A Secured, Multilevel Face Recognition based on Head Pose Estimation, MTCNN and FaceNet

Thai-Viet Dang [1], Hoai-Linh Tran [2*]
[1] School of Mechanical Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam
[2] School of Electrical and Electronics Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam
Email: [1] viet.dangthai@hust.edu.vn, [2] linh.tranhoai@hust.edu.vn
*Corresponding Author

Abstract—**Artificial Intelligence and IoT have always attracted a lot of attention from scholars and researchers because of their high applicability, which make them a typical technology of the Fourth Industrial Revolution. The hallmark of AI is its self-learning ability, which enables computers to predict and analyze complex data such as bio data (fingerprints, irises, and faces), voice recognition, text processing. Among those application, the face recognition is under intense research due to the demand in users' identification. This paper proposes a new, secured, two-step solution for an identification system that uses MTCNN and FaceNet networks enhanced with head pose estimation of the users. The model's accuracy ranges from 92% to 95%, which make it competitive with recent research to demonstrate the system's usability.**

Keywords—*Face Recognition, Head Pose Estimation, Computer Vision, MTCNN.*

I. INTRODUCTION

In the era of 4th Industrial Revolution, the rapid development of various new sophisticated hardware has enabled many solutions that can support the high computing power required for artificial intelligence (AI) models [1]. At the same time, scientific research discovered many interesting, powerful theoretical tools for automatic knowledge discovery and AI computing [2, 3]. For the last few years, computer vision (CV) has been a fast and widely field of application, that was developed and implemented in many smart devices and systems. Newest technological and algorithmic achievements play an increasingly important role in supporting systems that based on CV, such as object recognition and detection, image processing, and information extraction [4-7]. One of the important products of computer vision software development is the face recognition solutions, widely used in industrial and civil applications.

Despite the existence of other user identification solutions, such as RFID-based tag, that are quite convenient and accurate, they still have their disadvantages. For example, the standard ID tags can be easily copied, the newest, anti-theft tags are more expensive. The tags can be stolen or purposely unlawfully borrowed [8-11]. Biometrics based detector for fingerprint or iris patterns are quite expensive, hard to operate and maintenance and these devices have potential problem when the original information is unlawfully accessed [8, 9]. For example, the fingerprints database in the device can be stolen [10, 11]. Because of issues like those just mentioned, in this paper we propose to use facial recognition by using deep learning neural network

as a method of user identification, which would have high accuracy and security. Currently, there are many popular models available, for example ArcFace, DeepFace and FaceNet [12-23]. Although ArcFace and DeepFace provide high accuracy, they require larger computational resources than FaceNet, making them unsuitable for large-scale deployment. Therefore, in this paper, the FaceNet model was selected to be the base model for our solution. The general schematic of the system is shown in Fig. 1.
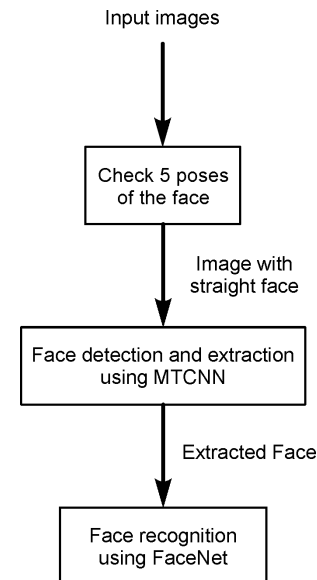

Fig. 1. The proposed schematic of the solution

First, in order to improve the fake images or masks, the user is asked to input 5 different poses of the face, including the standard straight pose, with the head turn left, right, up and down. The head pose estimation algorithm will check for the presence of all 5 poses (if one is missing, the user will be asked to retake the require image with the system's camera). Once all the poses have been achieved and recognized correctly, the straight pose image will be rotated and face extracted using Multi-tasked Cascaded Convolutional Networks (MTCNN) [13]. The extracted face only will be put into the FaceNet for final recognition of the user.

The numerical simulation has shown that due to the face rotation algorithm and the extraction of face area only the accuracy of the proposed model significantly increased to about 90-95% comparing to the accuracy of about 85% of the recognition when no treatment was introduced.

II.　THE PROPOSED SOLUTION FOR FACE RECOGNITION

The proposed system shown in Fig. 2 uses two neural networks: the MTCNN and the FaceNet. Their brief descriptions are presented in the following Section.

A. The Multi-tasked Cascaded Convolutional Networks

The Multi-tasked Cascaded Convolutional Networks (MTCNN) is a popular choice for use in Deep Convolutional Neural Networks (DCNN) for face detection and related facial feature recognition for the following reasons [24]:

1) Multi-tasking: MTCNN is a multi-tasking system capable of performing multiple tasks simultaneously such as face recognition, facial landmarks (eyes, nose, mouth) location or boundary boxes detection. This allows MTCNN to provide detailed information about the face and its components, facilitating tasks such as face recognition, classification, and expression analysis.

2) High reliability: MTCNN uses convolutional neural network (CNN) for face recognition and feature localization. CNN is a proven simple but powerful network for image processing tasks. This allows MTCNN to manage to achieve high reliability in detection and robustness against noise, scaling, and light changes.

3) High efficiency: MTCNN is designed to work fast and efficient. With the ability of sequential feature localization and detection stages, enriching with techniques such as convolutional networks and local awareness convergence, MTCNN can process images and videos of varying complexity with high speed.

4) Diverse applications: MTCNN can be applied in many different tasks, including facial recognition, age and gender classification, facial expression analysis and security monitoring system. The merge of MTCNN into DCNN opens up new opportunities for research and development of artificial intelligence applications related to image processing and face recognition.
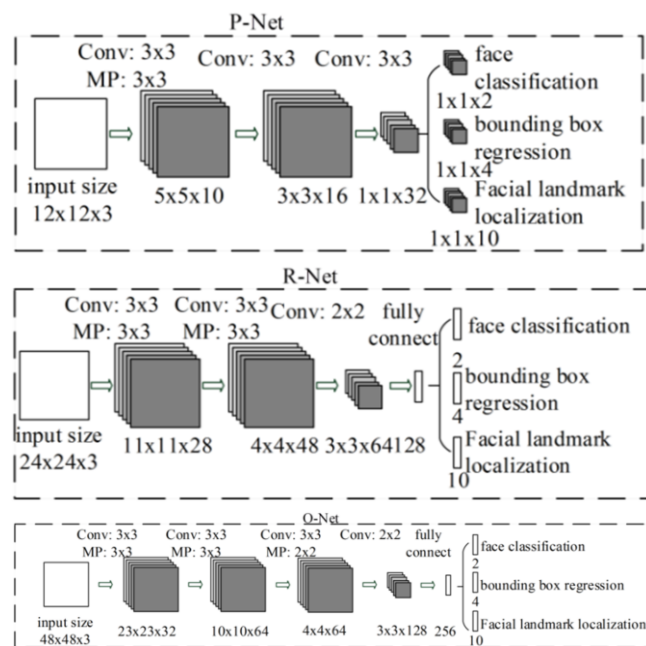
Face detection, landmark location, and accurate bounding box regression are the three steps that make up the MTCNN model. These steps are made so that faces can be found and identified in photographs of varied complexity, such as those with size, position, and lighting changes. MTCNN has shown promising results in both still image and video face identification and feature recognition. In the paper, the MTCNN was used to detect the face in the image to crop it for the user recognition purposes.

B. FaceNet Network Model

FaceNet is a deep learning neural network, that was proposed and used as detecting feature points on face images. FaceNet [24-31] was announced in 2015. It takes, as input, face images and generates at output a vector of 128 features representing the face. With this features vector, further computation blocks will work faster than computing on the whole image. In this paper, the FaceNet network was trained with following steps:

- Step 1. Randomly select an image as the so called *anchor image*.
- Step 2. Randomly select an image of the same person and mark it as a *positive image*.
- Step 3. Randomly select an image of *another* person and mark it as a *negative image*.
- Step 4. Adjust the parameters of the FaceNet, such that the positive image is closer to the anchor image than the negative image, or in term of a distance measure we have, can shown in by the equation (1) [32].

$$\forall f(x_i^a), f(x_i^p), f(x_i^n) \in T : \left\| f(x_i^a) - f(x_i^p) \right\|_2^2 + \alpha < \left\| f(x_i^a) - f(x_i^n) \right\|_2^2$$

$$(1)$$

where the parameter α is used to ensure that the distance between samples is far enough.

An example of one step of training FaceNet is demonstrated in Fig. 3, where there were images from two different persons. Once randomly selected, the positive sample is moved closer to the anchor and the negative sample is moved further from the anchor.

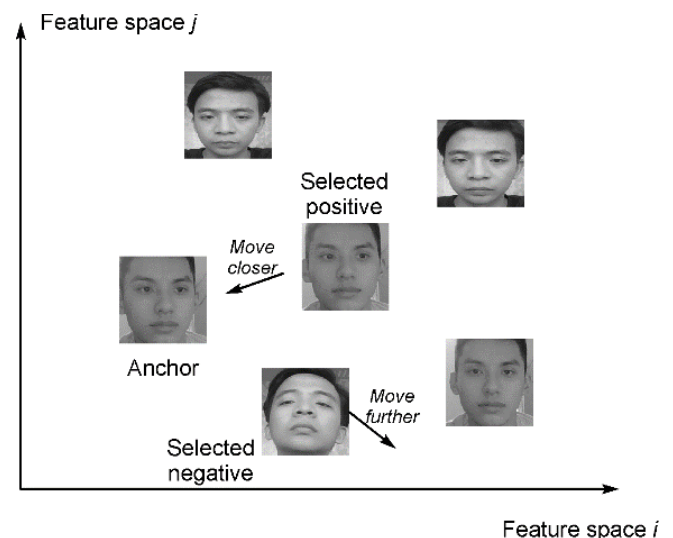Fig. 2. The MTCNN structure with 3 layers P-Net, R-Net and O-Net [24]

Fig. 3. A demonstration of FaceNet training on input images

This increases the accuracy of the classification system and reduces the dependence on unnecessary features. Then, the loss function is represented as equation (2).

$$L = \sum_i^N \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 \right] + \alpha \tag{2}$$

The process is continued until all images of the same human face were close together and images of different human faces were far apart. In order to achieve high training efficiency for the model, we need to adapt the network, such that the anchor and the negative images are as far apart as possible, when the anchor and positive images are as close together as possible, can be shown in equation (3).

$$\begin{cases} \left\| f(x_i^a) - f(x_i^p) \right\| \to min \\ \left\| f(x_i^a) - f(x_i^n) \right\| \to max \end{cases} \tag{3}$$

The effects of moving samples during FaceNet training are demonstrated in Fig. 4, where Fig. 4(a) demonstrated the result of one step and an example of final result is shown in Fig. 4(b) [33, 34].
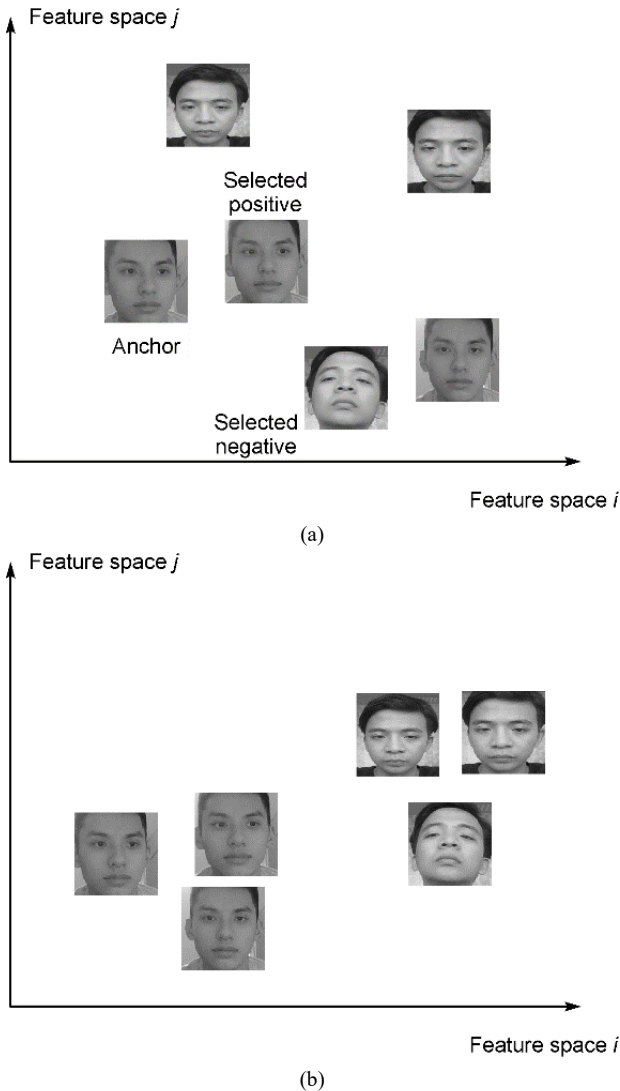

(a)

(b)

Fig. 4. (a) Example of moving one positive sample closer to the anchor and one negative sample further from the anchor and (b) example of final results after many repeated training epochs

C. Head Pose Estimation

Head pose refers to the direction or position of a person's head relative to the frame of reference [35-38]. It usually describes the rotation of the head along three axes: pitch, yaw, and roll (rotation around x, y, z). Three degrees of freedom (DoF) can be calculated using a spherical coordinate system to estimate surface normal n_f. Based on a set of fixed distance ratios from multiple individuals, the tilt angle is calculated based on any viewing direction. In this case, the projected length of the surface normal n_f is used in both coordinate axes. In addition, it was assumed that the main view point was fixed and the spherical coordinate system would be avoided if possible [39-44].

For the two images as seen in Fig. 5, let's start with the estimation of the head position using the 4-point method (f_a, f_b, f_c, f_d) in Fig. 5. First, f_a and f_d represent the two central points between the eyes and the corners of the mouth. Then f_b is the point lying on the axis perpendicular to the plane normal. Finally, f_c can be calculated using the following equation (4) while t is chosen in the range of 0.45.
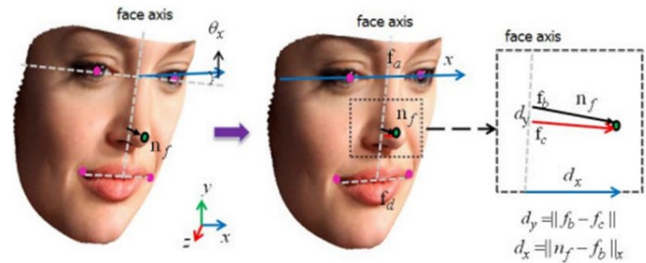

Fig. 5. Illustration of de-rotation to reliably compute surface normals from a face image

$$f_c = (1 - t) \cdot f_a + t \cdot f_d \tag{4}$$

The value of n_f in [7] was calculated using the four measures of data samples with relative as in the equation (5).

$$n_f = \frac{\|f_b - f_a\|}{\|f_a - f_d\|} \tag{5}$$

The general relative nose length ratio can be found by equation (6).

$$l_n = \frac{L_n}{\lfloor f_a' - f_d' \rfloor} \tag{6}$$

The rotations around the z and y axes are calculated using the relative observation distance divided by the maximum of nose height l_n, as in the equation (7).

$$\begin{cases} \theta_y = arcsin\left(\frac{d_x'}{l_n}\right) \\ \theta_z = arcsin\left(\frac{d_y'}{l_n}\right) \end{cases} \tag{7}$$

Based on Fig. 5 and Eq. (2) to (7), it's shown that the rotational angles of the human face can be represented as a function of head positions [45-51]. This will be used twofold in this paper. First, the formulas are used to check if the user has input all required poses of the head to reduce the risk of faking input images to the camera. Then for the straight face image, it will be rotated to have the face lined up with horizontal and

vertical axes. This image preprocessing will help the accuracy of face detection and recognition later [52-58].

III. SOLUTION IMPLEMENTATION AND DISCUSSIONS

3.1. The Experimental System

Fig. 6 presents a hardware realization of the proposed solution for face recognition used in this research. The system is based on an embedded computer (a 8GB Jetson Nano). It takes images from the camera IMX-219 and processes the images for head pose estimation and running the two neural networks (in inferencing/testing mode).
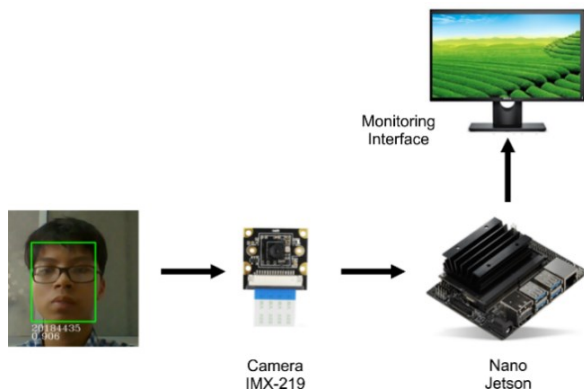

Fig. 6. The proposed hardware system for face recognition

The user is asked to allow the camera to capture 5 images with straight face and with the head turning left, right, up and down. A demonstration of input images and their positive checking for all the poses is shown in Fig. 7. Once the checking of all the poses are finished, the image with straight face will be rotated to line up the face and input into MTCNN. The MTCNN was train to detect the face area with the databases of rotated images generated using head rotation estimation. This preprocessing of the images help to improve the accuracy of MTCNN to 100% correct face detection for the testing images.

Next, FaceNet was trained with the cropped images of the face only. Fig. 8 shows an example of a crop image from Fig. 7(a) as the output of MTCNN and is used for training the FaceNet.

One advantage of the FaceNet is the down-sampling method from the input image to a vector of 128 components used to recognize the users. For the purpose of this paper, we only need the simple K nearest neighbor algorithm to classify the actual feature vector with the ones from the learning database. With those simplifications, when using Jetson Nano board as in the proposed system, the average processing throuIgput was from 21 to 25 frames per second (fps). The accuracy and the fps for the proposed solution and some comparisons are given in Table I.

TABLE I. COMPARATIVE RESULTS OF ACCURACY AND FPS FOR SELECTED METHODS

System	Accuracy (%)	Fps
LBP [59, 60]	84-88	19-21
DLIB [59, 60]	52-59	9-10
MTCNN+Facenet+Head Pose	90-95	21-25

Example results of face recognition are shown in Fig. 9 where the input images (shown on the left) are transformed by a trained FaceNet to vectors of 128 features and matched with the trained samples. The closest trained samples corresponding to the input images are shown on the right in Fig. 9.

(a) Straight face

(b) Image with head left

(c) Image with head right

(d) Image with head up

(e) Image with head down

Fig. 7. An example of checking of the different poses of an user ((a) straight face, (b) with head turnt left, (c) right, (d) up and (e) down)
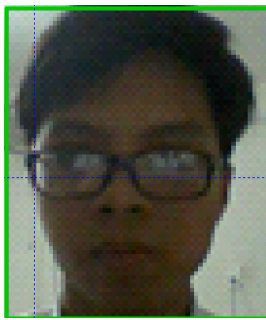
Fig. 8. The extracted face area from MTCNN will form the input into FaceNet for training for user recognition

Fig. 9. Example of face recognition using FaceNet: the extracted face area from MTCNN (on the left column) are put into FaceNet and the features vector is later matched with the closest trained samples (from the right)

IV.   CONCLUSION

This paper presented a new approach for face recognition using head pose estimation combined with MTCNN and Facenet networks to increased the recognition efficiency and to avoid faking input images. The head pose estimation was succesfully to detect 5 main poses of the head: straight face, head turning left/right/up/down to provide additional check on the user identity. The head pose information was also used to line up the straight face image to allow the MTCNN correctly identify the area containing the face on the images. The final user recognition was done by the FaceNet trained on the images generated from the MTCNN. The proposed model is simple enough to allow its running on a Nano Jetson board. When tested with an IoT system on the board, the accuracy achievied was 90-95%, with image processing speed of 21-25 fps.

CONFLICT OF INTEREST

The authors declare no conflict of interest

REFERENCES

[1] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, "Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook," *IEEE Access*, vol. 8, pp. 220121-220139, 2020.

[2] J. Huang, J. Chai, and S. Cho, "Deep learning in finance and banking: A literature review and classification," *Frontiers of Business Research in China*, vol. 14, pp. 1-24, 2020.

[3] R. Gautam, A. Gedam, and A. A. Mahawadiwar, "Review on Development of Industrial Robotic Arm," *Int. Res. J. Eng. Technol.*, vol. 4, no. 3, 2017.

[4] T.V. Dang and N.T. Bui, "Multi-Scale Fully Convolutional Network-Based Semantic Segmentation for Mobile Robot Navigation," *Electronics*, vol. 12, no. 3, p. 533, 2023.

[5] A. Gee and R. Cipolla, "Determining the gaze of faces in images," *Image and Vision Computing*, vol. 12, no. 10, pp. 639-647, 1994.

[6] N. L. P. Kurniawati, M. W. A. Kesiman, and I. M. G. Sunarya, "Recognition of Balinese Traditional Ornament Carving Images with Convolutional Neural Network and Discrete Wavelet Transform," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 8, no. 4, pp. 670-678, 2022.

[7] T. V. Dang and N. T. Bui, "Obstacle Avoidance Strategy for Mobile Robot based on Monocular Camera," *Electronics*, vol. 12, no. 8, p. 1932, 2023.

[8] T. V. Dang, "Smart Attendance System based on Improved Facial Recognition," *Journal of Robotics and Control*, vol. 4, no. 1, pp. 46-53, 2023.

[9] T. V. Dang, "Smart home Management System with Face Recognition based on ArcFace model in Deep Convolutional Neural Network," *Journal of Robotics and Control*, vol. 3, no. 6, pp. 754-761, 2022.

[10] F. Masalha and N. Hirzallah, "A students attendance system using QR code," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, pp. 75-79, 2014.

[11] O. Arulogun, A. Olatunbosun, O. Fakolujo, and O. Olaniyi, "RFID based students attendance management system," *Int. J. Sci. & Eng. Res.*, vol. 4, pp. 1-9, 2013.

[12] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815-823, 2015, doi: 10.1109/CVPR.2015.7298682.

[13] A. Wirdiani, P. Hridayami, and A. Widiari, "Face Identification Based on K-Nearest Neighbor," *Scientist J. Inform.*, vol. 6, no. 2, pp. 150-159, 2019.

[14] A. A. Süzen, B. Duman, and B. Şen, "Benchmark Analysis of Jetson TX2, Jetson Nano and Raspberry PI using Deep-CNN," *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1-5, 2020, doi: 10.1109/HORA49412.2020.9152915.

[15] S. Kurzadkar, "Hotel management system using Python Tkinler GUI," *Int. J. Comput. Sci. Mobile Comput.*, vol. 11, no. 1, pp. 204-208, 2022.

[16] Y. Nan, J. Ju, Q. Hua, H. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recognition," *Alexandria Eng. J.*, vol. 61, no. 6, pp. 4435-4444, 2021.

[17] J. Yu and W. Zhang, "Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4," *Sensor*, vol. 21, no. 9, p. 3263, 2021.

[18] A. Bochkovskiy, C. Y. Wang, and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.

[19] J. Jeong, H. Park, and N. Kwak, "Enhancement of SSD by concatenating feature maps for object detection," *arXiv preprint arXiv:1705.09587*, 2017.

[20] C. Angelo, Q. Gao, X. Zhang, and W. Pang, "Fast and Accurate Hand Visual Detection by Using a Spatial-Channel Attention SSD for Hand-

Based Space Robot Teleoperation," *International Journal of Aerospace Engineering*, vol. 2022, 2022.

[21] B. Huo, C. Li,J. Zhang, Y. Xue, and Z. Lin, "SAFF-SSD: Self-Attention Combined Feature Fusion-Based SSD for Small Object Detection in Remote Sensing," *Remote Sensing*, vol. 15, no. 12, p. 3027, 2023.

[22] H. Zhao *et al.*, "Local Binary Pattern-Based Adaptive Differential Evolution for Multimodal Optimization Problems," in *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3343-3357, July 2020, doi: 10.1109/TCYB.2019.2927780.

[23] S. Biswas, S. Kundu, and S. Das, "Inducing Niching Behavior in Differential Evolution Through Local Information Sharing," in *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 246-263, April 2015, doi: 10.1109/TEVC.2014.2313659.

[24] F. Caraffini, A. V. Kononova, and D. Corne, "Infeasibility and structural bias in differential evolution," *Infomation Science*, vol. 496, pp. 161–179, 2019.

[25] J. Wang, C. Zheng, X. Yang, and L. Yang, "EnhanceFace: Adaptive Weighted SoftMax Loss for Deep Face Recognition," in *IEEE Signal Processing Letters*, vol. 29, pp. 65-69, 2022, doi: 10.1109/LSP.2021.3125267.

[26] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 212-220, 2017.

[27] X. Li, D. Chang, T. Tian, and J. Cao, "Large-Margin Regularized Softmax Cross-Entropy Loss," *IEEE Access*, vol. 7, pp. 19572-19578, 2019.

[28] I. Goodfellow, Y. Bengio, and A. Courville, "Softmax Units for Multinoulli Output Distributions," *Deep Learning*, pp. 180–184, 2016.

[29] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-Margin Softmax Loss for Convolutional Neural Networks," *Proc. 33rd Inter. Conf. Mach. Learn.*, vol. 48, 2016.

[30] G.Bolin and P. Lacra, "On the Properties of the Softmax Function with Application in Game Theory and Reinforcement Learning," *arXiv preprint arXiv:1704.00805*, 2017.

[31] P. Reverdy and N. E. Leonard, "Parameter Estimation in Softmax Decision-Making Models With Linear Objective Functions," *in IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 54-67, Jan. 2016, doi: 10.1109/TASE.2015.2499244.

[32] C. Wu and Y. Zhang, "MTCNN and FACENET based access control system for face detection and recognition," *Automatic Control and Computer Sciences*, vol. 55, pp. 102-112, 2021.

[33] M. A. Falalu, U. Ibrahim, I. Amina, S. B. Abdulkadir, M. A. Baballe, and A. Ya'u, "A Smart Attendance System Based on Face Recognition: Challenges and Effects," *Global Journal of Research in Engineering & Computer Sciences*, vol. 3, no. 2, pp. 25-30, 2023.

[34] T. V. Dang, "Smart home Management System with Face Recognition Based on ArcFace Model in Deep Convolutional Neural Network," *Journal of Robotics and Control (JRC)*, vol. 3, no. 6, pp. 754-761, 2022.

[35] A. Gee and R. Cipolla, "Determining the gaze of faces in images," *Image and Vision Computing*, vol. 12, no. 10, pp. 639-647, 1994.

[36] H. Yuan, H. Li, J. Hou, and J. Xiao, "Single image-based head pose estimation with spherical parametrization and 3D morphing," *Pattern Recognition*, vol. 103, pp. 107316, 2020.

[37] N. Aunsri and S. Rattarom, "Novel eye-based features for head pose-free gaze estimation with web camera: New model and low-cost device," *Ain Shams Engineering Journal*, vol. 13, no. 5, pp. 101731, 2022.

[38] B. -C. Chen, C. -S. Chen, and W. H. Hsu, "Face Recognition and Retrieval Using Cross-Age Reference Coding With Cross-Age Celebrity Dataset," in *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 804-815, June 2015, doi: 10.1109/TMM.2015.2420374.

[39] B. Zhang and Y. Bao, "Age Estimation of Faces in Videos Using Head Pose Estimation and Convolutional Neural Networks," *Sensors*, vol. 22, p. 4171, 2022.

[40] R. Rothe; R. Timofte, and L. V. Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, vol. 126, pp. 1–14, 2016.

[41] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879-2886, 2012, doi: 10.1109/CVPR.2012.6248014.

[42] H. Zhou, F. Jiang, L. Xiong, and H. Lu, "An Intuitive and unconstrained 2D cube representation for simultaneous head detection and pose estimation," *arXiv preprint arXiv:2212.03623*, 2022.

[43] N. Ruiz, E. Chong, and J. M. Rehg, "Fine-Grained Head Pose Estimation Without Keypoints," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2155-215509, 2018, doi: 10.1109/CVPRW.2018.00281.

[44] E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607-626, April 2009, doi: 10.1109/TPAMI.2008.106.

[45] T. Martyniuk, O. Kupyn, Y. Kurlyak, I. Krashenyi, J. Matas, and V. Sharmanska, "DAD-3DHeads: A Large-scale Dense, Accurate and Diverse Dataset for 3D Head Alignment from a Single Image*," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20910-20920, 2022, doi: 10.1109/CVPR52688.2022.02027.

[46] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886-893, 2005.

[47] H. S. Dadi and P. G. Mohan, "Performance Evaluation of Eigen faces and Fisher faces with different preprocessed Data sets," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 4, no. 5, pp. 2110-2116, 2015.

[48] R. Fischer, M. Hödlmoser, and M. Gelautz, "Evaluation of Camera Pose Estimation Using Human Head Pose Estimation," *SN Computer Science*, vol. 4, no. 301, 2023.

[49] F. Camposeco, A. Cohen, M. Pollefeys, and T. Sattler, "Hybrid Camera Pose Estimation," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 136-144, 2018, doi: 10.1109/CVPR.2018.00022.

[50] Y. Xu, Y. -J. Li, X. Weng, and K. Kitani, "Wide-Baseline Multi-Camera Calibration using Person Re-Identification," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13129-13138, 2021, doi: 10.1109/CVPR46437.2021.01293.

[51] K. Khan, R. U. Khan, R. U., Leonardi, R., Migliorati, P., & Benini, S. (2021). Head pose estimation: A survey of the last ten years. *Signal Processing: Image Communication*, vol. 99, p. 116479, 2021.

[52] K. Khan, R. U. Khan, R. Leonardi, P. Migliorati, and S. Benini, "Head pose estimation: A survey of the last ten years," *Signal Processing: Image Communication*, vol. 99, p. 116479, 2021.

[53] T. Liu *et al.*, "GMDL: Toward precise head pose estimation via Gaussian mixed distribution learning for students' attention understanding," *Infrared Physics & Technology*, vol. 122, pp. 104099, 2022.

[54] D. Anitta and A. A. Fathima, "Human head pose estimation based on HF method," *Microprocessors and Microsystems*, vol. 82, p. 103802, 2021.

[55] B. G. D. A. Madhusanka, S. Ramadass, P. Rajagopal, and H. M. K. K. M. B. Herath, "Attention-Aware Recognition of Activities of Daily Living Based on Eye Gaze Tracking," *Internet of Things for Human-Centered Design*, vol. 1011, pp. 155–179, 2022.

[56] C. Bouras and E. Michos, "An online real-time face recognition system for police purposes," *2022 International Conference on Information Networking (ICOIN)*, pp. 62-67, 2022, doi: 10.1109/ICOIN53446.2022.9687212.

[57] S. Santhosh and S. V. Rajashekararadhya, "A Design of Face Recognition Model with Spatial Feature Extraction using Optimized Support Vector Machine," *2023 2nd International Conference for Innovation in Technology (INOCON)*, pp. 1-8, 2023, doi: 10.1109/INOCON57975.2023.10101149.

[58] H. Agobah, O. Bamisile, D. Cai, D. K. Bensah Kulevome, B. Darkwa Nimo, and Q. Huang, "Deep Facial Expression Recognition Using Transfer Learning and Fine-Tuning Techniques," *2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1856-1860, 2022, doi: 10.1109/EI256261.2022.10116540.

[59] H. S. Dadi and G. K. M. Pillutla, "Improved Face Recognition Rate Using HOG Features and SVM Classifier," *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, vol. 11, no. 4, pp. 33-44, 2016.

[60] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 765-781, Nov. 2011, doi: 10.1109/TSMCC.2011.2118750.