Abnormal video homework automatic detection system

Jinjiao Lin¹ · Yanze Zhao¹ · Chunfang Liu¹ · Haitao Pu^{2,3}

Received: 11 August 2020 / Accepted: 16 December 2020 / Published online: 6 January 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Automatic abnormal detection of video homework is an effective method to improve the efficiency of homework marking. Based on the video homework review of “big data acquisition and processing project of actual combat” and other courses, this paper found some student upload their videos with poor images, face loss or abnormal video direction. However, it is time-consuming for teachers to pick out the abnormal video homework manually, which results in prompt feedback to students. This paper puts forward the AVHADS (Abnormal Video Homework Automatic Detection System). The system uses suffix and parameter identification, Open CV, and the audio classification model based on MFCC feature to realize the automatic detection and feedback of abnormal video homework. Experimental results show the AVHADS is feasible and effective.

Keywords Video homework · Open CV · Abnormal detection · Audio classification

1 Introduction

Today’s rapid updating of knowledge expects a new learning mode that students shift from focusing on the acquisition of knowledge to skills of social adaptability (Jiang et al. 2016). Ways of evaluating students also shifted from the single knowledge to students’ ability and comprehensive quality. Homework is an important way to evaluate students, and traditional homework are generally in the form of text, sound, pictures, which cannot convey student’s status such as movement and expression etc. in a comprehensive way to teacher. Thus, video homework that combine text, pictures and audio to present more complete information has

become the choice of more teachers. During the COVID-19 epidemic, there was a spatial distance between students and teachers, and some teachers chose to learn about students’ learning status through video homework.

Video homework is a good way which integrated big data information acquisition and student autonomy project output (Wang 2019). Here, video homework refers to student’s record videos of related experiments, operations, presentations, or performances according to homework requirements. In the form of video homework, students are expected to explain how the experiment or operation process, or express the assignment theme through speech or performance after certain organization and design. Video homework is one of the operational forms that can promote personalized and proactive learning (Zhu 2019). Here is the advantages that other form of assessment do not have:

1. Video homework can help teachers to evaluate students better. An experiment on the sources of human information by the experimental psychologist Treicher showed that 83% of human information comes from sight, 11% from hearing, 3.5% from smell, 1.5% from touch, and 1% from taste (Xu 2003). Video homework contains text, pictures, sound, video and other rich information and content, which is helpful for students to express their learning content and thinking results in a more comprehensive way. Teachers can accurately evaluate students’ familiarity with knowledge through their tone, expres-

✉ Haitao Pu
pht@sdust.edu.cn

Jinjiao Lin
ljj@sdufe.edu.cn

Yanze Zhao
192106007@sdufe.edu.cn

Chunfang Liu
182106005@sdufe.edu.cn

¹ School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, China

² School of Information and Control Engineering, Qingdao University of Technology, Qingdao, China

³ College of Electrical Engineering and Automation, Shandong University of Science and Technology, Jinan, China

sion, sight, action in the video. In addition, Professor Jiang Dayuan pointed out that the purpose of learning is not to memorize knowledge, but to apply it (Jiang 2020). Video homework is a good way for students to apply their knowledge when solving problems and teachers can quickly evaluate how students can apply their knowledge by video homework and detect the problems of each student. Therefore, video homework is a conducive way to improving teaching.

2. Video homework can help students to master knowledge at a higher level. A famous experiment on memory persistence by the experimental psychologist Treicher showed that people tend to remember 10% of what they read, 20% of what they hear, 30% of what they see, 50% of what they hear and see, 70% of what they say in the communication (Wang 2011). Students present knowledge in the form of communication through video homework. It is a process of knowledge output through which students can achieve the lasting memory in the process of application of knowledge. In terms of content, video homework usually focus on complex problem solving, solution design, debate and discrimination, etc. It also contains the concept introduction, case evidence, hierarchical analysis, summarizes conclusions, etc. To ensure the quality of video homework, students will organize their presentations logically and ponder on the theme of homework carefully. Students might also evaluate themselves in the course of a lecture or performance. While recording of video assignments, students learn related technologies of video production and cultivate their ability to use information technology. In a word, video homework might improve students' cognition on memorization, understanding and application, and might also promote students' higher-level abilities such as synthesis, evaluation and innovation, which will realize Bloom's educational goal system (Xiang 2009) in the field of cognition.

2 Practical application

Video homework can achieve better learning results than what traditional homework can (Tu et al. 2017). This paper adopts the form of video homework in courses of "Actual practice of Big Data Collection and Processing Project" in Shandong University of Finance and Economics. The homework requires students to record the entire operation process of python experiment and explain the relevant knowledge with a purpose of getting to know the level of the students' knowledge familiarity in the six-level comprehensive education objective of Bloom, evaluating the students' learning situation and adjusting the later teaching. The homework ask the students to upload videos need which contain both input

and output process, successful run results and clear explanation. Students are supposed to show their faces in the videos instead of only recording audios and PPT.

There are 443 video homework in total, and 89 (nearly 21%) cannot satisfied the homework requirement. The unqualified videos are compressed files, no faces or unclear voices. Some cannot be reviewed online due to the direction of images. Teachers can only manually identify those problems while marking all the uploaded files, which is time consuming and hold back the homework correction cycle. As a result, students cannot receive feedback timely. When the unqualified homework are return to students, they are less motivated to resubmit video homework. Therefore, in order to improve the efficiency of review of video homework and save the labor of identifying the unqualified ones, Abnormal Video Homework Automatic Detection System (AVHADS) is proposed in this paper, which realizes the preliminary detection of uploaded video homework. Unqualified video homework is automatically sent back to students with sensible explanations of which students might see it as a prompt feedback. It might prevent students from academic pressure caused by long homework correction time.

3 Related work

Many scholars have been contributing research efforts on automatic homework detection and review. Some scholars focused on the automatic review of program homework. Martin et al. (2018) used the argument-based machine learning (ABML) to finish semiautomatic identification of typical approaches and the errors in student solutions. They believe that timely feedback can improve students' programming learning efficiency. Zhao et al. (2010) designed the program for the automatic detection and correction of student program work. It can screen the similar procedures and supervise and encourage students to finish the homework independently. On the review of the graphics homework, Yang et al. (2014) designed and implemented an automatic grading system for civil engineering drawing based on vector graphics platform ATVGP. Peiying (2001) used VC programming to realize the automatic correction of engineering drawing homework based on Web. Li et al. (2019) put forward a program for automatic recognition and rating of homework pictures taken by mobile phones, which has achieved good results. The above research on automatic review of homework focuses on images and texts, which cannot solve the problems of detection and correction of video homework. Here, we attempt to study and realize the abnormal detection of video homework.

At present, the research on abnormal detection of video mainly focus on the video content, such as traffic violations (Ye et al. 2012), people's behaviors testing (Lian

et al. 2002). These researches are quite different from the abnormal detection of video homework, so they cannot solve the problem in this paper. The problem of anomaly detection in this paper mainly includes four aspects: file type identification, face detection, video direction recognition and audio detection.

File type identification is relatively simple, it is generally implemented through methods based on statistical characteristic (Zheng et al. 2007) and content (McDaniel et al. 2020). The file types involved in this paper are relatively fixed, so we choose simple suffix name matching to realize file type identification. There are many studies on video face detection. Keke et al. (2008) conduct face detection by using the face detection function in Open CV. Goyal et al. (2017) conduct an in-depth study of face detection using open CV, Ma et al. (2018) use the multi-task cascaded convolutional neural networks to realize the frontal and the non-frontal face detection, this paper does not cover the non-frontal face detection, so we chose Open CV (2016) which is easy to implement face detection. The detection of video direction is rarely involved in other problems, and relevant researches are scarce. According to the specific problems in this paper, we choose to use the comparison of specific parameters to realize the detection of video direction. In this paper, it is difficult to solve the audio detection problem, and many researchers have done related researches. Muda et al. (2010) used Mel-Frequency Cepstral Coefficients (MFCC) to extract sound feature, and used the Dynamic Time Warping (DTW) to realize sound recognition. Yu et al. (2006) added Linear Prediction Cepstrum Coefficient (LPCC) on the basis of MFCC to describe sound feature, and used the methods of vectorization and DTW to realize speaker detection. Ali Technology (2018) makes use of the MFCC characteristics of human voice samples and non-human voices samples and the Inception-V3 model of CNN to realize the prediction of voice audio files. Based on MFCC + CNN, Wei et al. (2018) used random forest to classify audio, this method improved the accuracy of audio classification. Zhang (2019) replace CNN with ResNet to promote the accuracy of ESC recognition task. And Huang et al. (2020) used the multi-mode neural network to cluster the voices of different speakers (teachers and students) in the course audio, then realized the differentiation of multiple speakers by text matching. Most of these researches are aimed at specific kinds of problems; they are temporarily unable to solve the audio detection problem. The audio of video homework is different, as it mainly consists of the voices of other students in the student's living environment. Due to the big student's population and overlapping, this problem cannot be solved by matching voice print features. Therefore, we choose to use MFCC to describe

audio features of video homework and use the trained classifier to detect the sound clarity of video homework.

4 Abnormal Video Homework Automatic Detection System (AVHADS)

Based on the above questions and related researches, this paper puts forward the AVHADS (Abnormal Video Homework Automatic Detection System) which can realize the automatic detection and feedback of abnormal video homework.

4.1 System framework

The whole system is divided into four modules. The first module is file type Identification which adopts the file suffix (zip, rar etc.) matching to realize file type identification. In the module, after uploading the video file, the system will detect the direction of the video. The system adopts the comparison of video length and width to determine whether the video direction is in the normal landscape state. If the direction of the video is correct, it invokes Open CV face detection classifier to identify whether there are students in the video homework. If there is a face in video, the system will detect whether the audio of the video homework is clear voice or not. If the homework file passes the detection, it will be uploaded successfully. When a certain module is not satisfied, the system outputs a reason of failure to upload homework. The whole system framework is shown in Fig. 1.

As file type recognition and image direction detection can be realized by simple parameter comparison, the paper will not be further discussed in detail.

4.2 Face detection

There have been many studies about face detection. The Open CV (Open Source Computer Vision Library) (Bradski 2008), which is developed by Intel, is open source library of visual algorithm and image processing. It is quite mature and widely used in face detection and recognition. This paper chooses alt2 classifier (haarcascade_frontalface_alt2.XML) to realize video work face detection. And it proves that this classifier performs better in Open CV (Lian 2016). The file of the classifier contains Haar-like features describing various parts of the human body. It realizes the classification of faces and non-faces through Integral image, AdaBoost algorithm and cascade classifier.

Haar-like feature is a way of feature representation based on the differences between black and white pixels in gray images. It includes three forms: edge feature, linear feature,

Fig. 1 System framework

Fig. 2 Haar-like features based on human eye features

Fig. 3 Integral image

and center feature. Viola and Jones (2001, 2004) optimized it and applied it to face feature extraction. Lienhart and Maydt (2002) further expanded it and eventually applied it to Open CV classifier. Figure 2 shows Haar-like features based on human eye features, it is a kind of edge character, and the area of the black pixel represents the eye color darker than the surrounding area.

Integral Image is a fast method proposed by Viola and Jones (2001) to extract Haar-like features. It is a matrix representation method that can describe global information (Huang et al. 2005). It represents each point on the image as the sum of all pixels in the upper left of the point. Image feature representation can be realized by adding and subtracting pixel points between different rectangles. As shown in Fig. 3, point I can be expressed as the pixels sum of A, B, C, D, the pixel sum of D area may be expressed by the $I_b + I_c - I_a$. Integral Image improves the efficiency of Haar-like feature representation.

AdaBoost algorithm is a kind of adaptive algorithm, which seeks the optimal classifier through numerous loop iteration (Lin 2013). Different facial features represent different classifiers, namely weak classifiers (He and Cheng 2018). Through multiple training iterations, the weak classifier with better classification performance is selected to form a strong classifier, and the final classifier is formed through the cascade of strong classifiers.

The main process of Haar classifier is that using the sliding window and Integral Image to achieve rapid traverse of gray image and Haar-like features calculation. Next through AdaBoost algorithm based on the Haar-like features to train face weak classifier and build strong classifier, and then through multiple strong classifier combined enhancing the effect of classification. As a result, face classification can be realized. The process is shown in Fig. 4.

Here we use haar classifier of Open CV to realize face detection in video homework that has been trained through the above process. Specific face detection process is that the video homework is divided into frames of images. After image gray-processing, the detectMultiScale() function in the trained alt2 identifies whether there are faces in images. The process is shown in Fig. 5.

4.3 Audio detection

Noises are inevitable when students record video homework in the dormitory or classroom. Most of the noises are students' conversation voices and other noises around them. The existing research on audio classification and detection focuses on specific audio matching, speaker recognition and content extraction, which cannot solve the problem of audio detection of students' video homework. Therefore, we need to use the audio data from students' homework to train model to realize audio detection (Fig. 6).

Fig. 4 Face classifier

Fig. 5 Face detection in video homework

Fig. 6 MFCC extraction process of video homework

Firstly we select 14 samples out of the 89 unqualified video homework and set them as noise samples manually. Then, we choose 42 samples from qualified video homework and defined them as clear samples. And we select the audio detection model with higher accuracy through training.

Audio detection is divided into two parts: audio feature extraction and detection model training.

4.3.1 Audio feature extraction

Audio feature extraction is to identify the noticeable features in the audio and eliminate the rest of the redundant information. Basing on the relevant studies on audio processing and the need for the sound clarity of students' video homework, we select Mel-Frequency Cepstral Coefficients (MFCC) based on human ear perception characteristics (Chundong et al 2019) to describe the characteristics of clear audio and noisy audio. The process is as follows (Li et al. 2017 and Lingnizhan 2019).

(1) Pretreatment

The audio pretreatment includes pre-emphasis, framing and windowing.

The purpose of pre-emphasis is to highlight the high frequency formant. The filter coefficient is set as 0.97 in the pre-emphasis, and the formula is:

$$S(n)^* = S(n) - \partial \times S(n-1) \quad (1)$$

The audio is decomposed into shorter frames and processed as steady-state signals, and the smooth transition from frame to frame is realized through the partial overlap between each frame. Basing on the short-time stabilization characteristic of the audio of the student video homework, we frame the signal into 25 ms, set frame shift = 10 ms and $N = 512$. And each frame ($S_i(n)$) multiplied by the Hamming window ($W(n)$) to increase the continuity of left and right ends and reduce the leakage in the frequency domain, and we set $\alpha = 0.46$. The formula is:

$$W(n, \alpha) = (1 - \partial) - \partial \times \cos\left(\frac{2\pi n}{N-1}\right), n = 0, 1, 2, \dots, N-1 \quad (2)$$

$$S_i(n)' = S_i(n) \times W(n) \quad (3)$$

(2) FFT

The frequency domain signal $X_i(k)$ of each frame is obtained by the Discrete Fourier Transform. The formula is:

$$X_i(k) = \sum_{n=1}^N X_i(n) e^{-\frac{j2\pi kn}{N}}, \quad 1 \leq k \leq K \quad (4)$$

where i is the number of frames and K is the length of DFT.

(3) Mel filter bank

The power spectrum $E(I, k)$ is obtained by taking the square of the result of FFT operation. It was filtered through a filter to map the linear spectrum to the Mel nonlinear spectrum based on auditory perception. The conversion formula is as follows.

The formula for converting from frequency to Mel scale is:

$$\text{Mel}(f) = 2595 * \log_{10} \left(1 + \frac{f}{700} \right) \quad (5)$$

To go from Mel back to frequency:

$$f = 700 \left(10^{\frac{\text{Mel}(f)}{2595}} - 1 \right) \quad (6)$$

Then the energy of the power spectrum of each frame in the MEL filter is calculated:

$$SE(i, m) = \sum_{k=0}^{N-1} E(i, k) H_m(k), 0 \leq m \leq M \quad (7)$$

where i is the frame number, K is the spectral line k in the frequency domain, $H_m(k)$ is the frequency domain response of Mel filter, and M is the number of filters. In our experiment, we set $M=24$.

(4) Discrete Cosine Transform (DCT)

Logarithm of the energy obtained through the filter is taken decorrelation processing by DCT to obtain the MFCC.

$$C(i, n) = \sum_{m=0}^{N-1} SE(i, m) \cos \left(\frac{\pi n(m - 0.5)}{M} \right), 0 < n \leq L \quad (8)$$

where i is the frame number, M is the “ m th” filter, and L is the parameter order of MFCC. In this experiment, we set $L=12$.

Finally, we use the plt function and related parameters to obtain the MFCC spectrum diagram of audio.

MFCC feature were extracted with the above process to describe the audio features of the manually screened samples. The audio in 56 video homework selected manually were extracted. Then, first 3.5 s of the audio were intercepted. The integrity features of audio were preserved by pre-weighting, framing and windowing. After that, filter and DCT transformation, the MFCC feature of the audio in video homework was extracted through FFT. The extracted MFCC feature image will be used as the input of the audio detection classification model.

4.3.2 Model training and evaluating

The distinction between clear and noisy samples is essentially a dichotomy problem, so we can use the classification algorithm to realize the audio detection of video homework. Essentially, classification algorithm is to distinguish samples with different features, and the computer learns features to distinguish different categories. We choose more classical KNN, SVM and CNN models to train the manually screened data sets, and compare the training effect of using common spectrum features and MFCC features to describe the audio. Then, we select the training model with higher accuracy and apply it to the audio detection in the system.

(1) KNN

K Nearest Neighbor Classifier (KNN) was proposed by Cover and Hart (1967). KNN has been widely used in many fields because of its simplicity and high classification accuracy (Zhang et al. 2008). It calculates the adjacent sample of the predicting samples based on the distance function, and confirms the category of the predicting samples according to the category of the adjacent sample. The category of predicting audio samples of the video homework is determined by the number of the category of the nearest K samples. Among k adjacent samples, the predicting sample is regarded as a clear sample if there are more clear samples. The predicting sample is regarded as a noisy sample vice versa. Figure 7 shows an example of KNN classification when $K=6$. The experiment invokes the `KNeighborsClassifier()` function to realize the training of KNN.

(2) SVM

Cortes et al. (1995) proposed support vector machines (SVM), it has its unique advantages when solving small sample, nonlinear and high dimensional pattern recognition problems (Liu et al. 2003). SVM searches for the optimal classification surface based on two types of sample data, which not only enables the two types of samples to be separated without error, but maximizes the classification interval between the two types (Vapnik 1997). The experiment invokes the `svm.SVC()` function to realize the training of KNN.

Fig. 7 K nearest neighbor classifier

(3) CNN

With the development of deep learning, CNN has been widely used in many fields. Based on labeled sample data, it learns the sample features of different categories through iterative calculation to achieve classification. This experiment uses a five-layer convolutional neural networks to realize the classification training of video homework audio (Fig. 8), which includes two convolutional layers, two pooling layers and a full connection layer. The specific parameters of each layer are as follows:

Input: spectrum diagram of 128*128*3.

Layer1: convolution layer, convolution kernel size is (5,5), the number is 64, strides = 1;

Layer2: pooling layer with kernel size of (2,2);

Layer3: convolution layer, convolution kernel size is (5,5), the number is 128, strides = 1;

Layer4: pooling layer with kernel size of (2,2);

Layer5: full connection layer with 512 neurons.

We select ReLU as activation function. The features of sample spectrum diagram extracted by the convolution layer are inputted the ReLU function in the form of vector to nonlinear transformation. The ReLU function converges quickly and calculates easily. Problems are not detected in the gradient disappearance (Kutyniok 2019). Its function formula is:

$$\text{ReLU}(x) = \max(0, x) \quad (9)$$

Our experiment is actually a binary classification problem. Therefore, we choose the cross entropy function as the loss function. It can measure the effect of the model and is relatively easy to calculate. The prediction probability of the positive sample is p and the negative sample is $1-p$, and its calculation formula is (Ezail 2019):

$$L = -\frac{1}{N} \sum_i [y_i \times \log(p_i) + (1 - y_i) \times \log(1 - p_i)] \quad (10)$$

where $y_i = 1$ if i sample is positive and $y_i = 0$ if i sample is negative.

At the same time, we choose the Adam optimizer proposed by Kingma and Lei Ba to optimize the experiment, it has the advantages of high computational efficiency, simple implementation and small memory occupancy (Kingma and Ba 2014).

In the audio detection experiment, clear samples and noisy samples were input into the training model in the form of sample-label. The average accuracy of model was taken as the evaluation index to compare the effectiveness of the model in audio detection after lots of training. The experimental results are shown in Table 1.

Table 1 Results

Method	The average accuracy (%)
Spectrum + KNN	58.33
Spectrum + SVM	73.33
Spectrum + CNN	71.84
MFCC + KNN	92.38
MFCC + SVM	89.58
MFCC + CNN	81.17

According to Table 1, the average accuracy of MFCC + KNN reaches 92.38%, which is optimal in the comparison experiment and can be well applied to audio classification. Therefore, this system chooses MFCC + KNN method to realize audio detection of video homework (Fig. 9).

4.4 System application and evaluation

The final system framework is shown in Fig. 10.

We have tested the system with the collected video homework, the accuracy rate of the system reached 84.26%. The system detected 89 unqualified video homework within 5 min and 42 s, which is much faster than a few hours of manual screening. In the detection of individual homework, the system provides prompt feedback to students when they submit homework. Compared with the method of manual screening by teachers, this method is much more efficient.

5 Conclusions

Automatic detection of the unqualified video homework is getting more important as video homework becomes a more popular. This paper puts forward the AVHADS (Abnormal Video Homework Automatic Detection System), which is based on the problem of uploading video homework in “big data acquisition and processing project of actual combat” and other courses. Based on MFCC feature and CNN to realize the automatic detection and feedback of abnormal video homework, the system uses suffix and parameter identification, Open CV, and the audio classification model. It shows that the AVHADS is feasible and effective through experiments.

In conclusion, AVHADS can realize preliminary detection of the unqualified video homework, which is much more time-saving and can sent feedback to students promptly. Nevertheless, the accuracy of the system can be further improved through training. In addition, this system is only the preliminary form detection of students’ video homework, and it does not review the specific content of homework.

Fig. 8 Convolutional neural networks **Fig. 9** Audio detection **Fig. 10** Video homework detection

In the further researches, facial expression analysis, tone change, voice pause and other features can be combined to realize automatic review of video homework with artificial intelligence technology.

Acknowledgements Youth Innovative on Science and Technology Project of Shandong Province (2019RWF013), Postgraduate Education Reform Research Project of Shandong University of Finance and Economics (SCJY1911), Teaching Reform Research Project of Shandong University of Finance and Economics in 2020 (jy202011), Teaching Reform Research Project of Shandong Province (M2018X169, M2020283), “Course Ideological and Political” Teaching Reform Research Project of Shandong University of Finance and Economics (kcsz202085).

References

- Ali Technology (2018) Based on TensorFlow, how to realize voice recognition at the end. <https://developer.aliyun.com/article/59268>. Accessed 25 July 2020.
- Bradski G, Kaebler A (2008) Learning open CV: computer vision with open CV library. O'Reilly Media Inc, Newton, pp 222–264
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Cover TM, Hart PE (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
- Ezail_xdu (2019) Deep Understanding of the cross entropy loss function. https://blog.csdn.net/weixin_38526306/article/details/87831201?utm_medium=distribute.pc_aggpage_search_result.none-task-blog-2~all~first_rank_v2~rank_v25-1-87831201.nonecase. Accessed 15 July 2020.
- Goyal K, Agarwal K, Kumar R (2017) Face detection and tracking: using Open CV. In: International conference of electronics, pp 474–478.
- He Y, Cheng X (2018) Design of face detection system based on Open CV. *J New Ind* 008(006):83–89
- Huang W, Chen B (2005) An integral image method for fast image processing. *Comput Appl* 25(S1):266–268
- Huang GY, Chen J, Liu H et al. (2020) Neural multi-task learning for teacher question detection in online classrooms. https://doi.org/10.1007/978-3-030-52237-7_22
- Jiang D (2020) Structure is the key to curriculum development. Dissertation, Teacher said curriculum reform.
- Jiang Yu, Tao X (2016) Promote the comprehensive deepening reform of curriculum with the core literacy model. *Moral Educ China* 169(01):26–28
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. <http://arxiv.org/abs/1412.6980>.
- Kutyniok G (2019) Discussion of “Nonparametric regression using deep neural networks with ReLU activation function.” *Ann Stat* 48:1911–1915
- Lian D (2016) Algorithm optimization ii: how to improve the accuracy of face detection. <https://blog.csdn.net/u010402786/article/details/52261933>. Accessed 22 Oct 2020.
- Lian Y, Dai YD, Wang H (2002) Anomaly detection of user behaviors based on profile mining. *Chin J Comput* 03:325–330
- Lienhart R, Maydt J (2002) An extended set of Haar-like features for rapid object detection. In: Proceedings of the international conference on image processing in IEEE.

- Lingnizhan (2019) The principle explanation and Python implementation of mel-frequency cepstral coefficients (MFCC) of audio signal. Cnblog. https://blog.csdn.net/qg_34218078/article/details/101141066. Accessed 22 June 2020.
- Liu H, Li J, Cui G et al (2003) Writer identification using support vector machines and texture feature. *J Comput Aided Des Comput Graph* 15:1479–1484
- Li H, Xu X, Wu G et al (2017) Research on speech emotion feature extraction based on MFCC. *J Electron Meas Instrum* 031(003):448–453
- Li X, Yue T, Huang X et al (2019) BAGS: an automatic homework grading system using the pictures taken by smart phones. <http://arxiv.org/abs/1906.03767>.
- Martin M, Timotej L, Ivan B (2018) Identifying typical approaches and errors in Prolog programming with argument-based machine learning. *Expert Syst Appl* 112:110–124
- Ma M, Wang J (2018) Multi-view face detection and landmark localization based on MTCNN. In: 2018 Chinese Automation Congress (CAC).
- McDaniel M (2020) An algorithm for content-based automated file type recognition. Mel frequency cepstral coefficient (MFCC) tutorial. <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>. Accessed 22 June 2020.
- Muda L, Begam M, Elamvazuthi I (2010) Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *J Comput* 2(3). <http://arxiv.org/abs/1003.4083>
- Open Source Computer Vision (2016) <https://docs.opencv.org/3.2.0/index.html>. Accessed 20 June 2020.
- Peiyang He, Ming C (2001) The correcting system of engineering drawing assignment based on Web. *J Eng Graph* 04:156–161
- Tu J, Chen F (2017) Application of video work in higher vocational nursing training class. *Nurs Res China* 15:1867–1868
- Vapnik V (1997) The nature of statistical learning theory. Springer, New York, p 199513
- Viola PA, Jones MJ (2001) Rapid object detection using a boosted cascade of simple features. *Computer vision and pattern recognition, 2001. CVPR 2001*. In: Proceedings of the 2001 IEEE computer society conference on IEEE.
- Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vision* 57(2):137–154
- Wang J (2011) Let modern information technology into mathematics classroom. *Commodities and quality. Front Observ* 18–19.
- Wang L (2013) Research on key technology of face detection. Beijing University of Posts and Telecommunication, Beijing
- Wang M (2019) Feasibility study of video assignment in college English listening and speaking courses. *Inf Rec Mater* 020(007):223–225
- Wei Fu, Yang Y (2018) Audio classification method based on convolutional neural networks and random forest. *J Comput Appl* S2:58–62
- Xiang P (2009) On the construction of two-dimensional teaching objective system. *Shanghai Res Educ* (9):74.
- Xu S (2003) A brief discussion on the application of multimedia technology in mathematics teaching of higher vocational colleges. *China Adult Educ* 13(012):80–81
- Xu C, Jing Z, Dongwen Y et al (2019) Optimization of the extraction method of MFCC feature vectors for heart sound signals. *J Signal Process* 03:410–418
- Yang W, Wang N, Wang G (2014) Study and implement of assignment auto-checking system for civil engineering graphics. *J Graph* 35(4):613–618
- Ye R, Li Z, Chen Y et al (2012) A detection method for abnormal vehicle behaviorbased on space time diagram. *J TranspInfSaf* 030(004):89–92
- Yu D, Yuan Y, Dong H (2006) Text-dependent speaker recognition method using MFCC and LPCC features. *J Comput Appl* 04:883–885
- Zhang X, Peng Y (2019) An audio recognition method based on residual network and random forest. *Comput Eng Sci* 04:727–732
- Zhang Z, Huang Y, Wang H (2008) A new KNN classification approach. *Comput Sci* 035(003):170–172
- Zhao K, Peng T, Li B (2008) Method of anchor person shots detection in news video. *Comput Eng* 19:239–241
- Zhao C, Yan H, Song C (2010) A system for automatic assessment and plagiarism detection of student programs. In: International conference on e-business and e-government. IEEE.
- Zheng J, Luo J, Lu B (2007) Documents type identification based on statistical characteristic. *Comput Eng* 33(001):142–144
- Zhu H (2019) Exploration and practice of homework by video in advanced mathematics. *Educ Teach Forum* 414(20):103–104

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.