Received 19 July 2023, accepted 4 August 2023, date of publication 7 August 2023, date of current version 18 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3302919

RESEARCH ARTICLE

Chinese Face Dataset for Face Recognition in an Uncontrolled Classroom Environment

NIANFENG LI[1], XIANGFENG SHEN[2], LIYAN SUN[1], ZHIGUO XIAO[1], TIANJIAO DING[2], TIANSHENG LI[2], AND XINHANG LI[3]

[1]College of Computer Science and Technology, Changchun University, Changchun 130022, China
[2]College of Cyber Security, Changchun University, Changchun 130022, China
[3]School of Physical Science and Information Engineering, Liaocheng University, Liaocheng 252059, China

Corresponding author: Liyan Sun (2276632042@qq.com)

This work was supported in part by the Science and Technology Development Center of the Ministry of Education under Grant 2020HYB03002 and Grant 2020ITA07031, and in part by the Jilin Science and Technology Development Plan Project under Grant 20210201083GX.

ABSTRACT Since the position of the classroom surveillance camera is not fixed, the angle of the face captured through the surveillance video is also different. The deep learning-based face verification model has achieved good results in controlled environments, but there is still much room for improvement in the face verification ability in uncontrolled environments. The performance of the model depends not only on the results of the network but also on the quality and diversity of the dataset. The current Asian face dataset in an uncontrolled environment is insufficient; for this reason, this paper constructs a Chinese face dataset (UCEC-Face) in an uncontrolled classroom environment, which is collected by 35 real classroom surveillance videos. The UCEU dataset contains 7395 images of 130 subjects, including 44 males and 86 females. To verify that there is still room for improving the performance of existing face verification models for Asian face verification, we further utilize four models such as OpenFace and ArcFace for face verification, as well as the VGG-Face model for gender, expression, and age recognition on the UCEC-Face. The experimental results show that the UCEC-Face constructed in this paper is more challenging and difficult to verify in face verification tasks because it is closer to the real environment, and the best results obtained on the existing models only reach 69.7%, which is largely below the average accuracy of the identification results of other datasets.

INDEX TERMS UCEC-face dataset, face recognition, face verification, face detection.

I. INTRODUCTION

Face recognition has been an active topic in the field of computer vision. With the advancement of convolutional neural networks, deep learning-based methods have achieved great success in face recognition tasks. In face recognition tasks, the image is generally used as the input, and the output is obtained through face detection, face alignment, and face representation, for a total of three steps [1]. Face detection is the first stage of the face recognition task, and its job is to locate the position of the face in the image. Specifically, face detection aims to locate the position of faces in the image and output the coordinates of bounding boxes with confidence scores. Face detection is mainly divided into single-stage [2], dual-stage [3], [4], and multistage methods [5], [6], and in the experiments, each stage is demonstrated to reflect the inspection effect of different methods in a real environment. The second stage is face alignment, which crops the image to a normalized pixel size. based face alignment utilizes spatial transformation to calibrate faces to the predefined canonical layout by involving facial landmarks as the reference [7], [8] [9], [10]. Landmark-free Face [11], [12], on the other hand, processes the alignment transformation in DCNNs without using facial landmarks. Since we collected the dataset without landmark operations, the second face alignment method is used in the processing of the alignment. The third

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Jin.

step is the face representation phase, where the features are extracted from the aligned facial images, and the extracted features are used to calculate the similarity, thus completing the face recognition task. The current achievements in face representation are due to advances in deep learning network architectures, and in addition to traditional network architectures [13], [14] [15], [16] [17], [18], supervised training also plays an important role in the face representation task. In the supervised approach, we can divide the existing techniques into three subsets: classification, feature embedding, and hybrid methods. Classification methods focus on learning representations of faces by classifying them into different categories. They use an "N-way classification" objective, where each of the N classes represents an identity. The goal is to train the model to recognize and differentiate between these identities accurately. Feature embedding methods, however, aim to optimize the representation of facial features. They work by minimizing the distance between samples of the same identity (intraperson distance) while maximizing the distance between samples from different identities (interperson distance). Hybrid methods combine both classification and feature embedding approaches. These methods leverage the benefits of both strategies by jointly training the network. By doing so, they seek to improve the accuracy of identity classification while also optimizing the feature distances within and between identities.

The face recognition ability depends not only on the architecture of the network model but also on the dataset used for training. Most of the current datasets [19], [20] used to train state-of-the-art face recognition frameworks are limited to European and American faces or frontal poses [21], such as Annotated Thermal Faces in the Wild Dataset (TFW) [22], but none of these datasets are specifically designed for Asian face design. We address this issue here by designing a classroom dataset of Chinese students in which images of Chinese students' faces with different shadows, lighting, side faces, age and expression variations are explicitly collected in the dataset.

With the development of artificial intelligence, AI is gradually being used in various environments, and schools can also go for intelligent management through AI. Many excellent face recognition models [23], [24] [25], [26] and high-quality datasets have emerged in recent years, and although the recognition of facial images collected in controlled environments is good, the recognition of facial images collected under classrooms is still not high. Due to the irregular camera positions in the classroom and many variables in the camera capture process, including camera capture angle, face size, pose, expression, occlusion, age, and lighting, all of these can make face recognition difficult. To facilitate research on Asian face detection and face recognition, we propose a Chinese face dataset in an uncontrolled classroom environment. Moreover, in addition to the various challenges involved in state-of-the-art datasets, the images of the proposed dataset are captured in uneven and different backgrounds, which is lacking in most datasets. The dataset is available for download at https://github.com/Shensxf/Face-Dataset.

In this paper, our main innovations, compared to other studies, are as follows:

1) In this paper, we propose a facial dataset in an uncontrolled environment. We constructed a dataset of Chinese students' faces in uncontrolled classroom scenarios, which was extracted from classroom surveillance videos from 35 different schools. The extraction was performed by using a Python script to intercept images of the videos frame by frame, followed by manual filtering, resulting in 7395 facial images of 130 targets. The dataset we collected is diverse and contains a number of variations, including angle, lighting, glasses, occlusion, gender, age group, etc.

2) Improving the recognition capability of face recognition models. The results obtained by experimenting with four models, OpenFace, DeepFace, DeepID, and ArcFace, using our dataset show that the current face recognition models are not effective in uncontrolled environments and that the performance of these models can be improved even more by training them with the proposed dataset.

3) Improve the performance of facial feature model recognition. By using the VGG-Face model on our dataset and other datasets for gender, age, and expression recognition, it is demonstrated through experimental results that the VGG-Face model has poor gender, age, and expression recognition ability for Chinese faces, weak model generalization, and room for improvement in recognition ability.

The rest of the article is organized as follows: in Section II, we summarize common public face datasets, face detection algorithms, and related face recognition network models. In Section III, the dataset we propose is described, and the evaluation method is illustrated. In Section IV, we show the experimental results of face recognition on different models for the dataset and the experimental results of age recognition, emotion recognition and age recognition on the VGG-Face model for the dataset. Section V summarizes all the work.

II. RELATED WORK

In this section, we first briefly review common public face datasets and provide a brief summary of these datasets, which are mostly collected in a controlled environment. Second, we outline the face detection methods and face recognition network models used in this case. Next, we summarize the capture methods for the datasets. Finally, we provide a detailed description of the protocols used for the facial datasets.

A. PUBLIC FACE DATASETS

In 2004, one of the few datasets collected under realistic conditions, the BioID face dataset [27], was released. The BioID face dataset contains 1521 grayscale images with a resolution of 384×286 pixels. Each image is derived from 23 different frontal angles of the test subject's face. These images have various illumination levels, backgrounds, and facial dimensions. The dataset maps face images to manually labeled human eye position files. Although the size of the dataset is relatively small and the acquisition sample is unbalanced, the authors placed special emphasis on ''real world'' conditions in the creation of the dataset.

In 2004, the CASIA face dataset [28] was released by the Institute of Automation, Chinese Academy of Sciences. Although most face datasets are predominantly European and American, the CASIA dataset is one of the few Asian face datasets but was created in a specific controlled environment. CASIA-FaceV5 contains 2500 color face images from 500 subjects. CASIA-FaceV5 volunteers included graduate students, workers, waiters, and others. All face images are 16-bit color BMP files with an image resolution of 640*480. Typical intraclass variations include lighting, pose, expression, glasses, imaging distance, etc. The CASIA dataset suffers from the same drawbacks as the BioID dataset in that the CASIA shooting environment is limited to an indoor setting, while the sample is unbalanced and the people photographed have a minimum of one image and a maximum of 804 images.

In 2007, the LFW (LabledFacesintheWild) [29] dataset contained 5749 IDs and 13233 images from natural scenes in life. This face dataset is a common test set for face recognition, which is made more difficult due to factors such as pose, lighting and expression. The image size is 250 Ã— 250 and contains not only black and white images but also color images. In 2018, the CALFW (Cross-AgeLFW) [30] dataset and the CPLFW (Cross-PoseLFW) [31] dataset were released as revolutionary versions of the LFW dataset, where the CALFW data are based on the LFW dataset annotated with the cross-age dataset, containing 3000 pairs of face images with a large age span, and are used to evaluate the performance of face recognition algorithms in cross-age recognition. The CPLFW dataset is a cross-pose dataset annotated based on the LFW dataset. CPLFW is a crowdsourced collection of human images compared to LFW, and the positive pairs in CPLFW contain significant pose differences. The LFW dataset, in addition to sample imbalance, also has the problem of poor quality because the images are from the internet.

In 2009, the Public Face Database (PubFig) [32] was released. The PubFig database is a large, real-world face dataset consisting of 58,797 images of 200 people collected from the internet. Unlike most other existing face datasets, these images were taken from different subjects in a completely uncontrolled manner. As a result, there are significant differences in pose, lighting, expression, scene, camera, imaging conditions, and parameters. Some of the images in the PubFig dataset were captured in public, such as press conferences and TV shows. Such acquisition conditions may lead to a lack of diversity in the dataset to reflect real-world faces under different acquisition conditions.

In 2015, the large-scale CelebFacesAttributes (CelebA) dataset [33] was released. CelebFacesAttributesDataset (CelebA) is a large face attribute dataset with over 200,000 celebrity images, each with 40 attribute annotations. celebA has large diversity and rich annotations, including 10177 identity counts, 202,599 face image counts, and 5 landmark locations.40 The dataset can be used for face attribute recognition, face recognition, face detection, face part localization, and face editing and composition. Since the images in this dataset cover large pose variations and background clutter, there may be differences in image quality and resolution in the CelebA dataset, which may affect the training and performance of the model.

In summary, most face datasets are collected in a controlled environment where the person sits at a predetermined distance from the camera and against a neutral background. The illumination, room temperature, facial expressions, and pose are all predetermined [34]. However, in real-world environments, the number of faces, background environment, occlusion, illumination, scale and pose vary greatly from image to image. Meanwhile, most face datasets are collected in the European region, which is still a neglected part of the Asian face recognition domain in uncontrolled environments. Therefore, to address the above issues, we propose a dataset based on uncontrolled classroom environments, where images are intercepted frame by frame from real classroom surveillance videos and manually filtered to obtain a high-quality and more realistic face dataset.

B. FACE DETECTION ALGORITHMS AND FACE RECOGNITION NETWORK MODEL

We use the proposed dataset to evaluate the performance of different face recognition network models and test the performance of the dataset on the VGG-Face [35] model for predicting expression, age and gender.

Face recognition consists of two steps: face detection and face recognition. Early face detection relied on classifiers to achieve face detection by extracting face features, such as the Haar cascade classifier [36] and HOG (histogram of orientation gradient) [37]. Although these types of classifiers are able to detect faces in simple environments, the performance of the classifier degrades continuously as the environment becomes more complex. Subsequently, better results were obtained for face recognition using convolutional neural networks [38].

Currently, with the rapid development of deep learning, more network models for face detection and recognition are constantly being updated. The proposed dataset in this paper tests the performance of several models and concludes in the test results that the proposed dataset is more difficult in face recognition and determining expression and age. In the experimental section, we tested five face detection algorithms and four commonly used face recognition models, including the state-of-the-art network model.

Face detection algorithms include OpenCV, SSD, Dlib, RitinaFace, and Mediapipe. Face recognition network models include OpenFace, DeepFace, DeepID, and ArcFace. The detection results of the above different face detection algorithms on face images are shown in Fig1. The network structure of the four face recognition models is shown in Fig2.

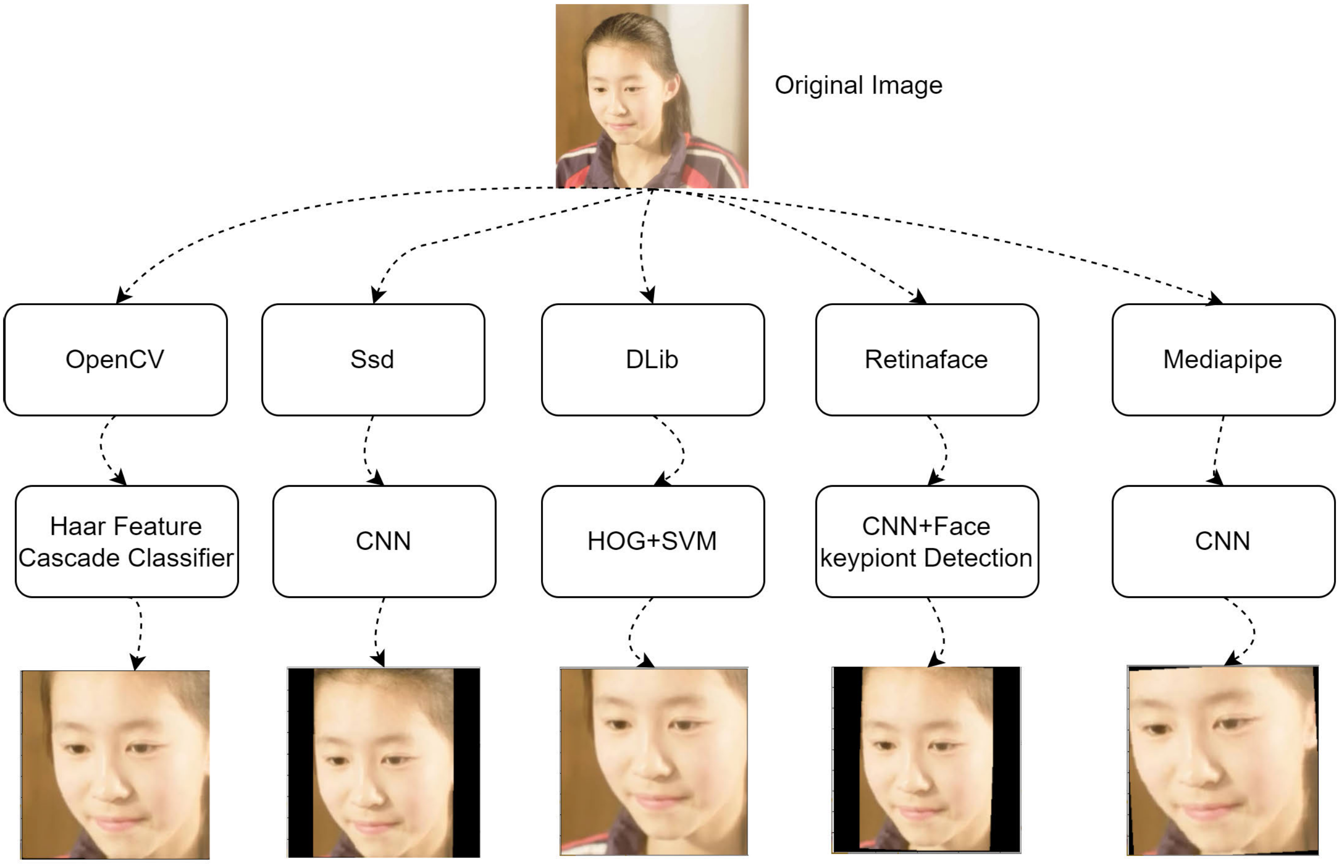

FIGURE 1. The effect of different face detection algorithms for extracting face regions in images. We used five face detection algorithms to extract facial regions in the images: OpenCV, SSD, DLib, Retinaface, and Mediapipe.

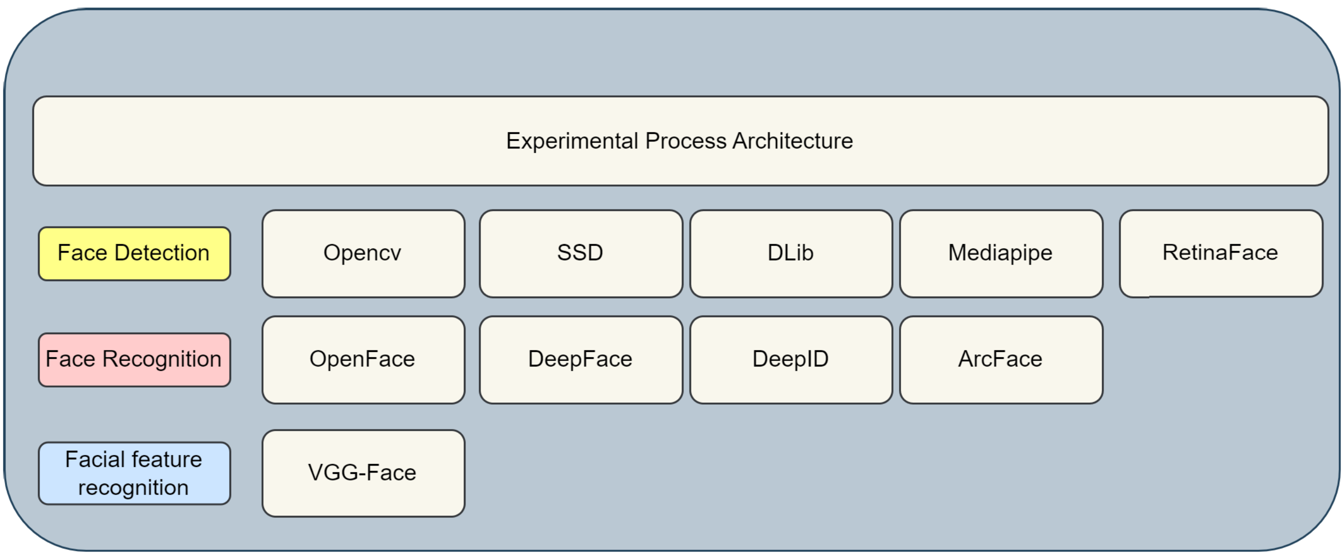

FIGURE 2. The architecture diagram of the experimental process. In the experimental part, we first use a face detection algorithm to extract the facial region in the image and then use face recognition models for face recognition. Face recognition models include OpenFace, DeepFace, DeepID, and ArcFace. At the same time, we also recognize the facial features of the dataset, including gender, age, and expression.

OpenFace [39] is a deep learning-based face recognition model developed by Adam Harvey and Alexander T. Berg at CMU. It uses convolutional neural networks (CNNs) to extract features from face images and map them to a 128-dimensional vector space. OpenFace can be used for face recognition, face tracking, expression recognition and other face analysis tasks. The advantage of OpenFace is that it can process a large number of images in a short time and can be trained in different environments. In addition, OpenFace can be used for video analysis, which can detect faces in videos and recognize people in videos.

DeepFace [40] is a deep learning model for face recognition developed by the Facebook AI Institute. It uses a convolutional neural network (CNN) [41] to extract features from face images and a three-layer deep neural network (DNN) [42] to recognize these features. Nose, mouth, and so on. These features can be used to recognize a person's face even under different lighting conditions. Joint modeling and adaptive modeling techniques are also used, which can combine face features from multiple photos, while the parameters of the model can be automatically adjusted according to the input image, thus improving recognition accuracy.

The DeepID [43] model was developed by Professor Xuantian Lin and his team at the Institute of Computing Technology, Chinese Academy of Sciences. The DeepID model is a multilayered neural network, and the main feature of the model is that it can extract a series of features from face images without manual annotation, and these features can be used to recognize faces. In addition, the DeepID model can learn automatically, thus improving recognition accuracy.

ArcFace [44] was invented by Dr. Hongbin Zhang, a depth scientist. The core idea of ArcFace is to project face features onto the hypersphere so that each face feature has a unique angle and thus it is easier to distinguish different faces. The model structure of ArcFace consists of three parts: a feature extractor, a projector and a classifier. The feature extractor is used to extract face features from the input image, the projector is used to project the face features onto the hypersphere so that each face feature has a unique angle, and finally, the classifier is used to classify the projected features to achieve face recognition.

We will test the proposed dataset on the above model to determine the complexity of the dataset. In addition, five face detection algorithms are tested in this paper, including opencv, ssd, dlib, mtcnn, retinaface, and MediaPipe.

The OpenCV face detection algorithm is a computer vision-based technique for recognizing and detecting faces in images. The detection algorithms work as follows: first, they use feature detection algorithms (e.g., Haar [45] features) to detect faces in images, and then, they use machine learning algorithms (e.g., support vector machines [46]) to recognize faces. It recognizes faces in images and can identify different faces based on their features (e.g., eyes, nose, mouth, etc.

SSD [47] (SingleShot MultiBox Detector) is a deep learning-based target detection algorithm that can detect multiple targets simultaneously in a single inference. Its main idea is to use a network structure to segment the input image into multiple boxes of different proportions and then use multiple convolutional layers to detect the targets in each box.

The Dlib [48] face detection algorithms detect a face from an image and return the position and size of that face. It can also detect multiple faces in an image and can detect faces from different angles, such as front, side and oblique. It uses a technique called a convolutional neural network (CNN) that can detect faces from an image and return the position and size of the face.

RetinaFace [49] was developed by the Tencent Youtu Lab and can achieve high-precision face detection with lower computing resources. Face detection algorithms were developed by the Tencent Youtu Lab and can achieve high-precision face detection with lower computing resources.

The Mediapipe [50] face detection algorithms use convolutional neural networks and sliding windows to detect faces in images and can also detect the expressions of faces, as well as the age and gender of faces. In addition, the Mediapipe face detection algorithms can detect the position of the eyes, nose and mouth in an image.

In the experimental part, we will evaluate the performance on our proposed dataset using five face detection algorithms paired with four network models. Since the face recognition model is already very mature and achieves good results on the prevalent face dataset, to verify that our proposed dataset is more challenging, we also evaluate it on three other datasets, and the results also show that our proposed dataset is more valuable.

C. DATASET CAPTURE METHODS

Common methods of constructing facial datasets include filming in a lab or filming studio, filming in natural scenes, collecting from the internet, etc. For datasets shot in labs or shooting studios such as the Yale Face Database and AR Face Database, the environment of the datasets constructed by this method is controlled and cannot reflect real face recognition scenarios. Datasets taken in natural scenes are labeled Faces in the Wild (LFW), CelebA. This method can collect more realistic facial data, but the control conditions are poor and may be affected by factors such as lighting, background and pose, and the data quality may be degraded. Datasets collected from the internet include the VGG Face Dataset, MS-Celeb-1M, and FaceForensics++ [51]. This method can collect a large amount of facial data, but the data quality is uneven, and there are obvious data biases, while more effort is needed to filter and clean the collected datasets. In this paper, a newer dataset construction method is used to obtain the dataset in the real environment by cropping, filtering and enhancing the video frame by frame to obtain the dataset in the real environment to obtain facial data closer to the real environment. To avoid the close facial features in the dataset images, when selecting the surveillance videos, we chose the classroom surveillance videos of elementary schools, middle schools, high schools, and universities as the resources, thus collecting the facial images of students in different age groups, which ensures the authenticity and

FIGURE 3. Sixteen images were randomly selected from the dataset for presentation. Subjects included elementary, middle school, high school, and college populations, and the images were diverse in nature, including changes in lighting, differences in angle, occlusion, and wearing glasses.

reliability of the image uncontrolled environment and ensures the diversity of the images.

D. PROTOCOLS FOR THE USE OF DATASETS

This UCEC-Face is made available under the Open Database License: http://opendatacommons.org/licenses/odbl/1.0/. Any rights in individual contents of the database are licensed under the Database Contents License: http://opendatacommons.org/licenses/dbcl/1.0/ Attention, images are for research purposes only. Third, not using face images in any such manner leads to embarrassment among the subjects involved in this database. Fourth, it is prohibited to provide the images to a second party. Fifth, the study was not used for any commercial purpose without prior approval from the authors.

III. THE PROPOSED DATASET AND EVALUATION METHOD

A. PROPOSED DATASET

The facial dataset of Chinese students in an uncontrolled classroom environment (UCEC-Face) was obtained by purchasing classroom surveillance videos from different schools, and we filtered the surveillance videos to obtain 35 surveillance videos that met the conditions of an uncontrolled environment. On this basis, manual filtering and data enhancement operations were obtained. The proposed dataset contains variations of different poses, lighting, expressions and scenes to make the dataset conform to the unconstrained environment. The proposed dataset consists of 130 individuals, including 44 males and 86 females. The UCEC-Face dataset is used for research purposes only. The image of the dataset part is shown in Fig3.

The students' classroom videos were obtained through https://so.vjshi.com/ [52] purchase. We took frame-by-frame screenshots of the 35 eligible surveillance videos. These videos contain various activities of the students, such as listening, singing, writing, and speaking. After the frame-by-frame screenshots, 189,369 images were obtained. Some of these images were invalid, and the invalid cases included no facial area in the intercepted images, very serious occlusion, duplication, etc. After manually filtering to remove these invalid images, we performed data enhancement on the filtered images and finally obtained 7,395 images. We then classified these images and finally classified 130 subjects. One subject's image contained up to 85 images and at least 25 images, of which 69.2% of the subjects contained 60 facial images (see Fig4 for statistics). Table 1 shows some of the features contained in the proposed dataset. Next, we illustrate the experiments performed on the proposed face dataset. The collection process of the dataset is shown in Fig5.

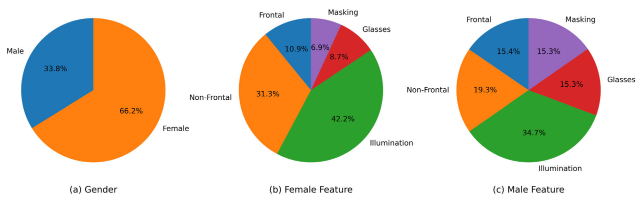

FIGURE 4. **Statistics of our fig feature. (a) shows the statistics of the percentage of males and females in the dataset, (b) shows the statistics of the characteristics of males in the dataset, and (c) shows the statistics of the characteristics of females in the dataset.**

TABLE 1. **A summary of the characteristics of the dataset, in which 48% of male images were frontal, 52% were non-frontal, 100% were light changes, 25% were subjects wearing glasses, and 16% were images of facial occlusions; 30% of female images were frontal, 70% were non-frontal, 100% were light changes, 24% were subjects wearing glasses, and 37% were images of facial occlusions.**

Feature	Frontal	Non-Frontal	Illumination	Glasses	Masking
MALE	48%	52%	100%	25%	16%
FEMALE	30%	70%	100%	24%	37%

B. EVALUATION METHOD

To validate the dataset capability, we analyzed the evaluation accuracy with the recognition rate obtained by each algorithm separately and used three distances, cosine distance [53], Euclidean distance [54] and Euclidean_L2 distance, to find the distance metric for the best performance for the proposed dataset. Additionally, to verify the significant differences between our proposed dataset and other datasets, we used the independent sample t test method [55] in our evaluation method.

1) DISTANCE METRICS

The cosine distance measures the similarity of two vectors by calculating the cosine of the angle between them. The closer the value of cosine distance is to 1, the more similar the two vectors are; the closer it is to 0, the less similar the two vectors are.

$$\text{Cosine distance(a,b)}=\frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\,\left\|\vec{b}\right\|} \tag{1}$$

where a and b are two vectors that represent the dot product of the vectors and denote the modal lengths of the two vectors, respectively. In face recognition, a face image is represented as a vector, where each element represents a feature of the face. If the cosine distance between two face images is less than a threshold, they are considered to be from different people; conversely, they are considered to be from the same person. The Euclidean distance and Euclidean_L2 distance are commonly used similarity measures to calculate the linear distance between two vectors.

$$\text{euclidean distance}(a, b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2}$$

$$\text{L2 distance}(a, b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{3}$$

where a and b are two n-dimensional vectors, and the vector representations of the face images are a and b, respectively. Then, the Euclidean distance between them can be calculated by the above equation. In practical applications, the Euclidean L2 distance is usually used to distinguish the magnitude of the distance between two vectors, while the Euclidean distance is more often used in classification tasks.

2) T TEST FOR INDEPENDENT SAMPLES

The t test for independent samples is a statistical method used to compare whether the means of two independent samples are significantly different from each other. Its basic principle is to compare the mean and variance of two samples and determine whether they are different enough to indicate the significance of the difference. In performing the independent sample t test, we compare the proposed dataset with other datasets separately and first verify the chi-square between the two datasets. Since the sample size of each dataset is 20, the normality needs to be verified, and if normality is satisfied, the formal independent sample t test begins; otherwise, a nonparametric version of the validation method, such as Mann-Whitney, is used. The workflow for conducting independent sample t tests in this paper is shown in Fig6.

IV. EXPERIMENTAL RESULTS

In the experimental results section, the performance metrics obtained by testing the above different detection algorithms with different models on the dataset will be listed, and it will be concluded by comparing other datasets that our proposed dataset is more difficult and challenging in terms of face recognition.

The average recognition rate of the model for face recognition on the proposed UCEC-Face dataset was used as the evaluation score of the model. The average recognition rate is calculated by taking the average of the recognition accuracy of all objects in the UCEC-Face dataset.Due to the data augmentation performed on the UCEC-Face dataset, in order to enhance the reliability of the comparative experiments, we also applied data augmentation to other datasets. To determine which distance is more suitable for the proposed UCEC-Face dataset, experiments were conducted using different distance measures, such as cosine distance, Euclidean distance, and Euclidean_l2 distance.

Table 2, Table 3, Table 4, Table 5 and Table 6 summarize the experimental results, which contain the average recognition rates of the UCEC-Face dataset, AT&T dataset, CASIA dataset, CELAB dataset, and MFace dataset using cosine distance in different models paired with different face detection algorithms, respectively.

From the experimental results, it is clear that ArcFace is the best performer among all hand-made descriptors, with an average recognition rate of 69.7% on the UCEC-Face dataset. Among these face recognition models, ArcFace is the best performing model, and OpenFace is the worst performing model because the training process of the OpenFace model

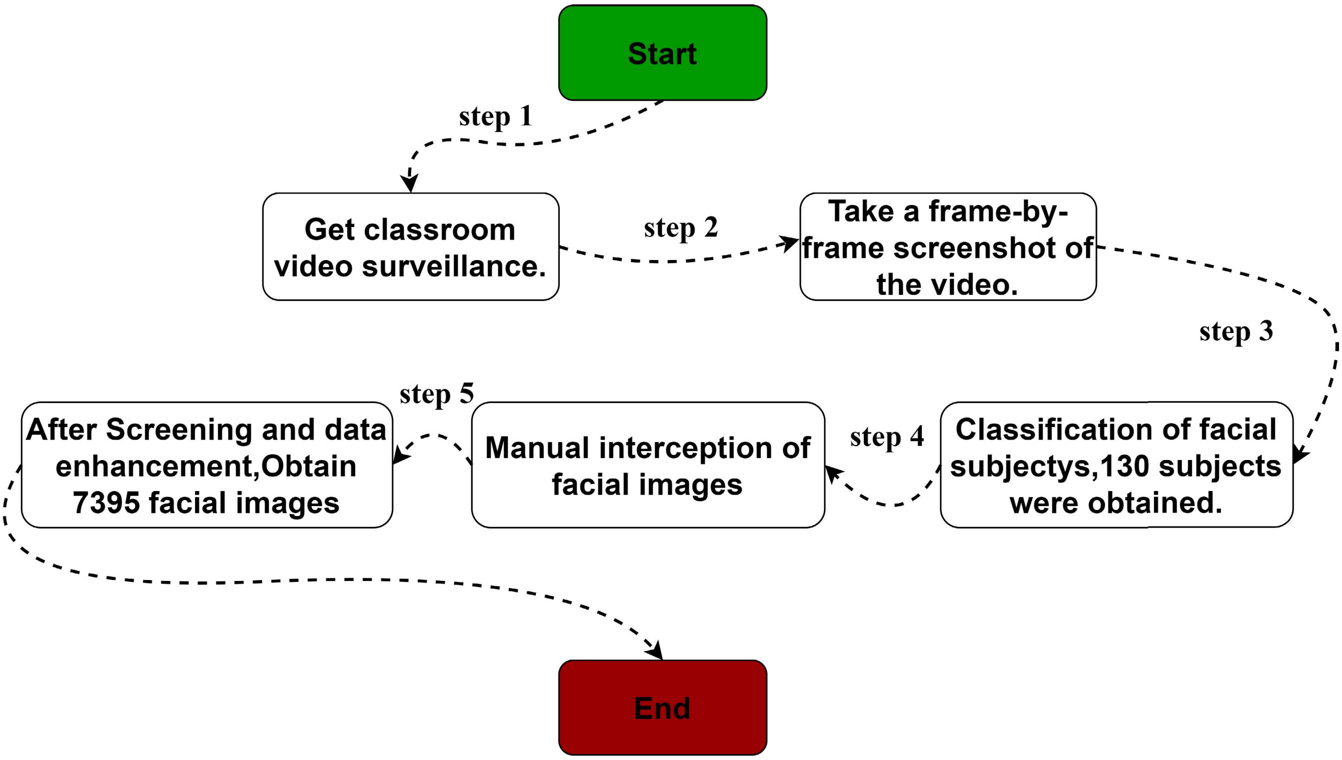

FIGURE 5. The dataset was collected by first cropping images by frame from the classroom surveillance video, manually classifying the images into a total of 130 subjects, followed by repeated filtering and data enhancement operations on the images, and finally obtaining 7,395 facial images.

TABLE 2. The UCEC-Face face recognition results at cosine distance. UCEC-Face dataset uses cosine distance for face recognition under four models after five face detection algorithms detect facial regions, where the ArcFace model has the best recognition effect and the OpenFace model has the worst recognition effect.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	11.9%	15.8%	41.4%	25.5%	15.8%
DeepFace	27.2%	35.3%	54.7%	62%	24.9%
DeepID	23.3%	33.1%	49.1%	59.8%	47.8%
ArcFace	62.4%	67.3%	69.5%	69.7%	67.6%

TABLE 3. The AT&T face recognition results at cosine distance. AT&T dataset uses cosine distance for face recognition under four models after five face detection algorithms detect facial regions, in which the ArcFace model has the best recognition effect and the OpenFace model has the worst recognition effect, and the overall recognition effect is better than the UCEC-Face dataset.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	Mediapipe
OpenFace	57.2%	53.3%	55.8%	56.1%	59.7%
DeepFace	65.3%	61.4%	68.6%	65.6%	59.7%
DeepID	56.7%	63.9%	68.3%	69.6%	58.6%
ArcFace	94.4%	99.2%	98.6%	96.7%	89.2%

does not take into account many complex factors, such as illumination, expression, angle, size, and age.

Additionally, we performed independent sample t test tests on the data separately, and the recognition results of the UCEC-Face dataset and AT&T on the OpenFace model were used as examples, and 10 experimental result data were added to participate in this experiment. UECE-dataset (n=5) and AT&T (n=5) were used by the OpenFace model for face recognition. The data showed that the UCEC-Face dataset recognition rate (M=22.08, SD=10.65) was significantly lower than the AT&T recognition rate (M=56.42, SD=2.08), t=6.1, p(two-tailed)=0.003, d=2.72, 95% CI [18.71:49.96].

TABLE 4. The CASIA face recognition results at cosine distance. CASIA dataset uses cosine distance for face recognition under four models after five face detection algorithms detect facial regions, among which the ArcFace model has the best recognition effect, the DeepFace model has the worst recognition effect, and the overall recognition effect is better than the UCEC-Face dataset.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	Mediapipe
OpenFace	61.5%	61.6%	64.8%	69.0%	60.9%
DeepFace	56.8%	56.4%	59.0%	56.9%	60.5%
DeepID	68.1%	62.3%	64.1%	62.9%	66.0%
ArcFace	92.4%	91.6%	93.7%	100.0%	80.8%

TABLE 5. The CELAB face recognition results at cosine distance. CELAB dataset uses cosine distance for face recognition under four models after five face detection algorithms detect facial regions, in which the ArcFace model has the best recognition effect and the OpenFace model has the worst recognition effect, and the overall recognition effect is better than the UCEC-Face dataset.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	Mediapipe
OpenFace	56.6%	54.6%	57.3%	57.6%	60.5%
DeepFace	60.4%	60.4%	61.6%	60.4%	60.2%
DeepID	67.7%	68.2%	60.8%	68.2%	66.3%
ArcFace	79.3%	73.0%	82.1%	72.3%	76.7%

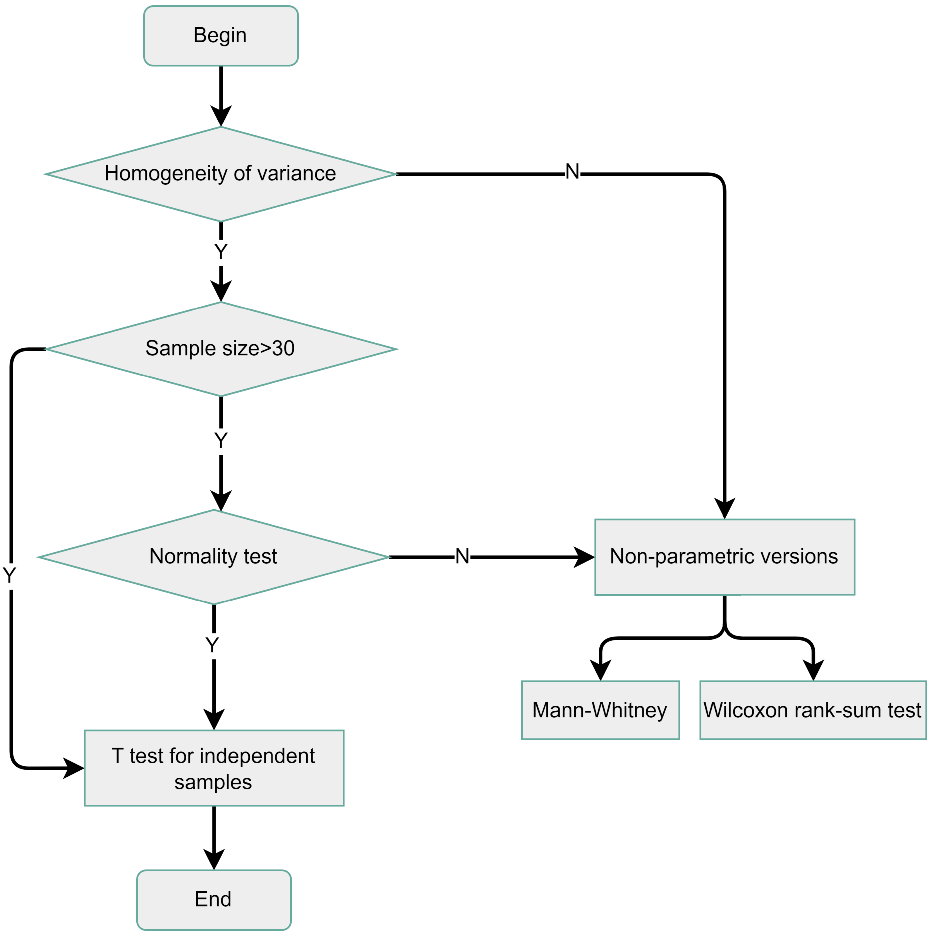

FIGURE 6. In the independent sample t test procedure, we performed an independent sample t test on the comparison results of the experiments. First, the data are verified by homogeneity of variance. If the data satisfy homogeneity of variance, then a normality test is performed; if not homogeneity of variance or normality test, then the nonparametric test method is used. After the data satisfied the normality test, then the t test for independent samples was performed. We do this to make a comparison between the results using other datasets and the proposed dataset under the same model and whether this result is significantly different.

TABLE 6. The MFace face recognition results at cosine distance. MFace dataset uses cosine distance for face recognition under four models after five face detection algorithms detect facial regions, where the ArcFace model has the best recognition effect and the OpenFace model has the worst recognition effect.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	54.6%	55.6%	66.3%	64.8%	71.1%
DeepFace	68.3%	67.9%	74.6%	73.1%	72.2%
DeepID	62.4%	71.9%	72.7%	72.5%	71.4%
ArcFace	59.6%	79.0%	63.9%	88.9%	75.3%

In the above results, we obtained the mean of the difference (M) equal to 34.34, standard deviation of the difference (SD) = 12.58, t-statistic = 6.10, p (one-tailed) value = 0.001, d-value = 2.72, 95% confidence interval (CI) of the difference = (18.719047702203355, 49.96095229779665). Since the p value is less than 0.05, we can reject the null hypothesis and obtain that the difference between the UCEC-Face dataset recognition results and AT&T's recognition results is significant and has a significant difference.

By an independent sample t test, we obtained the result that under cosine distance, the recognition rate of our dataset is significantly lower, closer to the real environment, and more difficult to recognize than the recognition rate of other datasets. More detailed results of the independent sample test can be found in the Appendix.

Table 7 summarizes the average recognition rates of different models with different face detection algorithms on the dataset when using Euclidean distance. We only show the experimental results for the UECE-Face dataset, and the results for other datasets will be shown in the Appendix.

We can observe that the ArcFace model with Retinaface face detection algorithms is still the best performing model with an accuracy of 68.4%, while the Opencv face detection algorithms with Openface model performs the worst with an average recognition rate of 19.4%, because the accuracy of the Opencv face detection algorithms is already low,

TABLE 7. **The UCEC-Face face recognition results at Euclidean distance. UCEC-Face dataset uses Euclidean distance for face recognition under four models after five face detection algorithms detect facial regions, where the DeepFace model has the best recognition and the OpenFace model has the worst recognition.**

Euclidean distance	OpenCV	SSD	Dlib	RetinaFace	Mediapipe
OpenFace	19.4%	26.2%	56.0%	42.1%	28.1%
DeepFace	34.8%	50.5%	65.0%	69.1%	57.3%
DeepID	20.4%	31.7%	40.9%	52.6%	44.1%
ArcFace	59.3%	61.5%	66.1%	68.4%	53.5%

TABLE 8. **The UCEC-Face face recognition results at Euclidean L2 distance. UCEC-Face dataset uses Euclidean L2 paradigm distance for face recognition under four models after five face detection algorithms detect facial regions, where the ArcFace model has the best recognition effect and the OpenFace model has the worst recognition effect.**

EuclideanL2 distance	OpenCV	SSD	Dlib	RetinaFace	Mediapipe
OpenFace	21.4%	26.2%	56.0%	42.1%	28.1%
DeepFace	23.7%	31.0%	50.3%	54.3%	21.5%
DeepID	22.9%	32.4%	48.3%	58.7%	46.7%
ArcFace	61.5%	65.8%	69.6%	68.6%	65.5%

TABLE 9. **We use the VGG-Face model [57] for facial feature recognition, facial features including emotion, age, and gender, and we use the UCEC-Face dataset to compare with a South Asian student facial dataset [58], and the recognition results are significantly lower than MFace, especially in the age recognition gap is larger.**

Dataset	Expression	Age	Gender
UECE-Face	34.7%	8.0%	49.8%
IIITSMFace	57.0%	66.0%	80.0%

and our proposed dataset contains variations in illumination, expression, angle, size, and age, all of which affect the recognition rate of this model [56].

Table 8 summarizes the average recognition rates of different models with different face detection algorithms on the dataset when using the L2 distance. We only show the experimental results for the UECE-Face dataset, and the results for other datasets will be shown in the Appendix.

As seen from Table 7, ArcFace is still the best performing model with a recognition rate of 69.0%, and OpenFace is the worst performing model with a recognition accuracy of only 21.4%.

We note that the cosine distance performs well in the ArcFace model. For the proposed dataset, the use of Euclidean distance is not recommended. Although we used L2 distance in other experiments, the best results (i.e., 69.7% accuracy) were obtained for the ArcFace model using cosine distance.

Table 9 summarizes the average recognition rates of expressions, ages, and genders recognized using VGG-Face on the proposed dataset and compares them with the average recognition rates of expressions, ages, and genders recognized on the IIITSMFace dataset, a dataset of student faces collected in a controlled environment, on this model.

The results show that the model is more likely to recognize the expression, age and gender of the subjects on the IIITSMFace dataset, i.e., the model is more likely

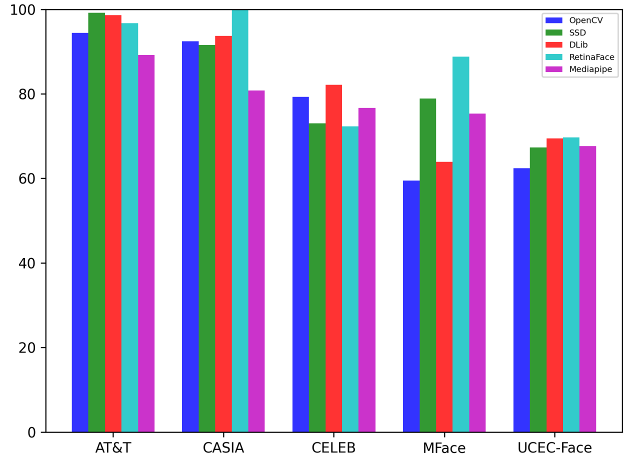

FIGURE 7. **We use the proposed dataset and four additional datasets, AT&T, CASIA, CELEB, and MFace, and use the cosine distance under the ArcFace model to obtain the results of face recognition, as shown in Fig.**

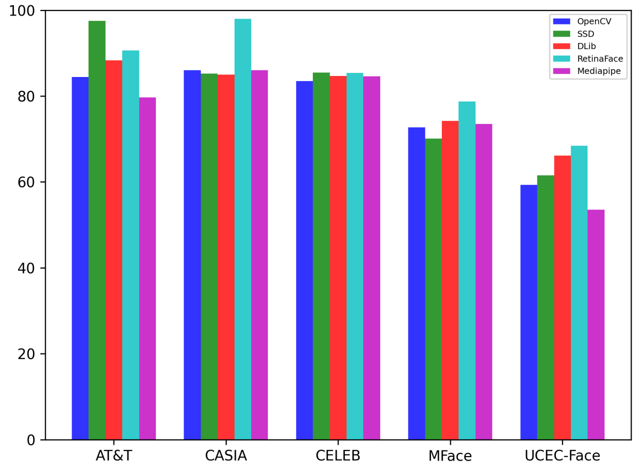

FIGURE 8. **We use the proposed dataset and four additional datasets: AT&T, CASIA, CELEB, and MFace, and perform face recognition under the ArcFace model, using Euclidean distance, to obtain the results of face recognition as shown in Fig.**

to recognize expression, age and gender in a controlled environment, and it can also be seen that the model is less accurate for female gender in our proposed UCEC-Face dataset because there are more female facial images in the dataset.

In summary, it can be seen that the ArcFace model has the best performance among the four models, and we will show the experimental results of our proposed dataset with other datasets on the Arcface model through pictures. 7, 8, and 9 show the experimental results of our proposed dataset using cosine distance, Euclidean distance, and Euclidean l2 distance under the ArcFace model, respectively, and the experimental results on the AT&T dataset, CASIA dataset, and CELEB dataset are also shown together for easy observation and comparison.

There are some challenging datasets in the literature, such as different side poses, obscured faces, and different light

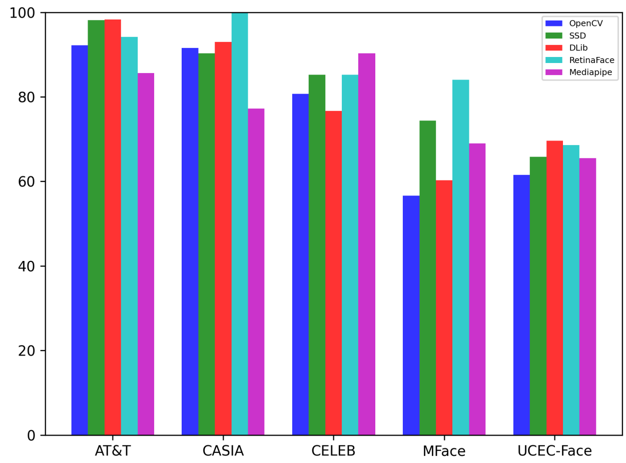

FIGURE 9. We use the proposed dataset and four additional datasets: AT&T, CASIA, CELEB, and MFace, and perform face recognition under the ArcFace model, extracting using the Euclidean L2 paradigm distance and obtaining the results of face recognition as shown in Fig.

TABLE 10. Comparison of our proposed dataset with the AT&T [16], PubFig and CASIA-FaceV5 datasets. Here, "Y", "N", and "P" represent the presence, absence, and partial presence of effects such as non-frontal (Non-Fro), occlusion (OCCL), illumination change (IllVar), and background change (BackVar). The last row shows the accuracy (%) measured using the ArcFace face recognition model in each database using cosine distance.

Features	AT&T	PUBFUG	CASIA	UCEC-Face
Non-Fro	N	Y	P	Y
OCCL	N	N	N	Y
IllVar	N	Y	N	Y
BackVar	N	Y	N	Y
ARCFACE Result(%)	96.7	95.3	100	69.7

intensities. These datasets are discussed in the introduction section. However, the proposed UCEC-Face dataset is more challenging than other existing face datasets, such as Yale, LFW, and PubFig, as shown in Table 10.

V. CONCLUSION & FUTURE DIRECTION

To address the lack of Chinese facial datasets in the field of face recognition, as well as the lack of real bad context factors in facial datasets, this paper constructs a facial dataset of Chinese students in an uncontrolled classroom environment. To verify the impact of facial images on the performance of face recognition models in an uncontrolled environment, we use four face recognition models, OpenFace, DeepFace, DeepID, and ArcFace, on the UCEC-Face dataset and experiment with four datasets, AT&T, CASIA, CELEB, and MFace, under the same conditions. The results are compared. Compared with the above four datasets, there are more variables in our dataset, such as angle of the captured face, image size, subject's pose, subject's facial expression, subject's facial occlusion, subject's age, lighting variation, etc. On the face recognition task, the highest accuracies of the AT&T, CASIA, CELAB, MFace, and UECE-Face datasets for face recognition in the OpenFace model were 74.2%, 71.1%, 75.9%, 76.3%, and 56.0%, respectively; the

TABLE 11. Comparison of UCEC-Face dataset test Face Recognition Algorithms with Different detection algorithms at cosine distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	11.9%	15.8%	41.4%	25.5%	15.8%
DeepFace	27.2%	35.3%	54.7%	62%	24.9%
DeepID	23.3%	33.1%	49.1%	59.8%	47.8%
ArcFace	62.4%	67.3%	69.5%	69.7%	67.6%

TABLE 12. Comparison of AT&T Dataset Test Face Recognition Algorithms with Different detection algorithms at cosine distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	57.2%	53.3%	55.8%	56.1%	59.7%
DeepFace	65.3%	61.4%	68.6%	65.6%	59.7%
DeepID	56.7%	63.9%	68.3%	69.6%	58.6%
ArcFace	94.4%	99.2%	98.6%	96.7%	89.2%

TABLE 13. Comparison of CASIA Dataset Test Face Recognition Algorithms with Different detection algorithms at cosine distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	61.5%	61.6%	64.8%	69.0%	60.9%
DeepFace	56.8%	56.4%	59.0%	56.9%	60.5%
DeepID	68.1%	62.3%	64.1%	62.9%	66.0%
ArcFace	92.4%	91.6%	93.7%	100.0%	80.8%

highest accuracies of face recognition in the DeepFace model were 69.4%, 99.7%, 76.6%, 74.6%, 69.1%, respectively; the highest accuracy rates of face recognition in the DeepID model were 69.6%, 78.8%, 77.7%, 72.7%, 68.1%; and the highest accuracy rates of face recognition in the ArcFace model were 99.2%, 100.0%, 90.3%, 88.9%, and 69.6%, respectively. Overall, although these models obtain better results in face recognition tasks in controlled environments, the accuracy of face recognition in our proposed uncontrolled environment dataset has not yet exceeded 70%. Meanwhile, facial feature recognition using UCEC-Face and MFace on the VGG-Face model, where the accuracy of expression recognition is reduced by 22.3%, the accuracy of gender recognition is reduced by 58%, and the gender recognition rate is reduced by 30.2%. The experimental comparison shows that the current model with excellent performance for face recognition tasks and facial feature recognition tasks does not achieve good results in the recognition of Chinese faces. I believe that in the future, I will further develop models with better performance for face recognition and facial feature recognition based on the Chinese face dataset in uncontrolled environments. At the same time, since this dataset provides face images in real scenarios, it can help the algorithm to better handle various situations, thus improving the recognition accuracy and facilitating the optimization and improvement of the algorithm, enabling researchers to train more effective face recognition algorithms. It is also hoped that the release of this dataset will promote the development of face recognition technology, contribute to the stability of society, and play an important role in more areas, such as finding missing persons and fighting crime.

TABLE 14. Comparison of CELAB Dataset Test Face Recognition Algorithms with Different detection algorithms at cosine distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	56.6%	54.6%	57.3%	57.6%	60.5%
DeepFace	60.4%	60.4%	61.6%	60.4%	60.2%
DeepID	67.7%	68.2%	60.8%	68.2%	66.3%
ArcFace	79.3%	73.0%	82.1%	72.3%	76.7%

TABLE 15. Comparison of MFace Dataset Test Face Recognition Algorithms with Different detection algorithms at cosine distance.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	54.6%	55.6%	66.3%	64.8%	71.1%
DeepFace	68.3%	67.9%	74.6%	73.1%	72.2%
DeepID	62.4%	71.9%	72.7%	72.5%	71.4%
ArcFace	59.6%	79.0%	63.9%	88.9%	75.3%

TABLE 16. Comparison of UCEC-Face Dataset Test Face Recognition Algorithms with Different detection algorithms at Euclidean distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	19.4%	26.2%	56.0%	42.1%	28.1%
DeepFace	34.8%	50.5%	65.0%	69.1%	57.3%
DeepID	20.4%	31.7%	40.9%	52.6%	44.1%
ArcFace	59.3%	61.5%	66.1%	68.4%	53.5%

TABLE 17. Comparison of AT&T Dataset Test Face Recognition Algorithms with Different detection algorithms at Euclidean distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	34.2%	30.0%	44.2%	33.9%	22.8%
DeepFace	58.6%	64.7%	73.9%	70.8%	71.7%
DeepID	63.1%	61.9%	62.2%	53.6%	53.9%
ArcFace	84.4%	97.5%	88.3%	90.6%	79.7%

TABLE 18. Comparison of CASIA Dataset Test Face Recognition Algorithms with Different detection algorithms at Euclidean distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	63.6%	67.6%	70.2%	69.5%	71.1%
DeepFace	81.3%	75.6%	79.0%	99.7%	68.8%
DeepID	74.0%	77.1%	70.1%	78.8%	77.5%
ArcFace	86.0%	85.2%	85.0%	98.0%	86.0%

TABLE 19. Comparison of CELAB Dataset Test Face Recognition Algorithms with Different detection algorithms at Euclidean distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	73.5%	73.6%	75.9%	73.6%	72.4%
DeepFace	76.0%	76.5%	74.2%	76.5%	73.9%
DeepID	75.4%	65.9%	61.0%	65.9%	64.7%
ArcFace	83.5%	85.5%	84.7%	85.4%	84.6%

TABLE 20. Comparison of MFace Dataset Test Face Recognition Algorithms with Different detection algorithms at Euclidean distance.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	69.8%	73.3%	71.7%	74.4%	72.8%
DeepFace	72.4%	71.9%	69.8%	76.4%	76.9%
DeepID	74.1%	72.0%	73.1%	74.3%	72.1%
ArcFace	72.7%	70.1%	74.2%	78.7%	73.5%

APPENDIX

In the appendix section, we present the detailed results of the experimental part through tables.

TABLE 21. Comparison of UCEC-Face Dataset Test Face Recognition Algorithms with Different detection algorithms at L2 distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	21.4%	26.2%	56.0%	42.1%	28.1%
DeepFace	23.7%	31.0%	50.3%	54.3%	21.5%
DeepID	22.9%	32.4%	48.3%	58.7%	46.7%
ArcFace	61.5%	65.8%	69.6%	68.6%	65.5%

TABLE 22. Comparison of AT&T Dataset Test Face Recognition Algorithms with Different detection algorithms at L2 distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	64.2%	70.0%	74.2%	73.9%	72.8%
DeepFace	69.4%	60.6%	60.8%	67.2%	66.4%
DeepID	66.4%	62.2%	67.8%	67.8%	65.6%
ArcFace	92.2%	98.1%	98.3%	94.2%	85.6%

TABLE 23. Comparison of CASIA Dataset Test Face Recognition Algorithms with Different detection algorithms at L2 distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	63.6%	67.6%	70.2%	69.5%	71.1%
DeepFace	79.8%	79.8%	74.6%	93.1%	79.0%
DeepID	67.2%	71.2%	72.9%	71.6%	65.0%
ArcFace	91.6%	90.3%	93.0%	100.0%	77.2%

TABLE 24. Comparison of CELAB Dataset Test Face Recognition Algorithms with Different detection algorithms at L2 distance.

	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	63.5%	63.6%	65.9%	63.6%	62.4%
DeepFace	70.2%	70.2%	73.9%	70.2%	70.2%
DeepID	77.2%	77.7%	70.7%	77.7%	75.9%
ArcFace	80.7%	85.2%	76.7%	85.2%	90.3%

TABLE 25. Comparison of MFace Dataset Test Face Recognition Algorithms with Different detection algorithms at L2 distance.

cosine distance	OpenCV	SSD	Dlib	RetinaFace	MediaPipe
OpenFace	69.8%	73.3%	70.7%	64.4%	72.8%
DeepFace	72.8%	77.0%	72.3%	65.9%	71.5%
DeepID	72.4%	71.9%	72.6%	72.3%	71.2%
ArcFace	66.6%	74.5%	70.3%	84.0%	69.3%

ACKNOWLEDGMENT

(Xiangfeng Shen is co-first author.)

REFERENCES

[1] H. Du, H. Shi, D. Zeng, X.-P. Zhang, and T. Mei, "The elements of end-to-end deep face recognition: A survey of recent advances," *ACM Comput. Surv.*, vol. 54, pp. 1–42, Jan. 2022, doi: 10.1145/3507902.

[2] Y. Liu, X. Tang, J. Han, J. Liu, D. Rui, and X. Wu, "Hambox: Delving into mining high-quality anchors on face detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13043–13051, doi: 10.1109/CVPR42600.2020.01306.

[3] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[4] Y. Wang, X. Ji, Z. Zhou, H. Wang, and Z. Li, "Detecting faces using region-based fully convolutional networks," 2017, *arXiv:1709.05256*.

[5] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5325–5334, doi: 10.1109/CVPR.2015.7299170.

[6] D. Zeng, H. Liu, F. Zhao, S. Ge, W. Shen, and Z. Zhang, *Proposal Pyramid Networks for Fast Face Detection*, vol. 495. Amsterdam, The Netherlands: Elsevier, 2019, pp. 136–149.

[7] Y. Xu, W. Yan, G. Yang, J. Luo, T. Li, and J. He, "Centerface: Joint face detection and alignment using face as point," *Sci. Program.*, vol. 2020, pp. 1–8, Jul. 2020, doi: 10.1155/2020/7845384.

[8] X. Huang, W. Deng, H. Shen, X. Zhang, and J. Ye, "PropagationNet: Propagate points to curve to learn structure information," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7263–7272, doi: 10.1109/cvpr42600.2020.00729.

[9] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021, doi: 10.1109/TPAMI.2020.2983686.

[10] Z. Liu, X. Zhu, G. Hu, H. Guo, M. Tang, Z. Lei, N. M. Robertson, and J. Wang, "Semantic alignment: Finding semantically consistent ground-truth for facial landmark detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3462–3471, doi: 10.1109/cvpr.2019.00358.

[11] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, 2015, pp. 2017–2025. [Online]. Available: http://papers.nips.cc/paper/5854-spatial-transformer-networks.pdf and https://www.bibsonomy.org/bibtex/214d7850ca8e1d4823e7f44c8b3beeac3/flodal

[12] Y. Zhong, J. Chen, and B. Huang, "Toward end-to-end face recognition through alignment learning," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1213–1217, Aug. 2017, doi: 10.1109/LSP.2017.2715076.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf and https://www.bibsonomy.org/bibtex/2886c491fe45049fee3c9660df30bb5c4/albinzehe

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds. San Diego, CA, USA, Jul. 2019. [Online]. Available: https://dblp.org/rec/journals/corr/SimonyanZ14a.bi and https://dblp.org

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[17] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[19] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2020, pp. 1–5, doi: 10.1109/ASYU50717.2020.9259802.

[20] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 67–74.

[21] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. s Alamitos, CA, USA: IEEE Computer Society, Jun. 2015, pp. 2892–2900. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298907, doi: 10.1109/CVPR.2015.7298907.

[22] A. Kuzdeuov, D. Aubakirova, D. Koishigarina, and H. A. Varol, "TFW: Annotated thermal faces in the wild dataset," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2084–2094, 2022.

[23] *Keras Insightface*, GitHub, Leondgarse, San Francisco, CA, USA, 2022.

[24] T. Mare, G. Duta, M.-I. Georgescu, A. Sandru, B. Alexe, M. Popescu, and R. T. Ionescu, "A realistic approach to generate masked faces applied on two novel masked face recognition data sets," 2021, *arXiv:2109.01745*.

[25] F. Liu, M. Kim, A. Jain, and X. Liu, "Controllable and guided face synthesis for unconstrained face recognition," in *Proc. ECCV*, 2022, pp. 701–719.

[26] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 18729–18738.

[27] B. GmbH. (2008). *The Bioid Face Database*. [Online]. Available: https://https://www.bioid.com/facedb/

[28] J. Dong, W. Wang, and T. Tan, "CASIA image tampering detection evaluation database," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Jul. 2013, pp. 422–426, doi: 10.1109/chinasip.2013.6625374.

[29] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognit.* Marseille, France: Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Oct. 2008, pp. 1–15. [Online]. Available: https://inria.hal.science/inria-00321923

[30] T. Zheng, W. Deng, and J. Hu, "Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments," 2017, *arXiv:1708.08197*.

[31] T. Zheng and W. Deng, "Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments," Beijing Univ. Posts and Telecommun., Beijing, China, Tech. Rep. 18–01, Feb. 2018.

[32] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 365–372.

[33] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," Tech. Rep., 2015.

[34] P. Panchal, P. Patel, V. Thakkar, and R. Gupta, "Pose, illumination and expression invariant face recognition using Laplacian of Gaussian and local binary pattern," in *Proc. 5th Nirma Univ. Int. Conf. Eng. (NUiCONE)*, Nov. 2015, pp. 1–6, doi: 10.1109/NUICONE.2015.7449622.

[35] A. L. Cambridge. (2002). *The AT&T Face Database*. [Online]. Available: https://www.kaggle.com/datasets/kasikrit/att-database-of-faces

[36] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2001, p. 1.

[37] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[39] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–10.

[40] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[41] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[42] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[43] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," 2014, *arXiv:1406.4773*.

[44] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.

[45] P. Viola and M. Jones, "Robust real-time face detection," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2001, p. 747.

[46] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[47] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 21–37.

[48] M. A. N. Reza, E. A. Z. Hamidi, N. Ismail, M. R. Effendi, E. Mulyana, and W. Shalannanda, "Design a landmark facial-based drowsiness detection using Dlib and openCV for four-wheeled vehicle drivers," in *Proc. 15th Int. Conf. Telecommun. Syst., Services, Appl. (TSSA)*, Nov. 2021, pp. 1–5.

[49] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot multi-level face localisation in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5202–5211.
[50] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A framework for building perception pipelines," 2019, *arXiv:1906.08172*.
[51] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, "FaceForensics++: Learning to detect manipulated facial images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1–11.
[52] (2021). *VJSHi—A Professional Video Material Sharing Platform for Creators*. Accessed: Feb. 18, 2023. [Online]. Available: https://so.vjshi.com/
[53] A. Abbad, K. Abbad, and H. Tairi, "Face recognition based on city-block and Mahalanobis cosine distance," in *Proc. 13th Int. Conf. Comput. Graph., Imag. Visualizat. (CGiV)*, Mar. 2016, pp. 112–114.
[54] M. D. Malkauthekar, "Analysis of Euclidean distance and Manhattan distance measure in face recognition," in *Proc. 3rd Int. Conf. Comput. Intell. Inf. Technol. (CIIT)*, Oct. 2013, pp. 503–507.
[55] Y. Chen and L. Li, "Research on the differential changes in the physical health level of general higher education institutions students in the epidemic era by data analysis model based on independent sample T-test," in *Proc. Int. Conf. Inf. Technol. Contemp. Sports (TCS)*, Jan. 2021, pp. 427–434.
[56] H. Sharma, S. Saurav, S. Singh, A. K. Saini, and R. Saini, "Analyzing impact of image scaling algorithms on viola-jones face detection framework," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2015, pp. 1715–1718.
[57] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–12.
[58] S. R. Dubey and S. Mukherjee, "A multi-face challenging dataset for robust face recognition," in *Proc. 15th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Nov. 2018, pp. 168–173.

NIANFENG LI is currently the Dean of the Computer Institute, Changchun University; the Secretary General of the Jilin Provincial Computer Professional Teaching Steering Committee; the Director of the Jilin Provincial Computer Society; a member of the China Computer Society; and the Head of the Provincial Experimental Teaching Demonstration. He has hosted or participated in more than 20 provincial and ministerial scientific research projects, published more than 20 articles on research results, holds 14 licensed patents and six software copyrights, and received one Second Prize of the Jilin Provincial Science and Technology Award and three academic achievement awards for natural sciences in Jilin Province. His main research interests include machine learning, image processing, and campus security technology.

XIANGFENG SHEN was born in Baicheng, Jilin, China, in 1997. He is currently pursuing the master's degree in cyberspace security with Changchun University. His main research interest includes computer vision.

LIYAN SUN was born in Changchun, China, in 1987. He received the B.S. degree in computer science from Jilin University, China, in 2010. His research interests include machine learning, artificial intelligence, big data, evolutionary algorithms, and bioinformatics.

ZHIGUO XIAO is currently a master student supervisor. In the past five years, he has presided more than three key projects and participated in six projects above the provincial level, published 11 SCI and EI retrieved papers, and obtained three authorized patents. His main research interest includes machine learning. He is a member of the Industrial Control Committee, Chinese Computer Society, a member of the Robotics Committee, Chinese Simulation Society, and a Certified Artificial Intelligence Engineer of Huawei.

TIANJIAO DING is currently pursuing the master's degree. Her research interests include machine learning and graph neural networks.

TIANSHENG LI was born in Changchun, Jilin, China. He is currently pursuing the master's degree. His research interest includes natural language processing.

XINHANG LI was born in Jinan, Shandong, China, in 2003. He is currently pursuing the degree in communication engineering with the School of Physical Science and Information Engineering, Liaocheng University. His research interests include communication technology and communication system network design. He has won the First Prize of the 8th Shandong University Students Microcontroller Application Innovation Design Competition, the First Prize of the 14th Shandong University Students Science and Technology Festival Physical Science and Technology Innovation Competition, and the Provincial Project Approval of Shandong University Students Innovation and Entrepreneurship Competition.