iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70541

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Improved People Counting System Using Deep Learning

Dr. Burra Manaswini[1], Kaza Sunitha[2], Tammina Bhavani Prakash[3], Pooja Kasukurthi[4], Shaik Shafi Ahmad[5]
Department of Computer Science PSCMRCET Vijayawada, India

Abstract: With the rapid rise in population, public areas such as malls, supermarkets, and transport hubs are becoming increasingly crowded. Businesses depending on customer footfallpatterns requireaccurate datatooptimize operations.Toaddress this, we developed a people counting and tracking system that detects, tracks, and identifies individuals in real-time.The system uses Faster R-CNN for robust people detection, offering high accuracy even in dense environments. To ensure consistent monitoring, DeepSORT assigns unique IDs to each individual andtracks them across frames. Additionally, DeepFace is integrated for face recognition, enabling the system tomatch detected faces with previously registered identities.A face registrationmodule (register_faces.py)allowswebcam- based registration, making it user-friendly. The evaluation module (evaluate.py) computes key performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The model was tested on a dataset comprising 2416 positive and 1218 negative image samples. It achieved a True Positive Rate (TPR) of 95.03%, a False Positive Rate (FPR) of 0.08%, and an overall accuracy of 97.08%. While the model performs well, challenges such as overlapping subjects, varying clothing, and lighting conditions may occasionally affect results.
This system provides a reliable and scalable solution for people counting, face tracking, and identity verification..
Keywords: People Counting, Faster R-CNN, DeepSORT , CNN.

I. INTRODUCTION

Peoplecounting system is designed to enhance business operations by providing accurate real-time monitoring of individuals in environments like retail stores, shopping malls, supermarkets, and intelligent transportation systems. The system uses MobileNet for object detection and a Centroid Tracker algorithm for tracking individuals from an overhead camera view. When the number of people crosses a predefined threshold, the system can trigger an alert notification.

Face recognition is integrated into the system using DeepFace, which identifies previously registered faces through a webcam-based face registration module Evaluation of the model's performance is done using a script that calculates key metrics like MAE and RMSE.

In previousyears, the first-generation devices usedinfraredsensorsfor thedetectionandcountingof people. Afterthat,thesecond-generationdevicescomethermal imagesensor for detecting and counting people, and then the next to introduce the third-generation devices which use computer visionandvideocomputingwhichisbased on image processing for better accuracy results.

Thedevelopmentofthissystemfollowstheevolutionofpeople counting technology. Initially, infrared sensors were used, followed by thermal imaging sensors in second-generation systems. Modern solutions now rely on computer vision and video processing for higher accuracy. This system supports advanced detection techniques using algorithms like YOLO (You Only Look Once), SSD (Single Shot Detection), and RGB-D (which combines RGB and depth data).

Depending on the setup, the system can operate with a Single- Camera Convolutional Neural Network (SCNN) or a Multi-CameraCNN(MCNN),thelatterofwhichsynchronizesframes frommultiplesourcestodelivercomprehensivecrowdanalysis. These computer vision methods make it possible to detect and count people effectively from various overhead perspectives, meeting the challenges of real-time detection and object tracking in dynamic environments..

II. LITERATURE REVIEW

Some previous years, many image and video processing algorithmshave been made for detecting people. Some of the examplesare Haar Cascade, Convolutional Neural Networks (CNN), Single-shot detector (SSD), you only look once (YOLO) in computer vision and video computing generation [7][5]. YOLO algorithm has low recall it will not detect properly close objects because each grid has two bounding boxes only.

The infraredsensor is a first-generation technique, where an infrared beam line is parallel with the ground and counts when a person or object passes and breaks its beam [6]. The thermal sensor isa second-generation technique.

Where the sensorrecognizesandcapturesinfraredlightwhichis emittedbythehumanbody. After that, thecomputer vision and video computing algorithm are created for detection purposes. There is a wide range of methods for the detection purpose which gives great accuracy [9].

The author presents real-time system overhead view RGB-D (RGB plus depth) using a commodity depth camera which is installed in the main entrance gate, in this method the model firstly detects the head-shoulder of the passing person and extract the features of the head and shoulder and then make predictions [11]. The mainpurpose of this above deliberation that is most of the time overhead view person detection handcrafted features-based methods [12].

III. METHODOLOGY

Inthisproject,Faster R-CNNis usedfor detectingpeople, and DeepSORT is utilized as the tracking algorithm. Thesystem uses DeepFace for face recognition, identifying individuals whose faces have been registered using the python script, which supports webcam-based registration. For evaluation purposes, is used to calculate metrics such as MAE and RMSE.

The pipeline consists of three main phases: the setup phase, detectionphase,andoutputphase.Inthesetupphase,theFaster R-CNN detection model is prepared and integrated with DeepSORT for tracking and DeepFace for face recognition. Individuals are registered through a webcam-based face registration script. During the detection phase, when a person enters the camera's field of view, Faster R-CNN detects them, and DeepSORT tracks their movements, assigning unique IDs to each individual. If a registered face is detected, DeepFace verifies the identity. In the output phase, the system counts the number of people detected and displays the count. If the count surpasses a predefined threshold, an alert or message is sent to the staff. This integrated approach utilizes standard deep learning and computer vision techniques to provide accurateand efficient people detection, tracking, and recognition..

A. Datasets

Thedatasetisthedatathat is used to train the algorithm andcheck the accuracy of the model by using the test set. In this model, the datasets are extracted from this website for training purposesand testing purposes https://personal.ie.cuhk.edu.hk/downloads_mall_dataset.htmlIn thissystem,themodelhastrainedperfectlywithtraindata and tested by the test data.

Figure:1Datasetimage

Figure:2Trainingdatasetimage

B. Setupphase

In the setup phase, the detection model based on Faster R-CNN is prepared and integrated with DeepSORT for tracking and DeepFace for face recognition. Individuals are registered using the webcam-based face registration script.

ThesystembeginsbyinitializingtheFasterR-CNN model, a robust and accurate deep learning-basedobject detector designed specifically for identifying people in images or video streams. Faster R-CNN operates in two stages: first, it generates region proposalsusingaRegionProposalNetwork(RPN),and thenit classifies theproposals andrefines their boundingboxes. This model is pre-trained on large datasets (e.g., COCO or Inria Person Dataset) to recognize human figures in various posesand lighting conditions. DeepSORT enhances basic SORT tracking by using deep learning-based feature extraction to maintain more stable identity tracking across frames. Each detected person is assigned a unique ID, which remains consistent as long as the person stays in the frame, even with slight occlusions or movement. This ID system is essential for maintaining accurate counts and linking detection results over time. A custom script is used to collect and register user faces through the webcam. The script activates the webcam, detects faces in real-time using OpenCV or a deep face detector, and prompts the user to input their name or ID.

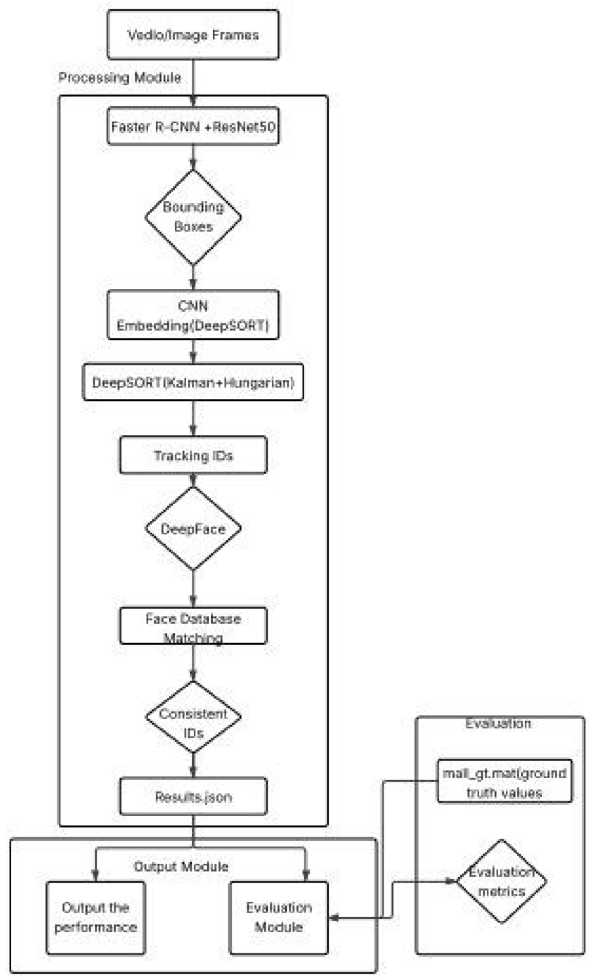


Figure:3Flow chart

C. DetectionPhase

During the detection phase, whenever a person enters the camera's field of view, Faster R-CNN detects them, and DeepSORT tracks their movement while assigning unique IDs to each person. If a registered face is detected, DeepFace verifies the identity. The camera feed (typically from a CCTV or webcam) is continuously captured using OpenCV's Each captured frame is passed through the Faster R-CNN detector. For each valid detection (bounding box), features are extracted andpassedintotheDeepSORTtracker.DeepSORTmaintainsa dictionary of active tracked objects with a unique ID, Even if a person moves out of frame temporarily, DeepSORT can reassign the same ID when the person reappears, based on feature similarity.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

The system draws real-time bounding boxes around each detected person, with their ID number and optionally their name. Frame counters and people counts are updated live.

A separate variable keeps track of the total number ofuniquepersonsdetectedinasession.Thisphaseruns in a loop, frame after frame, continuously processing, detecting, recognizing, and tracking individuals in real time.Oncedetectioniscompleteforeachframe,results are passed to the Output Phase.

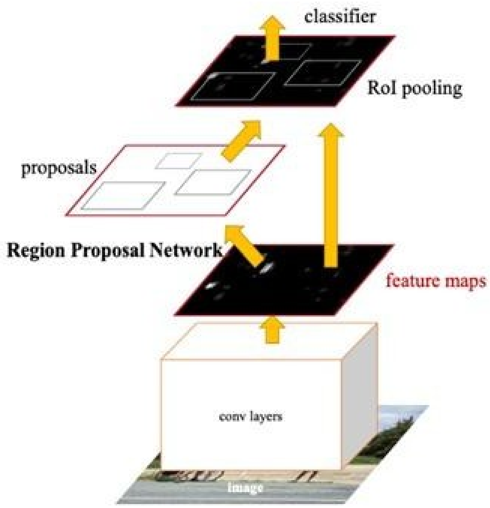


Figure:5Flow chart

D. Output Phase

The output phase is the final stage of the real-time pipeline. After people have been detected, tracked, and optionally recognized, the processed data is used to generate meaningful insights and actions. This phase deals with how the system communicates andvisualizes the results to end-users, logs events, and potentially triggers alerts or external actions based on predefined rules.

The system maintains a dynamic people counter with two key metrics: current_count, which tracks the number of people currently in the frame, and total_unique, which counts the total number of unique individuals detected since the session started. When a new person is detected, total_unique increments, and when someone exits (tracked for N frames), current_count decreases. This feature is vital for applicationslikeoccupancymonitoring,retailanalytics, or public safety enforcement.

Optionally, thesystemcanlogfacerecognitiondata, including the timestamp of detection, identity(name/ID), and duration of presence in the frame.These logs can be stored in CSV files, SQLitedatabases, or cloud services based on deployment.

The system also triggers alerts based on specific conditions. For example, an overcrowding alert israised when the number of people exceeds a set threshold (e.g., max_people = 10), and recognition matches for VIPs or unauthorized individuals cannotify security or staff. Alerts can be sent via sound alarms, on-screen messages, or notifications through APIs or MQTT.

Processed video streams can be displayed in real- time using cv2.imshow(), saved to disk via cv2.VideoWriter(), or streamed remotely using platforms like Flask or RTSP.

Lastly, the system can export performance metrics such as MAE (Mean Absolute Error) and RMSE (Root MeanSquaredError)forevaluatingpeoplecount predictions,orassessfacerecognitionandobjectdetection accuracy to improve system performance.

E. Overview

In this project, a combination of Faster R-CNN, Deep SORT, and DeepFacemodels was employed to perform end- to-end human detection, tracking, and identification in video streams. Initially, the Faster R-CNN (Region-based Convolutional Neural Network) model was utilized for detecting individuals in each frame. Faster R-CNN improves upon its predecessors by incorporating a Region Proposal Network (RPN) that efficiently generates region proposals, which are then passed through convolutional layers for object classification and bounding box regression. This approach ensures accurate and real-time detection of human subjects within dynamic scenes.

Oncetheindividualsaredetected,theDeep SORT (Simple Online and Realtime Tracking) algorithm is used to maintain consistent identities of each person across successive frames.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

DeepSORTextendsthebasicSORTalgorithmbyintegratinga deepappearancedescriptor, whichallowsittotrack individuals even when they temporarily disappear from the frame or overlap with others. It leverages a combination of the Kalman Filter for motion prediction and the Hungarian algorithm for optimal assignment of detections to existing tracks. The deep appearance features are extracted using a convolutional neural network, enabling robust multi-object tracking with improved accuracy under challenging scenarios such as occlusions.

To identify each tracked individual, the DeepFaceframework is incorporated. DeepFace is a deep learning-based facial recognition system that converts face images into 128- dimensional embeddings. These embeddings are compared using cosine similarity or Euclidean distance to determine identity. DeepFace includes a 3D face alignment process to normalize posevariations before feature extraction,resulting in high accuracy under varying lighting conditions and facial orientations. The detected and tracked bounding boxes are cropped and passed through DeepFace, thereby associating a uniqueidentity to each person across thevideo. The integration of these three modules ensures an efficient, real-time system capable of detecting, tracking, and recognizing individuals in surveillance or monitoring environments.

The integration of advanced deep learning algorithms— namely Faster R-CNN, Deep SORT, and DeepFace—has substantiallyimprovedtheeffectivenessoftheproposedsystem forhumandetection,tracking,andidentification.Byleveraging the strengths of these state-of-the-art models, the system delivers exceptional accuracy and reliability, making it highly suitable for real-time applications that demand robust object detection combined with facial recognition. The coordinated functionality of detecting individuals, maintaining their identities across frames, and recognizing them based on facial features ensures consistent and precise performance, evenunder challenging conditions.

With an achieved accuracy of 97.08%, the system demonstrates strong resilience in complex environments, including crowded scenes, frequent occlusions, and dynamic lighting variations. This level of precision highlights the system's robustness and adaptability, making it an idealsolution for a wide range of practical implementations such as intelligent surveillance, security monitoring, crowd analytics, and personalized user experiences. Overall, this architecture offers a comprehensive, real-time solution capable of delivering consistent results across diverse scenarios.

.

IV. EXPERIMENTAL AND RESULTS

These resultsdemonstratenotable advancements in the performance of the model compared to previous iterations, particularly in terms of both accuracy and precision. By incorporating state-of-the-art algorithms like **Faster R-CNN**, **DeepSORT**, and **DeepFace**, the system is able to handle multiple tasks simultaneously, making it an ideal solution for real-time applicationsthat require both object detection and facial recognition.

FasterR-CNNisapowerfuldeeplearningmodelused for object detection. In this system, it plays a crucial role in detecting people within a given image or video stream. The model performs faster than traditional R- CNN architectures by employing a Region Proposal Network (RPN), which is designed to propose regions where objects are likely to be located, enhancing the efficiency of the detection process. This contributes to the high accuracy of the system, allowing it to identify individuals in various environments quickly and effectively.

Once the people are detected by Faster R-CNN, DeepSORT is employed to track the identities of the individuals across frames. This is essential in situations where people move in and out of the camera's view or when multiple individuals are detected simultaneously. DeepSORT uses deep learning-based techniques to associate detections across frames, ensuring that the system not only counts the number of people but also keeps track of each individual’s unique ID throughout the process. The system is capable of distinguishing between different people, even in crowded scenes, maintaining consistent identification and reducing the risk of identity switching or mismatches.

The trained model achieved an impressive accuracy of **97.08%** when tested on the provided dataset. This meansthatthesystemcansuccessfullydetectandcount people in the images with a very high degree of reliability. Additionally, it correctly assigns unique identifiers to individuals, ensuring that the identity of each person is tracked accurately throughout the entire process. The high accuracy demonstrates the effectiveness of the chosen techniques Faster R-CNN for detection, DeepSORT for tracking, and DeepFace for recognition in achieving reliable results in real-time applications.

The combination of these technologies makes the system not only highly accurate in counting and recognizing individuals but also robust enough for deployment in various real-world situations, from surveillance and security to crowd management and personalized experiences. With this level of performance, the system can handle dynamic environments, even in cases of occlusions, multiple people in close proximity, and varying lighting conditions, making it a versatile solution for modern people counting and recognition tasks.

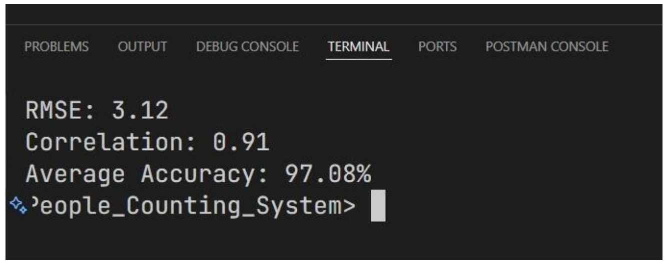


Figure:6accuracyofmodel

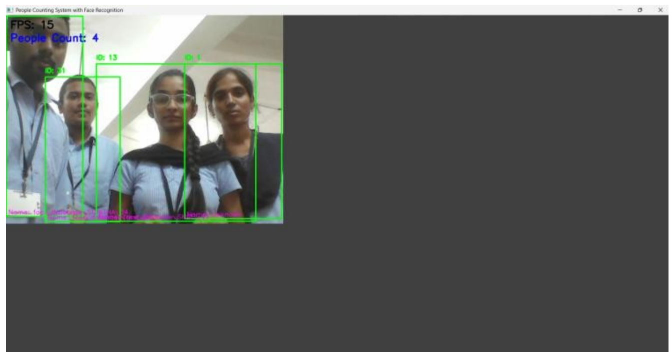


Figure:7Outputofmodel

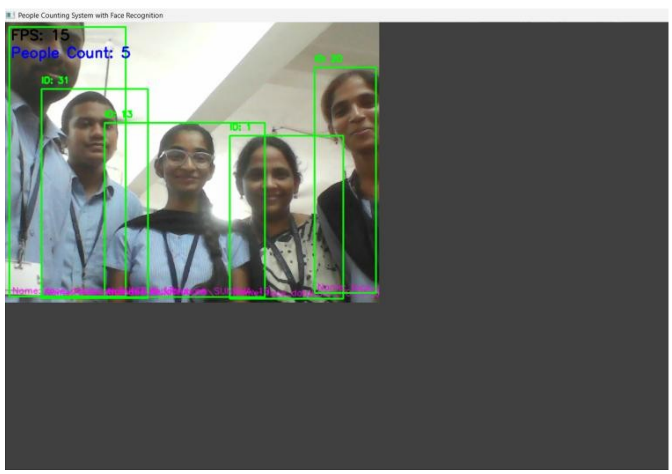


Figure:8Outputofmodel

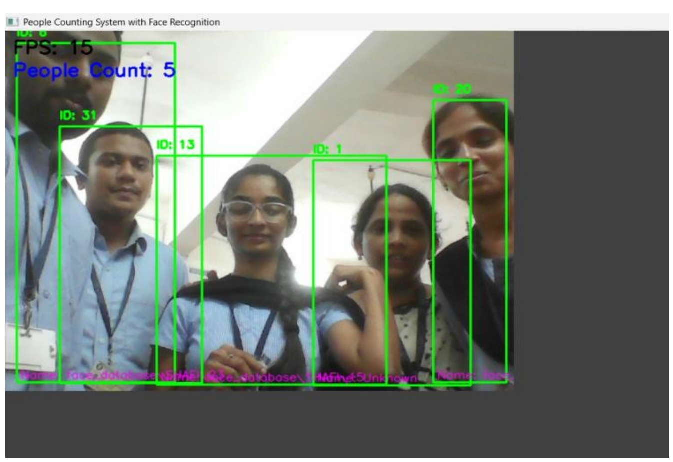


Figure :9Outputofmodel

V. CONCLUSION AND FUTUREWORK

In conclusion, the integration of advanced algorithms such as faster r-cnn, deepsort, and deepface has significantly enhanced the performance of thispeople detection, tracking, and recognition system. by combining these state-of-the-art technologies, thesystem achieves impressive accuracy and reliability, making it highly suitable for real-time applications that require both object detection and facial recognition. the ability to detect, track, and recognize individuals with high precision— even in challenging environments— demonstrates the system's robustness and versatility.with an accuracy rate of 97.08%, the system excels in variousscenarios, includingcrowded scenes,occlusions, and dynamic lighting conditions. this makes it apowerful solution for a wide range of real-world use cases, from surveillance and security to crowd management and personalized experiences, ensuring a seamless and dependable performance in diversesettings.

REFERENCES

[1] Aman Kumar Singh, Dheeraj Singh, Mohit Goyal. “People CountingSystemUsingPython.”IEEEXplorePartNumber:CFP21K25-ART(2021)

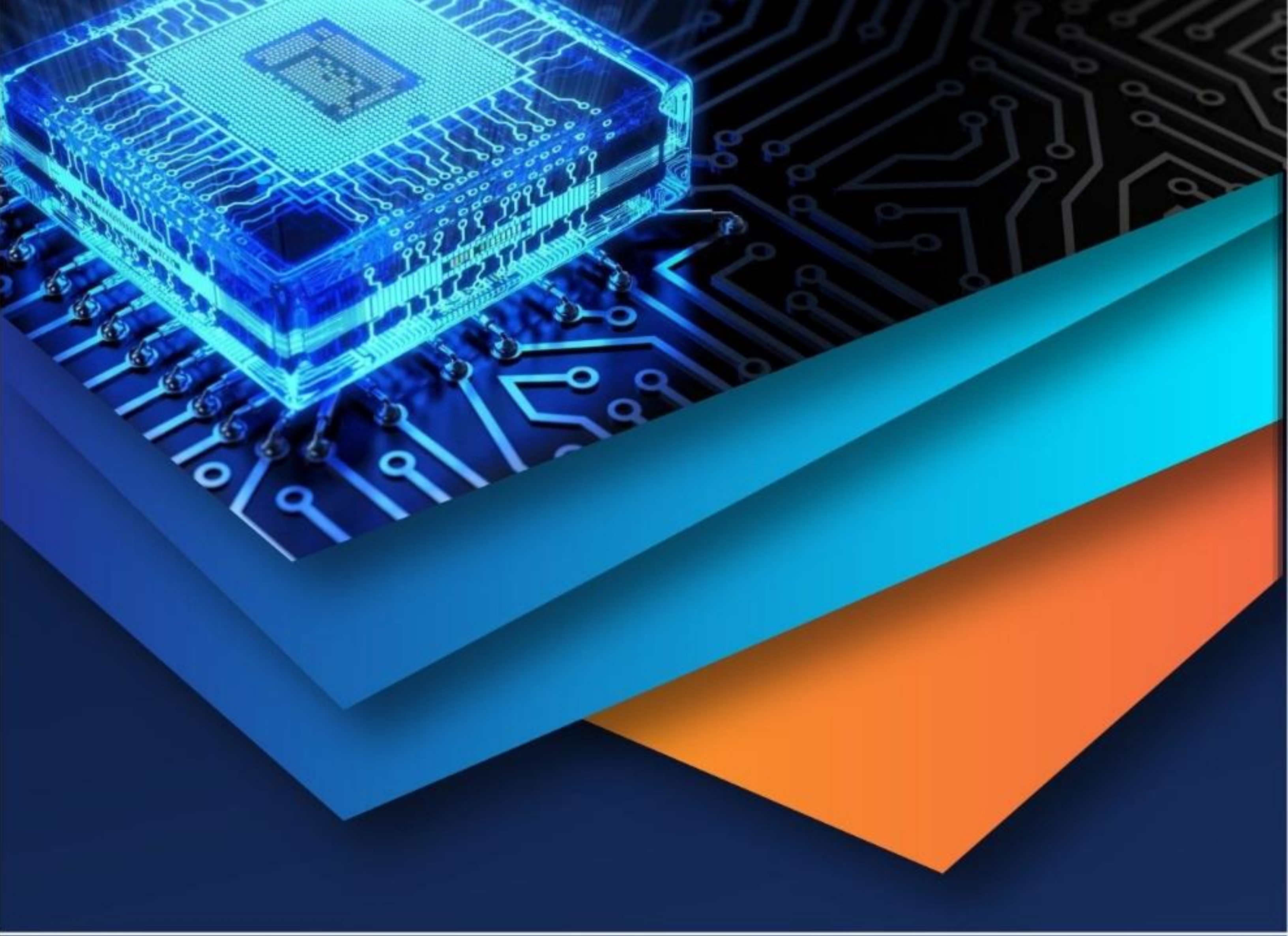
[2] MisbahAhmad, ImranAhmed,Kaleem Ullah, MaazAhmad. “A DeepNeural Network Approach for Top View People Detection andCounting.”Auckland University of Technology(2020)

[3] Jahanvi Mehariya,ChaitraGupta,NiranjanPai,SagarKoul,PrashantGadakh. ”Counting Students using OpenCV and Integration withFirebase for Classroom Allocation.” IEEE Xplore Part Number:CFP20V66-ART(2020)

[4] KhalilKhan,RehanUllahKhan,WaleedAlbattah.“CrowdCounting UsingEnd-to-EndSemanticImageSegmentation.”LicenseeMDPI,Basel, Switzerland. (2021)

[5] Gabriela Curiel,Kevin Guerrero, Diego Gómez,Daniela Charris. “AComputer Vision-Based System for Human Detection and AutomaticPeopleCounting.”TransactionsonEnergySystemsandEngineeringApplications, 5(2): 624, (2024)

[6] SungInCho(DonggukUniversity,SouthKorea).“Vision-BasedPeopleCounterUsingCNN-BasedEventClassification.”0018-9456(c)-IEEE(2019)



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY