Smart System to Recapitulate Student Attendance on Virtual Meeting Platforms During Covid-19

Cahya Rahmad[1(✉)], Arie Rachmad Syulistyo[1], Dimas R.H. Putra[1], Andrea Prati[2], Tomaso Fontanini[2], Rudy Ariyanto[1]
[1] State Polytechnic of Malang, Jawa Timur, Indonesia
[2] University of Parma, Parma, Italy
cahya.rahmad@polinema.ac.id

Abstract—Educators have problems conducting online learning, such as monitoring student attendance while presenting the material. This paper aims to predict student names who attend zoom video conferences with various lighting conditions and face angles by comparing two detection and two recognition methods. This paper proposes an intelligent system based on the use of a bot that will analyse a combination of face detection and recognition method for attendance systems using video conferencing applications to carry out online learning. The proposed system will use the best combination of two methods to recapitulate student attendance. The face detection system uses Haar Cascade and Multi-Task Cascaded Convolutional Neural Network (MTCNN), and the face recognition system uses ResNet and FaceNet. The tests were conducted on video zoom footage taken during online lectures. The results show that MTCNN and FaceNet get the highest accuracy, 93.23%.

Keywords—FaceNet, Facial Recognition, Haar Cascade, MTCNN, ResNet

1 Introduction

Recently, digital transformation has been swift and massive. This phenomenon is mainly due to the pandemic caused by the Covid-19 virus. In education, digital transformation occurs because schools or lectures are conducted online using LMS (Learning Management Systems) to manage teaching materials and video conferencing for student meetings with the teachers. Many video conferencing options can be used for online learning, for example, Zoom, Google meet, Jitsi, Webex, etc. Of these applications, the one with the most users is zoom because they are already familiar with it and the user interface is easy to use. Indeed, the advantages possessed by zoom are well described in [1]–[3]. The advantages are flexible working options for staff, process learning activities in schools and colleges can be carried out at home, especially in the Covid-19 affected areas, and the learning continues to happen. Zoom is also one modern application approach to teaching-learning. Previously, e-learning, distance educa-

tion and correspondence courses were popularly considered part of non-formal education. Still, they seem to gradually replace the formal education system if the circumstances persist over time.

Some problems arise when educators face online meetings. One of them is the difficulty of monitoring or supervising students when the teacher explains [4]. The main problem is that students must focus on listening to the presented material. While educators also know the condition of students, whether students are still in the meeting room or have left the room while in the middle of learning. However, this is not easy because of the large number of students to supervise [5], [6]. Furthermore, the teacher is explaining the material, and due to the limitations of the zoom participant's screen, it would be difficult for educators to check the attendance in the room [7].

An attendance system using facial recognition whose datasets are directly obtained from zoom meetings has never existed. In developing countries, especially in our country, Indonesia, there has yet to be a video conferencing-based attendance system based on the use of a bot, so this research is significant. The image output quality from this video conference has many challenges: varying lighting conditions, occlusions of the student's faces, and blurred facial conditions due to the quality of the software resolution.

Based on the described background, this research will be conducted to solve the problems educators face by using automation technology and algorithms in digital image processing and deep learning, comparing the results and choosing the best model to be used. For the face detection method, we will compare Haar cascade and MTCNN, and for face recognition, we will compare ResNet and FaceNet. Furthermore, this research aims to prove that the proposed algorithm can solve the problems encountered and implement the algorithm with high accuracy, light and angle invariant in the form of a monitoring system based on the use of a system based on a bot. Finally, this system can help educators to monitor student attendance during learning.

2 Research method

2.1 Previous research

There is much research that discusses facial detection and recognition. For example, K. Zhang et all [8] developed a technique that combines face alignment and detection in unrestricted contexts by cascading three deep convolutional network layers. It is known as the MTCNN, the most accurate and widely used face detection tool today. It is composed of 3 neural networks connected in a cascade. The prior face detection algorithm widely used was Haar Cascade, invented by Viola and Jones [9]. This algorithm can detect faces from an input video [10].

The novel technique and architecture to recognise the face using a deep neural network have become a hot topic in the last decade [11]. FaceNet [12] has the architecture to process a person's face image as input and compresses it into a vector of 128 numbers representing the essential features of the face. Residual Networks (ResNet) [13] is a classic and robust neural network used as a backbone for many usages in computer

vision tasks. The breakthrough with ResNet was that it could successfully train a deep neural network with more than 150 layers. The research continues to be developed by several researchers worldwide and produces many benefits in computer vision, especially for face detection and facial recognition. [14] uses the enhanced Viola-Jones algorithm so that the detection results increase by 12%, and the speed increases four times. Haar Cascade algorithm is also applied in facial emotion recognition by [15] with 74% accuracy, 73% precision, 76% recall, and computation time reaching 15 seconds per frame. [16], also, comparing the algorithm, which was developed by Viola-Jones, with another face detector method has better results. The basic principles of current mainstream face detection algorithms MTCNN and the YOLOV3 model were deeply analysed by [17], and MTCNN can handle the extreme case where many faces with small targets in pictures. [18] optimised the MTCNN face detection algorithm by modifying each cascading network module (PNet, RNet, and ONet) using the FDDB face test set. The detection speed result has increased by 70.1%, and the MTCNN face detection algorithm has good robustness for face pose changes and can work well in low computing power scenarios.

2.2 Research design

In this study, the educators used two computers to run zoom, as shown in Figure 1. The teacher will use the first device to present the material. At the same time, the educators will use the second computer to run the facial recognition algorithm on the students who attend.

The system based on the use of a bot runs on the second device and consists of 4 steps, as seen in Figure 2. During the first step, the system captures a screen containing the student's face, and then the system detects faces using one of the Haar Cascade or MTCNN methods. In the third step, image processing techniques are performed, the system crops the detected faces, and finally, facial recognition is performed using FaceNet and ResNet. The details of this process are explained in Figure 2.
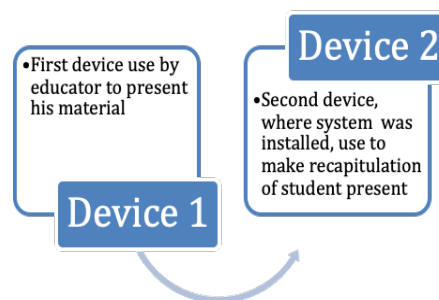

Fig. 1. General Idea of Attendance Monitoring System

Fig. 2. Illustration Process of Attandance Monitoring System

2.3 Data samples

One of the essential parts of conducting this research is data. The data used is a collection of student facial images from 5 classes taking online learning. It collects 494 data and 106 folders or class labels. The collection of face images has been labelled with the name [student_name][space][student_id_number] for the name folder. Each folder consists of 5 face images cropped according to various angles or head positions with the naming format [student_name][number].jpg. In the file name, the number represents each angle orientation. The figures and angles are as follows; the illustration of the folder name can be seen in Figure 3.

- Number 1 = front view
- Number 2 = turn to the left
- Number 3 = turn to the right
- Number 4 = looking up
- Number 5 = looking down

Fig. 3. Folder Naming Management

2.4 The tools and application

We use two computers, one computer to make presentations or learning processes, the second computer to carry out monitoring and attendance processes using a bot-based system. one full screen computer with presentation material, while two full screen computers with video images of all participants starting from the beginning of learning to the end of learning. the bot-based system will periodically make attendance, once every one minute, the results of facial recognition will be included in the csv file. which can be monitored directly by the teacher.

The tools we use are zoom meetings, screen captures, and systems that use the Python-based programming language. the output from our system is a csv file. which contains the attendance of students periodically, that is once a minute. for example, if we hold a meeting for 60 minutes, there will be 60 attendance times, so if a participant disappears or is not present in the middle of the learning process, the teacher will monitor it.

3 Experiment result and discussion

3.1 Experiment result

The system was experimented with in six classrooms. The system takes 15 screen captures in each class, so 1746 faces are captured. The system results applied to these conditions can be summarised as follows. Based on the performance in Table 1, MTCNN performs better than Haar Cascade performing face detection. This is because MTCNN has taken advantage of a more complex machine learning architecture. So that with difficult conditions, MTCNN can still detect students' faces, for example, poor lighting. The results based on Table 1 are system recognition capabilities per frame. In the recognition time interval of 30 minutes, these methods provide high accuracy results in recognising the faces of students in a class and the average accuracy of attendance success for every class and method by using 10-Fold cross-validation is shown in Table 2. The average ability of each technique in performing automatic attendance for 30 minutes using Haar+ResNet is 72.22%, Haar+FaceNet is 87.62%, MTCNN+ResNet is 78.62%, and MTCNN+FaceNet is 93.23%.

Table 1. Face Recognition Experiment Results per Frame

Architecture	Face Detected	Face Detection Success Percentage	Total True Recognition	True Recognition Percentage
Haar+ResNet	979	56.07%	6	0.34%
Haar+FaceNet	1011	57.9%	198	11.34%
MTCNN+ResNet	1710	97.94%	11	0.63%
MTCNN+FaceNet	1710	97.94%	202	11.57%

Experiments show that the best combination for face detection and face recognition is MTCNN for face detection and FaceNet for face recognition. This is because MTCNN works well even in low light conditions [19], and faces only partially appear. On the other hand, FaceNet is also robust in distinguishing the faces of each student who has been trained [20]. This happens because FaceNet is invariant in low-light conditions.

Table 2. Automatic Attendance Recognition Experiment Result

Class	Method	Average attendance success	Class	Method	Average attendance success
1	Haar+ResNet	70.3%	4	Haar+ResNet	69.6%
	Haar+FaceNet	85.3%		Haar+FaceNet	84.6%
	MTCNN+ResNet	77.6%		MTCNN+ResNet	76.3%
	MTCNN+FaceNet	93.3%		MTCNN+FaceNet	92.3%
2	Haar+ResNet	72.6%	5	Haar+ResNet	75.6%
	Haar+FaceNet	88.6%		Haar+FaceNet	91.3%
	MTCNN+ResNet	79.3%		MTCNN+ResNet	82.3%
	MTCNN+FaceNet	95%		MTCNN+FaceNet	94.6%
3	Haar+ResNet	71.6%	6	Haar+ResNet	73.6%
	Haar+FaceNet	87.3%		Haar+FaceNet	88.6%
	MTCNN+ResNet	75.6%		MTCNN+ResNet	80.6%
	MTCNN+FaceNet	93.6%		MTCNN+FaceNet	90.6%

4   Conclusion

This paper proposes a system that can help teachers to check and recapitulate the attendance automatically when online learning. Based on experiments, the system has challenges, namely low lighting, face that are not fully visible and the angle of the face when captured. Those problems can be solved by the system with the best accuracy using the MTCNN face detection method and FaceNet face recognition with an accuracy of 93.2%.

5   References

[1] T. W. Afrianty, I. G. L. S. Artatanaya, and J. Burgess, "Working from home effectiveness during Covid-19: Evidence from university staff in Indonesia," *Asia Pacific Manag. Rev.*, vol. 27, no. 1, pp. 50–57, 2022. https://doi.org/10.1016/j.apmrv.2021.05.002

[2] L. Mishra, T. Gupta, and A. Shree, "Online teaching-learning in higher education during lockdown period of COVID-19 pandemic," *Int. J. Educ. Res. Open*, vol. 1, no. June, p. 100012, 2020. https://doi.org/10.1016/j.ijedro.2020.100012

[3] S. Vandenberg and M. Magnuson, "A comparison of student and faculty attitudes on the use of Zoom, a video conferencing platform: A mixed-methods study," *Nurse Educ. Pract.*, vol. 54, no. June, p. 103138, 2021. https://doi.org/10.1016/j.nepr.2021.103138

[4] N. Nordin and N. H. M. Fauzi, "A web-based mobile attendance system with facial recognition feature," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 5, pp. 193–202, 2020. https://doi.org/10.3991/ijim.v14i05.13311

[5] J. Tao and X. Gao, "Teaching and learning languages online: Challenges and responses," *System*, vol. 107, no. May, p. 102819, 2022. https://doi.org/10.1016/j.system.2022.102819

[6] C. R. Vicente *et al.*, "The Joint Initiative for Teaching and Learning on Global Health Challenges and One Health experience on implementing an online collaborative course," *One Heal.*, vol. 15, no. June, p. 100409, 2022. https://doi.org/10.1016/j.onehlt.2022.100409

[7] E. Al Hajri, F. Hafeez, and N. V. Ameer Azhar, "Fully automated classroom attendance system," *Int. J. Interact. Mob. Technol.*, vol. 13, no. 8, pp. 95–106, 2019. https://doi.org/10.3991/ijim.v13i08.10100

[8] 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, 2016. https://doi.org/10.1109/LSP.2016.2603342

[9] P. Viola and M. J. Jones, "Robust real-time audiovisual face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004. https://doi.org/10.1117/12.545934

[10] J. Anitha, G. Mani, and K. Venkata Rao, "Driver Drowsiness Detection Using Viola Jones Algorithm," *Smart Innov. Syst. Technol.*, vol. 159, pp. 583–592, 2020. https://doi.org/10.1007/978-981-13-9282-5_55

[11] M. L. Prasetyo *et al.*, "Face Recognition Using the Convolutional Neural Network for Barrier Gate System," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 10, pp. 138–153, 2021. https://doi.org/10.3991/ijim.v15i10.20175

[12] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 815–823, 2015. https://doi.org/10.1109/CVPR.2015.7298682

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016. https://doi.org/10.1109/CVPR.2016.90

[14] J. Huang, Y. Shang, and H. Chen, "Improved Viola-Jones face detection algorithm based on HoloLens," *Eurasip J. Image Video Process.*, vol. 2019, no. 1, 2019. https://doi.org/10.1186/s13640-019-0435-6

[15] K. Candra Kirana, S. Wibawanto, and H. Wahyu Herwanto, "Facial Emotion Recognition Based on Viola-Jones Algorithm in the Learning Environment," *Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018*, pp. 406–410, 2018. https://doi.org/10.1109/ISEMANTIC.2018.8549735

[16] C. Rahmad, R. A. Asmara, D. R. H. Putra, I. Dharma, H. Darmono, and I. Muhiqqin, "Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 732, no. 1, pp. 0–8, 2020. https://doi.org/10.1088/1757-899X/732/1/012038

[17] N. Zhang, J. Luo, and W. Gao, "Research on face detection technology based on MTCNN," *Proc. - 2020 Int. Conf. Comput. Network, Electron. Autom. ICCNEA 2020*, pp. 154–158, 2020. https://doi.org/10.1109/ICCNEA50255.2020.00040

[18] Y. G. Xie, H. Wang, and S. H. Guo, "Research on MTCNN face recognition system in low computing power scenarios," *J. Internet Technol.*, vol. 21, no. 5, pp. 1463–1475, 2020, doi: 10.3966/160792642020092105020.

[19] N.-M.-Q. D. Trong-Nghia Pham, Nam-Phong Nguyen and T. Le, "Tracking Student Attendance in Virtual Classes Based on MTCNN and FaceNet," in *(eds) Intelligent Information and Database Systems. ACIIDS 2022. Lecture Notes in Computer Science()*, 2022, pp. 382–394. https://doi.org/10.1007/978-3-031-21967-2_31

[20] M. Shen, H. Yu, L. Zhu, K. Xu, Q. Li, and J. Hu, "Effective and Robust Physical-World Attacks on Deep Learning Face Recognition Systems," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 4063–4077, 2021. https://doi.org/10.1109/TIFS.2021.3102492

6   Authors

Cahya Rahmad is with The State Polytechnic of Malang, East Java, Indonesia (email: cahya.rahmad@polinema.ac.id).

Arie Rachmad Syulistyo is with The State Polytechnic of Malang, East Java, Indonesia.

Dimas R.H. Putra is with The State Polytechnic of Malang, East Java, Indonesia.

Andrea Prati is with The University of Parma, Parma, Italy.

Tomaso Fontanini is with The University of Parma, Parma, Italy.

Rudy Ariyanto is with The State Polytechnic of Malang, East Java, Indonesia.