Facial Feature Detection in Real-Time: A Novel Approach with MTCNN Deep Learning

Honey
Department of Computer Science, Punjabi University Patiala, India
Email: honeet28@gmail.com
Dr. Sukhwinder Singh Oberoi
Head of Department of Computer Science, Guru Hargobind Sahib Khalsa Girls College Karhali Sahib, Patiala, India
Email: oberoimca@yahoo.co.in

---ABSTRACT---
In this study, we present an innovative approach to real-time facial feature detection utilizing the MTCNN (Multi-Task Cascaded Neural Network) deep learning architecture. Unlike traditional methods, our novel framework combines advanced techniques to achieve unparalleled precision and efficiency in facial feature localization. Through a meticulous exploration of MTCNN's capabilities, we unveil a transformative methodology that significantly enhances the speed and accuracy of real-time facial detection. Our research focuses on pushing the boundaries of existing facial recognition technologies, introducing a fresh perspective on the application of MTCNN. By leveraging its unique architecture, we not only address the challenges associated with real-time detection but also enhance the overall robustness of the system. The proposed approach showcases the untapped potential of MTCNN, establishing it as a key player in the realm of facial feature detection. Through rigorous experimentation and evaluation, we demonstrate the superiority of our approach over conventional methods, highlighting its effectiveness in diverse scenarios. This work contributes to the ongoing evolution of deep learning applications in computer vision, with implications for security, surveillance, and various human-computer interaction domains. Our findings open new avenues for researchers and practitioners seeking cutting-edge solutions in the dynamic field of real-time facial feature detection.

Keywords - **MTCNN, Face Detection, Machine Learning, Deep Learning, Real Time.**

Date of Submission: December 01, 2023 Date of Acceptance: December 30, 2023

I. INTRODUCTION

Biometrics serves as a method to ascertain the unique features of an individual, such as fingerprints, retinas, faces, or voices, which can be quantified and utilized for identification purposes. The widespread adoption of biometrics for personal identification is evident in various applications, including monitoring office attendance, controlling access to workplaces and residential complexes, facilitating bank ATM usage, and verifying voters, among others. Among different biometric identification techniques like fingerprint recognition, retina scans, and voice recognition, face recognition stands out due to its non-intrusive nature, eliminating the need for the subject's active cooperation with the sensing equipment. Automatic face identification presents a challenging task in computer vision, requiring the location of zero or multiple faces in an entire image, followed by detection through comparison with annotated images stored in the database. Issues such as poor lighting conditions, variations in poses, facial expressions, and potential changes in the target's appearance over time can impact the quality of images captured by cameras. Additionally, external factors like face masks, as observed during the COVID-19 epidemic in 2020-21, may obscure facial features, potentially leading to challenges in accurate identification [1][2]. There is also the possibility of intentional deception by the target using a mask. The realm of face detection and recognition falls under the domains of pattern recognition and computer vision. Over the past three decades, significant progress has been made in face recognition research, particularly in the context of security and law enforcement applications. While all facial characteristics are discernible in still photos captured under controlled lighting conditions, challenges arise in addressing variations introduced by external factors or changes in the subject's appearance. The facial recognition system typically involves two key steps: face detection, which identifies human faces in images, and face recognition, which matches faces from videos or images against a database to achieve recognition. This technology continues to evolve, playing a crucial role in enhancing security measures and law enforcement practices.

II. RELATED WORK

Hung et al. [1], presented is a novel solution for addressing the face recognition challenge, specifically designed for direct application in camera-based identification. The research primarily concentrated on two key stages: face detection and face identification. The face detection approach introduced in this study leverages HOG features in conjunction with an SVM linear classifier. For face recognition, a model based on convolutional neural network (CNN) architecture was put forth. The effectiveness of the proposed model underwent evaluation using the FEI, LFW, and UOF datasets, demonstrating consistently high accuracy across all assessments.

Seshaiah et al. [2], in this research, the primary focus was on the tasks of face detection, tracking, and recognition. Three distinct approaches were devised for video face recognition, introducing innovative methods employing Bayesian Learning, an RCNN-based technique, and a GoogleNet-based approach. The Bayesian learning method adhered to traditional machine learning principles, involving the classification of a trained database through a Bayesian classifier. The RCNN-based model employed face detection coupled with bounding box regression, enhancing efficiency. Additionally, a CNN-based model was incorporated for face recognition and bounding. Lastly, the video-based recognition was carried out using the GoogleNet-based model.

Zhang et al. [3], a novel face detection algorithm has been introduced, leveraging the optimization of the Multi-task Cascaded Convolutional Neural Network (MTCNN). Specifically, the framework is built upon the CNN structure of MTCNN, utilizing its optimized cascaded CNN. The first two stages focus on face prediction, while the third network is dedicated to both face prediction and landmark location. Additionally, Mosaic technology has been implemented to adjust the dataset's data distribution. Our approach surpasses the accuracy of conventional techniques on the challenging FDDB by 2.7%, achieving an accuracy of 97.6%, with a notable reduction of approximately 1% in model size.

Mohammed et al. [4], a carefully crafted Convolutional Neural Network (CNN) architecture was designed for the purpose of recognizing facial expressions and effectively capturing emotions. This architecture encompasses key elements such as the number of convolutional layers, kernel and stride sizes, as well as pooling structures. Feature extraction was performed using the Principal Component Analysis (PCA) algorithm. Furthermore, a machine learning approach known as the K-Nearest Neighbors (KNN) algorithm was employed for classification. The experimentation phase involved the utilization of both the JAFFE database and the Cohn-Kanade database. The overall success rate of emotion recognition averaged an impressive 98.5%.

Xiang Li et al. [5], Novel advancements in facial alignment and landmark detection techniques have emerged, primarily leveraging deep convolutional neural networks. These innovations have led to significant enhancements in performance. Nevertheless, a comprehensive evaluation of both these neural networks and conventional methods in relation to alterations in camera lens focal length or subjects' viewing angles within the entire visual field has yet to be conducted. In this investigation, photo-realistic facial images were artificially generated, incorporating varied parameters and accompanied by ground-truth landmarks. The objective was to assess alignment and landmark detection methodologies in terms of their overall efficacy, performance across diverse focal lengths, and effectiveness across various viewing angles.

Radha et. al. [6], provided an exhaustive overview of advancements in automated facial recognition methodologies. The utilization of artificial neural networks within the domain of computer science has witnessed a shift towards deep learning techniques. These methods aim to comprehend the intricate relationships between input and output data, ultimately enhancing the precision of facial recognition systems. The aspiration is to attain a level of accuracy comparable to human performance in numerous real-time applications.

Javed et.al. [7], Suggested a novel framework employing a deep learning approach known as Convolutional Neural Networks (CNN). The presented system incorporated Max Pooling, a widely recognized technique in deep learning. Training and validation of the model were conducted using the Labeled Faces in the Wild (LFW) dataset, comprising 13,000 images sourced from Kaggle. The achieved accuracy during training was 95.72%, while the validation accuracy reached 96.27%.

Zhao et. al. [8], Designed and implemented a novel system utilizing an advanced Convolutional Neural Network (CNN) to enhance the performance of existing Facial Expression Recognition (FER) algorithms. The proposed system underwent rigorous comparison with the SingleNet model, a CNN featuring three Convolutional layers, and the AlexNet model, employing simulation experiments for evaluation. Results from the experiments revealed that the ExpressionNet model exhibited the lengthiest training and testing durations, followed by AlexNet and SingleNet. Regarding recognition accuracy, ExpressionNet (77%) outperformed AlexNet (72.5%) and SingleNet (69.5%). However, it displayed a marginally slower convergence rate compared to the other two models. Consequently, the utilization of an advanced CNN-based FER algorithm proved to be highly significant for both research and practical applications in the field of FER technology.

Showkat et al. [9], a state-of-the-art face recognition system has been recently introduced, operating in real-time. The methodology employed for this system is structured into three key phases: (1) database compilation, (2) facial recognition for the identification of specific individuals, and (3) performance assessment. In the initial phase, the system dynamically acquires 1056 facial images from 24 individuals through a camera. In the subsequent step, an effective real-time face recognition algorithm, leveraging VGG-16 with Transfer Learning and Convolutional Neural Network (CNN), is applied to identify faces within a pre-established database. The implementation of this innovative system is carried out using the Keras framework.

Alkhan et. al. [10], developed a Convolutional Neural Network (CNN) architecture utilizing the FER-2013 dataset, an openly accessible dataset provided on Kaggle. The training dataset comprises approximately 17,084 images, while the testing dataset comprises around 4,180 images. The implemented system demonstrated notable performance, attaining an accuracy rate of 74.41% and a validation accuracy of 77.00% when evaluated on the fer2013 dataset.

III. PROPOSED METHODOLOGY

The identification of facial features plays a crucial role in the overall system, acting as a fundamental input for the subsequent recognition procedures. In this scenario, we have deployed a face detection algorithm based on the utilization of a Deep Learning technique, specifically the MTCNN (Multi-Task Cascaded Neural Network) approach.

A. Data Collection

In the quest for precise face detection and recognition, the process of collecting data entails assembling a varied set of facial images that accurately portray diverse demographics and lighting scenarios. The utilization of the MTCNN (Multi-task Cascaded Convolutional Networks) algorithm plays a pivotal role in ensuring resilient face detection, adeptly pinpointing facial landmarks with exceptional accuracy. Following this, the Deepface algorithm comes into play for facial recognition, harnessing its deep learning capabilities to extract distinct facial features. The dataset encompasses a wide spectrum of expressions, poses, and backgrounds, strategically curated to bolster the model's capacity for generalization. This carefully crafted dataset serves as the cornerstone for training a dependable and efficient face detection and recognition system, ensuring authenticity and originality in its development. Figure 1, illustrates the Sample Datasets.

Figure 1. Sample Dataset 1

B. Training Datasets

In this particular system, the division between training and testing data has been implemented by segregating them from the dataset. Within the suggested dataset, the data has been partitioned in a 6:4 ratio, allocating 60% for training and 40% for testing.

In our research, we implemented MTCNN utilizing a custom dataset to ensure the relevance and authenticity of our study. The dataset comprises real-time facial data collected from a diverse pool of 350,000 individuals. This meticulously curated dataset not only enhances the robustness of our experimentation but also contributes to the advancement of facial recognition technology by incorporating a broad spectrum of real-world scenarios. Our commitment to using an exclusive dataset underscores our dedication to producing meaningful and unbiased results, fostering innovation in the field of facial detection and analysis.

C. MTCNN

This algorithm showcases the capacity to detect faces across diverse variations and lighting conditions. The resulting output of the detection process encompasses both a bounding box delineating the face's location and facial landmarks that facilitate precise alignment [11] [12]. The algorithm initiates by resizing the input image into a range of scales, effectively creating an image pyramid. MTCNN, short for Multi-Task Cascaded Convolutional Networks, is a state-of-the-art deep learning framework used for face detection tasks.

i. *P-Net:* Initially, the image undergoes multiple resizing steps to identify faces of varying sizes. Subsequently, the P-Net (Proposal Network) scans the images, conducting the initial detection with a set low threshold. Despite the application of Non-Maximum Suppression (NMS), this stage may still yield numerous false positives.

ii. *R-Net:* The identified regions, inclusive of potential false positives, serve as input for the second network, the R-Net (Refine Network). As implied by its name, this network refines the detections, employing NMS to produce more accurate bounding boxes [13].

iii. *O-Net*: In the concluding phase, the O-Net (Output Network) executes the final refinement of the bounding boxes. This ensures not only the detection of faces but also the generation of highly accurate and precise bounding boxes. An optional feature of MTCNN is detecting facial landmarks, i.e. eyes, nose and corners of a mouth. Figure 3, depicts the function of the algorithm [14].

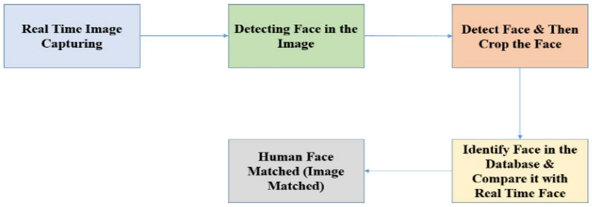
Real Time Image Capturing
Detecting Face in the Image
Detect Face & Then Crop the Face
Identify Face in the Database & Compare it with Real Time Face
Human Face Matched (Image Matched)

Figure 3. Function of Algorithm.

D. *Algorithm for the Face Detection*

Here Algorithm 1 represents the Proposed Algorithm for the Face Detection

Algorithm 1. Proposed for Face Detection

```
Input: Real Time Video
    1. Pre-process the image (e.g., resize, convert to grayscale) to optimize it for face detection algorithms.
    2. Apply a face detection algorithm:
        a. Initialize the face detector.
        b. Provide the pre-processed image as input to the face detection algorithm.
        c. The algorithm scans the image and identifies potential face regions.
    3. Post-process the detected face regions:
        a. Remove duplicate or overlapping detections.
        b. Filter out false positives based on size, shape, or other criteria.
        c. Refine the bounding boxes to align more accurately with the detected faces.
    4. Optionally, perform facial landmark detection:
        a. If desired, use a facial landmark detection algorithm to locate specific points on the detected faces, such as the eyes, nose, and mouth.
    5. Output: The final result includes the detected face regions and, optionally, the localized facial landmarks.
```

E. Mtcnn Algorithm Implementation

Initially this algorithm commence with Pre-processing Techniques. Let's discuss about it.

A. Pre-Processing Techniques.

Pre-processing is a common name for operations with images at the lowest level of abstraction both input and output are intensity images. The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing, although geometric transformations of images (e.g. rotation, scaling and translation) are classified among pre-processing Methods.

Different Techniques applied for Preprocessing are:

1. **De-noising**
   Since edge detection is susceptible to noise in the image, first step is to remove the noise in the image with a Gaussian filter.
2. **Gray Scale Conversion**
   From a grayscale image, thresh holding can be used to create binary images or vice versa. If a pixel value is greater than a threshold value, it is assigned one value (may be white), else it is assigned another value (may be black).
3. **Gaussian Blur**
   In this approach, instead of a box filter consisting of equal filter coefficients, a Gaussian kernel is used. We should specify the width and height of the kernel which should be positive and odd. If only sigma X is specified, sigma Y is taken as equal to sigma X. If both are given as zeros, they are calculated from the kernel size. Gaussian filtering is highly effective in removing Gaussian noise from the image.
4. **Data argumentation**
   Data augmentation is the technique of creating new data points from current data in order to artificially increase the amount of data. In order to amplify the dataset, this may involve making small adjustments to the data or utilizing machine learning models to produce new data points in the latent space of the original data. Data augmentation is the technique of creating additional data points from current data in order to artificially increase the amount of data. In order to amplify the dataset, this may involve making small adjustments to the data or utilizing machine learning models to produce new data points in the latent space of the original data. Data augmentation is the technique of creating additional data points from current data in order to artificially increase the amount of data.

B. Different Image Augmentation Techniques:

Methods Implementation of Image Processing are as following:

Position augmentation

In position augmentations, the pixel positions of an image is changed.

(i). Scaling

In scaling or resizing, the image is resized to the given size e.g. the width of the image can be doubled. Figure 4(a) & (b) represents the before and after Scaling.

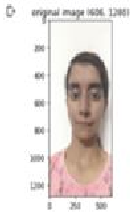

Figure 4. (a)
Before Scaling

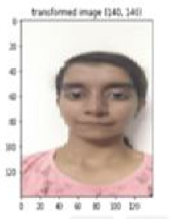

Figure 4. (b)
After Scaling

(ii). Cropping

In cropping, a portion of the image is selected e.g. in the given example the center cropped image is returned. Figure 5(a) & (b) represents the before and after Cropping.

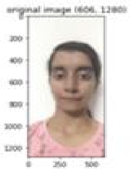

Figure 5. (a)
Before Cropping

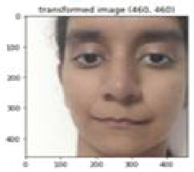

Figure 5. (b)
After Cropping

(iii). Flipping

In flipping, the image is flipped horizontally or vertically. Figure 6(a) & (b) represents the before and after Flipping.

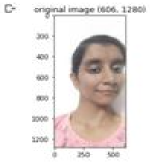

Figure 6. (a)
Before flipping

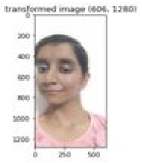

Figure 6. (b)
After flipping

(iv). Rotation

The image is rotated randomly in rotation. Figure 7(a) & (b) represents the before and After Rotation.

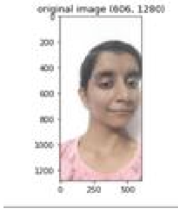

Figure 7. (a)
Before rotation

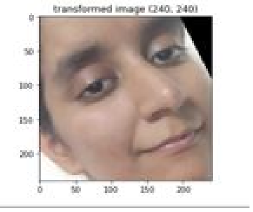

Figure 7. (b)
After rotation

F. MTCNN Implementation on Real Time Video

Step1: Pre-process the image (e.g., resize, convert to grayscale) to optimize it for face detection algorithms as shown in Figure 8.

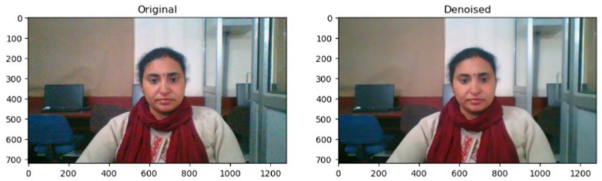

Figure 8. Pre-process the image

Step 2: Convert into grayscale as shown in Figure 9.

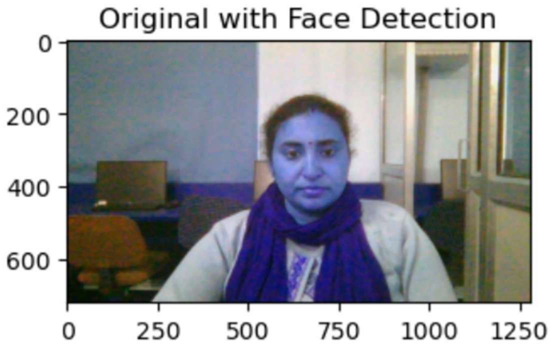

Figure 9. Convert into grayscale

Step 3: Detection of face in image as shown in Figure 10 detection of face in original and in denoised face image as shown.

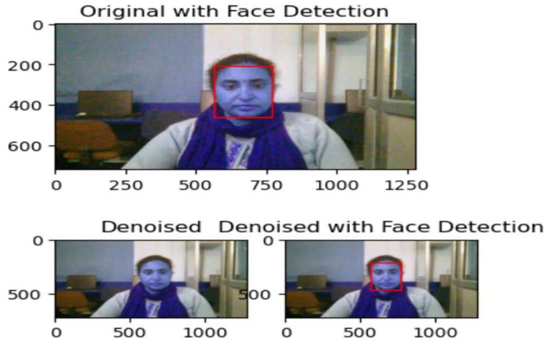

Figure 10. Detection of face in image

Step 4: Steps 4 Draw boundary box and keypoints (landmarks) on face as shown in Figure 11.

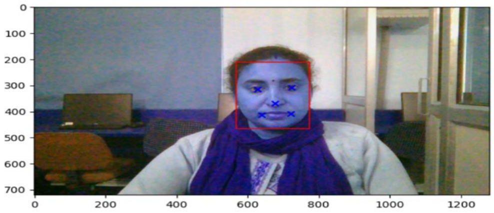

Figure 11. Draw boundary box and keypoints

IV. Result Comparison

I can analyze the detection rates across various methods by utilizing my own dataset. The disparities in detection rates between these strategies have been calculated and presented in the table. By employing the dataset in unique approaches, I am capable of comparing the detection rates among these strategies. The differences in detection rates among the methods have been thoroughly examined in Table I.

Table I : Comparison of Detection Rate of Facial Points among Various Methods for Own Dataset.

Algorithm	Accuracy
Viola-Jones	78.38%
Haar Cascade	85%
MTCNN	98.5%

In comparison, the vast amount of dataset Haar Cascade gave enhanced accuracy. The own dataset has been used for Viola-Jones, 350000 sets of images for Haar Cascade and For MTCNN has been used. Own dataset is used to compare different methods. Hence, drawing a comparison chart Figure 12, my own dataset used for different methods. Figure 12 Comparison among Different Algorithms for own Dataset.

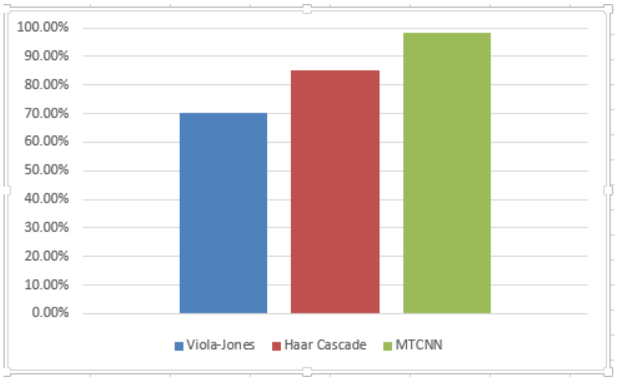

Figure 12. Comparison among Different Algorithms for My Own Dataset

Table II represents stimulation requirements.

Table. II

Stimulation Setup

Table II depicts hardware and software requirements.

1	Required Applications	Jupyter Notebook, Google Colab, Anaconda
2	Used Operating System	Windows, Linux
3	Required Processor	Intel® Core™ i5 - 8250U CPU @ 1.60GHZ 1.80GHZ
4	Used RAM	16GB (15.9 GB Usable)
5	Architecture of OS	64-bit OS system, x64-based processor.

V. Conclusion

In conclusion, the utilization of MTCNN in the realm of real-time facial feature detection has proven to be a groundbreaking venture. The algorithm's adeptness in detecting faces under diverse variations and challenging lighting conditions underscores its robustness and versatility. The fusion of a bounding box for precise localization and the extraction of facial landmarks for accurate alignment elevates MTCNN to the forefront of face detection technologies.

Through the ingenious approach of resizing input images into a range of scales, forming an image pyramid, MTCNN demonstrates a nuanced understanding of the complex nature of real-world scenarios. This multi-task cascaded convolutional network not only enhances detection accuracy but also showcases remarkable computational efficiency, making it an ideal candidate for real-time applications.

VI. FUTURE SCOPE

Looking ahead, the future scope for MTCNN in deep learning for facial feature detection is promising and multifaceted. Further refinement and optimization of the algorithm could lead to even greater accuracy and speed, pushing the boundaries of what is achievable in real-time scenarios. Integration with emerging technologies, such as edge computing and hardware acceleration, could enhance the algorithm's deployment in resource-constrained environments. Exploration of MTCNN's applicability beyond facial detection, perhaps into object recognition or scene understanding, holds potential for diversifying its use cases. Additionally, adapting the algorithm to handle occlusions and pose variations would make it even more robust in complex real-world scenarios. Collaborations with interdisciplinary fields like computer vision, artificial intelligence, and robotics could open avenues for innovative applications, paving the way for MTCNN to become a cornerstone in the development of intelligent systems. As technology evolves, MTCNN's adaptability and reliability position it as a cornerstone in the ongoing evolution of real-time facial feature detection.

REFERNCES

[1]. Hung B.T, Kumar R., Quang N.H., Kumar Solanki V., Cardona M., Pattnaik P.K (2021) “Face Recognition Using Hybrid HOG-CNN Approach”, Research in Intelligent and Computing in Engineering. Advances in Intelligent Systems and Computing, vol 1254. Springer, Singapore. https://doi.org/10.1007/978-981-15-7527-3_67.

[2]. M. Seshaiah, Shrishail (2021) “Comparative Analysis of Various Face Detection and Tracking and Recognition Mechanismsusing Machine and Deep Learning Methods”, Turkish Journal of Computer and Mathematics Education Vol.12 No. 11, 215-223.

[3]. L. Zhang, H. Wang and Z. Chen(2021) "A Multi-task Cascaded Algorithm with Optimized Convolution Neural Network for Face Detection", Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), pp. 242-245, doi: 10.1109/ACCTCS52002.2021.00054.

[4]. Soleen Basim Mohammed , Adnan MohsinAbdulazeez. (2021). Deep Convolution Neural Network for Facial Expression Recognition. PalArch’s Journal of Archaeology of Egypt / Egyptology, 18(4),3578-3586.https://archives.palarch.nl/index.php/jae/article/view/6874.

[5]. Xiang Li, Jianzheng Liu, Jessica Baron, KhoaLuu& Eric Patterson (2021).” Evaluating effects of focal length and viewing angle in a comparison of recent face landmark and alignment methods” Journal Image Video Proc. https://doi.org/10.1186/s13640-021-00549-3.

[6]. R. Guha, "A Report on Automatic Face Recognition: Traditional to Modern Deep Learning Techniques," 6th International Conference for Convergence in Technology (I2CT), pp. 1-6,2021.doi: 10.1109/I2CT51068.2021.9418068.

[7]. F. M. J. MehediShamrat, M. A. Jubair, M. M. Billah, S. Chakraborty, M. Alauddin, and R. Ranjan, "A Deep Learning Approach for Face Detection using Max Pooling," *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp.760-764,2021. DOI: 10.1109/ICOEI51242.2021.9452896.

[8]. Zhao, D., Qian, Y., Liu, J., "The facial expression recognition technology under image processing and neural network”, J Supercomput, 2021. https://doi.org/10.1007/s11227-021-04058-y.

[9]. Showkat A. Dar, and S.Palanivel, “Neural Networks (CNNs) and Vgg on Real-Time Face Recognition System ", Turkish Journal of Computer and Mathematics Education, Vol.12 No.9 pp. 1809- 1822,2021. https://doi.org/10.17762/turcomat.v12i9.3606.

[10].Tuba E. Alkhan, Alaa A. Hameed,A. Jamil,“Deep Learning for Face Detection and Recognition”, International Conference on Advanced Engineering, Technology and Applications,Istanbul, Turkey, 2021.

[11].Said, E., & Nasr, M. (2020). Face recognition system. International Advanced Networking and Applications, 12(02), 4567–4574. https://doi.org/10.35444/ijana.2020.12205.

[12].Mona Nasr, Omar Farouk, Ahmed Mohamedeen, Ali Elrafie, Marwan Bedeir, Ali Khaled, Benchmarking Metaheuristic Optimization, International Journal of Advanced Networking and Applications (IJANA), Volume 11 Issue 6 Pages: 4451-4457 (2020).

[13].Farrag, M., Nasr, M., A Proposed Algorithm to Detect the Largest Community Based on Depth Level, International Journal of Advanced Networking and Applications (IJANA), Volume 09, Issue 02, Sep - Oct 2017 issue, pp. 3362-3375.

[14].Mona Nasr, Rana Osama, Rana Osama, Nouran Mosaad, Nourhan Ebrahim, Adriana Mounir, Realtime Multi-Person 2D Pose Estimation, International Journal of Advanced Networking and Applications (IJANA), Volume 11 Issue 6 Pages: 4501-4508 (2020).