Journal of Applied Informatics and Computing (JAIC)
Vol. 9, No.1, February 2025, pp. 16~22
e-ISSN: 2548-6861

Analysis of the Use of MTCNN and Landmark Technology to Improve the Accuracy of Facial Recognition on Official Documents

Ferri Rama Chandra [1] *, Hajra Rasmita Ngemba [2] **, Odai Amer Hamid [3] *, Nouval Trezandy Lapatta [4] *, Syaiful Hendra [5] *, Deny Wiria Nugraha [6] * Syahrullah[7] ****
* Informatics Engineering, Faculty of Engineering, Tadulako University, Palu
** Information Systems, Faculty of Engineering, Tadulako University, Palu
*** College of Technical Management Mosul, Northern Technical University, Iraq
ferriramachandra@gmail.com [1], hajra.rasmita@gmail.com [2], oudayamer@ntu.edu.iq [3], nouval@untad.ac.id [4], syaiful.hendra.garuda@gmail.com [5], deny.wiria.nugraha@untad.ac.id [6], syahroellah.ms@gmail.com [7]

Article Info

Article history:

Received 2024-11-01
Revised 2024-11-11
Accepted 2024-11-14

Keywords:

Facial Detection,
Facial Landmark,
Face Recognition,
MTCNN.

ABSTRACT

A face recognition system consists of two stages: face detection and face recognition. Detection of features such as eyes and mouth is important in facial image processing, especially for official documents such as identity cards. To ensure identification accuracy, this research applies facial landmark extraction technology and MTCNN (Multi-Task Cascaded Convolutional Neural Network). The purpose of this research is to evaluate the accuracy of MTCNN in detecting facial features at the Department of Population and Civil Registration (dukcapil) Palu City, using facial landmarks and waterfall methods as an application development methodology. The evaluation results show that MTCNN has high face recognition accuracy and good positioning ability regardless of what GPU in use as long have right CPU and System Operation. In comparison, the Viola-Jones algorithm is effective for high-speed applications, while SSD offers balanced performance with GPU device requirements for optimal performance. While MTCNN proved to be effective, challenges still exist, such as false positives and false negatives, especially in poor lighting conditions and extreme poses. Image and camera quality, including resolution and facial expression, also affects detection accuracy. These findings suggest that the application of MTCNN can improve face recognition accuracy for official documents, although it requires addressing existing challenges. With this technology, it is expected that errors in facial recognition can be minimized, resulting in more reliable data that meets the standards for issuing identity documents. This research contributes to the development of a more accurate and efficient face recognition system for personal identification applications.

I. INTRODUCTION

A facial recognition system is a computational process that automatically uses digital images or video frames from a video source to automatically identify or authenticate a person [1],[2]. Comparing selected facial features from an image to a database of faces is one approach to this method [3]. A face is an object consisting of several elements that appear the same but are not. As a result, these characteristics form the basis for identifying an individual's identity. This requires detecting a face from the entire image containing relevant features. A facial recognition system is generally divided into two stages, namely a face detection system which is the initial stage (pre-processing) and continued with a face recognition system [4], [5]. Detecting facial features such as eyes and mouth is an important issue in processing facial images that will be used for many research areas such as emotion detection and facial identification [6], [7]. In the context of making official documents, such as Identity Cards (KTP), certain requirements such as a straight face, eyes not closed, and a closed mouth are crucial to ensure the accuracy of personal identification. The process of capturing images for official documents often requires adjustments to meet certain standards set by related institutions. In order to meet these requirements, the use of facial landmark extraction technology and MTCNN (Multi-Task Cascaded Convolutional Neural Network) is considered very relevant. Facial landmark extraction technology allows the system to accurately identify the position of the eyes and mouth in facial images. [8]. Meanwhile, MTCNN makes a major contribution in improving the accuracy of detecting eye, mouth, and head tilt conditions, producing facial images that meet standard criteria [9], [10]. By implementing this technology, it is expected to significantly reduce the possibility of errors or inaccuracies in facial recognition results. Success in ensuring that certain requirements are met during the capture of official documents such as ID cards will create more reliable data that complies with the norms for issuing identity documents.

Facial landmarks are key components in determining the position of the eyes, mouth, and other facial features with high accuracy [11]. The application of this technology helps ensure that facial images taken for official document purposes meet certain standards, such as the presence of uncovered eyes and a closed mouth. Based on the explanation above, the purpose of this study is to prove the accuracy of the algorithm used, namely MTCNN, against a case study at the Dukcapil of Palu City.

II. METHOD

A. Facial Landmark

Facial landmarks have been widely studied by researchers. Research on Multi-Task Learning of Facial Landmarks and Expressions can produce learning representations to predict the position and shape of facial *landmarks* in improving expression recognition from images [12], [13]. Furthermore, Fast Facial Landmark Detection and Applications: A Survey that Dense Facial Landmark is one of the key elements in facial processing [14]. This technology is used in various applications such as virtual face reenactment, emotion recognition, driver status tracking, and so on. Early approaches were only suitable for facial landmark detection in controlled environments, which was clearly insufficient. Neural networks have shown tremendous qualitative improvement in the problem of facial landmark detection in natural conditions [15], and are now widely studied by researchers in this field. Many clever ideas have been proposed, often complementing each other. Facial Landmark-Based Face Detection Using Opencv and Dlib, in his research using OpenCV and Dlib, this method aims to improve the accuracy of human face detection by focusing on facial landmark points [16]. The system test results showed success in the face detection process, especially when involving facial movement during processing. The location of facial landmark fiducial points around facial components and facial contours captures rigid and non-rigid facial deformations due to head movements and facial expressions [17]. Therefore, these locations are important for various facial analysis tasks. Over time, many facial landmark detection algorithms have been developed to automatically detect these key points, this paper reviews 3 facial landmark algorithms namely the Holistic Method, Constrained Local Model (CLM) methods, and regression-based methods.

B. Multi-Task Cascaded Convolutional Network Algorithm

The MTCNN algorithm in the face recognition-based authentication mechanism that MTCNN is a restructuring of the combination of the CNN model with 3 layers of networks, namely p-net, r-net, and o-net which uses candidate grouping and classification to obtain efficient and fast face detection results [18]. Face recognition using the MTCNN algorithm provides an accuracy rate of 76% and 84% for indoor and outdoor conditions, respectively [19]. These results are influenced by lighting factors measured by a lux meter, camera distance, and the tilt of the user's face (researcher, year). In addition, the authentication process with MTCNN takes 3.78224 seconds indoors and 3.56268 seconds outdoors, with the processing time influenced by the performance of the virtual machine used. In other studies, [9]this study identifies face detection as a crucial research area in target detection. This study examines the key steps in the task of face detection using deep learning, focusing on feature extraction as the most important aspect. Two-stage and one-stage detection models are compared, and their applications in face detection are analyzed. In addition, MTCNN is explored in depth, and its effectiveness in face detection tasks is validated through experiments, with a comparison of the model results with the yolov3 model on a wider face dataset [20].

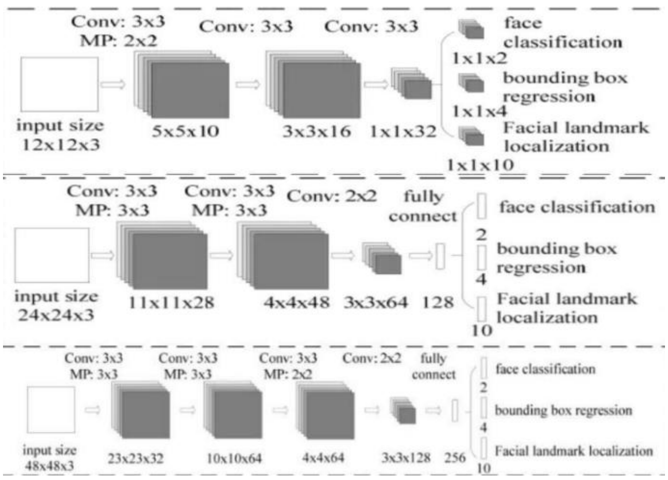

Figure 1. P-net structure, R-net structure, O-net structure on MTCNN

C. Research Stages

The Waterfall system development methodology is a structured approach to software development that follows specific steps in a step-by-step manner. A waterfall is called a waterfall because it flows from one stage to the next without returning to the previous stage, like water flowing downwards [21]. The stages of the waterfall method start from the selection The purpose of this requirement is to understand the system requirements from the user's perspective.

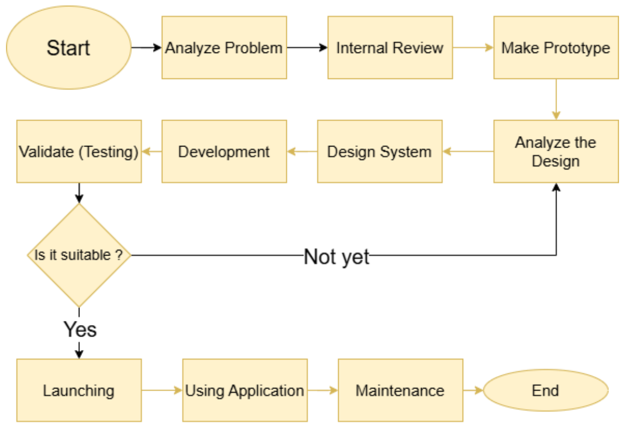

Figure 2. Research Stages

All requirements are recorded, analyzed, and documented in detail. The result of this stage is the specification of the requirements that form the basis for system development. The next stage is the construction and design of the system, which involves designing the system architecture and technical design based on the requirements and specifications collected. During this stage, decisions are made regarding data structures, system architecture, interface design, and other details related to implementation. After that, testing and launching, after the system is developed, testing is carried out to ensure that all features operate according to specifications. The purpose of the testing phase is to find and fix any errors (bugs) that may exist in the system. After testing is complete, the system will be launched to the user. The last stage is *Maintenance*, which is the system that is integrated into the user's operating environment. After implementation, the system enters the maintenance stage, where changes and improvements needed to maintain system performance are carried out periodically.

D. System Development Methods

The method used for development is waterfall, where the first thing to do is observation, namely collecting information, literature study which collects written sources, Needs analysis, which stage identifies problems and researchers are expected to find problems and solutions to these problems, then system design where at this stage the application interface design is carried out to be created, then system implementation, carried out according to the design and design of the application being built and coding for it. Then testing the previously planned application system is carried out at the testing stage to check whether there are errors or damage to the application that has been designed [22].

III. Results and Discussion

In this section, we will explain how the algorithm process is used and tested.

A. MTCN Architecture

The first stage is the P-Net (Proposal Network) Stage. P-Net is responsible for performing initial detection by generating a number of *Bounding Box candidates* that may contain faces. The process is in the form of *a Sliding Window* where the image is divided into small blocks which are then examined by P-Net. In addition to *the Bounding Box*, P-Net also estimates *Bounding Box Regression*, which helps refine the position of *the Bounding Box*. P-Net also provides suggestions in the form of *Bounding Boxes* and initial face *key points*. Next is the R-Net (Refinement Network) Stage. R-Net is used to filter *the Bounding Box* generated by P-Net. Here, the False *Positive Bounding Box* is removed. R-Net is more complex than P-Net and can provide more detailed detection. The R-Net *Bounding Box results* are more accurate because this network improves the P-Net proposal. This network will also estimate facial landmarks and further *Bounding Box Regression*. The last stage is the O-Net Stage (Output Network) O-Net is the last stage that receives *Bounding Box* from R-Net which O-NET will then refine. O-Net is more accurate in recognizing facial features such as the position of the eyes, nose, and mouth. O-Net provides the final results of the *Bounding Box* and landmarks [9].

B. Implementation Stages

1) Dataset

Dataset Selection: Common datasets used in facial recognition research are: LFW (Labeled Faces in the Wild): Contains over 13,000 facial images from over 5,700 people. CelebA (CelebFaces Attributes Dataset) Contains over 200,000 facial images in various poses, backgrounds, and lighting. *Custom Face* (Optional) can also create its own dataset by taking facial images with various expressions and positions (front, side, etc.) to test *Custom*. Dataset Division: The dataset is divided into two parts, namely *Training* (80%) and *Testing* (20%) so that the model can be trained and tested properly [23].

TABLE 1.
DATASET

Dataset Name	**Number of Images**	**Image Size**
LFW	13,000+	250 x 250
CelebA	200,000+	178 x 218
Custom	1,000	Free

2) Preprocessing

At this stage, *Resizing is done* where all images must be of uniform size. Commonly used sizes are 128 x 128 pixels or 160 x 160 pixels to reduce calculations. *Normalization*: Data is normalized so that pixel values range from 0 to 1 or -1 to 1. This helps improve the stability of the neural network during training. *Face Detection and Cropping*: Detecting and cropping faces from images using MTCNN. Only the face area is used to train the model, other areas (background) are ignored [13].

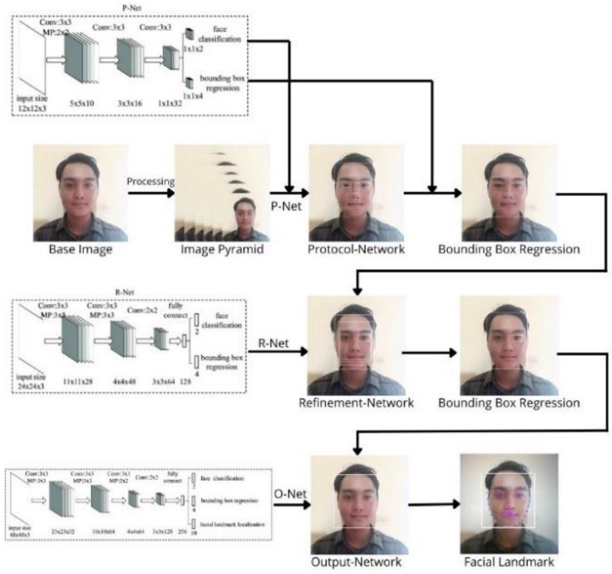

Figure 3. MTCNN Structure

3) *Training and Testing*

First, Hyperparameter Learning Rate is performed using a value of 0.001, then tuning is done if the results are less than optimal. Batch Size 32 or 64. Epochs start with 10-20 epochs adjusted to the size of the dataset and the loss results. Training Loop: At each epoch, the data set is fed into the model and the model updates the weights to minimize the loss. Loss and Accuracy Tracking: records the loss and accuracy at each epoch to ensure the model improves. Testing or testing the model on test data to verify its accuracy in detecting faces that were not seen during training.

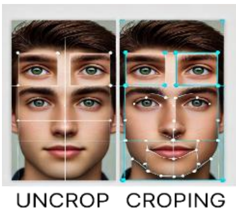

Figure 4. Examples of UNCROP and CROPPING

C. *Performance Evaluation*

1) *Face Detection Accuracy*

Precision or measuring the percentage of correctly recognized faces out of all the faces predicted by the model. Recall or measuring the percentage of correctly recognized faces out of all the faces that should be recognized. Also F1-score which is a metric to combine precision and recall to give an overall score to evaluate the performance of the model in face recognition.

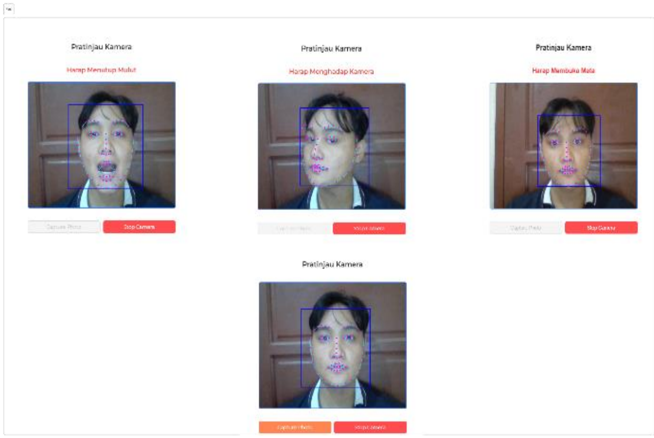

Figure 2. MTCNN Face Detection Results

2) *Face Alignment Evaluation*

Landmark Error is a measure of the distance between the average estimated location of a landmark (such as eyes, nose, mouth, etc.) and its actual location in the image. Common metrics such as L2 norm and Root *Mean Square Error* (RMSE) are used in evaluating the accuracy of face alignment. MSE (Mean Squared Error) is applied to assess landmark alignment, with lower values indicating better alignment results.

3) *Punctuality and Efficiency*

Detection Time Measure the time required to detect a face in an image. It is very important for real-time applications, such as facial recognition applications in security systems [24]. The test was conducted on four devices, with a testing distance of 30 cm, lighting at 50 Lux, and each device tested 10 times per device. From 10 trials, each test consisted of 4 conditions: when the eyes were closed, the mouth was open, the face was tilted, and the successful test for capturing. Testing was conducted with lighting at 50 Lux and a distance of 30 cm, lighting below 50 Lux made it difficult to detect faces, and a distance greater than 30 cm caused difficulty in detecting the tester's eyes.

Based on the results above, it can be observed that CPU performance and device resolution play a significant role in the final measurement time. Devices with a high core count and exceptional speed, such as the Acer Swift 3 Evo with an Intel Core i7-12700H, and Intel Iris XE GPU outperform other devices, achieving results around just 1 second. Meanwhile, the MacBook Air shows a unique performance—despite having a Dual-Core i5 processor with a speed of 3.6 GHz, it achieved a time result less than half that of the Acer E5-476G despite being inferior in term of CPU dan GPU, which has a Core i7-8550U running at 4.0 GHz and have Discete GPU.

TABLE 2.
RESULT TIME OF TESTING ON DIFFERENT DEVICES

Devices	CPU	GPU	Camera Resolution	Storage	Time to Detect
Acer AIO Veriton Z4	Core i3-10100 3.6 GHz	Intel UHD Graphic 630	HD 720P 30FPS 2 Megapiksel	512 GB SSD SATA	103.4 Seconds
Acer Aspire E5-476G	Core i7-8550u up to 4.0 GHz	MX150	Webcam 720P 30FPS 0.9 Megapiksel	512 GB SSD NVME Gen 3 by 2	23.7 Seconds
Macbook Air 2018	1.6GHz dual-core Intel core i5, Turbo Boost up to 3.6GHz, 4MB L3 cache	Intel UHD Graphic 617	720P FaceTime HD Camera	128 GB PCIe-Based SSD	9.363 Seconds
Acer Swift 3 Intel Evo 2022 Edition	Intel Core i7 - 12700H Up to 4.7 GHz	Intel Iris XE Graphic 96 EU	1080P 30FPS 2.1 Megapiksel	512 GB SSD NVME PCIE Gen 4	1.7 Seconds

The MacBook Air test results suggest the hypothesis that, despite lagging in core number and speed of CPU and GPU, the operating system plays an important role in efficiency and test time. On rare occasions, the feather duster was detected as a face, regardless of which device among the four was used.

D. Comparison with other methods

In this section, MTCN will be compared with other methods such as Viola-Jones and SSD.

TABLE 3.
COMPARISON OF EVALUATION RESULTS

Method	Precision (%)	Recall (%)	F1-Score	FPS (Frames per Second)	Face Alignment
MTCN	95.2	93.8	94.5	10	Yes
Viola-Jones	85.3	78.9	82.0	30	No
SSD	92.5	90.4	91.4	25	No

From the evaluation table, it can be said that MTCNN has excellent recognition accuracy and face positioning capabilities, and is suitable for face recognition applications that require high accuracy. Viola-Jones is fast and ideal for high-speed applications that do not require complex detection. SSD has a balanced performance, offering high speed and accuracy, but requires a GPU device for optimal performance [25].

This study shows that the use of *Multi-Task Cascaded Convolutional Neural Network (MTCNN) and landmark* technology can significantly improve the accuracy of face detection on official documents regardless of GPU As long as it has the right CPU and operating system, showing by result above. With MTCNN's ability to detect and align faces through three tiered networks—P-Net, R-Net, and O-Net—this study is able to minimize errors in face identification despite variations in pose, lighting, and expression. *Landmark technology then* completes this process by detecting important points on the face such as the eyes, nose, and mouth, ensuring consistent and accurate facial positioning. As a result, this technology not only improves the reliability of face recognition on official documents such as ID cards and passports but also supports a safer and more efficient identity verification process, which is essential in e-government systems and other digital identity applications.

IV. CONCLUSION

The use of *Multi-Task Cascaded Convolutional Neural Network* (MTCNN) and *landmark technology* can significantly improve the accuracy of face detection on official documents, however, MTCNN (Multi-task Cascaded Convolutional Network), still has several challenges and problems that can affect its performance. Below is an analysis of common errors and problems encountered when using MTCNN.

False Positives: MTCNN can recognize areas without faces as faces. False Negatives: MTCNN may not be able to detect a face that is present. This is especially true for extreme poses, images at unusual angles, or when the face is partially obscured by other objects. MTCNN may have difficulty recognizing faces in poor or uneven lighting conditions. Although MTCNN is designed to handle a wide range of facial poses, detecting faces at very oblique or extreme angles can affect accuracy. Image quality and resolution also affect MTCNN performance. Low-resolution or blurry images and low light also make it difficult to accurately recognize faces, and recognition may be unstable. Significantly different facial expressions can affect MTCNN performance. For example, if a face is smiling or has an extreme expression, the location of facial features may not be accurately recognized, which can affect subsequent analysis. System Operation or OS and CPU also play significant role about performance where the Mac OS have such high result regardless of CPU and GPU.

ACKNOWLEDGMENT

The researcher would like to express his deepest gratitude to Tadulako University, Lecturer, also especially MBKM and the Informatics Engineering Research Program. The researcher was able to complete this research as well as possible. The opportunity to participate in the MBKM Independent Study program provides researchers with valuable experience that broadens their knowledge and insight in the field of computer science.

REFERENCES

[1] M. Andrejevic, C. O'Neill, G. Smith, N. Selwyn, and X. Gu, “Granular biopolitics: Facial recognition, pandemics and the securitization of circulation,” *New Media Soc*, vol. 26, no. 3, pp. 1204–1226, March. 2024, doi 10.1177/14614448231201638.

[2] K. Pandey, R. Lilani, P. Naik, and G. Pol, “Human Face Recognition Using Image Processing.” doi: 10.17577/IJERTCONV2IS04051.

[3] R. Neha and S. Nithin, “Comparative Analysis of Image Processing Algorithms for Face Recognition,” in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, Jul. 2018, pp. 683–688. doi: 10.1109/ICIRCA.2018.8597309.

[4] KD Anggara, DP Kartikasari, and FA Bakhtiar, “Implementation of MTCNN Algorithm in Facial Recognition-based Authentication Mechanism,” 2023. [Online]. Available: http://j-ptiik.ub.ac.id

[5] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, “Past, Present, and Future of Face Recognition: A Review,” *Electronics (Basel)*, vol. 9, no. 8, p. 1188, Jul. 2020, doi: 10.3390/electronics9081188.

[6] M. Kim, Y. Cho, and S.Y. Kim, “Effects of diagnostic regions on facial emotion recognition: The moving window technique,” *Front Psychol*, vol. 13, Sept. 2022, doi: 10.3389/fpsyg.2022.966623.

[7] I. Kumaran *et al.*, "Face Recognition Using Measurement-Based Approach and Segmentation Method in Multiple Positions and Lighting," *FIDELITY: Jurnal Teknik Elektro*, vol. 3, no. 1, pp. 5–8, Jan. 2021, doi: 10.52005/fidelity.v3i1.85.

[8] F. Alanazi, G. Ushaw, and G. Morgan, “Improving Detection of DeepFakes through Facial Region Analysis in Images,” *Electronics (Basel)*, vol. 13, no. 1, p. 126, Dec. 2023, doi: 10.3390/electronics13010126.

[9] N. Zhang, J. Luo, and W. Gao, “Research on Face Detection Technology Based on MTCNN,” in *2020 International Conference on Computer Network, Electronic, and Automation (ICCNEA)*, IEEE, Sep. 2020, pp. 154–158. doi: 10.1109/ICCNEA50255.2020.00040.

[10] A. Ghofrani, RM Toroghi, and S. Ghanbari, “Realtime Face-Detection and Emotion Recognition Using MTCNN and miniShuffleNet V2,” in *2019 5th Conference on Knowledge-Based Engineering and Innovation (KBEI)*, IEEE, Feb. 2019, pp. 817–821. doi: 10.1109/KBEI.2019.8734924.

[11] Y. Wu and Q. Ji, “Facial Landmark Detection: A Literature Survey,” *Int J Comput Vis*, vol. 127, no. 2, pp. 115–142, Feb. 2019, doi: 10.1007/s11263-018-1097-z.

[12] D. Lydia and Z. Astuti, *Study of Facial Expression Recognition using PCA and CNN Methods*. 2018.

[13] T. Devries, K. Biswaranjan, and G. W. Taylor, “Multi-task Learning of Facial Landmarks and Expressions,” in *2014 Canadian Conference on Computer and Robot Vision*, IEEE, May 2014, pp. 98–103. doi: 10.1109/CRV.2014.21.

[14] K. Khabarlak and L. Koriashkina, “Fast Facial Landmark Detection and Applications: A Survey,” *J Comput Sci Technol*, vol. 22, no. 1, p. e02, Apr. 2022, doi: 10.24215/16666038.22.e02.

[15] C. Zhongshan, F. Xinning, A. Manickam, and V.E. Sathishkumar, “RETRACTED ARTICLE: Facial landmark detection using artificial intelligence techniques,” *Ann Oper Res*, vol. 326, no. S1, pp. 63–63, Jul. 2023, doi: 10.1007/s10479-021-04355-y.

[16] S. Abidin *et al.*, “Face Detection Using Webcam-Based Haar Cascade Classifier Method in Matlab”.

[17] Y. Said and M. Barr, “Human emotion recognition based on facial expressions via deep learning on high-resolution images,” *Multimed Tools Appl*, vol. 80, no. 16, pp. 25241–25253, Jul. 2021, doi: 10.1007/s11042-021-10918-9.

[18] RR Hajar *et al.*, “Facial Landmark Based Face Detection Using OpenCV and DLIB,” *Jurnal Teknologi Informasi (Journal of Information Technology)*, vol. 5, no. 2, 2021.

[19] Z. Zhou, H. Li, H. Liu, N. Wang, G. Yu, and R. Ji, “STAR Loss: Reducing Semantic Ambiguity in Facial Landmark Detection,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 15475–15484. doi: 10.1109/CVPR52729.2023.01485.

[20] MH Robin, Md. M. Ur Rahman, AM Taief, and Q. Nahar Eity, “Improvement of Face and Eye Detection Performance by Using Multi-task Cascaded Convolutional Networks,” in *2020 IEEE Region 10 Symposium (TENSYMP)*, IEEE, 2020, pp. 977–980. doi: 10.1109/TENSYMP50017.2020.9230756.

[21] NG Arsandy, J. Maulindar, and Moh. Muhtarom, "Implementation of Employee Attendance with Face Recognition using Waterfall Method in Solo Technopark," *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 4, no. 2, pp. 680–689, Aug. 2024, doi: 10.35870/ijsecs.v4i2.2617.

[22] IM Revina and WRS Emmanuel, “A Survey on Human Face Expression Recognition Techniques,” *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 619–628, Jul. 2021, doi: 10.1016/j.jksuci.2018.09.002.

[23] P. Chandran, D. Bradley, M. Gross, and T. Beeler, "Attention-Driven Cropping for Very High-Resolution Facial Landmark Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2020, pp. 5860–5869. doi: 10.1109/CVPR42600.2020.00590.

[24] NK Benamara *et al.*, “Real-time facial expression recognition using smoothed deep neural network ensemble,” *Integr Comput Aided Eng*, vol. 28, no. 1, pp. 97–111, Dec. 2020, doi: 10.3233/ICA-200643.

[25] RD Djohari, HR Ngemba, S. Hendra, DS Angraeni, NT Lapatta, and DW Nugraha, "Employee Attendance System with Facial Recognition Technology Using a Single Shot Detector (SSD) Algorithm," *Journal Of Informatics And Telecommunication Engineering*, vol . 7, no. 2, pp. 424–434, Jan. 2024, doi: 10.31289/jite.v7i2.10869.