Penerbit Akademia Baru

Journal of Advanced Research in Computing and Applications

Journal homepage:
https://akademiabaru.com/submit/index.php/arca/index
ISSN: 2462-1927

Face detection based on Haar Cascade and Convolution Neural Network (CNN)

Noor Afiza Binti Mohd Ariffin[1], Usman Abdul Gimba[2,*], Ahmad Musa[3]

[1] Department of Computer Science, Faculty of Computer Science and Information Technology, University Putra Malaysia, Serdang, 43400, Malaysia
[2] Department of Cyber Security, Faculty of Computing, Federal University Dutse, Dutse, Jigawa, Nigeria
[3] Department of Computing, Canterbury Christ Church University, Canterbury, United Kingdom

ARTICLE INFO

Article history:
Received 9 January 2025
Received in revised form 2 February 2025
Accepted 27 February 2025
Available online 10 March 2025

Keywords:
Face detection; CNN; Haar-Cascade; Viola-algorithm; deep learning

ABSTRACT

Face detection plays a crucial role in identifying individuals during suspicious activities, serving as a foundational component in various security applications. Modern face detection systems leverage machine learning algorithms to accurately identify human faces in images or videos, facilitating authentication in security contexts. This paper presents an innovative face detection system that integrates the Haar cascade method with Convolutional Neural Networks (CNNs), aimed at enhancing the accuracy of facial detection. The evaluation of the proposed system was carried out in a Python environment, utilizing real images from well-established public datasets, including Faces94, Faces95, Faces96, and the Grimace dataset, curated by Libor Spacek. The results obtained demonstrate the efficacy of the integrated approach, achieving accuracy rates of 98.37%, 97.22%, 97.52%, and 100% for the Faces94, Faces95, Faces96, and Grimace datasets, respectively. These findings indicate that the combination of Haar cascade and CNN-based methodologies significantly outperforms traditional machine learning face detection techniques, underscoring the potential for improved accuracy in real-world face detection applications. This research contributes to the ongoing advancements in facial recognition technology, with implications for enhanced security measures and intelligent human-computer interaction.

1. Introduction

The As information technology advances quickly, face detection technology has emerged as a key indicator of intelligence as mentioned by Zhang *et al.,* [1]. Core tasks in traditional computer vision comprises of object detection, processing, recognition, and classification of images as mentioned by some researchers [2,3]. In computer vision, object detection in particular is essential, as resolving this issue paves the way for addressing other issues like image segmentation as mentioned by Minaee *et al.,* [4]. Face detection is utilized to identify individuals during suspicious events as mentioned by Wibowo *et al.,* [5].

* *Corresponding author.*
E-mail address: gs64115@student.upm.edu.my (Usman Abdul Gimba)

https://doi.org/10.37934/arca.38.1.111

For systems involving intelligent or inspection vision-based human-computer interaction, face detection is a critical task. While humans can easily see and understand faces and objects, systems cannot do so without assistance as mentioned by Kumar *et al.,* [6]. Finding the human face as the region of interest inside a digital image that includes other aspects, such backdrop scenarios, is the basic notion behind human face detection as mentioned by Zhuang *et al.,* [7]. Face detection allows us to extract features from the human body using numerous techniques. Face identification is a visual task that is straightforward for humans to perform with ease, but it is not easy for computers. In fact, face detection has been considered the most difficult task in the field of computer vision due to its complexity as mentioned by Adouami *et al.,* [8]. This complexity arises from the significant intra-class variances brought on by variations to facial features, expression and lighting as mentioned by Kumar *et al.,* [9]. There are two categories of face detection methods: deep learning based approaches and traditional methods as mentioned by Zhang *et al.,* [1].

Traditional face detection methods can be divided into two primary categories as mentioned by Jiang [10]. The first category depends on characteristics that are constant, like color, face symmetry, and certain shape details. This approach struggles with formulating appropriate rules for face detection and is significantly impacted by dense scenes where organ symmetry and face shape vary greatly.

In the second category, which relies on template matching, face detection is accomplished by figuring out how closely a standard template and the area that has to be detected match. In situations where there are several faces, this method is less successful and requires a lot of computing power. Additionally, it has low resolution for small-scale faces, resulting in poor template matching. However, templates can incorporate some of the target object’s invariants characteristics, providing certain advantages for detection.

The deep learning-based method leverages statistical knowledge and machine learning theory as mentioned by Xie *et al.,* [11]. To determine the face classifier's model parameters, it gathers information from a vast number of face and non-face samples. This classifier then receives the image to be identified as input in order to distinguish between the two kinds of patterns, thus completing the face detection process.

Convolutional neural networks (CNNs) are extensively utilized in various fields due to their strong robustness and accuracy, especially in object detection and image classification. By integrating fully connected and convolution layers, CNNs exhibit exceptional at learning capabilities and extracting features as mentioned by Zheng *et al.,* [12]. They can extract high-level characteristics of models trained, classifications and images, making them particularly great fit for face detection tasks.

Kumar *et al.,* [9] in their paper utilized the Haar-Cascade algorithm for face detection and discussed various AI-based detection methods and their associated challenges. However, it did not present any experimental results or solutions on how to deal with these issues. Sikder *et al.,* [13] proposed using viola-jones Haar Cascade process with Principal Component Analysis (PCA) techniques to gain better accuracy for face detection and recognition. MATLAB environment was used to test on four different databases Faces94, Faces95, Faces96, and grimace. However, the accuracy can be improved and false rejection rate is high.

Viola-Jones face detection algorithm is one of the most popular choices according to Sumanto *et al*., [14]. This algorithm uses Haar-like features to represent the face region and a cascade classifier based on AdaBoost training to determine whether a region contains a face. While this approach achieves good accuracy, its performance significantly declines with low-resolution images. Qi *et al.,* [15] proposed a Multi-task Convolution Neural Network (MTCNN) algorithm model for face detection, to analyze the accuracy and speed. However, the maximum accuracy achieved was less than 91% which shows there is need for improvement.

Raj *et al*., [16] introduced a new face detection approach. It utilizes a combination of RGB and YCbCr color space boundaries are combined to segment the image's skin pixels for face detection. Following the elimination of noise, facial regions are detected using Haar cascade for the eyes and overall geometric features of human faces. However, the accuracy was very low to be used on real devices for face detection. Koshy and Mahmood [17] developed an innovative deep architecture for recognition of face liveness in video frames. This method classifies a video sequence as real or false by first diffusing the images, then using a deep Convolutional Neural Network (CNN) and an LSTM. While the combination is not new for these two LSTM and CNN, it is novel to integrate it with diffusion, which has shown to be most effective method for single live image. However, when multiple live faces are presented, it tends to give false result.

In Sambhe and Deorankar [18] research, the objective is to compare different facial recognition and detection techniques and offer a comprehensive answer for image-based facial recognition and detection. The aim is to achieve high accuracy and a faster response rate, serving as a preliminary step in video-based police investigations. However, no practical solution was shown in the paper. Thai and Nguyen [19] presented an effective technique for face detection in a surveillance system is their research. The suggested approach blends contemporary and conventional computer vision methods. Several datasets have been used to test the method's efficacy and accuracy in face localization. 26.85 frames per second and 96.66% are attained as a result.

In this paper we will use for face detection, Convolution Neural Network technique and Haar cascade, to bridge the gap in accurate identification, this research aims to significantly prevent unauthorized identification and impersonation. The goal of this study is to improve the accuracy of the face detection using Convolution Neural Network and Haar cascade. In the method, we firstly use a public face images datasets available online Faces94, Faces95, Faces96, and Grimace datasets. Pre-process the datasets in order to lessen the impact of occlusion and noise interference on the detection outcomes. Second, feature extraction for classification and detection is carried out using a face detection technique based on convolutional neural networks and Haar cascades. Lastly, to evaluate the robustness and accuracy of the suggested technique, an experimental validation is carried out using pictures from the Faces94, Faces95, Faces96, and Grimace datasets will be used to experiment on.

2. Methodology

This section describes the implementation of the Haar cascade and Convolutional Neural Network (CNN) models used for face detection. First, images for identification are converted to grayscale with a resolution of 128x128 pixels. Using the pre-trained Haar-cascade function (haarcascade_frontalface_default.xml), initial face detection is performed, followed by processing with a convolution kernel. The CNN model comprises five convolutional layers (Conv2D) with increasing filter sizes of 32, 64, 128, 256, and 512, each with a 3x3 filter size and a ReLU activation function, enabling the network to learn progressively complex features as depth increases.

Each Conv2D layer is followed by a MaxPooling layer, which down samples the output by selecting the maximum value within each 2x2 window, effectively reducing spatial dimensions by half. A Dropout layer follows each pooling layer, randomly setting 25% to 50% of input units to zero during training to prevent overfitting. The output is then passed through a Flatten layer, which transforms the 3D feature maps into a 1D vector for input to the subsequent fully connected (dense) layers. Three dense layers follow, with the first two containing 128 and 512 units, respectively, using ReLU activation to further refine the features. The final dense layer, equal in units to the number of target classes, applies softmax activation to yield probability distributions for classification.

2.1 Haar Cascade

The method known as the Haar cascade is intended to identify objects in images, irrespective of their size or location. It's a real-time, somewhat straightforward algorithm. Paul Viola and Michael Jones [20]presented the technique—known as Haar feature-based cascade classifiers. This method is based on machine learning, where a vast collection of positive and negative images is used to train a cascade function. Haar features may accurately capture the gray variance of a facial image since they are formed based on the difference in image pixel gray values by Zhang *et al.,* [21].

2.2 Convolution Neural Network Model (CNN)

A convolutional neural network (CNN) is a specialized kind of feed-forward neural network designed to process large images effectively by Panagakis *et al.,* [22]. It uses artificial neurons that reply to specific parts of their surrounding units. CNNs consist of multiple hidden layers, each composed of two-dimensional planes of neurons that operate independently by some researchers [23,24]. The input to a CNN is typically a two-dimensional image, and the architecture is built to include an embedded feature extraction module. The CNN input layer, convolutional layer, pooling layer, fully connected layer, and output layer comprise its fundamental architecture as shown in Figure 1.

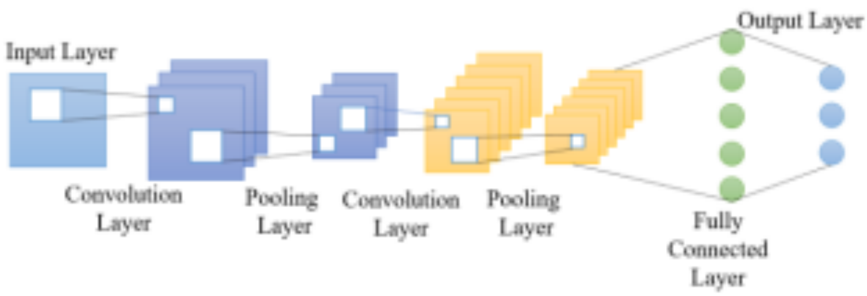

Fig. 1. Basic Convolution Neural Network architecture [23]

Each component in a Convolutional Neural Network (CNN) plays a critical role in the network's overall functionality and accuracy, contributing to the network's ability to recognize and classify complex patterns as mentioned by Wen and Abdullah [25]. The **Input Layer** initiates the process by receiving raw data, often images represented by pixel values, setting the stage for feature extraction. The **Convolutional Layer** is the core feature extraction component, where convolutional kernels (filters) slide across the input image to detect local patterns. Each kernel is designed to capture specific features—such as edges, textures, and shapes—and increasing the number of kernels enhances the network's capacity to capture intricate details within the data. Following this, the **Pooling Layer** further refines the feature maps produced by the convolutional layer, typically through max-pooling, which reduces dimensionality by selecting the maximum value within a specified region. This layer serves to retain essential features while minimizing spatial information, and most CNNs utilize at least two pooling and convolutional layers to achieve a comprehensive extraction of features.

The **Fully Connected Layer** (also known as dense layers) consolidates the high-level abstractions from the preceding layers, normalizing and integrating them to produce a probability distribution for various classes or categories. In this layer, each neuron is linked to all neurons in the preceding and subsequent layers, forming a fully connected network that allows the model to learn complex patterns and relationships within the data. Lastly, the **Output Layer** adapts its number of neurons to the specific classification task requirements, where each neuron corresponds to a class in the data. By calculating probabilities, this layer enables the network to make predictions or classifications with precision. The systematic interaction of these layers—from initial data input to final output—results in a robust model architecture capable of accurately detecting and classifying complex visual patterns, as shown in Figure 1.

2.3 Face Detection

The Faces94, Faces95, Faces96, and Grimace datasets are the sources of facial training and testing images. The Faces96 database has an image size of 196 by 196 pixels, whereas the Face94, Face95, and grimace datasets have an image size of 180 by 200 pixels. These datasets are used in the suggested idea's evaluation.

2.3.1 Pre-processing

The images in the datasets are converted to gray images of uniform size 128 × 128 pixels. Additionally, normalization is applied within the range of 0 -1 in the pixel values.

2.3.2 Face Detection Using CNN

There are six convolution neural networks in the general architecture by Li *et al.,* [26]. While the three CNN networks for calibration are used to adjust the frames of face regions, the main function of the other three of these networks is to categorize non-faces and faces. To get a detection frame, individual CNN network moves over the image with a particular step size. In real-world applications, the input image is scaled into a pyramidal shape, with the network model receiving input from each layer of the pyramid.

To ease the load on the next network layer, each Convolution Neural Network filters away a portion of the error detection frames. Three offset variables are used by each calibration Convolution Neural Network to adjust the detection frame: x_n for horizontal translation, y_n for vertical translation, and s_n for height to width scaling. The candidate box's expression is (x, y, w, h). (x, y) are the coordinates that indicates the upper left point, while the candidate frame's (w, h) indicates the width and height. This Eq. (1) is used to modify the detection frames' control coordinates:

$$x - \frac{x_n w}{s_n}, y - \frac{y_n h}{s_n}, \frac{w}{s_n}, \frac{h}{s_n} \tag{1}$$

The CNN is utilized to verify if the candidate's frames contain a face. If a face is detected, the network then performs key point detection and posture analysis on the identified face. For a face to be considered a human face all the features in Figure 2 needs to be identified.

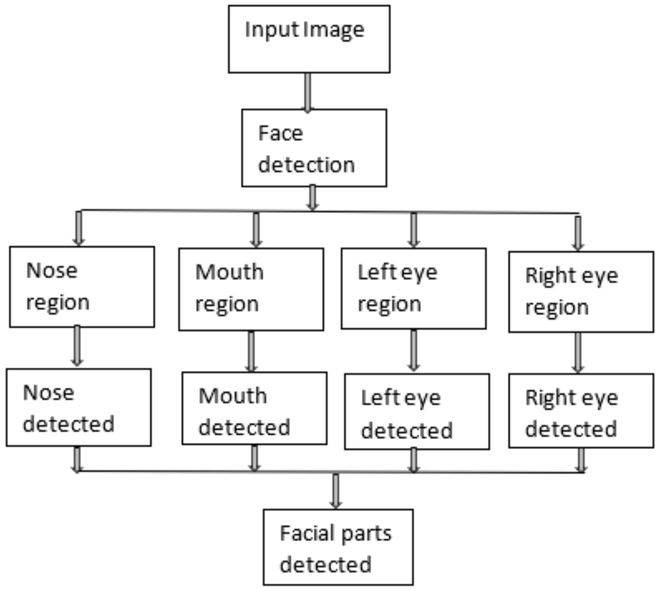

Fig. 2. Process of facial part detection [13]

2.3.3 Training Network

Three main tasks are performed by the cascaded neural network: facial landmark localization, bounding box regression, and face classification by Peng and Gopalakrishnan [27].

i. Face Classification

This task involves distinguishing between face and non-face samples, treated as a binary classification problem. Cross-entropy-loss is applied for each sample, calculated as Eq. (2)

$$l_{det} = -(y_{det}\log(p_i) + (1-y_{det})\log(1-p_i)) \quad (2)$$

where Pi indicates the likelihood that the sample contains a human face as predicted by the network, with true label represented as Eq. (3)

$$y_{det} \in \{0,1\} \quad (3)$$

ii. Bounding Box Regression

This task predicts the deviation of the proposal box from the true value. Regression analysis is performed and for each sample, the $smooth_{L1}$ loss is applied Eq. (4)

$$L_i^{box} = \sum_{i\in\{x,y,w,h\}} s\,mooth_{L1}\left(y_i^{box} - \hat{y}_i^{box}\right) \quad (4)$$

with the $smooth_{L1}$ loss defined as Eq. (5)

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & if|x|\leq 1 \\ |x| - 0.5 & otherwise \end{cases} \quad (5)$$

of these, the regression target's coordinate, derived from the network, is $\hat{y}_i^{box}$, while the real coordinate is y_i^{box}. Four variables make up the coordinates: the height and width from the upper left corner, as well as the point coordinates (x, y).

iii. Facial Landmark Localization

In this task, similar to bounding box regression, the problem is handled like a regression task, minimizing the Euclidean losses for each sample Eq. (6)

$$L_i^{landmark} = \left\|\hat{y}_i^{landmark} - y_i^{landmark}\right\|_2^2 \tag{6}$$

Here, the coordinated of the face acquired from the network are represented by, $\hat{y}_i^{landmark}$, while representation of the real coordinates, is $y_i^{landmark}$. Given that the face has five major points – corners of the mouth, two eyes and nose in Eq. (7)

$$y_i^{landmark} \in R^{10} \tag{7}$$

2.3.4 Optimizer

The CNN is trained jointly for multiple tasks using a joint loss function. The momentum-based stochastic gradient descent is used for overall optimization.

2.3.5 Learning Rate

The learning rate of the stochastic gradient descent algorithm is adjusted to modify the update rate of the network's weight parameters.

3. Results

The following information serves as the foundation for the research's experimental results:

3.1 Dataset

The public datasets used for this research have also been used by other researchers like Sikder *et al.,* [13] and Aminu and Ahmad [28]. 180 by 200 pixel photos of 153 different people are available in the Faces94 database [29]. The pictures have a green backdrop, different head slants and tilts, little face movement, and numerous stances. The 180 by 200-pixel photos in the Faces95 database [30] belong to 72 different people. These pictures have frontal positions because of flames and a red screen background. The 152 people's photos in the Faces96 database [31] have a 196 by 196-pixel resolution. Finally, the Grimace database [32] contains 180 by 200 pixel-resolution photos of 18 different people. All the images were scaled down to a uniform resolution of 128 by 128 pixels so as to lessen photos size while maintaining the quality of the images.

3.2 Haar Cascade

The Haar Cascade classifier will be used for the initial detection, to detect the region of interest in the images that are likely to have face, as it is efficient and lightweight. It stores the coordinates of the detected regions, and passes it to the CNN model.

3.3 CNN model

The identified regions by Haar Cascade, each pass through a pre-trained CNN model for face detection. It will then confirm whether region contains a face or not, discarding the false-negative and possibly providing a more precise bounding boxes, so as to increase the accuracy.

Training and testing: The training and testing dataset taken from the four dataset face94, face95, face96 and Grimace, 20% of each dataset is used for testing and 80% is used for the training. The Faces94, Faces95, Faces96, and Grimace databases were used in this work to conduct facial parts detection, which are described in Table 1. Eq. (8) for detection accuracy is presented below

$$Accuracy = 100 - \frac{FAR+FRR}{2} \tag{8}$$

Table 1
Face detection accuracy comparison

Dataset	FAR		FRR		Accuracy	
	[13]	proposed	[13]	proposed	[13]	proposed
Faces94	4.67%	0.01%	6.53%	1.63%	94.40%	98.37%
Faces95	0.00%	0.04%	7.41%	2.78%	96.29%	97.22%
Faces96	3.53%	0.02%	5.73%	2.48%	95.37%	97.52%
Grimace	0.00%	0.00%	7.19%	0.00%	96.41%	100%
Essex face database [28]	-	0.01%	-	1.72%	95.34% ±0.41	98.27%

Table 1 presents the False Acceptance Rate (FAR), False Rejection Rate (FRR), and overall accuracy of the proposed study. A comparison with the findings of Sikder *et al.,* [13] and Sambhe and Deorankar [28], who employed Haar Cascade and Principal Component Analysis (PCA) and Locality Preserving Partial Least Squares Discriminant Analysis (LPPLS-DA) respectively, indicates that the proposed study consistently outperforms their methodology across all evaluated metrics. The detection rates for each database are illustrated in the accuracy graph depicted in Figure 3.

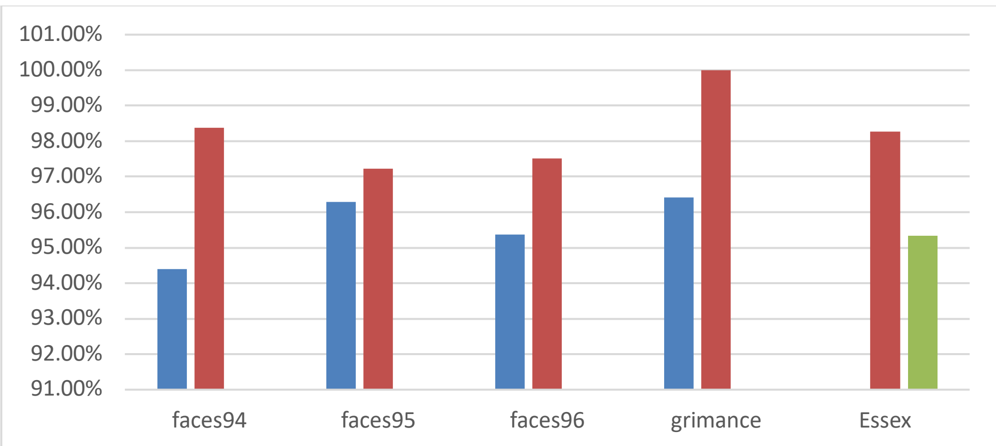

Fig. 3. Bar chart for accuracy

This bar chart visually contrasts the accuracy of the proposed method (highlighted in red) with that of Sikder *et al.,* (shown in blue) and Sambhe and Deorankar (shown in green), both of whom utilized the same public datasets (Faces94, Faces95, Faces96, and Grimace). The results clearly demonstrate that the proposed method achieves superior accuracy compared to both the Sikder *et al.,* approach, Sambhe and Deorankar and other traditional methods discussed in this paper.

4. Conclusions

In conclusion, the proposed face detection system utilizes Haar Cascade and Convolutional Neural Network (CNN) algorithms, which have been rigorously tested on four distinct databases: Faces94, Faces95, Faces96, and Grimace. This system has achieved impressive facial detection accuracy rates of 98.37%, 97.22%, 97.52%, and 100% for the Faces94, Faces95, Faces96, and Grimace databases, respectively. This research successfully addresses the critical gap in accurate identification, significantly mitigating the risks associated with unauthorized identification and impersonation. The findings demonstrate that the proposed system performs more effectively than both the Haar Cascade alone and the combination of Haar Cascade with PCA for face detection.

Looking ahead, there are plans to enhance the system's capabilities to include human facial expression detection and recognition, and to evaluate its performance on real mobile devices to further bolster security applications. While this research has focused primarily on face detection using Haar Cascade and CNN, future work will explore a broader range of algorithms and metrics, applying these advancements in real-world scenarios. Additionally, there is an intention to evolve the system into a comprehensive face recognition model that can be employed for authentication purposes, thereby preventing unauthorized access.

Acknowledgement
We would like to thank Petroleum Technology Development Fund (PTDF) under ministry of petroleum, Nigeria, for providing the funds and Universiti Putra Malaysia (UPM) for providing the equipment’s to make this research come to life.

References
[1] Zhang, Xue, Hailiang Zhao, and Yu Dai. "Small Scale Face Detection Method in Dense Scene." In 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), vol. 5, pp. 1826-1830. IEEE, 2021. https://doi.org/10.1109/IAEAC50856.2021.9391096
[2] Khalili, Sanaz, and Ali Shakiba. "A face detection method via ensemble of four versions of YOLOs." In 2022 International Conference on Machine Vision and Image Processing (MVIP), pp. 1-4. IEEE, 2022. https://doi.org/10.1109/MVIP53647.2022.9738779
[3] Jiang, Xinbei, Tianhan Gao, Zichen Zhu, and Yukang Zhao. "Real-time face mask detection method based on YOLOv3." Electronics 10, no. 7 (2021): 837. https://doi.org/10.3390/electronics10070837
[4] Minaee, Shervin, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. "Image segmentation using deep learning: A survey." IEEE transactions on pattern analysis and machine intelligence 44, no. 7 (2021): 3523-3542. https://doi.org/10.1109/TPAMI.2021.3059968
[5] Wibowo, Moh Edi, Ahmad Ashari, Ardacandra Subiantoro, and Wahyono Wahyono. "Human face detection and tracking using retinaface network for surveillance systems." In IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society, pp. 1-5. IEEE, 2021. https://doi.org/10.1109/IECON48115.2021.9589577
[6] Kumar, A. Senthil, N. Pranavi, and S. Gowri Priya Dharshini. "Emotions Based Voice Supportive Model Using SVM." In 2021 International Conference on Engineering and Emerging Technologies (ICEET), pp. 1-4. IEEE, 2021. https://doi.org/10.1109/ICEET53442.2021.9659702
[7] Zhuang, Yueting, Ruogui Xiao, and Fei Wu. "Key issues in video summarization and its application." In Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, vol. 1, pp. 448-452. IEEE, 2003. https://doi.org/10.1109/ICICS.2003.1292492

[8] Adouani, Amal, Wiem Mimoun Ben Henia, and Zied Lachiri. "Comparative study of face detection methods in spontaneous videos." In 2019 International Conference on Control, Automation and Diagnosis (ICCAD), pp. 1-6. IEEE, 2019. https://doi.org/10.1109/ICCAD46983.2019.9037883
[9] Kumar, A Senthil, P Sujith Sai, B. Parvathi, H.S. Manasa, and Kanishka. S. "Face Detection Approaches Using AI," In 2023 Int. Conf. Data Sci. Netw. Secur. ICDSNS 2023, pp. 1–6, 2023. https://doi.org/10.1109/ICDSNS58469.2023.10245563
[10] Jiang, Enjie. "A review of the comparative studies on traditional and intelligent face recognition methods." In 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), pp. 11-15. IEEE, 2020. https://doi.org/10.1109/CVIDL51233.2020.00010
[11] Xie, Rong, Qingyu Zhang, Enyuan Yang, and Qiang Zhu. "A method of small face detection based on CNN." In 2019 4th International Conference on Computational Intelligence and Applications (ICCIA), pp. 78-82. IEEE, 2019. https://doi.org/10.1109/ICCIA.2019.00022
[12] Zheng, Fang, Xiaoyan Wang, and Li Wei. "A Convolutional Neural Network based Method for Masked Face Detection." In 2023 4th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), pp. 379-383. IEEE, 2023. https://doi.org/10.1109/ICBASE59196.2023.10303057
[13] Sikder, Juel, Ratna Chakma, Rana Jyoti Chakma, and Utpol Kanti Das. "Intelligent face detection and recognition system." In 2021 International Conference on Intelligent Technologies (CONIT), pp. 1-5. IEEE, 2021. https://doi.org/10.1109/CONIT51480.2021.9498291
[14] Wijonarko, Bambang, Muhammad Qommarudin, Aji Sudibyo, Pudji Widodo, and Afit Muhammad Lukman. "Viola-Jones Algorithm for Face Detection using Wider Face Dataset." In 2022 10th International Conference on Cyber and IT Service Management (CITSM), pp. 1-4. IEEE, 2022. https://doi.org/10.1109/CITSM56380.2022.9935830
[15] Qi, Rong, Rui-Sheng Jia, Qi-Chao Mao, Hong-Mei Sun, and Ling-Qun Zuo. "Face detection method based on cascaded convolutional networks." IEEE Access 7 (2019): 110740-110748. https://doi.org/10.1109/ACCESS.2019.2934563
[16] Raj, Akhilesh, Soumay Gupta, and Nishchal K. Verma. "Face detection and recognition based on skin segmentation and CNN." In 2016 11th International Conference on Industrial and Information Systems (ICIIS), pp. 54-59. IEEE, 2016. https://doi.org/10.1109/ICIINFS.2016.8262907
[17] Koshy, Ranjana, and Ausif Mahmood. "Enhanced Anisotropic Diffusion-based CNN-LSTM Architecture for Video Face Liveness Detection." In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 422-425. IEEE, 2020. https://doi.org/10.1109/ICMLA51294.2020.00074
[18] Akhil Awdhutrao Sambhe, and A V Deorankar. "Face Detection And Recognization System." In Proc. - 2022 4th Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2022, pp. 1175–1179. 2022, doi: 10.1109/ICAC3N56670.2022.10074142. https://doi.org/10.1109/ICAC3N56670.2022.10074142
[19] Thai, Tin Trung, and Duc Tuan Nguyen. "Face detection method in surveillance systems using haar feature and deep neural network." In 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), pp. 434-438. IEEE, 2019. https://doi.org/10.1109/NICS48868.2019.9023868
[20] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I-I. Ieee, 2001. https://doi.org/10.1109/CVPR.2001.990517
[21] Zhang, Xinman, Dongxu Cheng, Pukun Jia, Yixuan Dai, and Xuebin Xu. "An efficient android-based multimodal biometric authentication system with face and voice." *Ieee Access* 8 (2020): 102757-102772. https://doi.org/10.1109/ACCESS.2020.2999115
[22] Panagakis, Yannis, Jean Kossaifi, Grigorios G. Chrysos, James Oldfield, Mihalis A. Nicolaou, Anima Anandkumar, and Stefanos Zafeiriou. "Tensor methods in computer vision and deep learning." *Proceedings of the IEEE* 109, no. 5 (2021): 863-890. https://doi.org/10.1109/JPROC.2021.3074329
[23] Gu, Hao, Yu Wang, Sheng Hong, and Guan Gui. "Blind channel identification aided generalized automatic modulation recognition based on deep learning." *IEEE Access* 7 (2019): 110722-110729. https://doi.org/10.1109/ACCESS.2019.2934354
[24] O'shea, Timothy, and Jakob Hoydis. "An introduction to deep learning for the physical layer." *IEEE Transactions on Cognitive Communications and Networking* 3, no. 4 (2017): 563-575. https://doi.org/10.1109/TCCN.2017.2758370
[25] Khaw, Li Wen, and Shahrum Shah Abdullah. "Mri Brain Image Classification Using Convolutional Neural Networks and Transfer Learning." *Journal of Advanced Research in Computing and Applications* 31, no. 1 (2023): 20-26. https://doi.org/10.37934/arca.31.1.2026
[26] Li, Haoxiang, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. "A convolutional neural network cascade for face detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5325-5334. 2015. https://doi.org/10.1109/CVPR.2015.7299170
[27] Peng, Bikang, and Anilkumar Kothalil Gopalakrishnan. "A face detection framework based on deep cascaded full convolutional neural networks." In *2019 IEEE 4th International Conference on Computer and Communication*

Systems (ICCCS), pp. 47-51. IEEE, 2019. https://doi.org/10.1109/CCOMS.2019.8821692
[28] Zhang, Xue, Hailiang Zhao, and Yu Dai. "Small Scale Face Detection Method in Dense Scene." In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, pp. 1826-1830. IEEE, 2021. https://doi.org/10.1109/IAEAC50856.2021.9391096
[29] "Libor Spacek's Facial Images Databases." *Center for Machine Perception*. Accessed: Jul 10, 2024. [Online]
[30] "Libor Spacek's Facial Images Databases." *Center for Machine Perception*. Accessed: Jul 11, 2024. [Online]
[31] "Libor Spacek's Facial Images Databases." *Center for Machine Perception*. Accessed: Jul 12, 2024. [Online]
[32] "Libor Spacek's Facial Images Databases." *Center for Machine Perception*. Accessed: Jul 13, 2024. [Online]