A Transfer Learning Approach for Face Recognition using Average Pooling and MobileNetV2

Sovon Chakraborty
European University of Bangladesh

M M Imran
Jahangirnagar University

Abdulla Not Applicable
Daffodil International University

Ankit Khater
Crystal Technology Bangladesh Ltd

A Transfer Learning Approach for Face Recognition using Average Pooling and MobileNetV2

F.M. Javed Mehedi Shamrat[1], Sovon Chakraborty [2], M. M. Imran[3], Abdulla[4], Ishtiak Ahmed[5], *Ankit Khater[6]

[1]Dept. of Software Engineering, Daffodil International University, Bangladesh
[2]Dept. of Computer Science and Engineering, European University of Bangladesh
[3]Cefalo Bangladesh Limited
[4]Dept. of Computer Science and Engineering, Daffodil International University
[5]Dept. of Computer Science and Engineering, North South University
[6]Dept. of Computer Science and Engineering, Jadavpur University, India

[1]javedmehedicom@gmail.com
[2]sovonchakraborty2014@gmail.com
[3]imranpollob.iitju@gmail.com
[4]abdulla15-11549@diu.edu.bd
[5]ishtiak.ahmed@northsouth.edu
[6]ankitkhaterak@gmail.com

Abstract. Facial recognition is a fundamental method in facial-related science such as face detection, authentication, and monitoring, as well as a crucial phase in computer vision and pattern recognition. Face recognition technology aids in crime prevention by storing the captured image in a database, which can then be used in a variety of ways, including identifying a person. With just a few faces in the frame, most facial recognition systems function sufficiently when the techniques have been tested under artificial illumination, with accurate facial poses and non-blurry images. in our proposed system, a face recognition system is proposed using Average pooling and MobileNetV2. The classifiers are implemented after a set of preprocessing steps on the retrieved image data. To compare the model is more effective, a performance test on the result is performed. It is observed from the study that MobileNetV2 triumphs over Average pooling with an accuracy rate of 98.89% and 99.01% on training and test data respectively.

Keywords: Face Recognition; CNN; Average pooling; MobileNetV2, Accuracy, Performance Comparison.

1 Introduction

The human face is a powerful factor that influences our daily social interactions, such as projecting an individual's personality. Face recognition is a biometric technology that uses mathematics to analyze facial features and then save them as a faceprint to

identify people. Due to a wide range of applications in law enforcement and other civilian industries, institutes, and organizations, biometric facial recognition technology has sparked a lot of interest in recent years. Facial recognition technology has an advantage over other biometric systems such as fingerprint, palmprint, and iris recognition since it is non-contact. For human face detection, Machine learning algorithms and Neural network models are widely used. Face detection using conventional image feature extraction algorithms is accurate and fast. Ma *et al.* [5] introduced an AdaBoost-based training method to achieve cascade classifiers of different function forms: Haar-like HOG for greater discrimination. Since there are too many bad classifiers, this requires a lot of computing power. A Bayesian framework-based algorithm [6] used the Omega form generated by a person's head and shoulder for head localization to resolve severe face occlusion. It detects faces with intense occlusion well in Automatic Teller Machines, but the scene is thin. In addition to AdaBoost-based methods, Mathias *et al.* [7] proposed face recognition utilizing deformable component models (DPM) and obtained positive results. However, the computational cost of this method is usually very large. Another technique based on DPM is proposed for identifying faces of occlusion [8]. Although only facial recognition representations are used in the tests, it has a low uniqueness and can minimize false-negative face recognition and identity error rates.
In the proposed system, face recognition is done using two neural network algorithms i.e. Average pooling and MobileNetV2. After a series of preprocessing steps on the retrieved image data, the classifiers are implemented. A performance evaluation of the outcome is done to compare which model is more efficient.

2 Research Methodology

The proposed study is based on two models for human face recognition. The average pooling and MobileNetV2 models are implemented in the image data. The whole suggested system diagram is shown in Fig. 1.

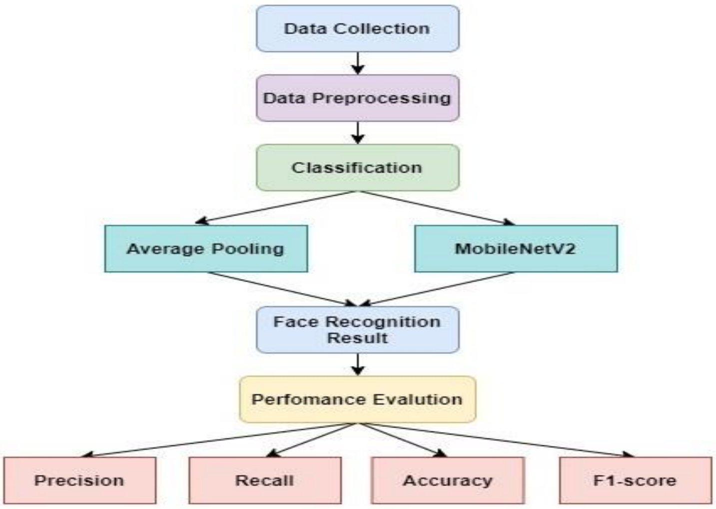

Fig. 1. System diagram of the process

2.1 Data Collection

This study makes use of the LFW (Face Recognition) dataset (https://www.kaggle.com/atulanandjha/lfwpeople). While the dataset includes over 13000 images, we only used 13000 for this analysis. Each image is appropriately labeled with the person's name. A total of 104 photographs were used to build the data set. Figure 2 contains the sample data from the retrieved dataset.

Fig. 2. Sample image data.

2.2 Preprocessing and augmentation of Data :

CNN performs better with a large volume of data. We've allowed zooming, sharing, and scaling in the ImageDataGenerator function. The photographs were first converted to 256 X 256 pixels. The ImageDataGenerator tool was used to expand the size of our current dataset. Numpy Arrays contain all the image data. We used 80% of our data along with size random state 42 for training purposes. We used the to_categorical function to binarize the labels when we have more than two outputs. We used the ImageDataGenerator feature from Keras to supplement images for better training. The following criteria are used to create our images:

2.3 Proposed Classifier:

1) ***Average Pooling:***
   To initiate, images are converted to 256 X 256 pixels and passed to the proposed model's first convolutional layer. There are a total of 128 hidden layers. After running the average pooling method, we converted all of the images to 128 X 128 pixels. The feature is then extracted by the second convolutional layer, which employs average pooling once more. The images are resized to 32 X 32 pixels in the final layer. I mages at this stage are converted into NumPy arrays to make measurements easier. Applying a connected layer is the final stage. In both convolution layers, we used the Relu activation feature, and in the output layer, we used the Softmax activation function. For finding the best results, Adam stochastic gradient descent is used. The following is the suggested algorithm 1:

Algorithm 1: Proposed Average Pooling

Step 1: Load Dataset
Step 2: Function Conv2D (matrix = 256 x 256){
Step 3: Activate RELU}
Step 4: Function AveragePooling2D (data, pool size)
Step 5: Function Conv2D (matrix = 128 x 128, padding){
Step 6: Activate RELU}
Step 7: Function AveragePooling2D (data, pool size)
Step 8: Function Conv2D (matrix = 64 x 64){
Step 9: Activate RELU)}
Step 10:Function AveragePooling2D (data, pool size)
Step 11: Function Conv2D (matrix = 32 x 32){
Step 12: Activate RELU)}
Step 13:Function AveragePooling2D (data, pool size)
Step 14: Reshape image, set list -> Flatten
Step 15: Activate Softmax
Step 16: Output Data Classification

2) MobileNetV2:

MobileNetV2 employs 28 Deep convolutional neural network layers. It is built on an inverted residual structure, with residual connections between bottleneck layers. For MobileNet Streamlined architecture, lightweight deep convolutional neural networks are built using depth-wise separable convolutions. MobileNet is a cost-effective model for mobile and embedded vision applications. The original completely convolution layer with 32 filters is followed by 19 residual bottleneck layers in the MobileNetV2 design. MobileNets are low-latency, low-power models that have been parameterized to satisfy the resource constraints of various use cases.

2.4 Evaluating performance using performance matrix:

After the training and testing process, we evaluated the performance using precision, recall, f1-score, and accuracy. Equations 1,2,3, and 4 are the formulas we used:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP + TN + FP + FN}} \tag{1}$$

$$\text{Sensitivity or recall} = \frac{\text{TP}}{\text{TP+ FN}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+ FP}} \tag{3}$$

$$\text{F1} - \text{score} = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

3 Results and Discussion

3.1 Performance of Average Pooling

The model will recognize a specific person's face based on their name. The dataset contains 13000 photos, including images from 1680 individuals. Training and research testing have been divided into 80/20 fractions. For training results, the model can detect images with 93.13% accuracy and validation accuracy reaches 93.65% at the top stage at 10 Epoch. The data loss during validation is a minimum of 6.89%. Table I shows the results of our training and validation sets.

Table I. Averager Pooling Classification Result of training and validation dataset.

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	47.34%	81.93%	19.74%	86.87%
2	15.13%	85.25%	15.57%	87.18%
3	10.25%	88.02%	11.73%	87.87%
4	10.02%	88.89%	9.28%	88.49%
5	9.47%	88.37%	9.54%	88.58%

6	9.29%	89.74%	8.90%	89.43%
7	9..21%	89.37%	8.58%	90.00%
8	7.78%	90.37%	7.36%	91.57%
9	7.25%	92.54%	7.28%	92.48.%
10	6.94%	93.13%	6.89%	93.65%

The model attained 91.78 percent training accuracy and 92.17 percent validation accuracy on the overall classification. After the evaluation of the dataset has been done, the performance evaluation is shown. In Table II, the performance measurements are stated.

Table II. Performance evaluation on average pooling.

Class	Precision	Recall	F1 - Score	Accuracy
Training Set	89.54	90.57	90.10%	91.78
Testing Set	90.89	91.35	90.85%	92.17

The classification of the model based on each class is done. The performance of the classification of the classes is described in Table III.

Table III: Performance evaluation of classes on average pooling.

Class	**Precision**	**Recall**	**F1 - Score**	**Accuracy**
Shamrat	92.74	90.59	91.78	93.53
Shongkho	92.05	91.56	90.11	92.89
Masum	93.35	90.29	91.84	89.47
Jubair	91.73	90.39	90.09	90.23

3.2 Performance of MobileNetV2:

The model can detect images with 98.92% accuracy in the training data, and validation accuracy reached 99.54% at 10 Epoch. A minimum of 3.59% of data is lost during validation. The outcomes of our training and validation sets are seen in Table IV.

Tabel IV: Accuray of MobileNetV2 on dataset

Epoch	**Training Loss**	**Training Accuracy**	**Validation Loss**	**Validation Accuracy**
1	6.47	96.46	5.84	97.05
2	5.83	96.84	5.25	97.67
3	5.73	96.92	5.94	97.85
4	5.49	97.02	5.67	97.58
5	3.87	98.59	5.37	97.04
6	4.49	97.99	4.39	98.49
7	3.78	98.73	3.93	98.64
8	3.57	98.79	3.68	98.95
9	3.98	98.62	3.73	99.10
10	3.76	98.92	3.59	99.54

The overall performance calculation of the MobileNetv2 model on the dataset is illustrated in table V and the performance based on class is shown in Table VI.

TABLE II: Performance of MobileNetV2.

Class	Precision	Recall	F1 - Score	Accuracy
Training Set	98.75	98.73%	98.28%	98.89
Testing Set	99.20	98.69%	98.74%	99.01

TABLE VI: Perfomance of each class in MobileNetV2

Class	Precision	Recall	F1 - Score	Accuracy
Shamrat	98.47	97.75	98.26	98.48
Shongkho	98.55	97.76	98.46	98.86
Masum	98.18	97.56	98.45	97.99
Jubair	98.54	97.82	97.28	98.43

3.3 Comparison of Models' classifications

In figure 3 shows the accuracy comparison graph of both the train and test set of the MobileNetV2 model and CNN average Pooling model. From the graph, it can be observed that the MobileNetV2 has a way more accuracy rate compared to the other model. In figure 4, the final output of the implemented MobileNetV2 model is given.

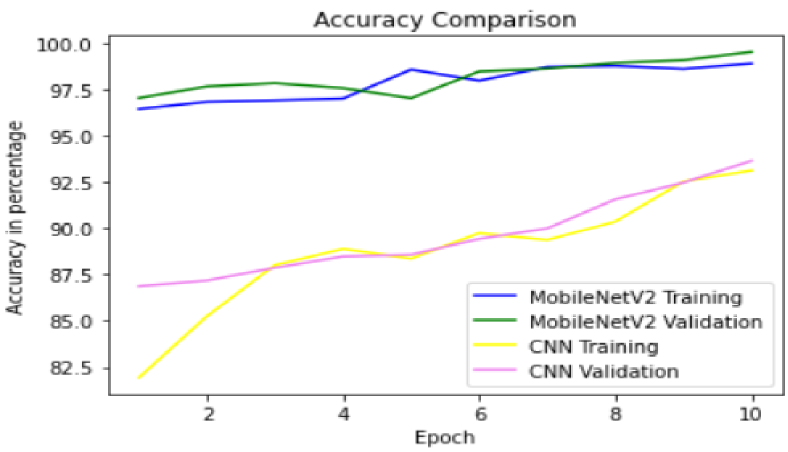

Fig. 3. Accuracy comparison of the models

Fig. 4. Face Recognition using MobileNetV2

4 Conclusion

Face recognition is a method of recognizing or verifying a person's identity by using his or her face. Face recognition has been utilized for a variety of uses, including an auto-management attendance control system, surveillance of restricted access areas. From the study, it is understood that the MobileNetV2 has a high accuracy rate compared to CNN average pooling. The model has an accuracy rate of 98.89% on training data and 99.01% on validation data. Furthermore, in the case of face recognition based on class, the model shows the highest accuracy up to 98.86%. in the future, we hope to implement more CNN models for face recognition to come up with the most accurate system.

References

1. T. Zhang, J. Li, W. Jia, J. Sun, and H. Yang, ''Fast and robust occluded face detection in ATM surveillance,'' Pattern Recognit. Lett., vol. 107, pp. 33–40, May 2018.
2. M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool, ''Face detection without bells and whistles,'' in Proc. Eur. Conf. Comput. Vis. Springer, 2014, pp. 720–735.
3. D. Marcetic and S. Ribaric, ''Deformable part-based robust face detection under occlusion by using face decomposition into face components,'' in Proc. 39th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO), May 2016, pp. 1365–1370.
4. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, Vanhoucke, and A. Rabinovich, ``Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1-9.

5. X. Li, Z. Yang and H. Wu, "Face Detection Based on Receptive Field Enhanced Multi-Task Cascaded Convolutional Neural Networks," in IEEE Access, vol. 8, pp. 174922-174930, 2020, doi: 10.1109/ACCESS.2020.3023782.
6. G. B. Huang, M. Mattar, T. Berg, and E. Learned–Miller, ‘‘Labeled faces in the wild: A database forstudying face recognition in unconstrained environments,’’ in Proc. Workshop Faces ‘Real-Life’ Images, Detection, Alignment, Recognit., Oct. 2008, pp. 1–11.
7. Y. Sun, D. Liang, X. Wang, and X. Tang, ‘‘DeepID3: Face recognition with very deep neural networks,’’ Feb. 2015, arXiv:1502.0087. [Online]. Available: https://arxiv.org/abs/1502.00873.
8. F. M. Javed Mehedi Shamrat, Z. Tasnim, P. Ghosh, A. Majumder and M. Z. Hasan, "Personalization of Job Circular Announcement to Applicants Using Decision Tree Classification Algorithm," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-5, doi: 10.1109/INOCON50539.2020.9298253.
9. S. Manlangit, “Novel Machine Learning Approach for Analyzing Anonymous Credit Card Fraud Patterns,” International Journal of Electronic Commerce Studies, vol. 10, no. 2, 2019.
10. F. M. Javed Mehedi Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298026.
11. P. Ghosh, F. M. Javed Mehedi Shamrat, S. Shultana, S. Afrin, A. A. Anjum and A. A. Khan, "Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm," 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Bangkok, Thailand, 2020, pp. 1-6, doi: 10.1109/iSAI-NLP51646.2020.9376787.
12. K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam, and D. Mathur, “Machine Learning Based PV Power Generation Forecasting in Alice Springs,” IEEE Access, pp. 1–1, 2021.
13. F.M. Javed Mehedi Shamrat, Md. Asaduzzaman, A.K.M. Sazzadur Rahman, Raja Tariqul Hasan Tusher, Zarrin Tasnim “A Comparative Analysis of Parkinson Disease Prediction Using Machine Learning Approaches” International Journal of Scientific & Technology Research, Volume 8, Issue 11, November 2019, ISSN: 2277-8616, pp: 2576-2580.
14. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, ‘‘DeepFace: Closing the gap to human-level performance in face verification,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1701–1708.
15. Y. Sun, X. Wang, and X. Tang, ‘‘Deep learning face representation from predicting 10,000 classes,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1891–1898.
16. F. M. Javed Mehedi Shamrat, Md. Abu Raihan, A.K.M. Sazzadur Rahman, Imran Mahmud, Rozina Akter, “An Analysis on Breast Disease Prediction Using Machine Learning Approaches” International Journal of Scientific & Technology Research, Volume 9, Issue 02, February 2020, ISSN: 2277-8616, pp: 2450-2455.
17. P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in IEEE Access, vol. 9, pp. 19304-19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
18. A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain “A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms” International Journal of Scientific & Technology Research, Volume 8, Issue 11, November 2019, ISSN: 2277-8616, pp: 419-422.

19. F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Imran Mahmud, Ms. Nusrat Jahan, Naimul Islam Nobel, “Application Of K-Means Clustering Algorithm To Determine The Density Of Demand Of Different Kinds Of Jobs”, International Journal of Scientific & Technology Research, Volume 9, Issue 02, February 2020, ISSN: 2277-8616, pp: 2550-2557.
20. A. Karim, S. Azam, B. Shanmugam, and K. Kannoorpatti, “Efficient Clustering of Emails Into Spam and Ham: The Foundational Study of a Comprehensive Unsupervised Framework,” IEEE Access, vol. 8, pp. 154759–154788, 2020.
21. P. Ghosh et al., “A Comparative Study of Different Deep Learning Model for Recognition of Handwriting Digits,” International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS 2020), pp. 857 – 866, January 19, 2021.
22. M. F. Foysal, M. S. Islam, A. Karim, and N. Neehal, “Shot-Net: A Convolutional Neural Network for Classifying Different Cricket Shots,” Communications in Computer and Information Science, pp. 111–120, 2019.
23. Junayed M.S., Jeny A.A., Neehal N., Atik S.T., Hossain S.A. (2019) A Comparative Study of Different CNN Models in City Detection Using Landmark Images. In: Santosh K., Hegadi R. (eds) Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2018. Communications in Computer and Information Science, vol 1035. Springer, Singapore. https://doi.org/10.1007/978-981-13-9181-1_48.
24. Biswas A., Chakraborty S., Rifat A.N.M.Y., Chowdhury N.F., Uddin J. (2020) Comparative Analysis of Dimension Reduction Techniques Over Classification Algorithms for Speech Emotion Recognition. In: Miraz M.H., Excell P.S., Ware A., Soomro S., Ali M. (eds) Emerging Technologies in Computing. iCETiC 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 332. Springer, Cham. https://doi.org/10.1007/978-3- 030-60036-5_12.
25. Javed Mehedi Shamrat F.M., Allayear S.M., Alam M.F., Jabiullah M.I., Ahmed R. (2019) A Smart Embedded System Model for the AC Automation with Temperature Prediction. In: Singh M., Gupta P., Tyagi V., Flusser J., Ören T., Kashyap R. (eds) Advances in Computing and Data Sciences. ICACDS 2019. Communications in Computer and Information Science, vol 1046. Springer, Singapore. https://doi.org/10.1007/978-981-13-9942-8_33
26. F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Naimul Islam Nobel, and Md. Razu Ahmed. 2019. An Automated Embedded Detection and Alarm System for Preventing Accidents of Passengers Vessel due to Overweight. In Proceedings of the 4th International Conference on Big Data and Internet of Things (BDIoT'19). Association for Computing Machinery, New York, NY, USA, Article 35, 1–5. DOI:https://doi.org/10.1145/3372938.3372973
27. Shamrat F.M.J.M., Nobel N.I., Tasnim Z., Ahmed R. (2020) Implementation of a Smart Embedded System for Passenger Vessel Safety. In: Saha A., Kar N., Deb S. (eds) Advances in Computational Intelligence, Security and Internet of Things. ICCISIoT 2019. Communications in Computer and Information Science, vol 1192. Springer, Singapore. https://doi.org/10.1007/978-981-15-3666-3_29.
28. A. Islam Chowdhury, M. Munem Shahriar, A. Islam, E. Ahmed, A. Karim, and M. Rezwanul Islam, “An Automated System in ATM Booth Using Face Encoding and Emotion Recognition Process,” 2020 2nd International Conference on Image Processing and Machine Vision, 2020.
29. F.M. Javed Mehedi Shamrat, Shaikh Muhammad Allayear and Md. Ismail Jabiullah "Implementation of a Smart AC Automation System with Room Temperature Prediction ", Journal of the Bangladesh Electronic Society, Volume 18, Issue 1-2, June-December 2018, ISSN: 1816-1510, pp: 23- 32.
30. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, ‘‘ArcFace: Additive angular margin loss for deep face recognition,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4690–4699.

31. D. Chen, C. Xu, J. Yang, J. Qian, Y. Zheng, and L. Shen, ''Joint Bayesian guided metric learning for end-to-end face verification,'' Neurocomputing, vol. 275, pp. 560–567, Jan. 2018.
32. M. H. Khan, J. McDonagh, and G. Tzimiropoulos, ''Synergy between face alignment and tracking via discriminative global consensus optimization,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3811–3819.
33. M. Drożdż and T. Kryjak, ''FPGA implementation of multi-scale face detection using HOG features and SVM classifier,'' Image Process. Commun., vol. 21, no. 3, pp. 27–44, Sep. 2016.
34. C. Ma, N. Trung, H. Uchiyama, H. Nagahara, A. Shimada, and R.-I. Taniguchi, ''Adapting local features for face detection in thermal image,'' Sensors, vol. 17, no. 12, p. 2741, Nov. 2017.