STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 2, 2024
May–August
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Research JoAIRA

Face Detection and Recognition Using MTCNN and FaceNet

Sahil Bisht[1,*], Sonal Fatangare[2], Apeksha Patil[1], Aakanksha Panadi[1], Dev Kulkarni[1]

Abstract

Face detection and face recognition are major tasks in the field of computer vision with several real-world applications and many products being developed in the same field. This study gives a detailed implementation of the product that is developed for accurate detection and recognition of faces along with audio output of the face detected. This development would act as a base for a few future products that can be developed for the visually impaired community. For the development of the product, a thorough survey was done from classical methods, such as eigenfaces and Fisher faces, to cutting-edge deep learning techniques, such as advanced convolutional neural networks analyzing various face detection and recognition algorithms and the challenges associated with them, like changing lighting scenarios, obstructed views, and pose fluctuations. Generally, Convolutional Neural Networks (CNNs) are employed in developing such products. However, instead of utilizing general CNN, this product utilizes specific algorithms, namely Multi-Task Cascaded Convolutional Neural network (MTCNN) for face detection and FaceNet for face recognition respectively which are based on CNN itself. The model is trained on images of different persons which helps in extracting the required facial features. Finally, when the program is run, the detected face is shown with a green bounding box and the name of the person detected at the right bottom of the bounding box. Also, an audio output is provided by the system giving the name of the detected person. If no person is detected, then "Unknown" is the audio output given. This product gives a high accuracy due to extra filter layers in both MTCNN and FaceNet which helps in refining the training process and improving the efficiency of the system.

Keywords: Face recognition, face detection, convolutional neural network (CNN), deep learning, MTCNN, FaceNet, industrial applications

INTRODUCTION

In recent years, the growth of digital imaging technologies and data have given rise to the development and deployment of a number of computer vision techniques. Among these techniques, face detection and recognition have been some of the most important and crucial components, playing a significant role in real-world applications. The automated identification and acknowledgment of faces in images have given rise to development of various solutions which are now helpful in industries such as law enforcement, healthcare, marketing, and entertainment. This product aims in utilizing specific algorithms based on Convolutional Neural Network (CNN) for face detection and recognition purposes which will help in overcoming various challenges, thereby contributing to the advancement of responsible applications in the field of computer vision.

***Author for Correspondence**
Sahil Bisht
E-mail: sahil.bisht0911@gmail.com

[1]Student, Department of Computer Engineering, Rasiklal M. Dhariwal Sinhgad Technical Institutes Campus, Pune, Maharashtra, India
[2]Assistant Professor, Department of Computer Engineering, Rasiklal M. Dhariwal Sinhgad Technical Institutes Campus, Pune, Maharashtra, India

Received Date: May 28, 2024
Accepted Date: July 05, 2024
Published Date: July 10, 2024

Citation: Sahil Bisht, Sonal Fatangare, Apeksha Patil, Aakanksha Panadi, Dev Kulkarni. Face Detection and Recognition Using MTCNN and FaceNet. Journal of Artificial Intelligence Research & Advances. 2024; 11(2): 132–140p.

Face Detection

Face detection is like breaking down a picture into two parts: one with faces and the other without. It scans the visual data like videos or images and locates the faces in them. It creates a bounding box on the faces which is like a boundary determining that the content inside the bounding box is a face. Face detection process firstly locates the eyes in a face as they are easier to locate as compared to other facial features like eyebrows, mouth or nose. Initially, in a face detection process, the algorithm first considers the visual data and preprocesses it, then it scans the visual data for faces, whether present or not. Finally, it detects the face in the picture. The detection process varies depending on the type of methodologies being employed by various face detection algorithms.

Face Recognition

Face recognition is the next step after face detection. It basically means to recognize whose face is detected based on the facial features of various individuals. Approaches utilized for face recognition are [1]:

Holistic Related Techniques

In this approach, entire face is given as input. Eigenfaces, PCA, LDA and Independent Component Analysis are some examples of holistic related techniques.

Structural Techniques

In this approach, unlike holistic techniques, specific facial features are extracted first. These features like eyes, nose and mouth are extracted and then their coordinates are fed into structural classifiers.

Hybrid Techniques

As the name suggests, these techniques are culminations of both holistic and structural techniques, especially being used in analyzing 3D images. Except for specific features, these techniques also consider facial contours and features which helps in improving the overall accuracy of these techniques and improving their efficiency for real-world applications.

Multi-Task Cascaded Convolutional Network (MTCNN)

MTCNN stands for Multi-task Cascaded Convolutional Networks. It is a renowned algorithm in the field of face detection and computer vision. It basically works in three steps: face detection, refining bounding box parameters, and locating specific facial features [2].

Stage 1: Face Detection

1. Initially, the input image is resized in order to help detect faces of different sizes.
2. The image is given input to P-Net, a compact convolutional neural network (CNN), which creates bounding boxes around face, each associated with a probability score.
3. There might be overlapping bounding boxes in the image. Only the most confident bounding box is considered in such cases along with the non-overlapping ones.

Stage 2: Bounding Box Regression

1. The bounding boxes found in the previous stage using P-Net are resized to a fixed size.
2. R-Net is a more complex CNN that takes resized images as input. It enhances the bounding boxes found using P-Net, increasing the accuracy of the bounding box coordinates. It then assigns a proper probability score to each enhanced bounding box.
3. Similar to Stage 1, Non-Maximum Suppression (NMS) is used to remove the extra bounding boxes that overlap, keeping only the most confident and distinct ones.

Stage 3: Facial Landmark Localization

1. Again, the bounding boxes found in the previous stage using R-Net are resized to a fixed size.
2. O-Net, also a CNN that takes resized images as input. It is used to locate specific facial features, giving the coordinates of eyes, nose, and mouth, inside each bounding box. It refines the bounding box coordinates and then assigns a proper probability score for the final prediction.

3. Lastly, again Non-Maximum Suppression (NMS) is used to remove the extra bounding boxes that overlap, keeping only the most confident and distinct ones.

FaceNet

FaceNet is a deep learning algorithm used for face recognition, known to generate precise embeddings of facial features. Similar to MTCNN, this algorithm is also based on a deep convolutional neural network (CNN) and uses a triplet loss function in its training process to generate face embeddings [3]. Following is the working of FaceNet [3]:

1. Initially, we use a large dataset of facial images. These images consist of various poses, lighting conditions and facial expressions.
2. In the next step, data preprocessing is done in which facial images are cropped as per required size and aligned so that the face remains the main focus of the image. Pixel values of images are normalized in order to make the data consistent for processing.
3. Feature extraction is carried out for which FaceNet uses the inception module which helps in determining the hierarchical features. Consistent embeddings are generated for each face detected in the image.
4. The triplet loss function plays a crucial role in training the model using FaceNet. It considers three images: anchor image, positive image (similar to anchor image) and a negative image (dissimilar to anchor image). Main aim is to minimize the distance between anchor and positive image and maximize the gap between anchor and negative image. The loss function as referred by Singh and Singh is formulated as [4]:

   $$L(A, P, N) = \max(0, ||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + margin)$$

   where f(.) represents the embedding function.

5. The FaceNet model is trained using the above triplet loss function and the input data in the next step. Good amount of labeled data must be used in order to get accurate facial representations. Hyperparameters, such as learning rate, batch size, and margin in the triplet loss can be used.
6. Validation of the FaceNet model should be done using distinct data to ensure that it can generalize faces that were not available during training of the model.
7. Once the training of the model is finished, this face data can be used to train the model using new facial images.
8. Evaluation metrics like accuracy, precision, recall, and F1 score can be used to check the model's accuracy and precision.

RELATED WORK

Hashmi and Aqib did their research on face detection in challenging conditions [5]. They proposed a framework which has three layers developed using convolutional networks. This helped them in identifying faces and facial features in various environments with challenging conditions.

Islam *et al*. have introduced integration of machine learning in face recognition techniques [6]. This helped them in proposing a system that can handle challenging conditions. The authors utilized data having 627 individuals from Bangladesh which consist of images clicked from four different angles. CNN, Harr Cascade, Cascaded CNN, Deep CNN, and MTCNN were the machine learning techniques used by them. After creating the model and executing it, MTCNN gave an accuracy of 96.2% with the training data, which was higher than any of the techniques mentioned above. Hence, the research showed how MTCNN can be used in improving face recognition accuracy, particularly in scenarios with hardware cost constraints.

Soni and Waoo explored face detection focusing on infants, the elderly, and people with darker skin, where existing techniques usually show some challenges or issues [7]. The paper shows the evolution of face detection algorithms, from the Viola-Jones, which has acted as a base for developments for many years due to Haar-like features and AdaBoost learning. The authors have also reviewed current

face recognition algorithms like the Convolutional Neural Networks (CNNs). CNNs are recommended in computer vision as they can recognize unlabeled patterns in data.

Rongrong *et al*. showed a system that uses MTCNN and FaceNet face recognition purposes [8]. The authors of this paper have addressed challenges like variations in poses, lighting conditions and images with low resolution. The authors proposed an architecture that combines face detection, key point location, and novel roll neural network algorithms. FaceNet maps images into a Euclidean space that represents similarities in facial features, which helps in easier implementation of face recognition and its validation. The developed system has achieved high accuracy but still the authors have found some imperfections.

In 2023, Dang and Tran developed a two-step identification system that uses MTCNN and FaceNet algorithms which are based on CNN [9]. The authors enhanced their system with head pose estimation which helped in improving face recognition. Their system is basically based on AI and its self-learning capabilities. The model developed by the authors shows accuracy in the range of 92 to 95%. Due to the high accuracy obtained, the system can be applied in industrial applications and for research purposes.

METHODOLOGY

The proposed system employs a combination of machine learning and deep learning algorithms to achieve face detection and recognition as shown in Figure 1.

The algorithm starts by loading the labelled data and training the model on it. The system then launches the camera for detecting faces present before it. Once detected, these faces are surrounded by a bounding box including special facial features that are used for further identification. The system then checks for a match in the facial recognition process. If the face is matched, the system gives the output of the label that was matched in the form of both text and audio. If the face is not matched, the system tries to find the nearest match or else give unknown as the output [10–15].

The training process is accomplished using a deep learning algorithm designed for accurately recognizing face data by finding patterns and similarities in various light conditions. The network learns to minimize the distance between embeddings of images depicting the same person while maximizing the distance between embeddings of images depicting different individuals. The network then learns from the input and provides output as the label name.

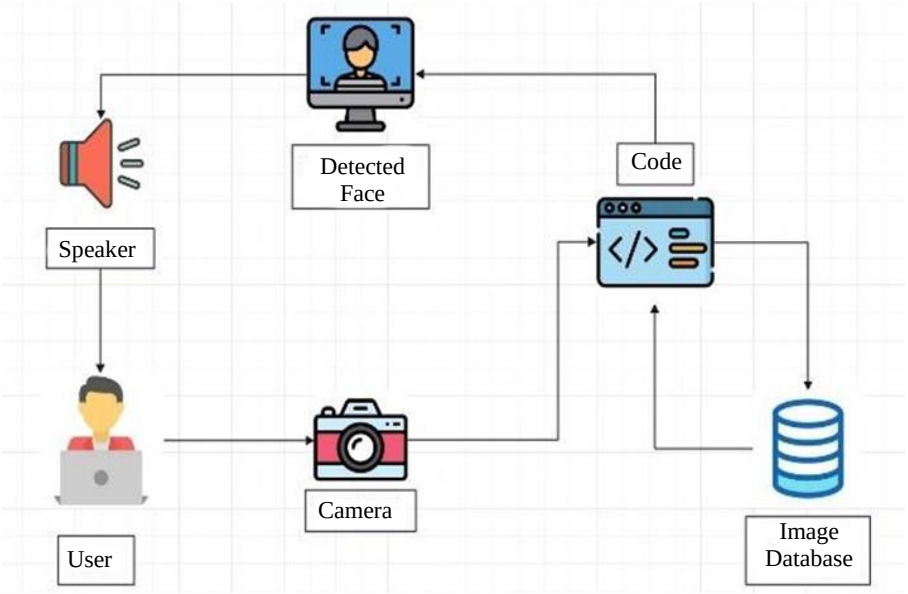

Figure 1. System architecture of face detection and recognition.

Algorithm

FaceNet is a model developed by Google researchers, which uses deep CNN architecture. During training, FaceNet learns to map faces into a high-dimensional feature space where similar faces cluster together, while dissimilar faces are far apart. This is achieved by minimizing the distance between embeddings of images depicting the same person and maximizing the distance between embeddings of images depicting different individuals. The model is pretrained on a large dataset of face images with corresponding identity labels, employing a loss function like triplet loss to optimize performance.

In operation, FaceNet generates embeddings for input faces, which are then compared to a database of known embeddings using similarity metrics like Euclidean distance or cosine similarity. A decision threshold is applied to these similarity scores to determine if the input face matches any identity in the database. If the similarity score exceeds the threshold, the system identifies the individual; otherwise, the face is considered unknown or rejected. FaceNet's success lies in its ability to learn discriminative features directly from raw pixel data, enabling it to achieve state-of-the-art performance in face recognition across diverse conditions such as pose, lighting, and facial expressions as shown in Figure 2.

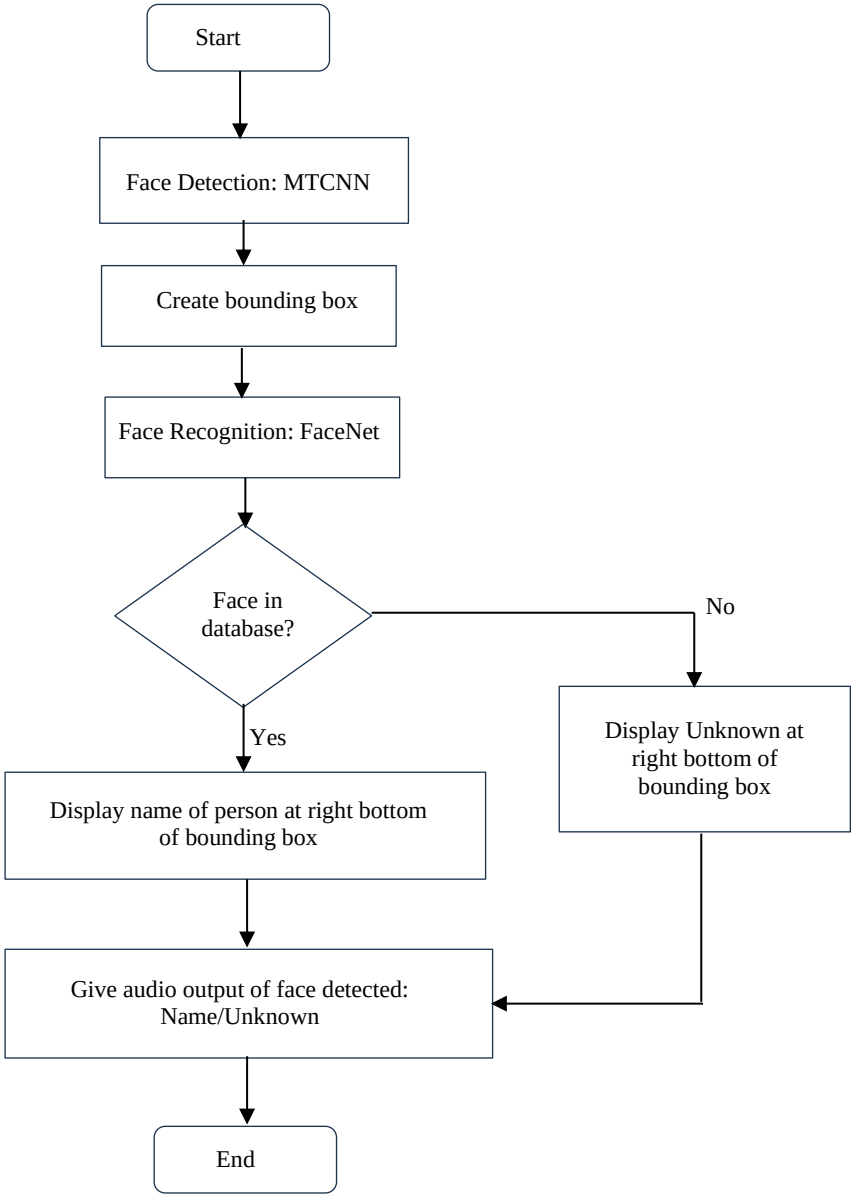

Figure 2. Flowchart of face database.

Triplet loss is a commonly used loss function in FaceNet training. Mathematically, it involves selecting triplets of face images: an anchor image of a specific identity, a positive image of the same identity (with similar facial features), and a negative image of a different identity (with dissimilar facial features) [16–20]. The loss function encourages the model to minimize the distance between the anchor and positive embeddings while maximizing the distance between the anchor and negative embeddings. This can be mathematically expressed as:

$$Triplet\ loss = max\ (0,\ d\ (A, P) - d\ (A, N) + \alpha)$$

where d represents a distance metric (e.g., Euclidean distance or cosine similarity) between embeddings, and α is a margin hyperparameter.

During training, FaceNet employs optimization algorithms such as stochastic gradient descent (SGD) or its variants (e.g., Adam optimizer) to minimize the chosen loss function (e.g., triplet loss). These algorithms iteratively update the model parameters (e.g., weights of the CNN layers) to improve the model's ability to generate discriminative embeddings.

Model Implementation

To start the system, we issue a wake-up command to the system. To register a user, we execute a manual command that instructs the face recognition module to train on the user's photos for the new labels that are to be added [21–26]. Face recognition systems are trained and implemented using the face recognition module. The face recognition module will take user photos, which will be input into convolutional neural networks (CNN) using feature descriptors. The user can now experiment with its features, by performing face recognition of both trained and new labels. The internal working of the model is depicted by a flowchart in Figure 2.

RESULTS

The results of our testing indicate that the developed system is a highly effective system that provides a convenient method for face detection and recognition. Earlier, different algorithms have been used for the purpose of face recognition like Local Binary Pattern (LBP), Viola-Jones, Haar-Cascade and DeepID, Deep Face [27–33]. Table 1 shows the accuracy score achieved by these systems along with the accuracy score achieved by the system developed using MTCNN and FaceNet algorithms.

To evaluate the effectiveness of the proposed facial recognition system, we compared its performance with existing facial recognition methods widely used in the field. The Table 1 shows the accuracy scores of different algorithms that were used in earlier systems.

The bar chart given in Figure 3 shows the visual comparison of the give accuracy scores of previously used algorithms for face detection and face recognition with MTCNN and FaceNet that are used in the proposed system mentioned in this study.

The proposed facial recognition system developed using MTCNN and FaceNet algorithms for face detection and face recognition respectively, demonstrated remarkable performance, achieving an accuracy score of 95%. This result highlights the effectiveness and reliability of the proposed system in accurately recognizing user's faces and validating their identities.

Table 1. Accuracy score of different algorithms.

Sr No.	Algorithms	Accuracy Score (%)
1	Haar-Cascade	68.16
2	Viola-Jones	90
3	LBP	84
4	DeepID, Deep Face	92
5	MTCNN, FaceNet	95

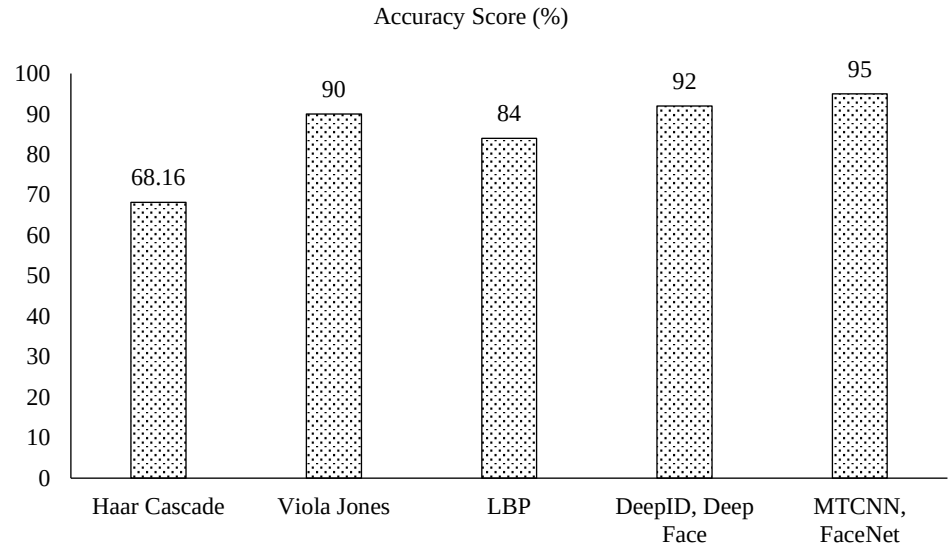

Figure 3. Comparison of algorithms.

CONCLUSION

In conclusion, this research work provides a detailed implementation of face detection and recognition, coupled with audio output functionality. By using state-of-the-art techniques such as Multi-Task Cascaded Convolutional Neural Network (MTCNN) for face detection and FaceNet for face recognition, the developed system offers enhanced performance compared to traditional technologies.

The utilization of specific algorithms for face detection and recognition instead of going for general Convolutional Neural Network (CNN), along with the extra filter layers provided by MTCNN and FaceNet, have helped in increasing the system's accuracy and improving its efficiency. The visual and auditory feedback provided by the system helps in improving the user experience enabling the system to be used by different user groups.

Lastly, the system not only gives a solution for face detection and face recognition but also shows the potential to be used in developing innovative solutions in the field of computer vision along with products that will help the visually impaired community.

Future Scope

Various opportunities lie in the future for the development of face detection and recognition systems. The system implemented in this study utilizes specific algorithms with high accuracy for the face detection and recognition purposes which makes it more accurate and precise than the traditional technologies.

The developed system can be further improved to integrate it in wearable technologies. Enhanced accessibility features like voice commands and gesture-based controls can be incorporated in the existing system which would lead to development of products that would be easily accessible to the visually impaired community. Further, the system can be modified such that it can be deployed on mobile devices which would make the system more user-friendly and users can benefit from its capabilities on-the-go.

Acknowledgement

Thanks to Mrs. Vina Lomte, HOD, RMDSSOE and Mrs. Sonal Fatangare, Assistant Professor, RMDSSOE our project guides, for providing us with continuous support, patience and motivation and with valuable guidance and insights throughout the research process.

REFERENCES

1. Tyagi R, Tomar GS, Baik N. A survey of unconstrained face recognition algorithm and its applications. Int J Secur Appl. 2016 Dec 1; 10(12): 369–76.
2. Ku H, Dong W. Face recognition based on mtcnn and convolutional neural network. Front Signal Process. 2020 Jan; 4(1): 37–42.
3. Manzoor S, Kim EJ, Joo SH, Bae SH, In GG, Joo KJ, Choi JH, Kuc TY. Edge deployment framework of guardbot for optimized face mask recognition with real-time inference using deep learning. IEEE Access. 2022 Jul 25; 10: 77898–921.
4. Singh AP, Singh V. Infringement of Prevention Technique against Keyloggers using Sift Attack. In 2018 IEEE International Conference on Advanced Computation and Telecommunication (ICACAT). 2018 Dec 28; 1–4.
5. Hashmi SA. Face Detection in Extreme Conditions: A Machine-learning Approach. arXiv preprint arXiv:2201.06220. 2022 Jan 17.
6. Islam MT, Ahmed T, Rashid AR, Islam T, Rahman MS, Habib MT. Convolutional neural network based partial face detection. In 2022 IEEE 7th International conference for Convergence in Technology (I2CT). 2022 Apr 7; 1–6.
7. Soni L, Waoo A. A Review of Recent Advances Methodologies for Face Detection. Int J Curr Eng Technol. 2023; 13(02): 86–92.
8. Jin Rongrong, et al. (2021). Face recognition based on MTCNN and Facenet. [Online]. Available from: https://jasonyanglu.github.io/files/lecture_notes/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0_2020/Project/Face%20Recognition%20Based%20on%20MTCNN%20and%20FaceNet.pdf
9. Dang TV, Tran HL. A Secured, Multilevel Face Recognition based on Head Pose Estimation, MTCNN and FaceNet. Journal of Robotics and Control (JRC). 2023 Jun 20; 4(4): 431–7.
10. Khan MZ, Harous S, Hassan SU, Khan MU, Iqbal R, Mumtaz S. Deep unified model for face recognition based on convolution neural network and edge computing. IEEE Access. 2019 May 23; 7: 72622–33.
11. Peralta B, Figueroa A, Nicolis O, Trewhela Á. Gender identification from community question answering avatars. IEEE Access. 2021 Nov 23; 9: 156701–16.
12. Wang Dalin, Rongfeng Li. Enhancing Accuracy of Face Recognition in Occluded Scenarios with OAM-Net. IEEE Access. 2023; 11: 117297–117307.
13. Li L, Liu M, Sun L, Li Y, Li N. ET-YOLOv5s: toward deep identification of students' in-class behaviors. IEEE Access. 2022 Apr 22; 10: 44200–11.
14. Tripathy R, Daschoudhury R. Real-time face detection and tracking using haar classifier on soc. International Journal of Electronics and Computer Science Engineering (IJECSE). 2014; 3(2): 175–84.
15. Hussain D, Ismail M, Hussain I, Alroobaea R, Hussain S, Ullah SS. Face Mask Detection Using Deep Convolutional Neural Network and MobileNetV2-Based Transfer Learning. Wirel Commun Mob Comput. 2022; 2022(1): 1536318.
16. Kaur G, Sinha R, Tiwari PK, Yadav SK, Pandey P, Raj R, Vashisth A, Rakhra M. Face mask recognition system using CNN model. Neuroscience Informatics. 2022 Sep 1; 2(3): 100035.
17. Mo H, Kim S. A deep learning-based human identification system with wi-fi csi data augmentation. IEEE Access. 2021 Jun 25; 9: 91913–20.
18. Coe J, Atay M. Evaluating impact of race in facial recognition across machine learning and deep learning algorithms. Computers. 2021 Sep 10; 10(9): 113.
19. Rajyalakshmi V, Lakshmanna K. Intelligent face recognition based multi-location linked IoT based car parking system. IEEE Access. 2023 Aug 7; 11: 84258–84269.
20. Raghavendra M, Neha R, et al. Missing Child Identification using Convolutional Neural Network. Int J Res Appl Sci Eng Technol (IJRASET). 2024; 186(16): 380–384.
21. Samatha Naidu DJ, Lokesh R. Missing Child Identification System using Deep Learning with VGG-FACE Recognition Technique. International Journal of Computer Science and Engineering (IJCSE). 2022; 9(9): 1–11.

22. Zahid SM, Najesh TN, Ameen SR, Ali A.s A Multi Stage Approach for Object and Face Detection using CNN. In 2023 IEEE 8th International Conference on Communication and Electronics Systems (ICCES). 2023 Jun 1; 798–803.
23. Osorio-Roig D, Rathgeb C, Drozdowski P, Busch C. Stable hash generation for efficient privacy-preserving face identification. IEEE Trans Biom Behav Identity Sci. 2021 Jul 28; 4(3): 333–48.
24. Viola P, Jones MJ. Robust real-time face detection. Int J Comput Vis. 2004 May; 57(2): 137–54.
25. Ibrahim AA, Nisar K, Hzou YK, Welch I. Review and analyzing RFID technology tags and applications. In 2019 IEEE 13th international conference on application of information and communication technologies (AICT). 2019 Oct 23; 1–4.
26. Liu X, Xie X, Zhao X, Wang K, Li K, Liu AX, Guo S, Wu J. Fast identification of blocked RFID tags. IEEE Trans Mob Comput. 2018 Jan 15; 17(9): 2041–54.
27. Hegde N, Preetha S, Bhagwat S. Facial Expression Classifier Using Better Technique: FisherFace Algorithm. In 2018 IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2018 Sep 19; 604–610.
28. Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In Computer Vision—ECCV'96: 4th European Conference on Computer Vision Cambridge, UK, April 15–18, 1996 Proceedings. Springer Berlin Heidelberg. 1996; 43–58.
29. Taheri S, Vedienbaum A, Nicolau A, Hu N, Haghighat MR. Opencv.js: Computer vision processing for the open web platform. In Proceedings of the 9th ACM multimedia systems conference. 2018 Jun 12; 478–483.
30. Viola P, Jones M. 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) 2001 Dec 8; 1: I–I.
31. Lu WY, Ming YA. Face detection based on viola-jones algorithm applying composite features. In 2019 IEEE International Conference on Robots & Intelligent System (ICRIS). 2019 Jun 15; 82–85.
32. Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell. 1989 Jul; 11(7): 674–93.
33. Mulyono IUW, Susanto A, Rachmawanto EH, Fahmi A. Performance Analysis of Face Recognition using Eigenface Approach. In: 2019 IEEE International Seminar on Application for Technology of Information and Communication (iSemantic). 2019; 12–16.

