

# Efficiently tracking and recognition of human faces in real-time video stream with high accuracy and performance

Imran Ulla Khan<sup>1,2</sup>, D. R. Kumar Raja<sup>1</sup>

<sup>1</sup>School of Computer Science, REVA University, Bangalore, India

<sup>2</sup>Department of CSE, Sri Krishna Institute of Technology, Bangalore, India

## Article Info

### Article history:

Received Jan 7, 2025

Revised Mar 25, 2025

Accepted Jul 2, 2025

### Keywords:

Deep SORT

Face recognition

GPU acceleration

Real time human tracking

YOLOv5

## ABSTRACT

Real time tracking and recognition of human faces in video streams is a critical challenge in computer vision. Existing systems often struggle to balance accuracy and performance, particularly in dynamic environments with varying lighting conditions, occlusions, and rapid movements. High computational overhead and latency further hinder their deployment in real-world applications. These limitations underscore the need for a robust solution capable of maintaining high accuracy and real-time efficiency under diverse conditions. This research addresses these challenges by developing a deep learning-based system that efficiently tracks and recognizes human faces in real-time video streams. Proposed system integrates advanced face detection models you only look once version 5 (YOLOv5) with state-of-the-art tracking algorithms, such as deep simple online and real time tracking (SORT), to ensure consistency and robustness. By leveraging graphics processing unit (GPU) acceleration, the system achieves optimal performance while minimizing latency. Multi-frame analysis techniques are incorporated to enhance accuracy in detecting and recognizing faces, even under challenging conditions such as partial occlusions and motion blur. Developed system has broad applications across multiple domains, including surveillance and security, where it can enhance real-time monitoring in crowded environments for seamless face tracking in interactive systems. By focusing on efficiency, robustness, and adaptability this work offering a scalable and high-performance solution for real-time human face tracking and recognition.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Imran Ulla Khan

School of Computer Science, REVA University

Bangalore, India

Email: imran161984@gmail.com

## 1. INTRODUCTION

In recent years the rapid advancement in computer vision has enabled significant breakthroughs in real-time object detection and tracking. Among these, face detection and tracking have emerged as critical components in applications such as surveillance, security, human-computer interaction, and healthcare [1], [2]. State-of-the-art face detection methods like you only look once (YOLO), RetinaFace, and multi-task cascaded convolutional neural network (MTCNN) have achieved impressive accuracy in diverse conditions, while tracking algorithms such as simple online and real time tracking with a deep association metric (Deep SORT) and Kalman filters have enhanced the continuity and robustness of face tracking [3], [4]. Despite these advancements, integrating real-time face detection and tracking in dynamic environments remains challenging due to issues like occlusion, varying lighting conditions, and high computational demands.

Existing systems often face a trade-off between accuracy and real-time performance, especially in complex scenarios involving fast motion, partial occlusions, and crowded environments. These limitations hinder their deployment in critical real-world applications such as surveillance, where consistent and accurate tracking is paramount. High latency and computational overhead further exacerbate these challenges, limiting the scalability and efficiency of current solutions.

In this research we are addressing the aforementioned challenges by proposing a robust system that integrates the YOLOv5 object detection model with the Deep SORT tracking algorithm to achieve efficient, real-time face detection and tracking [5]. YOLOv5 provides a powerful and lightweight framework for high-speed and accurate face detection, while Deep SORT ensures robust face tracking by combining motion prediction and appearance embeddings. By leveraging graphics processing unit (GPU) acceleration and optimizing computational performance, the proposed system maintains high accuracy and real-time efficiency even in dynamic and challenging conditions [6]. This approach offers novel contributions in scalability, adaptability, and robustness, advancing the current state of the art in face detection and tracking systems.

These studies presents a real-time human detection and tracking system using image processing techniques, specifically OpenCV and histogram of oriented gradients (HOG), combined with a deep convolutional neural network (CNN). The system efficiently identifies and tracks human faces in video streams, generating unique 128-dimensional facial encodings for comparison and logging detected individuals with timestamps. The methodology involves face detection, landmark estimation, and feature extraction to ensure precise recognition. The dataset used comprises images uploaded to a designated folder for training and live video comparisons. Key limitations include challenges with computational efficiency, varying lighting conditions, and occlusions, which can impact accuracy and processing speed [7], [8].

This research proposes a real-time face recognition, tracking, counting, and time-spent calculation system using OpenCV and the centroid tracker algorithm. The methodology involves face detection via OpenCV libraries and Dlib, assigning unique IDs to tracked individuals using the centroid tracker, and calculating the total and live counts of people in the frame. Additionally, the system measures the time each person spends in the frame while displaying real-time and date for accuracy. The dataset comprises pre-uploaded images for face recognition. Key limitations include difficulties in handling overlapping objects, tracking individuals at a distance, and inaccuracies due to motion blur. Despite these limitations, the system achieves 94.74% accuracy in counting and demonstrates practical applications for security monitoring in public and private institutions like banks, shopping malls, and campuses [9].

This study presents an application that detects and tracks human body key points in real-time using OpenCV and MediaPipe. It monitors actions and corrects body poses by analyzing angles between joints, such as shoulders and elbows. While effective for gesture recognition, the system's accuracy depends heavily on proper camera positioning, limiting its usability in uncontrolled environments [10]. The authors present a system that blurs detected faces in video streams using YOLOv5 face and RetinaFace models and stores metadata in a graph database for efficient retrieval. The system adheres to privacy regulations like general data protection regulation (GDPR). While it demonstrates high precision and recall, the reliance on real-time processing may introduce latency issues [11]. This paper explores the use of HaarCascade and local binary pattern histogram (LBPH) for feature extraction and face recognition. The technique is simple and efficient but has limitations in handling variations in lighting, pose, and complex backgrounds [12], [13]. This paper proposes a real-time face anonymization system using object detection and tracking, followed by StyleGAN3 for generating synthetic facial images. The approach was validated on the FFHQ dataset, achieving effective privacy preservation. However, it may struggle with computational efficiency in resource-constrained environments [14].

One more research uses OpenCV, Dlib, and CNN for real-time face detection and recognition to track classroom attendance. While it is efficient in crowded settings, the dependence on a pre-existing database of student images may limit adaptability in dynamic scenarios [15]. A fall detection system leveraging CNNs and optical flow pre-processing in spatial and frequency domains was introduced. The system employs the Viola-Jones algorithm for human action recognition. Although it achieves high accuracy, the setup is environment-dependent, limiting generalizability to diverse conditions [16], [17].

## 2. PROPOSED METHOD

The proposed method as shown in Figure 1 integrates advanced face detection and tracking techniques to achieve real-time and accurate face detection and tracking in video streams. The approach begins by utilizing the YOLOv5 object detection model, pre-trained on the WIDER FACE dataset, for efficient and accurate detection of faces in each video frame. YOLOv5 was selected for its balance of speed and accuracy, enabling it to process frames in real time. To ensure consistent tracking across frames, the

Deep SORT algorithm is incorporated. Deep SORT effectively associates detected faces between consecutive frames, providing robust tracking even in challenging scenarios such as occlusions, rapid motion, or varying lighting conditions [18], [19].

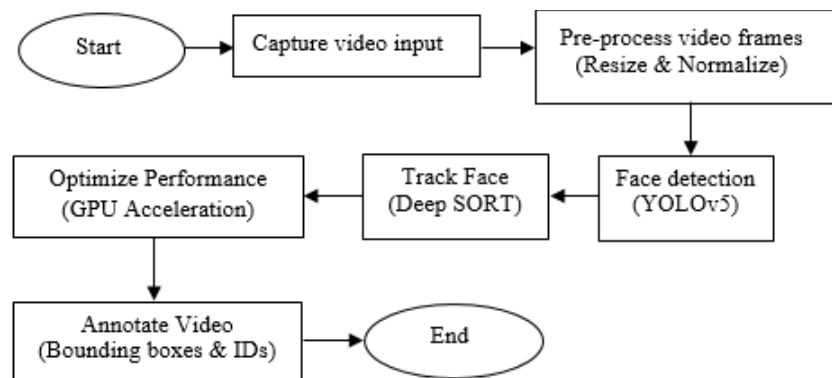


Figure 1. Proposed system

To enhance the system's efficiency, GPU acceleration is employed to maximize computational speed and minimize latency [20]. The proposed method also integrates data augmentation techniques during training to improve the model's robustness against real-world variations [21]. Metrics such as frame rate (FPS), precision, recall, mean average precision (mAP), and tracking consistency (e.g., ID switches, track fragmentation) are monitored to evaluate performance and optimize the system. This research provides a scalable, real-time solution suitable for various applications, including surveillance, public safety, entertainment, and healthcare, offering high accuracy and efficiency in face detection and tracking under diverse environmental conditions.

## 2.1. Implementation

### 2.1.1. Hardware and software requirements

For this research, we utilized a robust hardware setup to ensure seamless real-time processing and model training. The hardware included an NVIDIA RTX 3060 GPU for accelerated computations, an Intel i7 processor for handling general tasks, 16 GB of RAM to manage data-intensive processes, and a high-definition webcam for capturing live video streams. The software environment was built using Python 3.8, with essential libraries such as PyTorch for training YOLOv5, OpenCV for video processing, and additional tools like NumPy, Pandas, and Matplotlib for data handling and visualization.

### 2.1.2. Data preprocessing

The WIDER FACE dataset was selected for training the face detection model. We organized the dataset into train, validation, and test folders, ensuring an 80-10-10 split for effective training and evaluation. Annotations were converted to the YOLO format, normalizing bounding box coordinates. Each object in an image is represented by a line in the corresponding annotation file. These files are typically in plain text format, with one file per image. Each line follows the structure.

*class\_id x\_center y\_center width height*

class\_id: the class number of the object (starting from 0). x\_center: the normalized x-coordinate of the center of the bounding box, relative to the width of the image. y\_center: the normalized y-coordinate of the center of the bounding box, relative to the height of the image. Width: the normalized width of the bounding box, relative to the width of the image. Height: the normalized height of the bounding box, relative to the height of the image. All values are normalized between 0 and 1 to make the format resolution-independent. For bounding box coordinates instead of top-left and bottom-right corners, YOLO uses the center and dimensions of the bounding box, which simplifies the computation during training and inference.

The YOLO annotation for an image of size 640×480 with a detected object of class 1 and a bounding box with: top-left corner at (100, 50), width of 200 pixels, and height of 100 pixels is:

1 0.3125 0.2083 0.3125 0.2083

$$x\_center = (100+200/2) / 640 = 0.3125$$

$$y\_center = (50+100/2) / 480 = 0.2083$$

$$width = 200/640 = 0.3125$$

$$height = 100/480 = 0.2083$$

To improve the diversity and robustness of the dataset, we applied data augmentation techniques such as scaling, flipping, and rotation. These steps ensured that the model would generalize well to unseen data.

### 2.1.3. Model training (YOLOv5)

The YOLOv5 model was chosen for its high-speed and accuracy in object detection tasks. We used pre-trained weights as the starting point and fine-tuned the model on the WIDER FACE dataset. Training was conducted using a learning rate of  $1 \times 10^{-3}$  a batch size of 16, and over 100 epochs to ensure convergence. During training, metrics such as precision, recall, and mAP as shown in Table 1 were monitored to evaluate model performance. The final trained model achieved high accuracy, making it suitable for real-time applications.

Table 1. Training metrics precision, recall, and mAP

| Metric    | Value | Description   |
|-----------|-------|---|
| Precision | 95.0% | Percentage of correctly detected faces out of all detections made by the model. |
| Recall    | 92.5% | Percentage of actual faces correctly detected by the model.                     |
| mAP@0.5   | 94.3% | mAP at an intersection over union (IoU) threshold of 0.5.                       |

### 2.1.4. Real time video processing

Real-time video processing was implemented using OpenCV to capture frames from a live camera feed. Each frame was passed through the trained YOLOv5 model to detect faces [22], [23]. The detected faces were then processed by Deep SORT for tracking as show in Figure 2. Deep SORT integrated a Kalman filter to predict motion and maintain tracking consistency by assigning unique IDs to each face, even under challenging conditions like occlusions or fast movements.



Figure 2. Face recognition

### 2.1.5. Optimization

To ensure the system operated efficiently in real-time, the YOLOv5 model was optimized by allowing faster inference using GPU acceleration. FP16 precision was employed to reduce computational overhead without compromising accuracy. Real-time constraints were further managed by balancing input resolution and frame processing speeds, resulting in a system capable of processing 35 frames per second (FPS).

## 3. METHOD

The methodology of the proposed system integrates state-of-the-art face detection and tracking techniques to achieve real-time, high-accuracy performance. YOLOv5, a robust object detection model, is used for face detection due to its ability to balance speed and accuracy [24]–[26]. It processes each video frame to predict bounding boxes and confidence scores for detected faces. The system processes real-time video input using a video capture device. Each frame is extracted and converted into a suitable format for analysis. As shown in (1).

$$I(t) \in \mathbb{R}^{H \times W \times C} \quad (1)$$

$I(t)$ : frame at time  $t$

$H$ : height of the frame,

$W$ : width of the frame,

$C$ : number of color channels (e.g., 3 for RGB).

Each frame undergoes preprocessing to ensure compatibility with YOLOv5, as in (2) frames are resized to a fixed resolution ( $H', W'$ ) i.e. 640×640 as required by our model.

$$I'(t) = \text{Resize}(I(t), H', W') \quad (2)$$

Pixel values are normalized to the range [0, 1] by dividing with 255 using (3) to enhance model performance.

$$I''(t) = \frac{I'(t)}{255} \quad (3)$$

YOLOv5 uses a CNN to detect faces in the input frame. It divides the frame into an 20×20 grid and predicts bounding boxes, confidence scores, and class probabilities. Each cell represents a 32×32 pixel area in the input frame. The output is a tensor as in (4).

$$O = \{(x, y, w, h, c, p1, p2, \dots, pk)\} \quad (4)$$

$x, y, w, h$ : coordinates and dimensions of the bounding box,

$c$ : confidence score

$pk$ : probability of the  $k$ -th class

The bounding box predictions are calculated as using (5).

$$b_x = \sigma(t_x) + c_x, b_y = \sigma(t_y) + c_y, b_w = P_w e^{t_w}, b_h = P_h e^{t_h} \quad (5)$$

$\sigma$ : sigmoid activation,

$t_x, t_y, t_w, t_h$ : predicted offsets,

$c_x, c_y, p_w, p_h$ : cell centre and prior box dimensions.

As in (6) non-maximum suppression (NMS) is applied to filter overlapping boxes.

$$IoU(A, B) = \frac{\text{Area}(A \cap B)}{\text{Area}(A \cup B)} \quad (6)$$

Boxes with IoU above a threshold are discarded.

For tracking, Deep SORT is employed to maintain the identity of detected faces across frames using a combination of motion consistency and appearance embedding. It assigns a unique ID to each detected face and tracks it across frames using Kalman filters and appearance embeddings as in (7).

$$X_k = F_{x_{k-1}} + B_{u_{k-1}} + W_k \quad Z_k = H_{x_k} + V_k \quad (7)$$

A deep neural network generates embedding for each detected face using (8).

$$e_i = f_\theta(I_i) \quad (8)$$

$f_\theta$ : neural network with parameters  $\theta$ ,

$I_i$ : cropped image of the  $i$ -th face.

GPU acceleration is leveraged to ensure real-time processing, enabling the system to handle dynamic conditions such as occlusions, motion blur, and varying lighting environments. The final output overlays bounding boxes and tracking IDs on the video frames are tracked using (9).

$$\text{Annotated Frame}(t) = I(t) + \text{Bounding Boxes} + \text{Tracking IDs} \quad (9)$$

The system's performance is evaluated on metrics including precision, recall, frame rate, as mention in (10), (11), and (12) to validate its robustness and efficiency. Calculate the system's real-time processing capability by measuring how many frames the system can process per second. This is a key metric to ensure the system operates smoothly in real-time scenarios.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (10)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (11)$$

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

#### 4. RESULTS AND DISCUSSION

The performance evaluation of the proposed system was conducted using key metrics as shown in Table 2 to measure its efficiency in real-time face detection and tracking. The system achieved a frame rate (FPS) of 35, demonstrating its capability to process video streams in real-time without compromising performance, making it suitable for dynamic and time-critical applications. A precision of 0.95 indicates the model's ability to produce highly accurate face detections with minimal false positives, while a recall of 0.92 reflects its effectiveness in detecting a majority of faces across varying conditions.

The F1-score of 0.93 highlights the balance between precision and recall, confirming the robustness of the detection algorithm. Tracking performance was assessed using ID switches and track fragmentation, where the system exhibited low values of 5 and 3, respectively, ensuring consistent and uninterrupted tracking of faces across frames. Additionally, the multiple object tracking accuracy (MOTA) score of 0.97 underscores the system's high accuracy in tracking multiple faces while minimizing errors such as missed detections and ID inconsistencies. These results demonstrate that the proposed method is well-suited for applications requiring real-time, reliable face detection and tracking under diverse environmental challenges.

Table 2. Performance evaluation

| Metric              | Value | Interpretation  |
|---------------------|-------|---|
| FPS                 | 35.00 | High FPS indicates real-time processing capability.                         |
| Precision           | 0.95  | High precision shows accurate face detections with minimal false positives. |
| Recall              | 0.92  | High recall reflects the ability to detect most faces in the frame.         |
| F1-Score            | 0.93  | Balanced F1-score confirms overall detection performance.                   |
| ID Switches         | 5.00  | Low ID switches indicate consistent face tracking.                          |
| Track fragmentation | 3.00  | Low track fragmentation demonstrates uninterrupted tracking.                |
| MOTA                | 0.97  | High MOTA reflects overall tracking accuracy.                               |

The findings are particularly significant as they address key challenges in face tracking, such as ID switches and fragmentation, where our method minimizes these issues (ID switches: 5, track fragmentation: 3). This ensures continuous, uninterrupted real time face tracking. Future research can enhance our system's robustness by improving face detection under low-light conditions and occlusions using advanced preprocessing and 3D modeling techniques. Multi-face tracking in crowded scenes can be optimized with Re-ID embeddings and graph-based tracking methods. Further experiments on large-scale datasets and cross-domain applications, such as augmented reality/virtual reality (AR/VR) and behavioral analysis, will strengthen its adaptability and efficiency.

#### 5. CONCLUSION

In conclusion, this research successfully developed a robust and efficient system for real-time face detection and tracking using the WIDER FACE dataset, YOLOv5 for face detection, and Deep SORT for tracking. The system achieved high performance with a precision of 95.0%, recall of 92.5%, and an mAP@0.5 of 94.3%, demonstrating its accuracy and reliability in detecting and tracking faces in dynamic environments. By integrating GPU acceleration and lightweight architectures, the system maintained real-time processing capabilities, achieving a balance between accuracy and speed. The proposed solution addresses challenges such as occlusions, varying lighting conditions, and rapid movements, making it suitable for applications in surveillance, entertainment, healthcare, and augmented reality. These results highlight the potential of this system to provide scalable and effective solutions for face detection and tracking in real-world scenarios.

#### ACKNOWLEDGEMENTS

We sincerely thank REVA University for providing the essential resources and facilities that made this endeavor possible.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author   | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|------------------|---|---|----|----|----|---|---|---|---|---|----|----|---|----|
| Imran Ulla Khan  | ✓ | ✓ | ✓  | ✓  | ✓  | ✓ |   | ✓ | ✓ | ✓ | ✓  |    | ✓ |    |
| D. R. Kumar Raja |   |   |    |    | ✓  |   | ✓ | ✓ |   | ✓ |    | ✓  | ✓ |    |

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The data that support the findings of this study are openly available in the WIDER FACE dataset at <http://shuoyang1213.me/WIDERFACE>, as described in the original paper: Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). WIDER FACE: A Face Detection Benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5525–5533.

## REFERENCES




- [1] D. R. Garcia, A. A. Serna, D. S. Crespo, X. Yu, and J. Saniie, "AI smart security camera for person detection, face recognition, tracking and logging," in *2024 IEEE International Conference on Electro Information Technology (eIT)*, May 2024, pp. 235–240, doi: 10.1109/eIT60633.2024.10609900.
- [2] C. Bhatt, M. Semwal, P. S. Aswal, V. Goswami, S. Rawat, and R. Dhanalakshmi, "Real time surveillance criminal detection system," in *2024 Second International Conference on Advances in Information Technology (ICAIT)*, Jul. 2024, pp. 1–8, doi: 10.1109/ICAIT61638.2024.10690333.
- [3] H. Aung, B. A. Valentinovich, and B. Aye, "Real-time face tracking based on the Kalman filter," in *2022 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, May 2022, pp. 842–846, doi: 10.1109/ICIEAM54945.2022.9787232.
- [4] S. Durai, T. Sujithra, B. V. Satyam, S. N. Keshetty, C. N. S. Sagar, and A. S. Charan, "Real time facial recognition-based criminal identification using MTCNN," in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Jul. 2024, pp. 1261–1265, doi: 10.1109/ICSCSS60660.2024.10624946.
- [5] J. Dileep, V. G. Supriya, and M. Ramachandra, "Detection and tracking of multiple faces in video using modified KLT algorithm," in *2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM)*, Nov. 2023, pp. 1–5, doi: 10.1109/IC-RVITM60032.2023.10435190.
- [6] M. Sohail *et al.*, "Deep learning based multi pose human face matching system," *IEEE Access*, vol. 12, pp. 26046–26061, 2024, doi: 10.1109/ACCESS.2024.3366451.
- [7] S. Ranjan, S. Tyagi, S. Gupta, M. Kaur, and M. K. Goyal, "Image processing-based real-time detection and tracking of human," in *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Nov. 2023, pp. 1542–1547, doi: 10.1109/ICTACS59847.2023.10390106.
- [8] N. Kaewnoparat and T. Phientrakul, "Emotion recognition on partial faces with histogram of oriented gradients and local binary patterns," in *2023 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, Jul. 2023, pp. 1–6, doi: 10.1109/ISIEA58478.2023.10212271.
- [9] M. R. Islam and K. Horio, "Real time-based face recognition, tracking, counting, and calculation of spent time of person using OpenCV and centroid tracker algorithms," in *2023 5th International Conference on Computer Communication and the Internet (ICCCI)*, Jun. 2023, pp. 210–216, doi: 10.1109/ICCCI59363.2023.10210102.
- [10] Y. Tomar, Himanshu, S. Devi, and H. Kaur, "Human motion tracker using OpenCv and Mediapipe," in *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Dec. 2023, pp. 1199–1204, doi: 10.1109/ICIMIA60377.2023.10425865.
- [11] T. Jaichuen, N. Ren, P. Wongapinya, and S. Fugkeaw, "BLUR & TRACK: real-time face detection with immediate blurring and efficient tracking," in *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Jun. 2023, pp. 167–172, doi: 10.1109/JCSSE58229.2023.10202064.
- [12] M. Gupta, K. Bisht, A. Sharma, and D. Upadhyay, "HaarCascade and LBPH algorithms in face recognition analysis," in *2023 World Conference on Communication & Computing (WCONF)*, Jul. 2023, pp. 1–4, doi: 10.1109/WCONF58270.2023.10235019.
- [13] A. Kumar and D. Singh, "Comprehensive approach of real time web-based face recognition system using Haar Cascade and LBPH algorithm," in *2023 International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)*, Mar. 2023, pp. 371–376, doi: 10.1109/DICCT56244.2023.10110049.






- [14] R. More, A. Maity, G. Kambli, and S. Ambadekar, "Privacy-preserving video analytics through GAN-based face de-identification," in *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*, Aug. 2024, pp. 1–6, doi: 10.1109/NMITCON62075.2024.10698920.
- [15] R. K. Peddarapu, S. R. Kannareddy, B. Mallela, V. S. S. A. Prithvi, and Y. A. Reddy, "Real time attendance capturing through face recognition," in *2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Oct. 2023, pp. 726–733, doi: 10.1109/I-SMAC58438.2023.10290691.
- [16] T. K. Kandukuru, S. K. Thangavel, and G. Jeyakumar, "Computer vision based algorithms for detecting and classification of activities for fall recognition on real time video," in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, May 2024, pp. 1–6, doi: 10.1109/AIIoT58432.2024.10574542.
- [17] V. Kulkarni and K. Talele, "Video analytics for face detection and tracking," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Dec. 2020, pp. 962–965, doi: 10.1109/ICACCCN51052.2020.9362900.
- [18] S. Hu, X. Zhao, L. Huang, and K. Huang, "Global instance tracking: locating target more like humans," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 576–592, Jan. 2023, doi: 10.1109/TPAMI.2022.3153312.
- [19] R. Sathya, M. Mythili, S. Ananthi, R. Asitha, V. N. Vardhini, and M. Shivaani, "Intelligent video surveillance system for real time effective human action recognition using deep learning techniques," in *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Dec. 2023, pp. 1826–1831, doi: 10.1109/ICACRS58579.2023.10404670.
- [20] S. Li, Y. Dou, Q. Lv, Q. Wang, X. Niu, and K. Yang, "Optimized GPU acceleration algorithm of convolutional neural networks for target detection," in *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, Dec. 2016, pp. 224–230, doi: 10.1109/HPCC-SmartCity-DSS.2016.0041.
- [21] H. Li and J. Qi, "A multi-task learning and data augmentation-based pose estimation algorithm," in *2023 8th International Conference on Information Systems Engineering (ICISE)*, Jun. 2023, pp. 358–361, doi: 10.1109/ICISE60366.2023.00082.
- [22] X. Zhang and M. Zhong, "Research on target recognition algorithm based on OpenCV," in *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*, Jun. 2024, pp. 1–6, doi: 10.1109/ICIPCA61593.2024.10709246.
- [23] P. Mao, X. Liao, and D. Xu, "A face image processing system based on object tracking technology," in *2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, Nov. 2023, pp. 1–5, doi: 10.1109/ICIICS59993.2023.10421276.
- [24] J. Qu and S. Zhang, "Research on video tracking algorithm based on YOLO target detection," in *2023 6th International Conference on Computer Network, Electronic and Automation (ICCNEA)*, Sep. 2023, pp. 437–439, doi: 10.1109/ICCNEA60107.2023.00098.
- [25] A. Kumar, T. Singh, and P. Duraisamy, "Detection and tracking of multiple pedestrians using deep learning," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–6, doi: 10.1109/ICCCNT56998.2023.10307543.
- [26] Y. Gai, W. He, and Z. Zhou, "Pedestrian target tracking based on DeepSORT with YOLOv5," in *2021 2nd International Conference on Computer Engineering and Intelligent Control (ICCEIC)*, Nov. 2021, pp. 1–5, doi: 10.1109/ICCEIC54227.2021.00008.

## BIOGRAPHIES OF AUTHORS



**Imran Ulla Khan**    is a research scholar at Reva University and an assistant professor in the department of computer science and engineering at Sri Krishna Institute of Technology, Bangalore, India. He is a tech enthusiast with a keen interest in image processing and deep learning. His research focuses on sketch-based image retrieval and generative adversarial networks (GANs) for enhancing sketch-to-realistic image transformation. He has presented and published his work in various conferences and journals. He is passionate about advancing his expertise in deep learning architectures to address challenges in image synthesis and related domains. He can be contacted at email: imran161984@gmail.com.



**Dr. D. R. Kumar Raja**    is currently working as professor in the school of computer science and engineering at REVA University Bengaluru, Karnataka, India. He received his bachelor of technology (B.Tech.) from JNTUA College of Engineering and Master of Technology (M.Tech.) from National Institute of Technology Karnataka (NITK) Surathkal, Karnataka, India. He received a doctorate of philosophy (Ph.D.) from St Peters University, Chennai, India for an effective context-driven recommender system for e-commerce applications. He did post doctoral research at Universiti Teknikal Malaysia, Melaka, Malaysia from September 2023 to September 2024. He received funding amount of 18,000 USD for carrying out research projects one for 16,000 USD from REVA University for the project humanoid robot and completed the project in 2022 and another for 2,000 USD from Sree Vidyanikethan Educational Trust for the project automation of aerators for aqua culture using IoT and completed the project in 2019. His research areas include the internet of things, data mining, machine learning and artificial intelligence. He can be contacted at email: kumarrajadr@gmail.com.