International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Human Detector & Counting

Joglekar Narendra*1, Kumbhar Akshay*2
Guide- Prof. Khomane K.Y.
Savitribai Phule Pune University, Pune. Department of Computer Engineering.
Shivnagar Vidya Prasarak mandal's
College of Engineering, Malegaon (BK) Tal:-Baramati, Dist :- Pune

---***---

Abstract - Human detection and counting are crucial tasks in computer vision, with applications in security surveillance, crowd monitoring, retail analytics, and smart city planning. This paper explores various approaches to human detection and counting, including traditional image processing techniques and modern deep learning-based methods such as Convolutional Neural Networks (CNNs) and You Only Look Once (YOLO). Challenges such as occlusion, varying lighting conditions, and real-time processing constraints are addressed. The study also highlights the integration of human detection models with edge computing and IoT devices to improve efficiency and scalability. Experimental results demonstrate the effectiveness of deep learning models in achieving high accuracy and real-time performance.

Key Words: Human Detection, People Counting, Computer Vision, Deep Learning, YOLO, CNN, Object Detection, Surveillance, Crowd Monitoring, Edge Computing.

1.INTRODUCTION

The Real-Time Human Detection & Counting System is a cutting-edge technological advancement designed to accurately identify and quantify human presence in various environments. Its primary objective is to automate the process of human detection, offering enhanced efficiency and reliability across multiple sectors. This system leverages advanced algorithms and machine learning techniques to process visual data captured from cameras, enabling it to detect and count individuals in real time. The importance of automating human detection cannot be overstated, particularly in domains such as security, retail analytics, and crowd management. In the security sector, the ability to accurately monitor and respond to human activity in sensitive areas can significantly enhance safety protocols. For instance, in public spaces or high-security locations, the system can trigger alerts in the event of unusual crowd behavior, thereby allowing for timely intervention and prevention of potential threats. In the retail environment, understanding foot traffic patterns is crucial for optimizing store layouts and improving customer experience. The Real-Time Human Detection & Counting System can provide retailers with valuable insights into customer behavior, enabling them to make informed decisions about product placement, staffing, and marketing strategies. By quantifying customer interactions, businesses can tailor their offerings to better meet consumer needs. Furthermore, in crowd management scenarios, such as events or public gatherings, real-time monitoring of attendee numbers ensures that safety regulations are maintained. This system can help manage crowd density, thereby preventing overcrowding and ensuring a safer experience for all participants. Overall, the Real-Time Human Detection & Counting System represents a significant leap forward in the realm of automated monitoring, providing invaluable support across various applications

2.RELATED WORKDONE

Human detection and counting systems have been implemented in a wide range of real-world scenarios, delivering measurable benefits across various industries. In the retail sector, stores employ these systems to monitor customer footfall, analyze shopping patterns, and improve store layouts. Supermarkets and malls use them to track peak hours and manage staffing effectively. In transportation hubs like airports and metro stations, real-time crowd monitoring enables authorities to maintain safety, prevent congestion, and adjust schedules based on passenger density.

Smart city initiatives have adopted human detection technologies for public safety and urban planning. Municipalities use them to manage large crowds during festivals, protests, or sporting events, reducing the risk of stampedes and ensuring emergency response readiness. In workplaces and academic institutions, the systems are integrated into surveillance infrastructure to monitor attendance, track movement, and enhance security through intrusion detection.

Technically, models such as YOLOv8 and Deep SORT are often used in combination for accurate detection and tracking. Python, OpenCV, and deep learning frameworks like PyTorch or TensorFlow form the backbone of most implementations. These systems can be deployed on local servers, edge devices like Raspberry Pi or Jetson Nano, or in cloud environments, proving their flexibility and effectiveness in diverse, real-life conditions.

3. METHODOLOGY

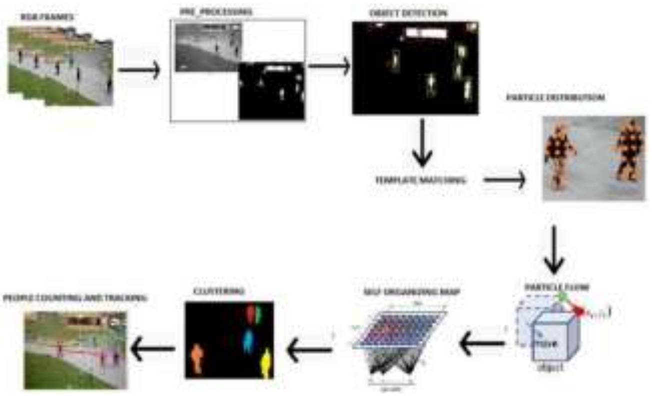


Fig -1: System Architecture

A. System Architecture Design:Modular Approach: The system is designed in a modular fashion, where each module (e.g., preprocessing, pose estimation, segmentation, masking) operates independently but communicates effectively with others. This ensures ease of maintenance and scalability.

B. Data Collection and Annotation:Data collection and annotation are critical steps in training an effective human detection model. Images and videos are gathered from various sources such as CCTV footage, public datasets (e.g., COCO, PASCAL VOC), or custom camera setups to ensure diversity in environments, lighting, and crowd density. Each image is manually annotated using tools like **LabelImg**, **CVAT**, or **VGG Image Annotator**, where bounding boxes are drawn around every visible person. These annotations are saved in formats like XML (Pascal VOC) or JSON (COCO) to serve as ground truth for training and evaluating the model.

C. Training the Models :

Training a human detection model involves feeding it large datasets of labeled images where humans are annotated with bounding boxes. Commonly used datasets include **COCO**, **PASCAL VOC**, and **Open Images**, which contain thousands of annotated images with people in various poses, backgrounds, and lighting conditions.

The process begins with **data preprocessing**, which includes image resizing, normalization, and data augmentation (like flipping, scaling, and rotation) to increase diversity and robustness. A deep learning model such as **YOLO**, **SSD**, or **Faster R-CNN** is then selected. These models use convolutional neural networks (CNNs) to extract features from images and predict object locations and classes.

The model is trained using a **loss function** (e.g., classification + bounding box regression loss) and optimized using algorithms like **Stochastic Gradient Descent (SGD)** or **Adam**. Training is performed on GPUs for faster computation, using frameworks like **PyTorch** or **TensorFlow**.

After training, the model is validated and tested on separate datasets to evaluate accuracy and generalization. If the model performs well, it is saved and used for inference on real-world inputs. Fine-tuning pre-trained models (transfer learning) is also common, reducing training time and improving performance with limited data.

D. User Interaction Flow

The user uploads an image, video, or connects a live camera feed through the interface. The system processes the input using a detection model, displays bounding boxes around detected humans, and shows the total count. Users can view results in real-time and export data or visual output if needed.

4. PROPOSED SOLUTION

The proposed solution uses a YOLOv8-based deep learning model combined with Deep SORT for real-time human detection and accurate counting. It processes inputs from images, videos, or live feeds, displays results with bounding boxes, and provides total human count. The system is scalable for edge or cloud deployment.

- **Input:** The input includes images, video files, or live camera feeds captured by devices like CCTV, webcams, drones, or smartphones.

- **User's Image:** A photo ,video of the user.

- **Preprocessing:** Preprocessing involves resizing images, normalizing pixel values, and applying data augmentation techniques like flipping and rotation to improve model robustness. It also includes noise reduction and frame extraction from videos for real-time analysis. Proper preprocessing enhances detection accuracy and ensures consistent input quality for the human detection model.

- **Components after preprocessing:**

- Components After Preprocessing

Once input images or video frames are preprocessed (resized, normalized, and cleaned), the system processes the data through several key components to detect, track, and count humans accurately.

- **1.FeatureExtraction:**

This module uses convolutional neural networks (CNNs) to extract important visual features from the input. CNN backbones like ResNet or Darknet transform raw pixel data into feature maps that highlight human shapes and characteristics. Effective feature extraction is essential for the detection model to distinguish humans from backgrounds and other objects, even under challenging conditions like occlusion or varied lighting.

- **2.Object.Detection:**

Using the extracted features, object detection models such as YOLO, SSD, or Faster R-CNN predict bounding boxes around humans along with confidence scores. The model scans the image once or uses region proposals to locate all human instances. The output is a list of bounding boxes indicating where humans are detected.

- **3.Post-Processing:**

To refine results, Non-Maximum Suppression (NMS) removes overlapping boxes by keeping only the most confident detections. Confidence thresholding discards weak detections. This step cleans up the outputs to ensure accuracy.

- **4.Tracking.and.Re-Identification:**

Tracking algorithms like Deep SORT maintain consistent IDs for individuals across video frames, preventing multiple counts of the same person. They use motion prediction and appearance features to associate detections over time, handling occlusions and movement.

- **5.Counting.Module:**

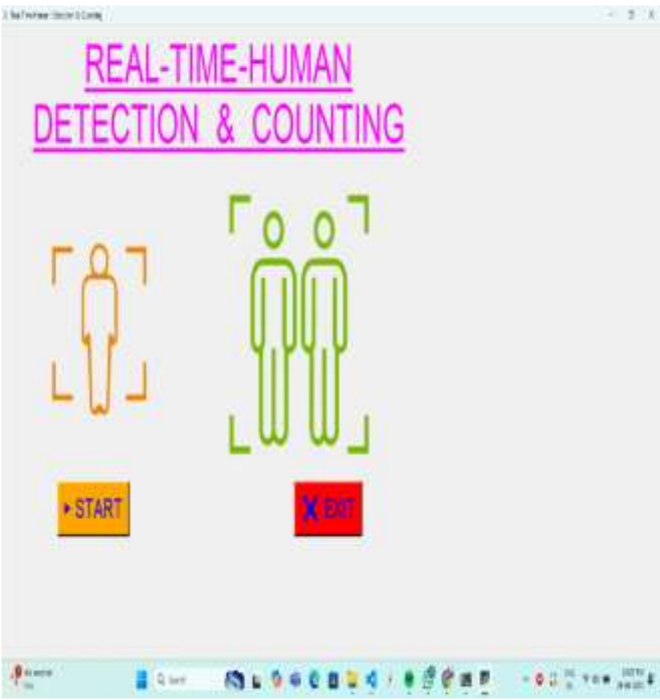
Instead of counting boxes per frame, the system counts unique tracked IDs, ensuring accurate person counts. It manages entries and exits to keep real-time totals.

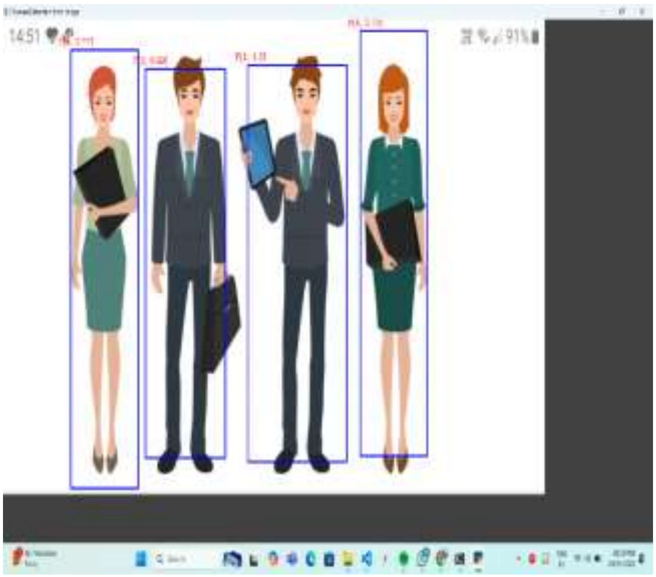
- **6.Visualization.and.Reporting:**

Results are displayed with bounding boxes and counts overlaid on frames, and data can be logged or exported. This helps users monitor crowds or occupancy effectively.

- Together, these components ensure reliable detection, tracking, and counting of humans for applications like surveillance, retail analytics, and public safety.

5. RESULTS





6. CONCLUSIONS

Recent advancements in deep learning, especially with YOLO and CNN-based architectures, have improved the accuracy and efficiency of human detection and counting. However, challenges such as occlusion, scale variations, and real-time performance optimization remain active research areas. This project aims to build on existing works and develop a robust and scalable human detection system for real-world applications.

REFERENCES

[1] **S. Lamba and N. Nain, "Crowd monitoring and classification: A survey," in Advances in Computer and Computational Sciences (Advances in Intelligent Systems and Computing). 2017**

[2] Q. Wang, J. Gao, W. Lin, and X. Li, "NWPU-Crowd: A large-scale benchmark for crowd counting and localization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 6, pp. 2141–2149, Jun. 2021.

[3] S. Elbishlawi, M. H. Abdelpakey, A. Eltantawy, M. S. Shehata, and M. M. Mohamed, "Deep learning-based crowd scene analysis survey," J. Imag., vol. 6, no. 9, p. 95, Sep. 2020, doi: 10.3390/ JIMAGING6090095.

[4] W.Albattah, M. H. K. Khel, S. Habib, M. Islam, S. Khan, and K. A. Kadir, "Hajj crowd management using CNN-based approach," Comput., Mater. Continua, vol. 66, no. 2, pp. 2183–2197, 2021, doi:10.32604/cmc.2020.014227.

[5] S. Habib ,A. Hussain, "Towards efficient detection and crowd management for law enforcing agencies,"

Apr. 2021 .