 for the Key points and corresponding naming.

By leveraging abundant feature point information, we can improve the accuracy of common facial information and extract emotional states. Specifically, we combine the Active Shape Model algorithm with the Convolutional Neural Network algorithm.(6) This approach allows us to determine human emotional information by analyzing the position and shape of facial features such as eyebrows, lips, and eyes.

2.3 System design

The experimental platform was developed using the C# programming language and is composed of three components, namely, student side, teacher side, and server side. To enhance client-side responsiveness and alleviate server-side pressure, WinForm programming was adopted for the student-side and teacher-side interfaces. The server side was developed using



0	Cheek_right_1 cheek_right_1	30	Eye_left_corner_right eye_left_corner_right	60	Mouth_lip_upper_outer_6 mouth_lip_upper_outer_6	90	Eye_right_eyelid_upper_5 eye_right_eyelid_upper_5	120	Mouth_lip_upper_outer_4 mouth_lip_upper_outer_4
1	Cheek_right_3 cheek_right_3	31	Eye_left_eyelid_upper_2 eye_left_eyelid_upper_2	61	Mouth_lip_upper_outer_9 mouth_lip_upper_outer_9	91	Eye_right_eyelid_upper_7 eye_right_eyelid_upper_7	121	Mouth_lip_upper_outer_5 mouth_lip_upper_outer_5
2	Cheek_right_5 cheek_right_5	32	Eye_left_eyelid_upper_4 eye_left_eyelid_upper_4	62	Mouth_corner_left_outer mouth_corner_left_outer	92	Eye_right_eyelid_lower_7 eye_right_eyelid_lower_7	122	Mouth_lip_upper_outer_7 mouth_lip_upper_outer_7
3	Cheek_right_7 cheek_right_7	33	Eye_left_eyelid_upper_6 eye_left_eyelid_upper_6	63	Mouth_lip_lower_outer_9 mouth_lip_lower_outer_9	93	Eye_right_eyelid_lower_5 eye_right_eyelid_lower_5	123	Mouth_lip_upper_outer_8 mouth_lip_upper_outer_8
4	Cheek_right_9 cheek_right_9	34	Eye_left_corner_left eye_left_corner_left	64	Mouth_lip_lower_outer_6 mouth_lip_lower_outer_6	94	Eye_right_eyelid_lower_3 eye_right_eyelid_lower_3	124	Mouth_lip_upper_outer_10 mouth_lip_upper_outer_10
5	Cheek_right_11 cheek_right_11	35	Eye_left_eyelid_lower_6 eye_left_eyelid_lower_6	65	Mouth_lip_lower_outer_3 mouth_lip_lower_outer_3	95	Eye_right_eyelid_lower_1 eye_right_eyelid_lower_1	125	Mouth_lip_upper_outer_11 mouth_lip_upper_outer_11
6	Chin_2 chin_2	36	Eye_left_eyelid_lower_4 eye_left_eyelid_lower_4	66	Mouth_lip_upper_inner_3 mouth_lip_upper_inner_3	96	Eye_right_eyeball_right eye_right_eyeball_right	126	Mouth_lip_lower_outer_11 mouth_lip_lower_outer_11
7	Cheek_left_11 cheek_left_11	37	Eye_left_eyelid_lower_2 eye_left_eyelid_lower_2	67	Mouth_lip_upper_inner_6 mouth_lip_upper_inner_6	97	Eye_right_eyeball_left eye_right_eyeball_left	127	Mouth_lip_lower_outer_10 mouth_lip_lower_outer_10
8	Cheek_left_9 cheek_left_9	38	Eye_left_eyeball_center eye_left_eyeball_center	68	Mouth_lip_upper_inner_9 mouth_lip_upper_inner_9	98	Eye_left_eyelid_upper_1 eye_left_eyelid_upper_1	128	Mouth_lip_lower_outer_8 mouth_lip_lower_outer_8
9	Cheek_left_7 cheek_left_7	39	Eyebrow_left_corner_right eyebrow_left_corner_right	69	Mouth_lip_lower_inner_9 mouth_lip_lower_inner_9	99	Eye_left_eyelid_upper_3 eye_left_eyelid_upper_3	129	Mouth_lip_lower_outer_7 mouth_lip_lower_outer_7
10	Cheek_left_5 cheek_left_5	40	Eyebrow_left_upper_2 eyebrow_left_upper_2	70	Mouth_lip_lower_inner_6 mouth_lip_lower_inner_6	100	Eye_left_eyelid_upper_5 eye_left_eyelid_upper_5	130	Mouth_lip_lower_outer_5 mouth_lip_lower_outer_5
11	Cheek_left_3 cheek_left_3	41	Eyebrow_left_upper_3 eyebrow_left_upper_3	71	Mouth_lip_lower_inner_3 mouth_lip_lower_inner_3	101	Eye_left_eyelid_upper_7 eye_left_eyelid_upper_7	131	Mouth_lip_lower_outer_4 mouth_lip_lower_outer_4
12	Cheek_left_1 cheek_left_1	42	Eyebrow_left_upper_4 eyebrow_left_upper_4	72	Cheek_right_2 cheek_right_2	102	Eye_left_eyelid_lower_7 eye_left_eyelid_lower_7	132	Mouth_lip_lower_outer_2 mouth_lip_lower_outer_2
13	Eye_right_corner_right eye_right_corner_right	43	Eyebrow_left_corner_left eyebrow_left_corner_left	73	Cheek_right_4 cheek_right_4	103	Eye_left_eyelid_lower_5 eye_left_eyelid_lower_5	133	Mouth_lip_lower_outer_1 mouth_lip_lower_outer_1
14	Eye_right_eyelid_upper_2 eye_right_eyelid_upper_2	44	Eyebrow_left_lower_3 eyebrow_left_lower_3	74	Cheek_right_6 cheek_right_6	104	Eye_left_eyelid_lower_3 eye_left_eyelid_lower_3	134	Mouth_lip_upper_inner_1 mouth_lip_upper_inner_1
15	Eye_right_eyelid_upper_4 eye_right_eyelid_upper_4	45	Eyebrow_left_lower_2 eyebrow_left_lower_2	75	Cheek_right_8 cheek_right_8	105	Eye_left_eyelid_lower_1 eye_left_eyelid_lower_1	135	Mouth_lip_upper_inner_2 mouth_lip_upper_inner_2
16	Eye_right_eyelid_upper_6 eye_right_eyelid_upper_6	46	Eyebrow_left_lower_1 eyebrow_left_lower_1	76	Cheek_right_10 cheek_right_10	106	Eye_left_eyeball_right eye_left_eyeball_right	136	Mouth_lip_upper_inner_4 mouth_lip_upper_inner_4
17	Eye_right_corner_left eye_right_corner_left	47	Nose_right_contour_1 nose_right_contour_1	77	Chin_1 chin_1	107	Eye_left_eyeball_left eye_left_eyeball_left	137	Mouth_lip_upper_inner_5 mouth_lip_upper_inner_5
18	Eye_right_eyelid_lower_6 eye_right_eyelid_lower_6	48	Nose_right_contour_2 nose_right_contour_2	78	Chin_3 chin_3	108	Nose_bridge_1 nose_bridge_1	138	Mouth_lip_upper_inner_7 mouth_lip_upper_inner_7
19	Eye_right_eyelid_lower_4 eye_right_eyelid_lower_4	49	Nose_right_contour_3 nose_right_contour_3	79	Cheek_left_10 cheek_left_10	109	Nose_bridge_2 nose_bridge_2	139	Mouth_lip_upper_inner_8 mouth_lip_upper_inner_8
20	Eye_right_eyelid_lower_2 eye_right_eyelid_lower_2	50	Nose_right_contour_4 nose_right_contour_4	80	Cheek_left_8 cheek_left_8	110	Nose_bridge_3 nose_bridge_3	140	Mouth_lip_upper_inner_10 mouth_lip_upper_inner_10
21	Eye_right_eyeball_center eye_right_eyeball_center	51	Nose_right_contour_6 nose_right_contour_6	81	Cheek_left_6 cheek_left_6	111	Nose_right_contour_5 nose_right_contour_5	141	Mouth_lip_upper_inner_11 mouth_lip_upper_inner_11
22	Eyebrow_right_corner_right eyebrow_right_corner_right	52	Nose_left_contour_6 nose_left_contour_6	82	Cheek_left_4 cheek_left_4	112	Nose_right_contour_7 nose_right_contour_7	142	Mouth_lip_lower_inner_11 mouth_lip_lower_inner_11
23	Eyebrow_right_upper_2 eyebrow_right_upper_2	53	Nose_left_contour_4 nose_left_contour_4	83	Cheek_left_2 cheek_left_2	113	Nose_left_contour_7 nose_left_contour_7	143	Mouth_lip_lower_inner_10 mouth_lip_lower_inner_10
24	Eyebrow_right_upper_3 eyebrow_right_upper_3	54	Nose_left_contour_3 nose_left_contour_3	84	Eyebrow_right_upper_1 eyebrow_right_upper_1	114	Nose_left_contour_5 nose_left_contour_5	144	Mouth_lip_lower_inner_8 mouth_lip_lower_inner_8
25	Eyebrow_right_upper_4 eyebrow_right_upper_4	55	Nose_left_contour_2 nose_left_contour_2	85	Eyebrow_right_upper_5 eyebrow_right_upper_5	115	Nose_middle_contour nose_middle_contour	145	Mouth_lip_lower_inner_7 mouth_lip_lower_inner_7
26	Eyebrow_right_corner_left eyebrow_right_corner_left	56	Nose_left_contour_1 nose_left_contour_1	86	Eyebrow_left_upper_1 eyebrow_left_upper_1	116	Mouth_corner_right_inner mouth_corner_right_inner	146	Mouth_lip_lower_inner_5 mouth_lip_lower_inner_5
27	Eyebrow_right_lower_3 eyebrow_right_lower_3	57	Nose_tip nose_tip	87	Eyebrow_left_upper_5 eyebrow_left_upper_5	117	Mouth_corner_left_inner mouth_corner_left_inner	147	Mouth_lip_lower_inner_4 mouth_lip_lower_inner_4
28	Eyebrow_right_lower_2 eyebrow_right_lower_2	58	Mouth_corner_right_outer mouth_corner_right_outer	88	Eye_right_eyelid_upper_1 eye_right_eyelid_upper_1	118	Mouth_lip_upper_outer_1 mouth_lip_upper_outer_1	148	Mouth_lip_lower_inner_2 mouth_lip_lower_inner_2
29	Eyebrow_right_lower_1 eyebrow_right_lower_1	59	Mouth_lip_upper_outer_3 mouth_lip_upper_outer_3	89	Eye_right_eyelid_upper_3 eye_right_eyelid_upper_3	119	Mouth_lip_upper_outer_2 mouth_lip_upper_outer_2	149	Mouth_lip_lower_inner_1 mouth_lip_lower_inner_1

Fig. 4. (Color online) Key points and corresponding naming.

.NET Core technology, featuring a highly available architecture to ensure system stability and reliability. The service is hosted on Alibaba Cloud servers. Refer to Fig. 5 for the system design.

The student side retrieves raw data by accessing the camera of the Windows system. However, during live sessions, the camera is typically occupied by live streaming software. To address this issue, the system creates a virtual camera to share the hardware and acquire the data. If the student does not have a camera enabled during a live session, the available camera is called for data acquisition.

The student-side program is designed to optimize system performance and reduce server workload. The program first checks the camera functionality and then proceeds with image acquisition, preprocessing, face detection, live recognition, and feature code extraction. Once the program confirms the validity of the detected face, it sends the face information data and timestamp to the server side using WebAPI, and stores the result locally. The program captures camera images every 30 s by default and stores them locally for additional information.

The server side is built using .NET Core technology and is hosted on a cloud server with high-availability architecture to ensure system stability and reliability. It receives data sent by each client in real time, performs operations such as cleaning, validity confirmation, classification, archiving, and data analysis, and filters events on the basis of data content. It provides forensic authentication through JWT and supports cross-platform deployment. The server side has important functions such as content analysis and message pushing, which filter events on the basis of data content and assign different levels of positivity. Warning and notification types are the two types of event with different priority levels. Warning-type events require immediate notification to the teacher side and include instances where the student does not appear in front of the camera within the specified time or when someone other than the certified person is in front of the camera. Notification-type events include instances where more than 20% of the students are absent or more than 30% of the students show expressions of frustration or boredom. All other data contents are saved as a record type in the database.

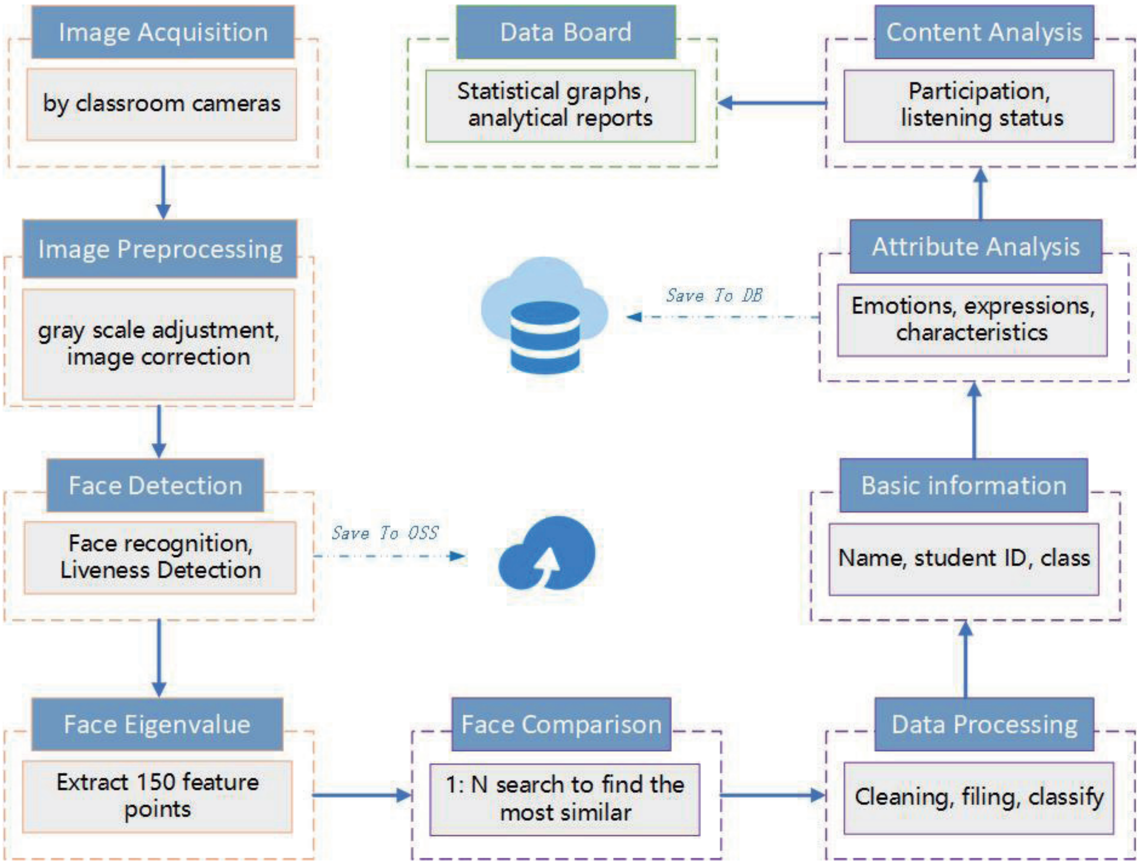


Fig. 5. (Color online) System design.

The teacher side is a standard desktop-type application that receives data transmitted from the server side in real time and has functions such as display, reminder, parameter setting, statistical analysis, import, and export. It operates silently during the live course and displays a mandatory pop-up reminder when the data type is warning and a reminder in the form of a special color text pop-up for notification types. The management side can set different types of trigger times and notification forms. At the end of the lesson, the management terminal automatically generates different dimensional letter report styles to help teachers review the teaching effect of the whole lesson. Teachers can view the statistics under different dimensions in detail by selecting students, time period, class status, and so forth.

2.4 Face detection

We detect the presence of human faces in the data source using a camera. If a face is detected, it is identified, and if multiple faces appear simultaneously, the first clear face detected is selected as the experimental object. The sorting rule for face detection is based on the area occupied by the face in the image, from largest to smallest. The face that occupies the largest position in the image is selected as the experimental object. During the detection process, transparent eyeglasses, masks, and similar items are allowed, but major obstructions of facial features are not permitted. We use the following parameter range settings (Table 1) to ensure the accuracy of face recognition. If no face information is detected in the data source, a warning will be issued to the server. Please refer to Fig. 6 for relevant reference information.

2.5 Face matching

As shown in Fig. 7, during the face recognition process, the program will compare the acquired face feature values with the feature values in the face library one by one and calculate the similarity between them. If the face in the image to be detected is not obscured, the program will select the target image with the highest similarity and a similarity score greater than 95% as the best result and return it. However, if the face is partially blocked in the image to be detected, the program will select the target image with the highest similarity and a similarity score greater than 85% as the best result and return it. This is because partial occlusion can affect the accuracy of face recognition, and a lower similarity threshold may be necessary to ensure accurate matching.

After ensuring the validity of the face, the client-side program sends the acquired face information data and timestamp to the server side using WebAPI. The server side then compares the acquired face feature values with the feature values in the face library to determine the

Table 1
Key parameters and values for face detection.

Index	Explain	Value range
Occlusion range	Value range: [0–1], 0 is no occlusion, 1 is complete occlusion	<0.5
Ambiguity range	The value range is [0–1], where 0 is the clearest and 1 is the most ambiguous	>0.5
Illumination range	Value range: [0–255], facial gray scale value, the larger the value, the clearer the image	>60
Face integrity	Whether to allow faces to overflow image boundaries	NO
Face size	Size of the face portion	$\geq 100 \times 100$ pix

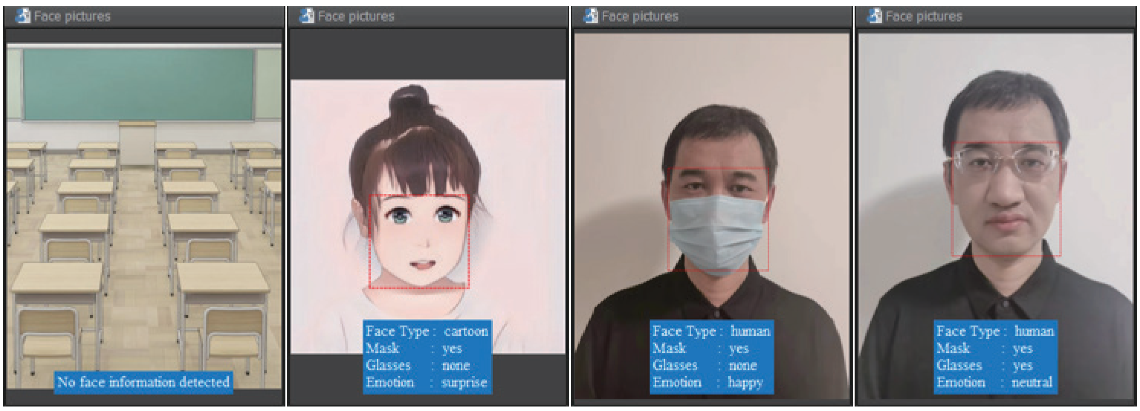


Fig. 6. (Color online) Face detection.

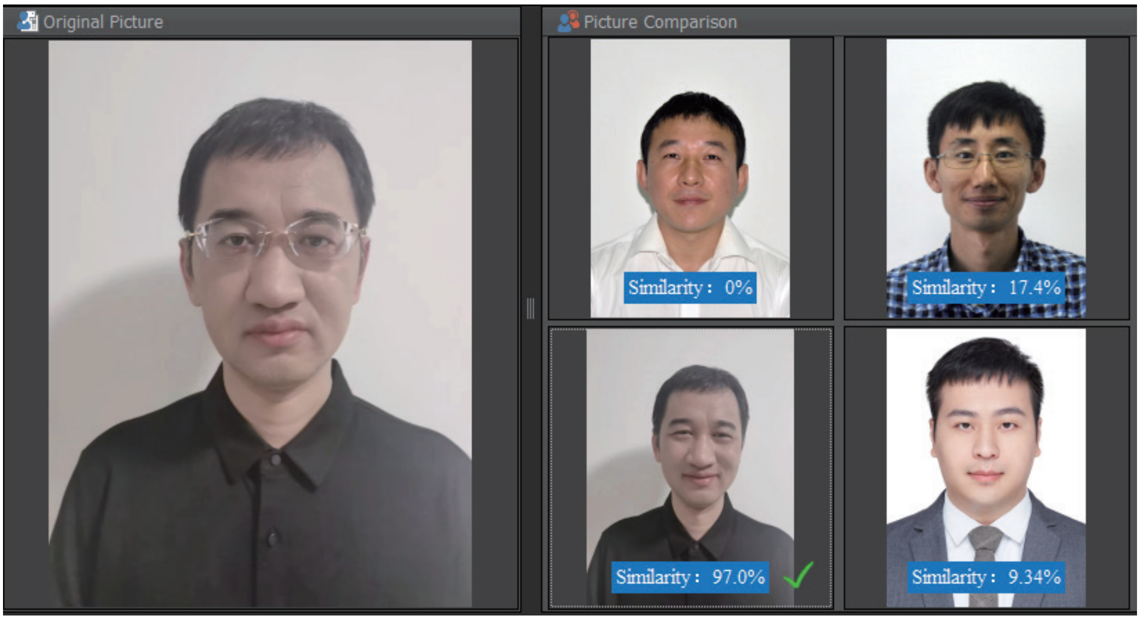


Fig. 7. (Color online) Face search comparison.

identity of the person. If the person is confirmed to be a valid attendee for the session, their attendance is recorded and their basic information is stored locally on the client side. If the person is not confirmed to be a valid attendee, a warning alert is sent to the server side and synchronized to the management side.

2.6 Face attribute recognition

Face attribute recognition can be divided into two parts. The first part involves identifying the authenticity, expression, and emotions of individuals based on the feature code information in the image, as shown in Fig. 8.

Second, by calling the server interface to obtain the registration information of personnel in the registration database, including class, name, student number, and related contents, the combination of the two can help teachers better understand the learning situation of students and provide reference and guidance for teaching, as shown in Fig. 9.

This part of the data is an important reference source for the results of this experiment. Response time is a critically important performance metric in academic research. The response time in this design comprises two main components, namely, image loading time and data analysis time.

On the basis of multiple experiments, we have concluded that the current system architecture pattern has achieved significant stability in both of these aspects, allowing the response time to be maintained at a very short duration of around 200 ms, which results in an excellent user experience. Refer to Fig. 10 for the time test info.

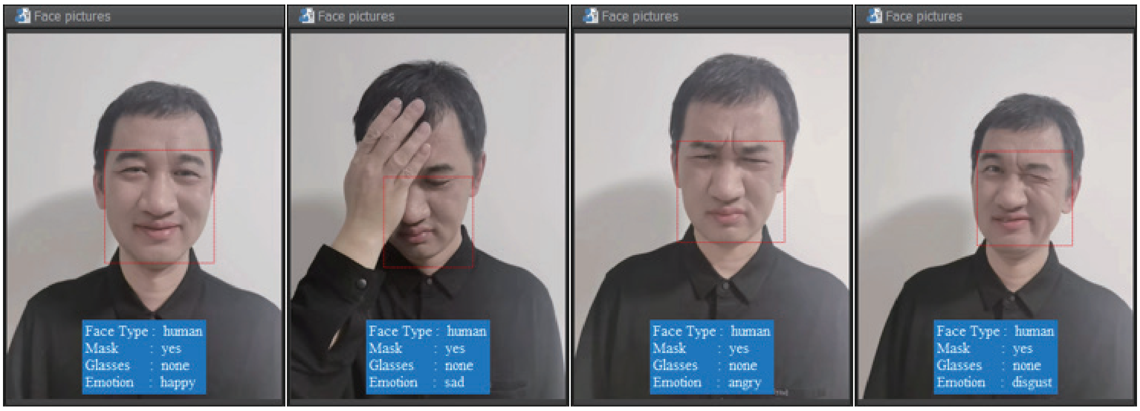


Fig. 8. (Color online) Expression analysis.

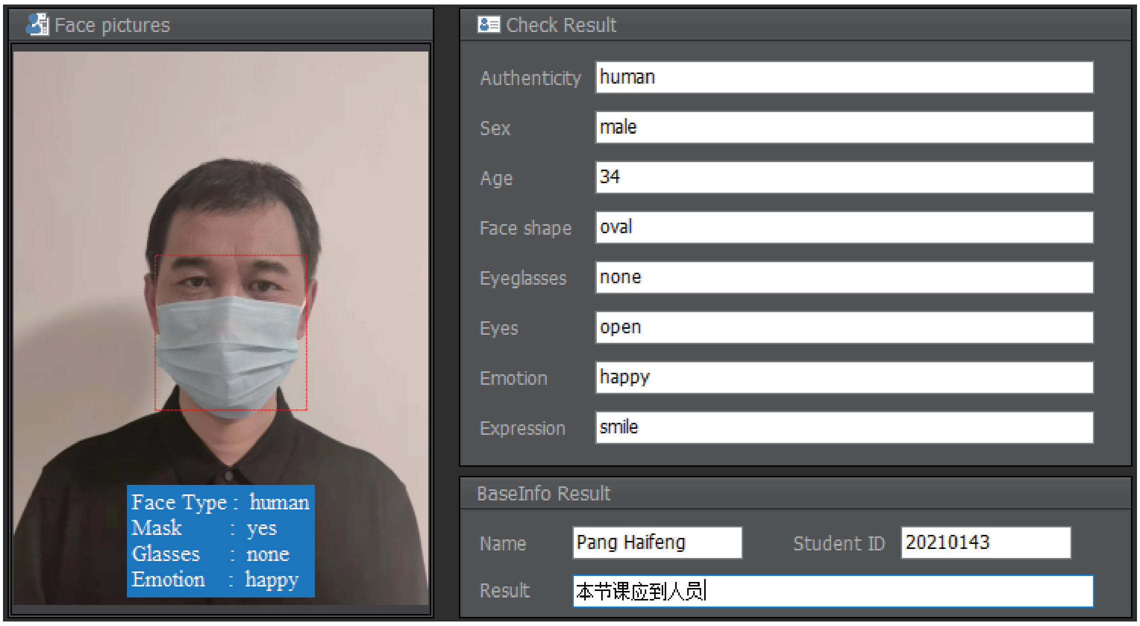


Fig. 9. (Color online) Comprehensive data properties.

2.7 Experimental scenario and process

The experiment was conducted during a 90-min lecture of a brand new course, with 25 students and one teacher participating in the entire process. The students were not notified in advance about the experimental content to ensure the authenticity and effectiveness of the experiment. We recorded and analyzed various states of the students to examine their classroom discipline and understanding of the course at different time periods. The lecture consisted of five parts, namely, attendance, review, teaching of new knowledge points, questioning, and assignment. We used Tencent Meeting to live stream the lecture. Refer to Table 2 for the Experimental environment requirements.

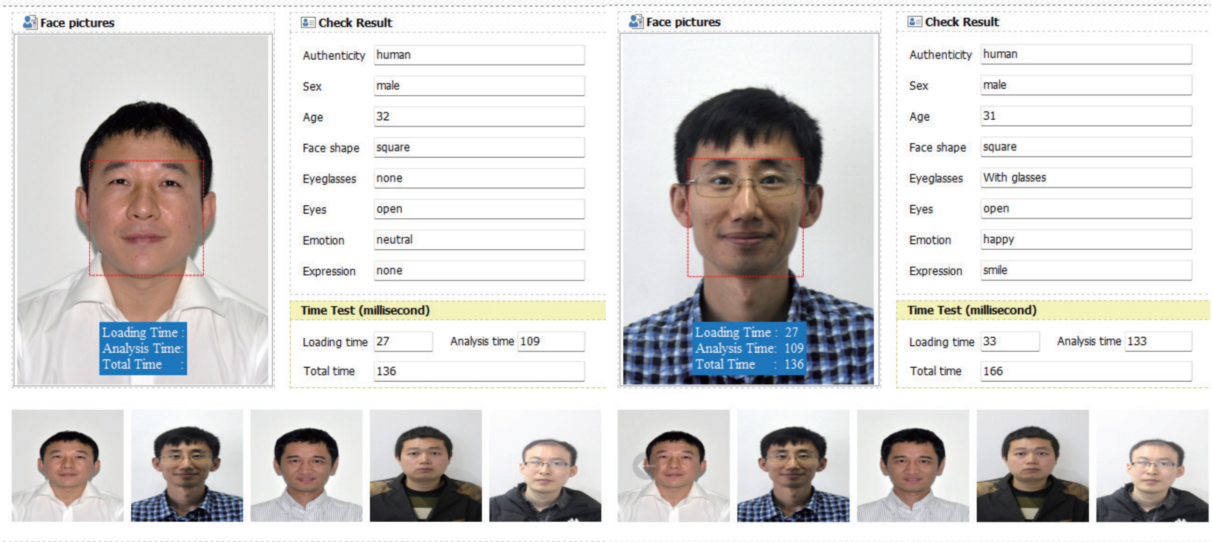


Fig. 10. (Color online) Time test info.

Table 2
Experimental environment requirements.

Configuration Items	Client	Server
Software Platform	Windows 7 or later	Windows server 2016
Support Environment	Net 4.0 or later	Net Core 6
Network Requirements	2M or later	10M or later
Hardware and Peripherals	Physical camera	None
Database	None	Sqlserver 2019
Image Storage	Local File System	OSS (Object Storage Service)

The course's overall difficulty was not high, and it was closely related to the content of the previous class. If the students did not understand the previous class's content thoroughly, they may find it difficult to learn the new knowledge points in the simple part of this class. However, the difficult part of the new knowledge points in this class was relatively independent, and it could be quickly understood with careful listening. The questioning and interactive part was conducted by randomly selecting answerers. The lecture started promptly at 14:00 pm, and the teaching content of each time period and the experimental focus are shown in Table 3.

When examining individual students, our focus is on whether they arrive on time for class, whether they are registered attendees, whether they leave during the class, whether they listen attentively throughout the class, and their overall behavior during the lecture. We pay special attention to their facial expressions and emotions, including when they smile, laugh, appear sad or dejected, and their expressions during the review and new knowledge teaching segments, as well as whether they frequently leave the camera view.

When analyzing the overall situation, we statistically analyze the distribution of various facial expressions and emotions over time. During the question-and-answer segment, we also

Table 3
Experimental process arrangement.

No.	Content	Time	Testing items
1	Call the roll	13:55–14:00	Attendance status, in person or not
3	Review of the previous lesson	14:01–14:20	Expressions, emotions
4	New knowledge points (easy)	14:31–14:50	Expressions, emotions, attendance status
5	New knowledge points (difficult)	14:51–15:10	Expressions, emotions, attendance status
6	Ask questions and interact	15:11–15:20	Expressions of the questioned
7	Class summary	15:21–15:26	Expressions, emotions, attendance status
8	Assignment of homework	15:27–15:30	Expressions, emotions
9	Class over	15:30	None

pay attention to the facial expression changes between the person being questioned and those who are not, in order to determine the relative difficulty of the question and the stress level it creates for the students. Through this data, we can better understand students' comprehension and learning attitude toward the course, which helps us continually optimize the course design and teaching methods to improve teaching effectiveness.

In addition, we also pay attention to students' participation and feedback. We observe whether students are actively answering questions, whether they have the desire to ask questions or discuss, and whether they engage in interactive communication with other classmates. We also pay attention to their language expression and logical thinking, as well as their attention concentration and depth of thinking during class. Observing and analyzing these aspects help us objectively reference the authenticity and accuracy of the entire experiment.

3. Results and Discussion

Since the front-end strategy was to capture camera data every 30 s, a total of 4750 data samples were obtained for a 95-min class attended by 25 students (including 5 min of pre-class roll call). From the data samples, it can be observed that the number of students attending the class increased sequentially during the pre-class roll call period, and all of them arrived before the formal class began, consistent with the teacher's roll call. At a few key time points, only 23 and 24 students were in front of the camera at 14:30 and 14:45, respectively, and all 25 students were online at the end of the course at 15:30.

During the sign-in phase, students exhibited predominantly relaxed and smiling facial expressions. This may be attributed to the fact that sign-in marks the beginning of the classroom activities, and hence students are generally at ease, not yet in a learning mode. During the review of the previous lesson, most students remained expressionless, whereas a few displayed relatively relaxed expressions, possibly due to familiarity with the content or a less stressful task. Additionally, students may have perceived reviewing as a non-stressful task, hence their relaxed expressions.

In the phase of teaching simple concepts of the new lesson, students exhibited even more relaxed expressions than during the roll-call phase. It is likely because the content was relatively straightforward, and students could comprehend it with ease. Alternatively, students may have been interested in the new knowledge, hence their pleasant expressions. However, during the complex knowledge stage, students' expressions became complicated, and they displayed expressions such as nervousness and surprise. Nonetheless, most students remained relatively expressionless, possibly due to their attentive listening and deep thinking, or difficulty grasping the new knowledge. Alternatively, students may have restrained their emotions in class.

As shown in Fig. 11, during the interactive segment of the class, the majority of students appeared tense, with only a few exhibiting a relaxed demeanor. This could be attributed to students' lack of confidence in answering questions or the fear of being ridiculed for giving incorrect responses. Additionally, students' tense expressions could be due to their desire to perform at their best during class. Finally, during the final assignment stage, students relaxed once more, with some even making faces. This could be because students felt relieved having completed their assignments for the day, or they wanted to unwind before the end of the class to prepare for the next one.

Facial accessory attributes can indeed provide valuable information in understanding facial expressions and emotions. The observations and analyses mentioned in the study shed light on the potential impact of environmental and situational factors on the collection of facial information, as well as the changing attitudes and perceptions of students towards the current epidemic situation.

The first phenomenon observed, where a student was flagged as suspicious owing to incomplete facial information during pre-class roll call, highlights the importance of careful observation and analysis to ensure accurate data collection. It is possible that some students may deliberately or accidentally interfere with the collection of facial information, and this should be taken into consideration during the analysis.

The gradual decrease in the number of students wearing masks during the session is also a significant finding. It may suggest that students are becoming more complacent about the epidemic situation or have developed a more relaxed attitude towards preventive measures. This observation may have implications for public health policy and interventions aimed at promoting adherence to preventive measures.

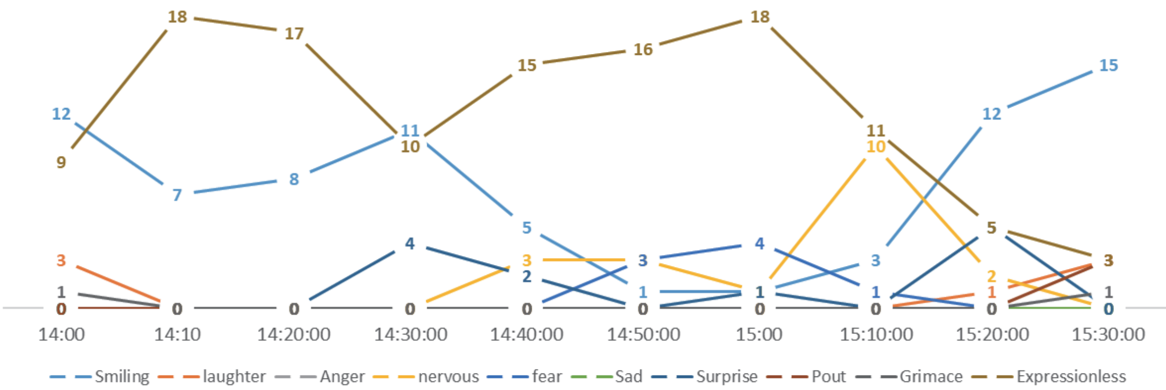


Fig. 11. (Color online) Classroom mood and time analyses.

The last phenomenon observed, where the number of students wearing glasses increased during the teaching of new knowledge, suggests that students may require more visual support during the learning process. This finding may have implications for the design of educational materials and instructional methods that take into account the visual needs of students. Refer to Fig. 12 for the facial attributes and time analyses.

Overall, the study highlights the importance of considering not only facial expressions but also facial accessory attributes in understanding human behavior and emotions. It also underscores the need for careful observation and analysis to ensure accurate data collection and interpretation.

Upon reflection, our experiment revealed that nearly half of the students (49%) exhibited an expressionless state, while 30% displayed a smiling expression and 8% showed signs of nervousness; the full data can be found in Fig. 13. The remaining 13% demonstrated a variety of

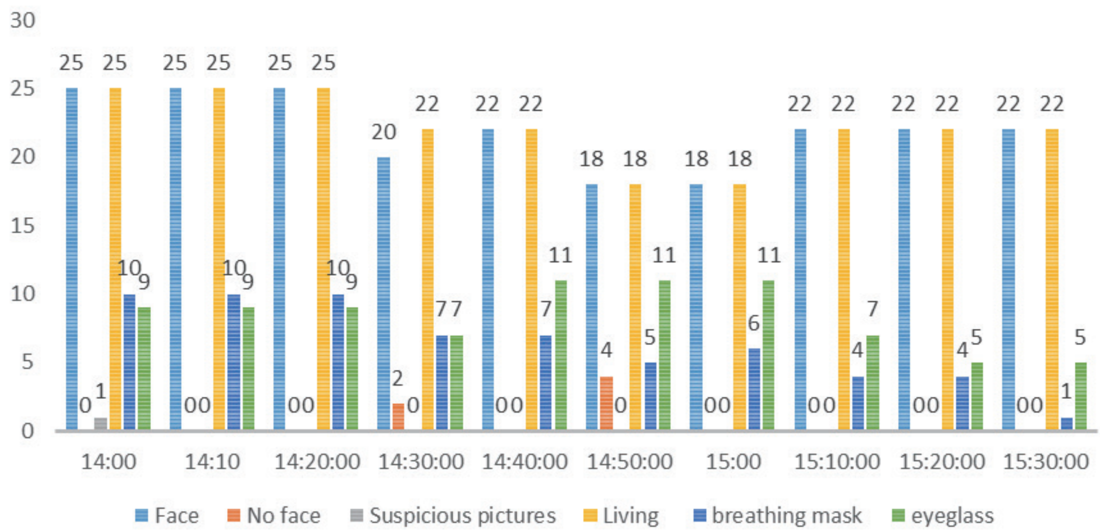


Fig. 12. (Color online) Facial attributes and time analyses.

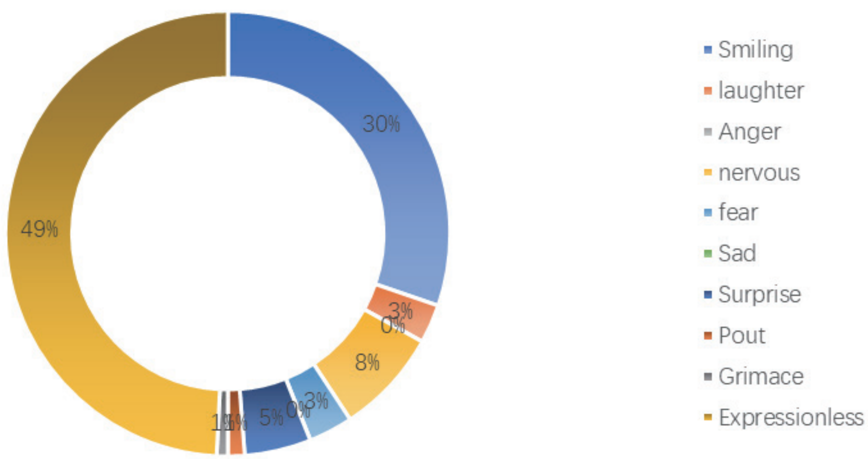


Fig. 13. (Color online) Emotion scale.

other emotional states. The data indicates a relatively uneventful classroom atmosphere and a lecture content that lacked engagement, which is consistent with the lecturer's post-class evaluation. These findings highlight the importance of considering students' emotional states, as they can significantly impact learning effectiveness and performance. To address different emotional states, tailored measures such as providing guidance during new knowledge delivery or encouragement during questioning sessions can enhance learning motivation and effectiveness.

3.1 Application outcomes

In a series of repeated teaching experiments, using the same course, the same teaching staff, and the same student cohort, we observed that facial recognition technology significantly enhances the quality of education across multiple dimensions. Notable impacts included the following:

1. Attendance Rate: The experimental group, which utilized facial recognition technology, exhibited a substantially higher attendance rate than the control group. After implementing facial recognition detection, the students' attendance consistently reached 100%, indicating the technology's efficacy in reducing absenteeism.
2. Student Engagement: Students in the experimental group demonstrated heightened levels of engagement, displaying a proclivity for active participation in classroom activities, posing questions, and sharing viewpoints. Relative to the control group, student engagement in the experimental group improved by 20%.
3. Student Focus: Students in the experimental group exhibited increased levels of focus. They were less susceptible to external distractions, allowing them to concentrate more effectively on their studies. Student focus in the experimental group improved by approximately 25%.
4. Teacher Classroom Control: Teachers employing facial recognition technology found it easier to maintain control over the classroom, as they could better discern students' emotions and needs. This led to more efficient teaching and improved student management, resulting in a 30% increase in teacher classroom control.
5. Online Learning Duration: Students in the experimental group spent more time engaged in online learning activities, potentially owing to their increased engagement and focus. On average, their online learning duration increased by 40%.
6. Post-class Exercise Accuracy: Students using facial recognition technology performed better in post-class exercises, with a 10% increase in exercise accuracy. This improvement may be attributed to their enhanced attentiveness and engagement during class, enabling them to better absorb the knowledge.

The utilization of facial recognition technology enables the real-time monitoring of individual students' classroom statuses, enabling instructors to focus more intensively on teaching activities and providing the means to contemporaneously apprehend diverse aspects of student engagement. Consequently, it surpasses, or at least equals, the advantages of traditional in-person instruction, thereby significantly enhancing educational quality. This technology contributes to the enhancement of student attendance, enthusiasm, attentiveness, teacher classroom management, extended online learning durations, and heightened post-class exercise accuracy. The findings underscore the latent potential of facial recognition technology in the domain of education, fostering an environment that is more conducive to student learning.

3.2 Limitations in practical application

In practical application, several challenges have arisen. First, concerns are related to the high network demands, which exert substantial pressure on network bandwidth. This strain primarily emanates from the real-time transmission of video data and the processing of facial recognition, potentially leading to latency and a degradation in video quality.

A second challenge manifests as sporadic anomalies in camera sharing. Users may encounter difficulties accessing images due to platform or software-related issues when employing cameras for facial recognition.

Furthermore, user experience has emerged as a critical consideration. Users are not only required to access the live teaching platform but also a distinct facial recognition monitoring program. This dual-access requirement results in a fragmented user experience.

To address these issues, our forthcoming research endeavors will encompass the following upgrades.

First, our focus will be on network optimization. This entails the exploration of various techniques, including data compression, optimized video encoding algorithms, and cloud-based solutions. These measures are aimed at alleviating the network load, thereby facilitating a more efficient data transmission. It is anticipated that this initiative will alleviate the strain on network resources and enhance system performance.

Second, we will tackle the challenges surrounding the camera sharing mechanism. A refinement of the current camera sharing method is imperative to resolve these issues. We will adjust the existing approach to camera access to ensure a reliable and seamless camera usage.

Third, our concentration will be on the integration of facial recognition technology to enhance the user experience. By directly integrating facial recognition technology into the live platform, we aim to enhance user-friendliness and create a seamless experience. This integration will spare users from the need to simultaneously utilize multiple applications, ultimately resulting in an improved overall user experience.

In summary, our implementation of these solutions is designed to enhance system stability, reliability, and user experience, aligning with the objectives of scientific research.

4. Conclusion

Compared with prior endeavors, our current study underscores numerous significant advantages. The incorporation of facial recognition technology into our platform represents a groundbreaking step in the field of distance learning. In contrast to previous approaches that often relied on conventional methodologies and standards, our system offers a comprehensive and detailed insight into learners' emotional and interactive responses.

One of the most noteworthy benefits of our work lies in its real-time monitoring capabilities. Traditional methods for assessing student engagement and emotional states frequently relied on self-reporting and retrospective analyses. In contrast, our platform provides immediate and continuous feedback, enabling educators to intervene promptly and significantly enhance the overall educational experience.

Furthermore, our research has achieved a remarkable level of granularity in the identification of emotional states. We have gained an in-depth understanding of the full spectrum of emotions displayed by students at various stages of their learning journey. This invaluable insight equips educators with a potent tool for customizing their guidance to cater to the individual needs of each student, particularly during synchronous online instructional sessions. The potential impacts on students' success and overall satisfaction with online education are profound. Our analysis has unveiled the pivotal role of emotional states and classroom ambiance in evaluating course quality. This comprehensive perspective greatly contributes to the assessment of pedagogical effectiveness, pinpointing areas in need of improvement and highlighting strengths that can be harnessed to enhance the learning experience.

While our current study has made substantial strides in data collection and analysis during online instructional delivery, we anticipate future research endeavors that will focus on incorporating advanced statistical data analysis features to provide valuable insights and actionable recommendations. Furthermore, we intend to implement interface services that will facilitate the seamless transfer of data to external systems through a variety of communication channels, thereby enhancing the platform's overall user-friendliness and adaptability.

In summary, our research stands out for harnessing the capabilities of facial recognition technology to revolutionize online education. Real-time monitoring, nuanced emotional insights, and improved course evaluation are clear indications of the superiority of our approach. We eagerly anticipate future inquiries that will continue to drive progress in this field, meeting the evolving demands of online education and creating new avenues of innovation in the educational sector.

References

1 J. F. Xu and Y. Yang: Foreign Lang. Educ. **44** (2023) 54. https://doi.org/10.16362/j.cnki.cn61-1023/h.2023.01.015
2 R, F. Qi and C. Y. Z: Mod. Educ. Technol. **31** (2021) 42. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=XJJS202112007&DbName=CJFQ2021
3 Y. Sun, L. Hu, and X. L. Feng: Comput. Eng. **49** (2023) 269. https://doi.org/10.19678/j.issn.1000-3428.0066497
4 M. C. Huang, I. P. Chen, H. L. Huang, S. F. Sung, and A. G. Wang: Sens. Mater. **33** (2021) 1787. https://doi.org/10.18494/SAM.2021.3203
5 X. Tang, D. K. Du, Z. He, and J. T. Liu: Computer Vision – ECCV 2018. (ECCV,2018) 812–828. https://doi.org/10.1007/978-3-030-01240-3_49
6 C. P. Chueh, K. J. Hu, Y. S. Chang, and S. L. Kao: Sens. Mater. **34** (2022) 3751. https://doi.org/10.18494/SAM4046
7 S. Y. Lin, C. M. Wu, S. L. Chen, T. L. Lin, and Y. W. Tseng: Sens. Mater. **32** (2020) 3243. https://doi.org/10.18494/SAM.2020.2863
8 S. Y. Deng and K. K. Fan: Sens. Mater. **33** (2021) 3379. https://doi.org/10.18494/SAM.2021.3479
9 A. Kawabe, R. Haga, Y. Tomioka, J. Shin, and Y. Okuyama: Electronics **12** (2023) 3932. https://doi.org/10.3390/electronics12183932
10 N. Kim, J.-H. Kim, and C. S. Won: Electronics **11** (2022) 875. https://doi.org/10.3390/ELECTRONICS11060875

11 K. Xiao, W. Guo, and X. Huang: Neural Comput. Appl. **35** (2022) 973. https://doi.org/10.1007/S00521-022-07823-Z
12 M. C. Zhong, J. L. Zhang, Y. B. Lan, and Y. H. He: Comput. Sci. **47** (2020). 196. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKibYlV5Vjs7i8oRR1PAr7RxjuAJk4dHXoinag5hSqNS9SnQL7968uhIpi7rar1E_LYrmFtEZmtoW&uniplatform=NZKPT
13 X. K. Liu, Z. Y. You, Y. X. He, S. Bi, and J. Wang: Pattern Recognit. **125** (2022) 108520. https://doi.org/10.1016/J.PATCOG.2022.108520
14 W. Z. Shao, H. M. Hu, J. Y. Li, and H. S. Deng: J. Nanjing University of Posts and Telecommunications (Natural Science Edition) **43** (2023) 5. https://doi.org/10.14132/j.cnki.1673-5439.2023.01.001
15 L. Q. Lan, Q. Y. Liu, and S. H. Lu: J. Beijing University of Aeronautics and Astronautics **48** (2022). 147. https://doi.org/10.13700/j.bh.1001-5965.2020.0507
16 Z. H. Li, Z. L. Zhang, and H. Liu: Chin. J. Distance Educ. **10** (2020) 9. https://doi.org/10.13541/j.cnki.chinade.2020.10.002
17 M. Yu, Y. X. Zhong, and Y. Wang: Comput. Eng. **49** (2023) 1. https://doi.org/10.19678/j.issn.1000-3428.0065790
18 X. X. Li, C. Zhao, and Z. L. Yan: Exp. Technol. Manage. **37** (2020) 172. https://doi.org/10.16791/j.cnki.sjg.2020.06.037
19 X. C. Chen, X. W. Zheng, K. Sun, W. L. Liu, and Y. Zhang: Inf. Sci. **634** (2023) 206. https://doi.org/10.1016/J.INS.2023.03.105
20 Y. Y. Liu, J. Y. Peng, W. Dai, J. B. Zeng, and S. G. Shan: Pattern Recognit. **139** (2023) 109496. https://doi.org/10.1016/J.PATCOG.2023.109496

About the Authors

Mingyang Liu received her B.S. degree from Harbin Normal University, China. Since 2017, she has been an associate professor in the College of Arts and Information Engineering, Dalian Polytechnic University, and is mainly engaged in education informatization. She has participated in many international conferences in related fields, has published several academic papers, and holds several related patents. (57965289@qq.com)

Haifeng Pang graduated from Heilongjiang University in China. He is a software engineer and has been working in software development since 2002. He has participated in the development of many large-scale application software and has developed his own open-source software. He is currently employed at Hi-Think Technology, Corp. in Dalian, focusing on system architecture and software development research. (214634166@qq.com)