An Incremental Training on Deep Learning Face Recognition for M-Learning Online Exam Proctoring

Asep Hadian Sudrajat Ganidisastra
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
hadithea@students.itb.ac.id

Yoanes Bandung
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
bandung@stei.itb.ac.id

Abstract*—The ability to provide an academic resource for a remote student has increased the use of m-learning in distance education. Online exams as a tool to measure the student's outcome need a proctoring method to detect cheats. Several methods had been proposed to fulfill these needs, from a no-proctoring exam to automatic online supervision. A visual verification during an online exam is required to verify a student took the exam, therefore a CNN-FR is used to do it. The problem that exists in face recognition is the system invariant against pose and lighting variations. In some proposed methods, an additional process such as image equalization and SURF features is executed to overcome the problems. In this paper, we proposed an incremental training process on face recognition training, so there will be no need to add another process so it will reduce the computation cost and time. To acquired high accuracy we've analyzed four different face detectors, which are Haar-cascade, LBP, MTCNN, and Yolo-face, as in face recognition a Facenet model was tested. The evaluation of the proposed method shows that a deep learning face detector has overcome the others, on the other hand, an incremental training of facenet model results in a smaller dataset size by 1% with a faster training time of 7% on Yolo-face face detector and 64% on MTCNN compared to batch training. The proposed method results in an equally high accuracy rate as in batch training (98%).

Keywords: automated online exam proctoring, continuous user verification, online examination, CNN-FR, incremental training, facenet, yolo-face, mtcnn.

I. Introduction

The forms of remote education such as m-learning can provide academic resources for students who hardly had access to campus. A student can perform the academic activities and access learning content provided through their own devices with an internet connection. An online exam as a tool to measure how far the student understands the material given in a remote education will need a method to detect the occurrence of cheating that always exists in an exam. As in conventional exams, a proctoring process during the exam is needed to verify whether the student who took the exam is valid.

Several methods had been proposed to fulfill the needs of proctoring an online examination. The methods to supervise an online exam can be categorized into a) no supervising; b) remote supervisor; c) using desktop robot; and d) fully automatic using face recognition [31]. In this work, we will be focused on the last category to supervised online exams. In an online exam, the system has to be able to make sure that the student taking the exam is a valid one. Therefore we have to visually verify the user continuously based on the student's face is provided. Researches on continuous user verification based on the ability to verify a user's face states that the problem that remains exists is the invariant against the variations of pose and lighting.

The proposed method in this work is by applying incremental training on the deep learning face recognition training process. The images used for training are acquired from the lecture sessions where the user attends. The method is expected to keep the invariant on pose and light variations while reducing the training time and disk space dataset size which can reduce the computation load on the server.

The Design Research Methodology (DRM) [23] was adopted in this research, where we can repeat the research stages until the objective defined is achieved. In this work, the evaluation is performed by measuring the training time, the accuracy rate, average face detection time, and disk space datasets size.

The rest of the paper will be presented as follow. Researches related will be presented in the first section. Then the system description will be delivered in the second section. The proposed method will be presented afterward, while in the fourth section the experimental result will be described. Then the conclusion will be delivered in the last part.

II. Related Works

Methods to supervise an online exam had been proposed. In [3], strict rules are applied to decrease the possibility of cheating to happens, while in [5] a method is proposed by sending images of exam classrooms to the server continuously. Then in [6], audio-video streaming, and the user's screen snapshots are send to the server continuously and the student is verified by the supervisor based on the image captured on sign-in. In [6], the attached 360^o camera desktop robot with movement sensor records the exam room when certain events are detected then sends it to the server for further evaluation. Then in [1], an automated online exam supervising was introduced to predict the cheating behavior of participants in an online exam continuously during the exam. The cheating behavior detected is including impostors, where a student is replaced by another in completing the exam.

The type of user verification that is mostly used is using a PIN and password. The other approach is using token or OTP, then another approach is becoming a trend to increase privacy. The approaches are using biometrics which offers more accuracy and effectiveness in the user's identification or verification [7], [8], [9]. In online exam matters, what's become the consideration is the deal between a valid student with the unauthorized person to replacing on taking the exam partially or for the entire session. For that reason, an online exam proctor should be able to verify whether the student taking the exam are valid during the exam session.

Continuous user verification is a process of verifying a user for the complete login session [13]. The use of small devices for verification such as tokens or cards are susceptible to losing, that's why biometrics is more secure because it disposes of the use of such devices. Several works mentioned that a uni-modal biometric is not adequate for user verification because a failure while taking the sample could happen anytime, for example, face verification could be failed when the system failed to detect a face in the image because of the low light condition. An effort to solve this problem [18] suggests the use of multi-modal biometrics to increase accuracy. Multi-modal biometrics means integrating different types of inputs, such as integrating the hard biometric like a fingerprint with soft biometrics such as clothes or keystroke dynamics.

Several works have been proposed a continuous user verification method. In [14], a camera is capturing a face image that will be verified. The concern is that there are possibilities that no face is provided by the user to the camera. Then in [11], a camera is recording the user's video during the login session and used it as input then the face features and movement are estimated based on the frames. In [19] multimodal biometrics (fingerprint, user's face, and keystroke dynamics) is captured and verified continuously in the server. Then in [20], a camera is capturing the user's image in a period to create the user's template, and face detection and verification are performed by using PCA and eigenface by using a 100x100 pixels image. The remaining is the invariant against variations on poses and lighting. A SURF feature is used to negotiate on the pose variations [25], while eye localization was done to overcome the pose variations [26]. While some methods proposed to cope with the lighting variations are typically by applying additional preprocess steps by executing histogram equalization, gradient, gamma correction, v histogram specification, or the combination of those methods [24].

In this work, we will be using face biometric that is captured using the user device throughout the online lecture attended for the training dataset and exam session for the verification dataset. Continuous user verification is important in an online exam because the student taking the exam needs to be verified continuously during the exam.

III. System Description

The face recognition process needs extra computation load when another pre-processing process was applied, such as histogram equalization, pose alignment, etc., which reduces the speed performance relatively. Limitation on mobile devices memory requires the use of algorithms that employs smaller memory to run. The proposed method is performed to satisfy those needs without disposing of the need for accurate and relatively fast performance that is suitable to be applied for online exam supervising on m-learning media.

Face Detection

The general face recognition process performs three main steps, which are: (1) face detection; (2) feature extraction; and (3) face recognition. Many methods had already been proposed in each step to increase accuracy. In face detection, several methods proposed that are commonly used are viola-jones/haar-cascade method, Local Binary Pattern (LBP) method, Multi-Task Cascaded CNN (MTCNN), and last but not least is YOLO-face. The first two methods are representing the traditional method in face detection, while the last two representing the deep learning area. Viola-Jones formerly a very famous method for face detection. This method uses three main ideas: the integral image, Adaboost classifier learning, then the attentional cascade structure [28]. On the other hand, LBP using the texture description to build local descriptions of the face and combining them into a global description. Major challenges in LBP are posing and light variation. The next one is the MTCNN method. MTCNN consists of three stages, first, it creates candidate windows through a shoal CNN. Then, another CNN improves the windows to reject non-faces windows. Finally, it uses powerful CNN to improve the result. MTCNN can exceed many face detection while preserving real-time performance [27]. The other algorithm use to detect a face in this research is YOLO-face. YOLO-face is a one-stage deep learning algorithm that was able to detect faces under various scales. YOLO-face is a face detection method developed based on YOLOv3. It includes the use of anchor bounding boxes of a face that more suitable for face detection and uses a more precise regression loss function. The Yolo-face accuracy increase significantly increased and still keep its fast detection speed [29].

Face Verification Method

The face verification process was evaluated using two different method that recently claims their popularity due to their high accuracy during the recognition process. Those methods are Facenet and mobileFacenet. Both methods are in the deep learning area performing a Convolutional Neural Networks (CNN) machine learning architecture.

The verification processes were performed using the same dataset produced by two different face detection methods (MTCNN and YOLOface) which show high accuracy results in detecting the face of actual dataset images under various light and pose conditions (above 90%).

A Facenet algorithm was used in this research to provide the system capability in verifying a user's face. FaceNet is a one-shot model, that maps the face images to a euclidean space that will be feed into the network after the feature extracted using FaceNet embeddings as feature vectors. The training method is using triplet loss which groups the images of the same user by reducing the distance between the positive sample and the anchor image, while the distance between the anchor image and a negative image sample is enlarged.

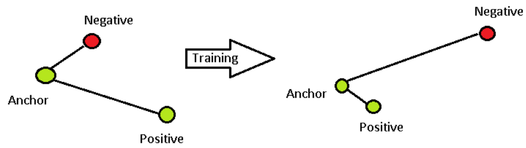

Fig. 1. Triplet loss before and after training

Higher Facenet accuracy can be slightly achieved when a good performance of a face detection algorithm is applied. Therefore we test the prior existing face detection method to complement the Facenet algorithm. Facenet also has its robustness against image compression (JPEG Image compression down to the quality of 20), besides that each face is represented by a 128-dimensional byte vector, smaller embeddings are possible at a minor loss of accuracy and could be run on mobile devices[30].

Evaluation

The evaluation process is performed to measure the result of the incremental training method by using test images under the various condition on pose and light conditions. The parameters to be evaluated are the time performance on the face detection process and FR training process, accuracy rate on face verification, and the image datasets size used in the training process. The accuracy rate is done by using the formula shown in (1), while to calculate the decrease of the training time we simply use the formula in (2).

$$Accuracy = \frac{Number\ of\ valid\ result}{Number\ of\ test} \times 100\% \quad (1)$$

$$\%Decrease = \frac{Original\ Value - New\ Value}{Original\ Value} x100\% \quad (2)$$

IV. Proposed Design

The proposed design aimed is to provide an accurate and fast verification process with a smaller dataset size and keep the robustness against the variations of pose and lighting and also applicable on mobile devices. The complete design of this method is explained in previous work [31], which consists of five main modules as shown in Fig. 2.

In this work, we proposed an incremental training for the faces of users is applied to the face recognition process. The incremental states that the training of user's faces are performed gradually, by using the user's face image acquired from the lecture session. Each time a new collection of user's image is acquired, the previous dataset saved in the server storage that has already been trained is deleted and replaced by the new dataset. By using this method, the usage of the server storage will not increase rapidly. Furthermore, the size of the dataset used for training will be reduced depends on the image quantity acquired in each session. This method will reduce the server's process load on behalf of a smaller dataset size compare to the one-batch training process but still maintain the accuracy of the face detection and face verification with different pose and lighting variations. The verification is using the FaceNet model for face recognition which has high accuracy in user verification and identification.

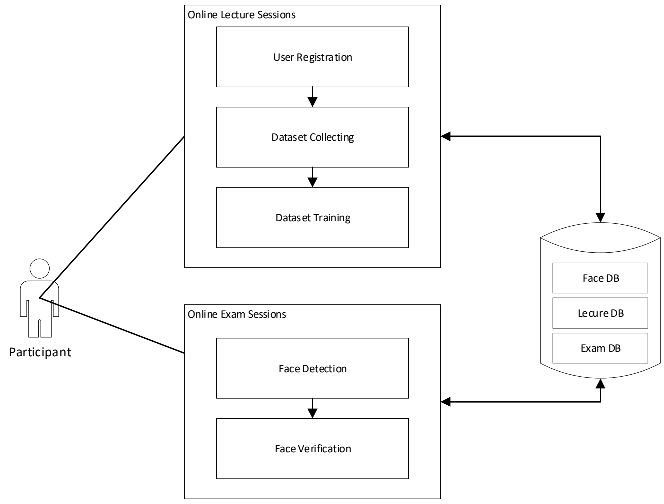

Fig. 2. The architecture of Online Exam Supervision System [31]

The first module is registering a user's identity using the user's smartphone and capture the outset user's face with various pose and lighting conditions as described in Fig. 3.

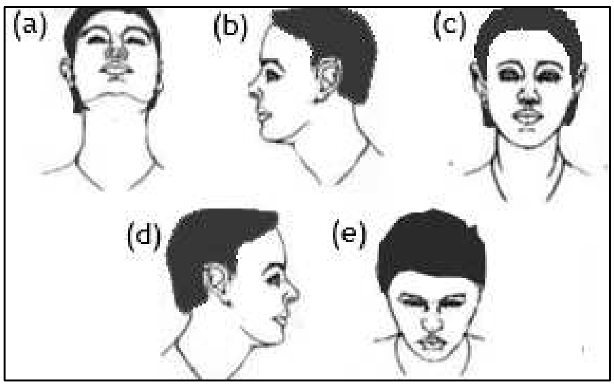

Fig. 3 Initial face poses a) half lookup view; b) left view; c) frontal view; d) right view; e). half-bow view

The image dataset will be acquired gradually on every user's online lecture session and used for the incremental training which will be performed every time a user's session ended. The image is resized into smaller sizes in consideration of the computational cost and user's internet capability. On dataset collecting process, continuous face detection is performed each time an image is captured on random time, while on supervision task, three different conditions will be detected: a) no-face; b) multiple-face; c) single face; then a continuous user verification will be performed each time a single face is detected on the captured image [31].

The system's prototype is developed to evaluate the results of the proposed method. The incremental training is applied to the face detection and face embedding stage before they are fed into the Facenet model. Bigger input image size is not linear compare to the face detection accuracy, because the image captured by the user's smartphone will be resized to 320x244 pixels, while the face extracted from the image is resized to 160x160 pixels based on the input size used by the Facenet model. The face detection, face verification, and training process are developed under Python and Tensorflow environments. The OpenCV library is used to support image processing such as resizing image size and developing face detection using haar-cascaded and LBP. The pre-trained model of MTCNN, YOLO-face, and Facenet are used in developing the evaluation.

V. EXPERIMENTAL RESULT

In this research, we compare the four face detection methods mentioned earlier. The image size uses 320x244 pixels as typically uses in image processing[28]. A larger image size will only cause a higher load on the system and may cause false detections. The comparison result is presented in Table 1 and Table 2. Based on evaluation data on both tables, we can prove the previous statement that higher image resolution doesn't always result in better performance in face detection. Table 2 shows the evaluation result of face detection using the recommended image size. The results show a better accuracy compared to the others.

A. *Evaluation Scenario*

As shown in Fig. 4, the face detection method is loaded alternately each time after all sample dataset images are processed, for instance, detect faces of all sample dataset images using LBP after finished detect faces of all sample dataset images then repeat the process by using MTCNN, and so on until the four face detection method is evaluated.

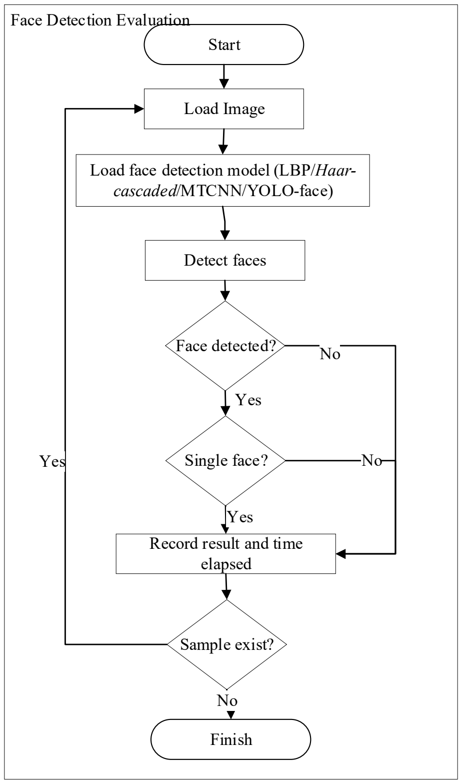

Fig. 4 Flowchart of face detection evaluation

The preview of the experimental stage for evaluating the face detector is shown in Fig. 5. The existence of the face is detected from all the image datasets. The face detection method was evaluated alternately after all the images are processed. This study uses 1295 images of four different users that can be classified based on different categories as depicted in Table 1.

Table 1. Images categories used in the research

Variation Type	Description	Quantity
Pose	Frontal view	944
	Left view	103
	Right view	96
	Half-bow view	49
	Lookup-view	103
Lighting	Range 0-93	247
	Range 94-186	1023
	Range 187-255	25
Occlusion	With Medical Mask	90
	No Medical Mask	1205
Expression	Normal	1070
	Other expressions	225
Scale/Distance	< 60 cm	911
	60 - 300 cm	279
	300 -500 cm	105

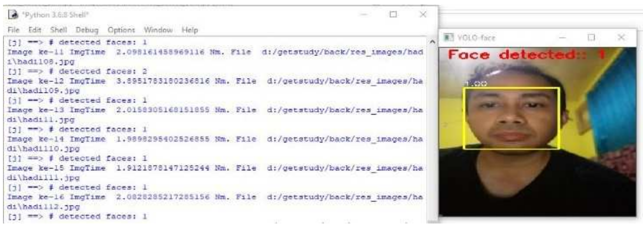

Fig. 5 Preview of face detections evaluation

The evaluation scenario for face recognition using one-batch training methods is shown in Fig. 7. The face detection method is loaded alternately each time after all sample dataset images are processed, for instance, detect faces of all sample dataset images using YOLO-face after finished detect faces of all sample dataset images using MTCNN. The next process perform is the face embedding on each face dataset resulted from YOLO-face and MTCNN methods. Once the face embeddings finish, the face embeddings dataset is then fed into the Facenet model to be trained. After the training process is finished then the test images are verified using the trained model. The parameters recorded are the time of detection, embedding, training and testing, the dataset size, and the accuracy rate. The preview of face verification is shown in Fig. 6.

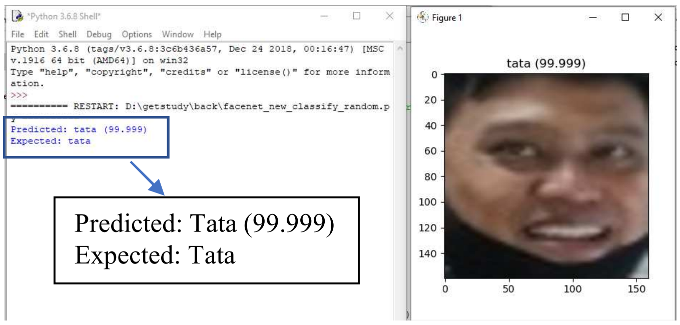

Fig. 6 Preview of face verification evaluation

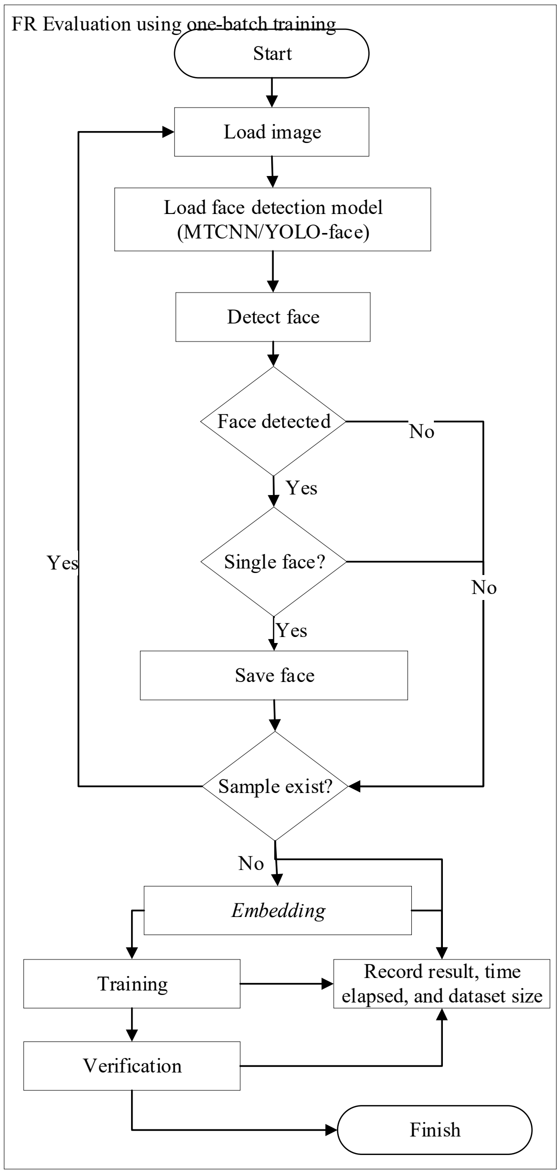

Fig. 7 Flowchart of FR using one-batch training evaluation

The evaluation scenario for face recognition using incremental training methods is shown in Fig. 8. The initial process of this method is equal to the previous, the difference is the process executed after the face embeddings. In incremental training, after the face embeddings finish, then the existing embeddings dataset is updated using the new face embeddings. Once the face embeddings update is finished, the new face embeddings dataset is then fed into the FaceNet model to be trained. After the training finished then the test images were then verified using the trained model.

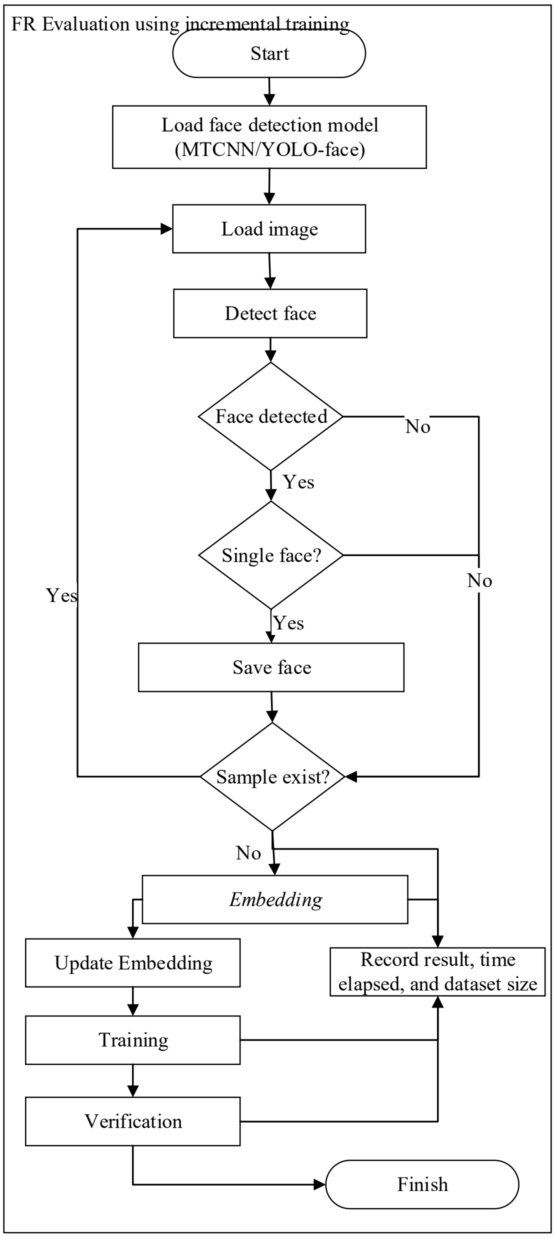

Fig. 8 Flowchart of FR using incremental training evaluation

B. Evaluation Results

Table 2. and Table 3. show the evaluation result of face detection with two datasets with different image sizes. Table 2. shows the evaluation result using datasets of images with the original size captured from the smartphone's camera (4068x3458 pixel), while Table 3. shows the evaluation result using the resized image size (320x244 pixel).

Table 2. Evaluation Result of Face Detection Methods with High-Resolution Image

Method	Image Qty	Average Detection Time (sec)	Accuracy Rate (%)
LBP	1295	4.89	22.70
Viola-Jones	1295	5.87	55.86
MTCNN	1295	11.29	80.46
YOLO-face	1295	1.49	95.83

Table 3. Evaluation Result of Face Detection Methods with Recommended Resolution Image (320x244 pixel)

Method	Image Qty	Average Detection Time (sec)	Accuracy Rate (%)
LBP	1295	0.08	43.71
Viola-Jones	1295	0.16	52.89
MTCNN	1295	1.47	93.28
YOLO-face	1295	1.42	96.06

Table 4. shows the evaluation result of face recognition with two different training methods, which is batch training where all the image dataset is trained at one time, while the other is incremental training where the image was trained incrementally, divided into several training sessions based on different lessons scenario. The prototype is using a pre-trained Keras FaceNet model provided by Hiroki Taniai The face image fed into the network resulted from two different face detection methods, YOLO-face and MTCNN. The results show that The accuracy rate is equal on both training methods, but the other parameter shows better results. The training time was measured by subtracting the end time by the start time on the dataset training session. The total training time of incremental training on YOLO-face-Facenet results in 7% faster than the one-time training, while on MTCNN-Facenet it results in 64% faster. The last disk space dataset size describes the disk space size used to save the image dataset on behalf of the incremental training because the dataset images were replaced by the newer datasets on each training session while the total disk space dataset size described the summary of the size of the dataset used in every training session. The disk space size of the image dataset of incremental training also decreased by 1% smaller than the other. Smaller memory usage is expected to be performed due to the smaller dataset size on each training process

Table 4. Evaluation Result of Face Recognition Methods

Model	Face Detection	Training Method	Accuracy Rate (%)	Training time (min)	Last disk space dataset size (MB)	Total disk space dataset size (MB)
Face net	Yolo-face	One time	98.22	34.82	61.1	61.1
		Incremental	98.22	32.21	1.87	60.42
	MTCNN	One time	98.63	62.87	63.4	63.4
		Incremental	98.63	22.37	1.97	62.76

VI. Conclusion

The evaluation results have shows us that incremental training has a better performance compared to batch training in speed and dataset size. The decrease of training speed and dataset size is not giving a negative influence on the accuracy rate, on the contrary, the proposed method will result in smaller storage space, smaller memory usage, and faster training speed.

On the other hand, the face detection method can result in better face recognition accuracy. The deep learning (MTCNN and YOLO-face) method shows better performance than the traditional one (Viola-Jones and LBP). Therefore a deep learning method can be a better consideration for using in face recognition. While the MTCNN and YOLO-face comparison show a slightly different evaluation result. Based on the datasets used, YOLO-face has overcome MTCNN both on detection accuracy rate and detection speed. The method with better performance will be qualified to fulfill the expectation in the development of a good proctoring system with high accuracy.

A good performance of accuracy and memory usage of face recognition will impact a better user verification and vast device compatibility when applied on mobile online exam proctoring system. So that in future research, we can analyze other novel methods or algorithms in the face detection area such as Ultra Light Fast Generic (ULFG) face detector and retinaFace and ArcFace, and another face recognition method to produce a better proctoring system performance. An implementation of the online exam proctoring by applying the proposed face detection and face recognition method on the smartphones will also be an appropriate agenda to evaluate whether the proposed method would fulfill the expectation under smartphone limitations.

REFERENCES

[1] Y. Atoum, L. Chen, Alex X. Liu, Stephen D. H. Hsu, and X.Liu, "Automated Online Proctoring", IEEE Transaction on Multimedia, 2015.

[2] G. Cluskey Jr, C. R. Ehlen, and M. H. Raiborn, "Thwarting Online Exam Cheating Without Proctor Supervision", Journal of Academic and Business Ethics, 4:1–7, 2011.

[3] Wahid, Y. Sengoku, and M. Mambo, "Toward Constructing a Secure Online Examination System", Proceeding of the 9th International Conference on Ubiquitous Information Management and Communication, pp. 95. ACM, 2015.

[4] P. Guo, H. Feng Yu, and Q. Yao, "The Research and Application of Online Examination and Monitoring System", IT in Medicine and Education, IEEE International Symposium, pp. 497–502, 2008.

[5] Jung and H. Yeom, "Enhanced Security for Online Exams Using Group Cryptography. Education", IEEE Transaction, 52(3):340–349, 2009.

[6] W. Rosen and M. Carr, "An Autonomous Articulating Desktop Robot for Proctoring Remote Online Examinations", Frontiers in Education Conference, IEEE, pp. 1935–1939, 2013.

[7] M. Chihaoui, A. Elkefi, W. Bellil, and C. B. Amar, "A Survey of 2D Face Recognition Techniques", Multidisciplinary Digital Publishing Institute (MDPI) Open Access Journal, 2016.

[8] N. M. Agashe and S. Nimbhorkar, "A Survey Paper on Continous Authentication by Multimodal Biometrics", International Journal of Advanced Research in Computer Engineering and Technology, 2015.

[9] S. Z. S. Idrus, E. Cherrier, C. Rosenberger, and J. J. Schwartzmann, "A Review on Authentication Methods", Australian Journal of Basic and Applied Sciences, 7 (5), pp. 95-107, 2013.

[10] H. C. Gibbs, N. Gupta, C. Northcutt, E. Cutrell, and W. Thies, "Deterring Cheating in Online Environments", ACM Trans. Comput.-Hum. Interact. 22, 6, Article 28, 2015.

[11] S. Prathish, A. Narayanan, and K. Bijlani, "An Intelligent System For Online Exam Monitoring", International Conference on Information Science (ICIS), Kochi, pp. 138-143, 2016.

[12] Karim, N. A., and Shukur, Z., "Review of User Authentication Methods in Online Examination", Asian Journal of Information Technology, 14(5), pp. 166-175, 2015.

[13] P. Mahale and N. L. Bhale, "Continuous and Clear User Identity Verification for Secure Web Services: A Survey", International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, no. 11, 2015.

[14] T. Sim, S. Zhang, R. Janakiraman, and S. Kumar, "Continuous Verification Using Multimodal Biometrics", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, 2007.

[15] Harshal A. Kute dan D. N. Rewadkar., "A Survey on Continuous User Identity Verification Using Biometrics Traits for Secure Internet Services", International Journal of Science and Research (IJSR), 2012.

[16] Rajat Kumar Das, Sudipta Mukhopadhyay dan Puranjoy Bhattacharya, "Continuous Multimodal Biometric Authentication for PC and Handheld Devices", IETE Journal of Education, vol. 52 no. 2, 2011.

[17] K. Toh, X. Jiang, and W. Y. Yau, "Exploiting Global and Local Decisions for Multimodal Biometrics Verification", IEEE Transactions On Signal Processing, vol. 52, no. 10, 2004.

[18] T. Sim, S. Zhang, R. Janakiraman, and S. Kumar, "Continuous Verification Using Multimodal Biometrics", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, 2007.

[19] E. Schiavone, A. Ceccarelli, A. Bondavalli and A. M. B. R. Carvalho, "Usability Assessment in a Multi-Biometric Continuous Authentication System", Seventh Latin-American Symposium on Dependable Computing, 2016.

[20] C. Shen, H. Zhang, Z. Yang, and X. Guan, "Modeling Multimodal Biometric Modalities for Continuous User Authentication", IEEE International Conference on Systems, Man, and Cybernetics, 2016.

[21] D. S. Trigueros, L. Meng, and M. Hartnet. "Face Recognition: From Traditional to Deep Learning Methods". arXiv:1811.00116, 2018.

[22] D. Scherer, A. Muller, S. Behnke. "Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition", 20th International Conference on Artificial Neural Networks (ICANN), 2010.

[23] Blessing, L. T., & Chakrabarti, "A. DRM, a Design Research Methodology". Luxembourg and Bangalore: Springer, 2009.

[24] S.Anila and Dr. N. Devarajan, **"**Preprocessing Technique for Face Recognition Applications under Varying Illumination Conditions**".** Global Journal of Computer Science and Technology Graphics & Vision**,** 2012.

[25] K. Cui, H. Cai, Y. Zhang, and H. Chen." A Face Alignment Method Based on SURF Features". International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. 2017.

[26] Y. Zhou, H. Tao, Y. Gong, G. Zhang, and Y. Zhao. "Eye Localization Based on Face Alignment". International Conference on Intelligent Human-Machine Systems and Cybernetics. 2016.

[27] Zhang, Kaipeng et al. "Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks." IEEE Signal Processing Letters 23.10 : 1499–1503, 2016.

[28] Viola P. and Jones M.J. "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137–154, 2004.

[29] Chen W., Huang H., Peng S., Zhou C. and Zhang C. "YOLO-face: a real-time face detector", Springer Nature 2020 on The Visual Computer. 2020.

[30] Schroff F., Kalenichenko D., and Philbin J. "FaceNet: A Unified Embedding for Face Recognition and Clustering", IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015.

[31] Hadian S. G. Asep and Y. Bandung, "A Design of Continuous User Verification for Online Exam Proctoring on M-Learning," International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, Indonesia, 2019, DOI: 10.1109/ICEEI47359.2019.8988786.