TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 3, June 2020, pp. 1389~1396
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i3.14790

Vision: a web service for face recognition using convolutional network

Akino Archilles, Arya Wicaksana
Department of Informatics, Universitas Multimedia Nusantara, Indonesia

Article Info

Article history:

Received Jul 21, 2019
Revised Jan 20, 2020
Accepted Feb 21, 2020

Keywords:

Computer vision
Convolutional neural network
Face recognition
Web service

ABSTRACT

This paper proposes a face recognition module built as a web service. We introduce a novel design and mechanism for face recognition on a web platform and to memorize most recent users for the user. This web service is called Vision and developed using the Flask and TensorFlow deep learning framework. The face recognition process is powered by FaceNet deep convolutional network model. The face recognition process done by Vision could also be utilized for user authentication and user memorization, both done in on a web platform. As a demonstration of concept and viability, in this study, Vision is integrated into a web-based voice chatbot. The testing and evaluation of Vision's face recognition process show an overall F-score of one for all test scenarios.

Corresponding Author:

Arya Wicaksana,
Department of Informatics,
Universitas Multimedia Nusantara,
Scientia Boulevard St., Gading Serpong, Tangerang-15810, Banten, Indonesia.
Email: arya.wicaksana@umn.ac.id

1. INTRODUCTION

A chatbot is a conversational agent where a computer program is designed to conduct a conversation (textual or auditory) [1]. In the development of chatbots, natural language processing and deep learning are the two main technologies of artificial intelligence that allows the advancement [2]. Nowadays, face identification and recognition are also already an established application of computer vision [3-9]. Although its performance is still not as good as the human eyes [10], face identification has already been used widely as a non-intrusive biometric technique [11]. This is due to its convenient nature for authenticating users without requiring any physical contact with the device [12]. In this study, we propose a novel way of doing face recognition built as a web service. The availability of a face recognition module as a web service would benefit many websites that would want to explore this technology. Web services allow the application to be platform and technology independent. In addition to that, the processing of the face image would also increase user experience. Thus, as proof of concept, the web service is developed and called Vision, and it is designed and implemented on Jacob.

Jacob [13] is a web-based voice chatbot that is powered by Wit.AI and programmed to provide information about Universitas Multimedia Nusantara joint-degree Informatics program information. Jacob works with sound (audio) input and translates it into text using the web speech API. It is then sent to Wit.AI to obtain the intent (the goal of the user is coming to the chatbot) and entities (important variable in intent that helps add relevance to an intent) of the text. The obtained intent and entities are checked and compared with the knowledge database to return with the appropriate response. The response which is in text format is translated into a voice using Speech Synthesis of the web speech API. Jacob recognizes three user roles: administrator, super administrator, and basic user. The role of an administrator and super administrator is to update the knowledge database with the latest information to ensure that the content is up-to-date. The authentication process is done using a password [13]. The addition of a face recognition module (Vision) to Jacob would enhance is interactivity and enrich its capabilities.

The main algorithm used for Vision is the deep convolutional network. As explained in [14, 15], convolutional neural network (CNN) performs better than multilayer perceptron and more robust in complex pattern recognition such as distortion, translation, scaling, and rotation. Simple CNN explained in [16] consists of three main layers and could be optimized according to needs. As there are various CNN architectures, the one that is utilized by Vision is FaceNet. FaceNet is a system that directly learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity [17]. FaceNet [17] achieved 99.63% accuracy on the Labeled Faces in the Wild database and 95.12% on YouTube Faces, which outperforms some other CNN models such as DeepID, DeepID2, and GaussianFace [18]. Thus, FaceNet is chosen as the core technology for the face recognition process in Vision. The face recognition of Jacob users is performed between 1 and 2 meters [19] and the lighting conditions are taken into account with intensity between 250 and 400 lumens/m^2 [20].

The implementation of a face recognition module (Vision) into a voice chatbot (Jacob) would allow the expansion of its features, in this work, such as the authentication process and user memorization feature. The authentication process for Jacob users is carried on using face recognition done as a background process. The user memorization feature allows Jacob to greet returning users, and develop a connection between the chatbot and the users. The authentication process to login into the system is designed to have a false positive rate less than or equal to 0.001% and false-negative rate less than or equal to 1% [21]. The performance evaluation of Vision would be measured according to [21], where F-measure is used to measure the F-score [22] of the authentication (face recognition) process for administrator and super administrator user roles.

2. RESEARCH METHOD

This section consists of requirement analysis, design, and implementation of Vision. The next section discusses the result and analysis of Vision, including the test scenarios and performance evaluation of the face recognition and authentication process.

2.1. Requirements analysis

Based on the preliminary studies on Jacob voice chatbot, Vision takes inputs of user details such as name, role, and face images. The face images are used for training the model (classifier). These face images are captured by Vision in the background. The number of training images is set to 20, 50, and 100 for testing purposes. Jacob authenticates the administrator role and the super administrator role by using the password during login. Thus, by using Vision, a registered administrator and super administrator could be authenticated using face recognition by Jacob. This new authentication process is proposed to replace the old-fashioned password-based user authentication process. Upon recognizing the face of a basic user, Vision memorizes by keeping the face images into the filesystem. This allows the memorization feature on Jacob as Vision could memorize up to 10 most recent users.

Considering the training process of the CNN model could take a lot of time, thus for the face recognition process, Vision uses the pre-trained model. Furthermore, instead of re-training the whole network, Vision only re-trains the classifier. The pre-trained model is trained using the VGGFace2 dataset which is based on the Inception-ResNet-v1 model used in the FaceNet as the classifier [23, 24]. Pre-processing step is done by using FaceNet's Multi-task Cascaded Convolutional Networks (MTCNN) to detect and align faces [25]. The classifier is trained using Support Vector Machine (SVM) as in [24].

For the memorization feature, as the number of basic users increases, the training time would also increase accordingly. This is where the limitation of n most recent users comes to exist. The Least Recently Used algorithm [26, 27] is used for this feature to memorize only n most recent users. The algorithm works by replacing the oldest memory of the user's face with the most recent one. The limit n is set to 10 persons for demonstration purposes.

The integration part between the Vision web service with Jacob web app is designed and implemented using a client-server architecture. The CNN engine is run on the server-side, thus ensuring the face recognition process to have a reliable performance with adequate processing power. It is then connected with the application programming interface (API) that is built using Flask 1.0.2 framework. On the other hand, Jacob's user interface runs on the client-side as well as the Vision sub-module that works for capturing face images.

2.2. Designs

The design of the web service model for Vision is presented in this section. The web service model consist of three resources: AlignTrain, IdentifyFace, and RegisterFace with two resource paths: IdentifyFace and RegisterFace using the same resource path. IdentifyFace and RegisterFace are separated in the API by the parameter received. AlignTrain resource is used to pre-process the registered training images and to train the classifier in background. The registered users could only be identified after the classifier has been training completely. All of the resources could be accessed through the given resource path using the HTTP protocol. The return value of the API is in JSON data format. Before starting Vision as a web service, there's a need to check the filesystem for duplication and unsuccessful registrations to be removed. A thread process is then created just before the Vision is started and it is scheduled to run every two hours (non-blocking). Vision web service model as shown in Figure 1.

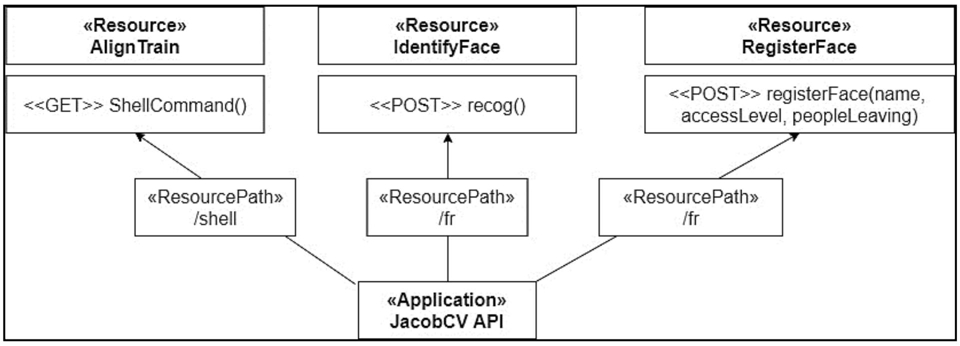

Figure 1. Vision web service model

In the registration process, the request is categorized as RegisterFace if the parameter contains: image, name, access level, and people leaving. It returns with code, message, number of images written, and number of images must be taken. If the registration process is interrupted, the registration process is cancelled automatically. Upon a successful registration, the AlignTrain resource is triggered to run to align the images and train the classifier. This process is carried out in background. All users have to pass through the registration process to be successfully identified by Vision. Roles such as administrator and super administrator are required to login into Jacob upon accessing the administration settings. In the identification process, request will be categorized as RegisterFace if the parameter contains only images. It returns with: code, message, name, access level and confidence level. Vision uses two thresholds: 0.9 and 0.8 for the confidence level. Confidence level below 0.8 is considered not recognize by Vision.

2.3. Implementation

Vision UI is placed in the center of Jacob homepage with a camera interface. The process of capturing image makes the camera interface blinks as feedback to the users. Login to the administrator page could take place at this step. After the image is captured, a request is sent to the server with the image encoded in base64 format. The image is then converted back to binary and stored in the filesystem. Vision processes the image by reading it from the filesystem, converting it to RGB, and reading the number of detected faces. Vision will only continue to process if there is only single face detected, Vision does not support identification for multiple faces. The image is then flipped and cropped to extract the features from it. The prewhiten feature extraction is used in this step. The result is then returned to the client.

If the result shows that a person is identified successfully, Jacob greets the users. The users could enter the registration process when Vision could not identify the face for two consecutive times. The users have to register 20 images into Vision. The information required in the user registration process is the name of the users. The registration process starts immediately upon receiving the name. In the event of failed authentication for administrators, login could be done by requesting face verification to Jacob. This replaces the old-fashioned password-based login. The registration page consist of basic requirements such as name, and role. Administrator has 100 face images captured for the registration process.

2.4. Testing

Testing is done under three test scenarios to show that the face identification process works with 13 users: 2 super administrator roles, 1 administrator role, and 10 basic user roles. The sample face images of the 13 users are displayed in Figure 2. The first test scenario is registering a basic user in Vision. The testing variables are shown in Table 1 and sample of training image shown in Figure 3. After the training process is done, the User3's face image as shown in Figure 4 is captured and sent to the server for identification. The identification result shows that the captured images of User3 gives a confidence level of 0.9371. This result passed the high threshold of 0.9 and Vision recognizes the person as User3. The second test scenario is unregistered access done by a non-registered user. The testing variables are given in Table 2 and the sample of the train face images of the registered users are shown in Figure 5. Figure 6 presents the face image sample of the non-registered user. Here the identification result shows that the non-registered user is not recognize by confidence level of 0.3281. Since the non-registered user is not recognize, so the user is registered automatically by the Vision when the conversation between the user and Jacob take place. After the registration process is finished, this previously non-registered user is now become a basic user and the face images are saved in the filesystem as shown in Figure 7.

Table 1. First test scenario variables

Testing Variables	Condition
Number of testing images	100
Face image input	User3

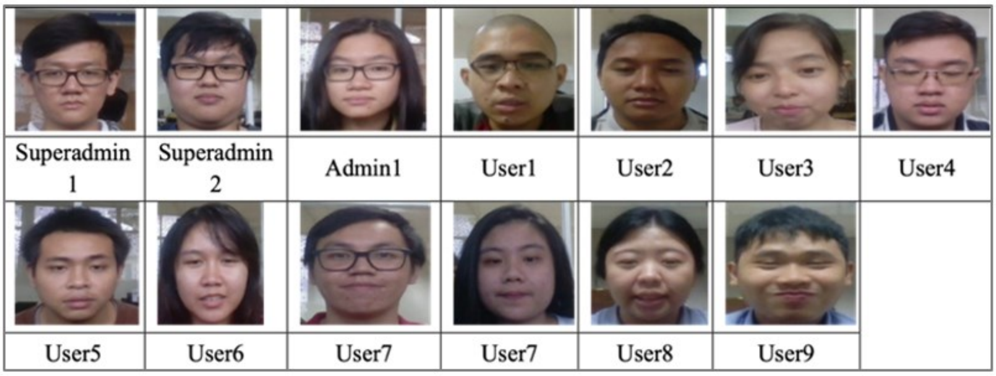

Figure 2. Registered users

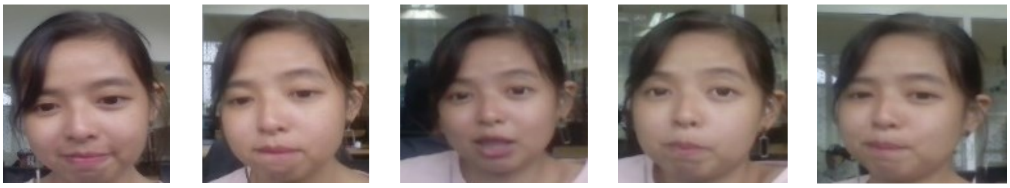

Figure 3. Face images sample of User3 for training

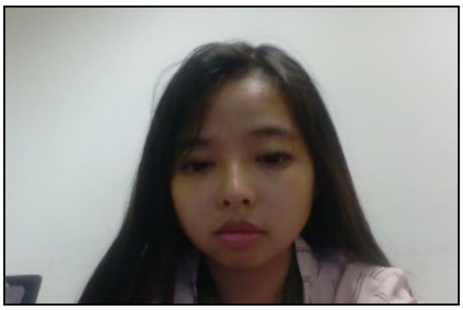

Figure 4. Face image sample of User3 for identification

Table 2. Second test scenario variables

Testing Variables	Condition
User status	Non-registered user
Input image	Non-registered user
Number of registered users	3 persons
Number of super administrator	1 person
Maximum basic users	3 persons
Number of registered basic users	3 persons
Least recently used algorithm	Running

(User4) (User5) (User 6)

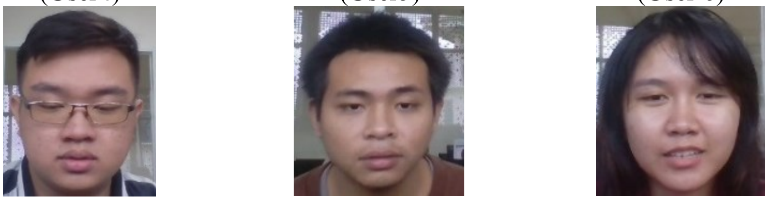

Figure 5. Face image sample of registered users

Figure 6. Face image sample of a non-registered user

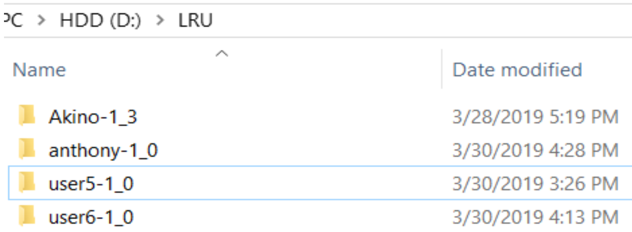

Figure 7. List of registered users

The third test scenario is a Superadministrator2 that has been registered in Vision uses Jacob. The testing variables shown in Table 3, train image sample of super administrator and administrator users shown in Figure 8, and captured image of superadministrator2 shown in Figure 9. Identification result shows that captured image of Superadministrator 2 achieve confidence result of 0.9287 which passed the high threshold of 0.9 and identified as Superadministrator2.

Table 3. Third test scenario variables

Testing Variables	Condition
User status	Superadministrator2 (registered)
Image input	Superadministrator2
Number of registered users	13 persons
Number of super administrator and administrator	3 persons
Maximum basic users	10 persons
Number of basic users	10 persons

(Superadministrator1) (Superadministrator2) (Administator1)

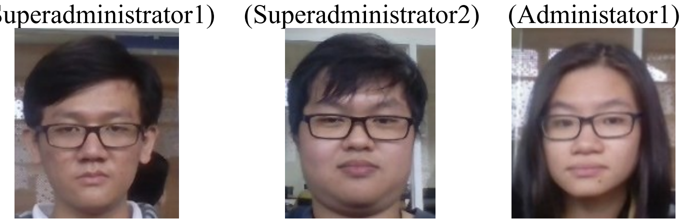

Figure 8. New list of registered users

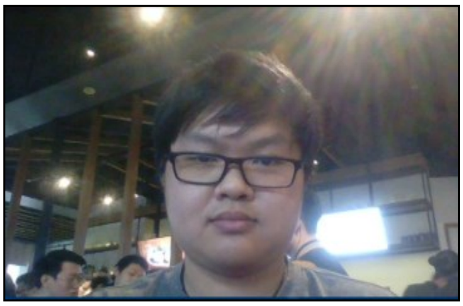

Figure 9. Face image sample of Superadministrator2

3. RESULTS AND ANALYSIS

The result of identification is recorded and evaluated using F-score. The F-score evaluations uses samples of 20, 50, and 100 training images for comparison. The evaluation refers to two types of threshold and it is evaluated from confidence level of 0.8 to 1. The testing is done 15 times for each number of training image per administrator. Table 4 shows the test result of Superadministrator1 named Steven. Based on the test results in Table 4, the F-score for Superadministrator1 is one. Table 5 shows the test result of Superadministrator2 named Akino. Based on the test results in Table 5 the F-score for Superadministrator2 is one. Table 6 shows the test result of Administrator1 named Octa. Based on the test result in Table 6, the F-score for Administrator1 is one.

Table 4. Test result of superadministrator1

Test No.	Steven / Superadministrator1					
	20 training images		50 training images		100 training images	
	Result	Confidence	Result	Confidence	Result	Confidence
1	Fail	0.4831256707306246	Steven	0.8486887517390871	Steven	0.8622850226792862
2	Fail	0.3968983503374864	Fail	0.7912751960836923	Steven	0.8963097771048575
3	Fail	0.3179188825009534	Fail	0.6719356101835616	Steven	0.9037249256249601
4	Fail	0.3293581751071452	Steven	0.8639671038567101	Steven	0.8915614756108579
5	Fail	0.2915295482988687	Steven	0.8393610777619160	Steven	0.9189174449524873
6	Fail	0.4780330066048949	Steven	0.8590177719030167	Steven	0.9025752084515857
7	Fail	0.6974929480961017	Fail	0.7684912796957375	Steven	0.902824734897516
8	Fail	0.5121738310615238	Fail	0.7839105555391056	Steven	0.8729555917501845
9	Fail	0.3891884170737012	Steven	0.8028572961083176	Steven	0.8750267206717658
10	Fail	0.3088813885614176	Steven	0.859101863673292	Steven	0.8539075108751068
11	Fail	0.3950073892569106	Steven	0.8201367392671069	Steven	0.9347298547209619
12	Fail	0.5725319619748922	Steven	0.8491761038671391	Steven	0.8986286678563803
13	Fail	0.3729178738557185	Fail	0.7892751613960386	Steven	0.8931622371331799
14	Fail	0.5294899956082455	Steven	0.8693717639617396	Steven	0.8609438571719787
15	Fail	0.6741913922990106	Steven	0.8193716666671937	Steven	0.8821074197564109

Table 5. Test result of superadministrator2

Test No.	Akino / Superadministrator2					
	20 training images		50 training images		100 training images	
	Result	Confidence	Result	Confidence	Result	Confidence
1	Fail	0.6661146706715512	Akino	0.8863968622124053	Akino	0.9206484151507499
2	Fail	0.6778212481999762	Akino	0.8609945286854576	Akino	0.9178415651142144
3	Fail	0.5895838655488481	Fail	0.779757496477541	Akino	0.9038368187576951
4	Fail	0.4613156360024367	Akino	0.8807935552948459	Akino	0.9241616389217459
5	Fail	0.6045864703957678	Akino	0.8708410150050343	Akino	0.9120598170852047
6	Fail	0.5672349222067348	Akino	0.8626607619631554	Akino	0.9206716336519734
7	Fail	0.5589413011192211	Fail	0.7642900203205136	Akino	0.9343422888699267
8	Fail	0.5848885179716995	Fail	0.769700825844427	Akino	0.93998698474758
9	Fail	0.5244985230576841	Akino	0.8063666092347131	Akino	0.9489666968462456
10	Fail	0.6316123637696031	Akino	0.821100171232869	Akino	0.9032190915032803
11	Fail	0.6868446788660262	Fail	0.7465601333419117	Akino	0.9251503592486126
12	Fail	0.6097396524445506	Akino	0.8791109825362003	Akino	0.924164182810119
13	Fail	0.6538067013367	Akino	0.8593193087326183	Akino	0.9261999943154746
14	Fail	0.6598772619444839	Akino	0.8737677891362153	Akino	0.9182874201435122
15	Fail	0.6331449059470585	Fail	0.7364376994226135	Akino	0.9063418785474986

Table 6. Test result of administrator1

Test No.	Octa / Administrator1					
	20 training images		50 training images		100 training images	
	Result	Confidence	Result	Confidence	Result	Confidence
1	Fail	0.4916995410624423	Fail	0.6861709553650175	Octa	0.9032324016883357
2	Fail	0.4709441801770915	Fail	0.7752983754925702	Octa	0.8599574309961089
3	Fail	0.3154287038263585	Fail	0.7890147588808165	Octa	0.9092814684895162
4	Fail	0.4299015659472068	Octa	0.8597819086345175	Octa	0.8950211711769138
5	Fail	0.4018626540552488	Octa	0.8147589534720676	Octa	0.8911998076413719
6	Fail	0.388377999272349	Fail	0.7499175489710957	Octa	0.8642403121384548
7	Fail	0.4605011443261597	Fail	0.6935552910483629	Octa	0.8460847480790611
8	Fail	0.4079790608866885	Octa	0.8507625601745617	Octa	0.885948001506788
9	Fail	0.5120193717264274	Fail	0.7826302746185624	Octa	0.8575463693227686
10	Fail	0.4017370697546190	Fail	0.7819305176010607	Octa	0.9150287792570698
11	Fail	0.4150443852165001	Fail	0.7498707777169365	Octa	0.9087485643158923
12	Fail	0.4158483228853045	Octa	0.8650198573562094	Octa	0.9078024077700945
13	Fail	0.3041688954042459	Octa	0.8689157043610945	Octa	0.9039452230759157
14	Fail	0.3316467842354488	Octa	0.8109856266988097	Octa	0.8363597838864007
15	Fail	0.6101372636222694	Fail	0.7947365108563923	Octa	0.8785901336080615

4. CONCLUSION

The proposed face recognition mechanism and web service (Vision) have been successfully implemented and tested with the use of a web-based voice chatbot (Jacob). This proves that the deep convolutional network could be used for face recognition for a real-world application such as the Jacob voice chatbot. The performance evaluation delivers overall F-score for Superadministrator1 and Superadministrator2, and Admininistrator1 of one. Based on the testing process, it is shown that 100 training images give a better success rate rather than the 20 and 50 training images. The number of images used for training affects the recognition confidence rate.

REFERENCES

[1] R. P. Mahapatra, et al., "Adding interactive interface to E-Government systems using AIML based chatterbots," *2012 CSI Sixth International Conference on Software Engineering (CONSEG)*, pp. 1-6, 2012.
[2] H. N. Io and C. B. Lee, “Chatbots and conversational agents: A bibliometric analysis,” in *Proc. IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 215-219, 2017.
[3] V. Wiley and T. Lucas, “Computer Vision and Image Processing: A Paper Review,” *International Journal of Artificial Intelligence Research*, vol. 2, no.1, pp. 29-36, June 2018.
[4] E.G. Learned-Miller, “Introduction to Computer Vision,” (2011, Jan.), University of Massachusetts. Accessed on January 20, 2019. [Online]. Available: https://people.cs.umass.edu/~elm/Teaching/Docs/IntroCV_1_19_11.pdf
[5] O. H. Babatunde, et al., “A Computer-Based Vision Systems for Automatic Identification of Plant Species Using Knn and Genetic PCA,” *Journal of Agricultural Informatics,* vol. 6, no. 1, pp. 61-71, 2015.
[6] S. Matiacevich, et al., “Quality Parameters of Six Cultivars of Blueberry Using Computer Vision,” *International Journal of Food Science*, 2013. doi:10.1155/2013/419535
[7] S. Ullman, et al., “Atoms of Recognition in Human and Computer Vision,” in *Proceedings of the National Academy of Sciences*, vol. 113, no. 10, pp. 2744-2749, 2016.
[8] V. Kurniawan, A. Wicaksana, and M. I. Prasetiyowati, “The implementation of eigenface algorithm for face recognition in attendance system,” in *International Conference on New Media Studies*, pp. 118-124, Nov. 2017.
[9] K. Alexander, A. Wicaksana, and N. M. S. Iswari, “Labeling Algorithm and Fully Connected Neural Network for Automated Number Plate Recognition System” in *Studies in Computation Intelligence*, vol. 847, pp 129-145, August 2019.
[10] F. Zhao, X. Xie and M. Roach, “Computer Vision Techniques for Trancatheter Intervention,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 3, pp. 1-31, June 2015.
[11] A. Majumdar and R.K. Ward, “Pseudo-Fisherface Method for Single Image Per Person Face Recognition,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 989-992, 2008.
[12] C. Zhen and Y. Su, “Reseach about human face recognition technology,” in *International Conference on Test and Measurement*, pp. 420-422, 2009.
[13] S. Wijaya and A. Wicaksana, “Jacob Voice Chatbot Application Using Wit.ai for Providing Information in UMN,” *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6S3, pp. 105-109. 2019.
[14] Y. Lecun, et al., “Gradient-Based Learning Applied to Document Recognition,” in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, November 1998.
[15] S. Lawrence, et al., “Face recognition: a convolutional neural-network approach,” *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98-113, January 1997.
[16] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” arXiv:1511.08458v2 [cs.NE], December 2015.
[17] F. Schroff, D. Kalenichenko and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823, 2015.

[18] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823, 2015.
[19] H. M. Moon and S. B. Pan, "The PCA-Based Long Distance Face Recognition using Multiple Distance Training Images for Intelligent Surveillance System," in *1st International Conference on Information and Communication Technology*, pp. 534-539, 2013.
[20] The National Optical Astronomy Observatory, "Recommended Light Levels (Illuminance) for Outdoor and Indoor Venues," America, 2019. [Online]. Available: https://www.noao.edu/education/QLTkit/ACTIVITY_Documents/Safety/LightLevels_outdoor+indoor.pdf. Accessed: February 3, 2019.
[21] P. Charles and S. L. Pfleeger, "Analyzing Computer Security: A Threat/vulnerability/countermeasure Approach," 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2012.
[22] C. D. Manning and H. Schutze, "Foundations of Statistical Natural Language Processing," 2nd ed. London, The MIT Press. 1999.
[23] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in *AAAI Conference on Artificial Intelligence*, February 2017.
[24] D. Sandberg. "Face Recognition using Tensorflow," GitHub. Accessed on 23 January 2018. [Online]. Available: https://github.com/davidsandberg/facenet
[25] K. Zhang, et al., "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters Society*, vol. 23, no. 10, pp. 1499-1503, October 2016.
[26] A. Khosrozadeh, et al., "Presenting a Novel Page Replacement Algorithm Based on LRU," *Journal of Basic and Applied Scientific Research*, vol. 2, no. 10, pp. 10377-10383, 2012.
[27] E. J. O'Neil, P. E. O'Neil and G. Weikum, "An optimality proof of the LRU-K page replacement algorithm," *Journal of the ACM*, vol. 46, no. 1, pp. 92-112, January 1999.

BIOGRAPHIES OF AUTHORS

Akino Archilles is graduated from Universitas Multimedia Nusantara as Sarjana Komputer (S. Kom.). A new researcher enthused in neural networks. Currently working as a software engineer and AI researcher in one of the biggest Indonesia retail company.

Arya Wicaksana is graduated from Universiti Tunku Abdul Rahman in VLSI Engineering (M.Eng.Sc.) and Universitas Multimedia Nusantara in Computer Science (S. Kom.). Research interests and works are: quantum computing and computational intelligence.