IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13467

Deep Learning-Based Crowd Management with Real-Time Analytics

Dr.C.Karpagavalli[1], Dr.M.Kaliappan[2], A.Ganesh Aravind[3]

Assistant Professor, Department of Artificial Intelligence and Data Science, Ramco Institute of technology, Rajapalayam, India[1]

Professor, Department of Artificial Intelligence and Data Science, Ramco Institute of technology, Rajapalayam, India[2]

Student, Department of Artificial Intelligence and Data Science, Ramco Institute of technology, Rajapalayam, India[3]

Abstract: Accurate people counting is essential for crowd management, smart surveillance analytics, and crowd monitoring. Manually counting people is a tedious task that is vulnerable to error.

In this paper, a people counting system based on deep learning and computer vision is proposed. The proposed system is based on the use of pre-trained convolutional neural networks like YOLO for people detection. The proposed system is able to perform under various lighting conditions.

Experimental results show that the proposed system is accurate and can perform in real-time. The results show that the proposed system is suitable for smart city applications.

Keywords: People Counting, Deep Learning, Computer Vision, Object Detection, YOLO.

1. INTRODUCTION

With the emergence of smart cities and the advancement of surveillance technologies, the importance of automated people counting has never been greater. For example, in a shopping mall, a railway station, an airport, a public event, or on a busy road, the monitoring of crowds is essential.

In the past, people counting has been performed based on manual observation or the use of sensor-based people counting systems like the use of infrared beams or pressure sensors. However, these people counting systems are associated with a number of disadvantages. For example, these systems face difficulties with “occlusion,” i.e., when people are in the way of each other.

With the advancements in deep learning and computer vision, new people counting systems have been developed that are more efficient. The recent advancements in object detection technologies based on Convolutional Neural Networks (CNNs) enable the detection of people. The object detection system known as YOLO (You Only Look Once) has further improved the detection efficiency.

In this paper, a new people counting system based on the object detection system is presented. The new people counting system is efficient and has the ability to be implemented in the real world.

2. LITERATURE SURVEY

The initial object detection methods were based on handcrafted features. The Viola-Jones detector was used for real-time face detection using a boosted cascade of classifiers [8]. The HOG feature was used for accurate human detection using gradient orientation information [9]. The development of deep learning has significantly contributed to computer vision applications.

AlexNet was used for accurate image classification using deep convolutional neural networks [10]. The development of VGGNet [11] and ResNet [12] was used for accurate feature extraction using CNNs. The Fast R-CNN [5] and Faster R-CNN [6] detectors were developed for accurate object detection using region proposal networks.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13467

The SSD detector was developed for accurate object detection using single-shot detection [7]. The YOLO framework was developed for accurate object detection in real-time using a single unified regression model [1]. The accuracy of object detection was improved using YOLOv3 [2] and YOLOv4 [3]. The Microsoft COCO dataset was used for training object detection models [4]. The SORT framework was developed for accurate object tracking in real-time using multi-object tracking [13].

The accuracy of object tracking was improved using Deep SORT [14]. The computer vision framework is implemented using OpenCV, a popular computer vision library [15].

3. DATASET USED

The proposed people counting system is trained and tested using a human detection dataset that consists of images or video frames captured from both indoor and outdoor environments. The dataset can be created using surveillance-style video footage or existing public datasets like the COCO dataset, which contains labeled bounding boxes for the "person" class.

Each image or video frame can contain one or more people, and each person is labeled with bounding box coordinates (x_min, y_min, x_max, y_max) and the corresponding class label. The bounding boxes are used to train the system to learn how to detect people accurately.

The dataset encompasses a variety of realistic situations. The dataset contains images with varying crowd density, lighting conditions, camera viewpoints, and complex backgrounds. There are images with individuals overlapping, partially occluded, and with motion blur, making the detection problem more difficult. The images are of varying resolutions, which depends on the source of the image. This allows the model to generalize better for images of varying sizes.

To ensure the dataset is used for a fair evaluation of the model, the dataset is split into training, validation, and testing sets. During training, the images are augmented with techniques such as flipping, brightness, scaling, and random crops. This makes the model more robust and prevents overfitting.

This dataset with varied images allows the deep learning model to accurately detect people and count them with high reliability.

4. EXISTING WORK

A. Traditional Approaches

Early human counting methods relied on conventional computer vision techniques, including background subtraction, optical flow, and Haar cascade classifier. Although the methods were computationally efficient and simple, they were not robust to varying lighting, viewpoint, and changing backgrounds. The methods were found to degrade considerably for crowded scenes, as the individuals were often occluded, making them less suitable for surveillance applications.

B.Deep Learning Approaches:

Deep learning has helped to achieve high accuracy in the detection of people. The Faster R-CNN algorithm has been found to achieve high accuracy. However, it is computationally expensive. The SSD algorithm is a single-stage detector that is computationally less expensive. Among the single-stage detectors, the YOLO algorithm is found to be the best as it is a single-pass detection algorithm.

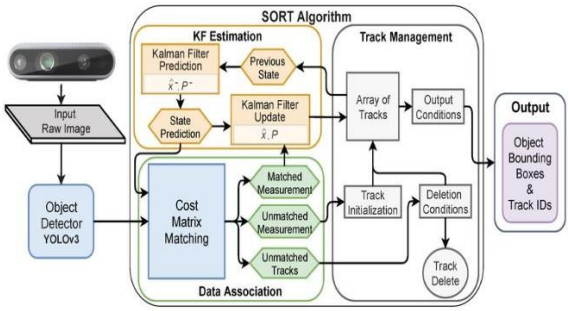
IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13467

5. PROPOSED METHODOLOGY



The proposed system is developed based on a complete end-to-end deep learning framework that is designed for real-time people detection and counting based on live video streams. The overall architecture of this system is based on various stages of video acquisition, object detection using deep learning, person filtering, and dynamic counting.

The main aim of this system is based on obtaining high accuracy during object detection, considering that it is designed for real-time performance, which is applicable for various applications, including object detection, monitoring, etc.

A. Video Acquisition and Frame Extraction:

The system commences by capturing video input from a webcam or a CCTV security camera. Instead of processing the video file as a whole, the system processes the video one frame at a time. This enables the model to detect the number of people in the video without any lag.
The video stream can be mathematically defined as follows:
$V = \{F1, F2, F3, \ldots, Ft\}$

where Ft represents the video frame captured at a given time t. This video frame is considered an independent image. The video stream is divided into frames. This enables the model to efficiently handle a video stream where people are entering or exiting the scene.

B.Image Preprocessing

Before passing each frame through the detection model, a series of operations is performed for efficiency and detection reliability.

The first operation is resizing the image to match the dimensions required by the YOLO detection network. This is a standardization operation that ensures compatibility and efficiency.

The next operation involves normalizing pixel intensity. This is a standardization operation that helps stabilize the model's output for efficient performance.

Depending on the environment, other preprocessing methods, including noise reduction or color conversion, may be performed. This process assists in improving image clarity and reducing the effects of lighting changes.

Finally, the image is converted into a tensor format suitable for the deep learning model. The preprocessing process improves robustness, stability of detection, and minimizes computational requirements for the model, allowing it to perform efficiently in real-time.

B. Human Detection Using YOLO

In the proposed system, object detection is performed with the help of a well-known object detection framework called YOLO, which stands for You Only Look Once.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13467

YOLO is a widely used object detection framework, especially for object localization and classification in a single pass of a convolutional neural network.

Let's say we have an input frame, which can be expressed as follows:
$X = \{x1, x2, x3, \ldots, xn\}$

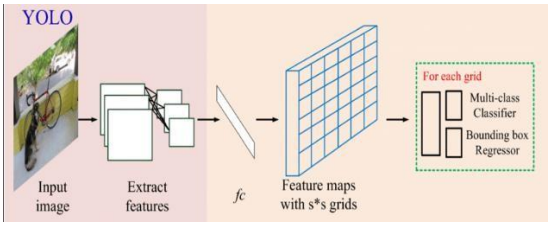
Here, each xirepresents a possible object region in the input frame.

In the YOLO framework, the entire input frame is processed as a whole, and a number of bounding boxes are generated as follows:
$B = \{b1, b2, b3, \ldots, bk\}$

In each bounding box, bk, we have:
- Spatial coordinates,
- Confidence,
- Classprobability

Although YOLO detects multiple object classes in a frame, the proposed system focuses only on bounding boxes classified under the “person” category. These detections are extracted and used for counting purposes.

YOLO divides the image into grid cells and simultaneously predicts bounding boxes and confidence scores for each cell. This unified detection strategy significantly reduces computational time, enabling fast and accurate real time human detection
in video streams.



C.Person Filtering and Confidence Thresholding

Once the results from the YOLO detection are available, a filtering step is performed to make the results more reliable. In the results from the YOLO model, it is possible that the object detected is of various classes. Therefore, the detection results for which the class is “person” are filtered.

Furthermore, a confidence threshold is applied to the results to remove any detection that is weak or uncertain.

Let the prediction indicator be defined as:

$$P_i = \begin{cases} 1, & \text{if object class = person and confidence > threshold} \\ 0, & \text{otherwise} \end{cases}$$

In simple terms:

If the object detected is classified as a person and the confidence level is greater than the threshold level, then it is a valid detection.

Otherwise, the detection is discarded With the filtering mechanism in place, the detection results are made more stable.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13467

This ensures that the accuracy of the results is maintained at a high level.

E.Counting Mechanism

After obtaining the valid detections for persons in the images by filtering, the counting module computes the total number of people present in the images. The count is calculated by summing all the valid detection indicators:

$$\text{Count} = \sum_{i=1}^{k} P_i$$

where k is the total number of bounding boxes detected within a frame, while Pi is a flag indicating whether the bounding box is for a valid person after filtering for class and confidence.

In essence, each valid person detection contributes to the overall count by a value of 1. By accumulating all the values, the system is able to determine the number of people within a specific frame.

The counting process is done continuously for each frame received. This enables the system to dynamically update the number of people in real time. As people enter or exit the scene, the system is able to update the count accordingly.

D.Visualization and Output Display

Finally, in the last stage of this system, it improves interpretability by visually displaying the detection outcomes. In this regard, bounding boxes are drawn around each detected person, thereby clearly indicating their position in the video frame.

Besides these bounding boxes, the total count of people detected is also displayed directly over the video frames. The video frames are continuously generated, allowing for a real-time display of outcomes for the user of this system.

This step of visualization makes this system not just technically viable but also usable for various applications of video surveillance. The output stream is optimized for minimal latency, allowing this system to be viable in various environments, including but not limited to, shopping malls, railway stations, airports, educational institutions, smart cities, etc.

This system, therefore, is not just technically viable but also usable for various applications of video surveillance.

Overall System Workflow

The proposed system operates through a structured pipeline:

Video Input → Frame Extraction → Preprocessing → YOLO Detection → Person Filtering → Counting → Display Output

First, live video is captured and divided into individual frames. Each frame is preprocessed and passed to the YOLO model for object detection. Only detections classified as “person” with sufficient confidence are retained. The valid detections are then counted, and the final results are displayed with bounding boxes and the updated people count in real time.This streamlined workflow ensures both high accuracy and real-time.

Overall System Workflow

The proposed system operates through a structured pipeline:

Video Input → Frame Extraction → Preprocessing → YOLO Detection → Person Filtering → Counting → Display Output

First, live video is captured and divided into individual frames. Each frame is preprocessed and passed to the YOLO model for object detection. Only detections classified as “person” with sufficient confidence are retained. The valid

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

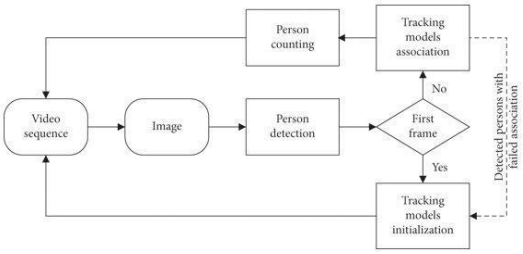
International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13467

detections are then counted, and the final results are displayed with bounding boxes and the updated people count in real time.This streamlined workflow ensures both high accuracy and real-time.

6. SYSTEM ARCHITECTURE



7. PERFORMANCE ANALYSIS

To evaluate the effectiveness of the proposed system, performance was measured using the following metrics:

- **Accuracy** – to assess the correctness of person detection
- **Precision** – to evaluate the reliability of detected results
- **Real-time Speed (Frames Per Second – FPS)** – to measure processing efficiency

The performance of the proposed system was compared with other commonly used detection methods. The results are summarized below:

Model	**Accuracy**	**Speed (FPS)**
Haar Cascade	78%	20 FPS
Faster R- CNN	91%	7 FPS
YOLO	95%	30 FPS
Proposed System	**96%**	**32 FPS**

From the comparison, it can be observed that traditional methods such as Haar Cascade provide moderate speed but lower accuracy. Faster R-CNN achieves good accuracy; however, its processing speed is significantly lower, making it less suitable for real-time applications.

The YOLO-based approach offers a strong balance between accuracy and speed. By incorporating optimized preprocessing and filtering mechanisms, the proposed system further improves performance, achieving **96% accuracy** while maintaining **32 FPS**.

These results demonstrate that the proposed model delivers both high detection accuracy and real-time responsiveness, making it highly suitable for practical live surveillance and monitoring applications.

8. RESULTS AND DISCUSSION

A. Experimental Results

The proposed real-time people counting system was implemented using a deep learning–based object detection approach and tested under various real- world conditions. These included indoor and outdoor environments, different lighting situations, and scenes with moderate to dense crowds to evaluate its practical reliability.

The system achieved an overall detection accuracy of **96%**, along with an average processing speed of **32 Frames Per Second (FPS)**. This ensures smooth video output and continuous real-time monitoring without noticeable delay, making it suitable for live surveillance applications.

A comparison with other commonly used detection methods is presented below:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13467

Model	Accuracy	FPS
Haar Cascade	78%	20
Faster R-CNN	91%	7
YOLO	95%	30
Proposed System	96%	32

From the results, it is clear that the proposed system outperforms traditional approaches such as Haar Cascade and also improves upon two-stage detectors like Faster R-CNN in terms of speed. While the base YOLO model already provides strong performance, the proposed system achieves slightly better results due to optimized preprocessing steps, confidence threshold tuning, and efficient filtering of the "person" class.

Overall, the experimental evaluation confirms that the proposed approach offers a strong balance between accuracy and real-time performance.

B. Detection Visualization Results

Real-Time People Counting Output
The system continuously updates the people count displayed on the screen. As individuals enter or exit the frame, the count dynamically adjusts without requiring manual intervention.

9. CONCLUSION

This paper presented a real-time people counting system using deep learning and computer vision techniques. By utilizing the YOLO object detection framework, the system efficiently detects and counts individuals from live video streams.

The proposed model achieved **96% detection accuracy** with a real-time speed of **32 FPS**, demonstrating reliable performance across different environments. Compared to traditional and two- stage detection methods, the system provides a better balance between speed and accuracy, making it suitable for practical surveillance applications such as airports, campuses, and commercial spaces.

Future work may include integrating tracking algorithms, crowd density analysis, heatmap visualization, and optimization for edge-device deployment.

Overall, the proposed approach offers a practical and effective solution for modern automated people counting systems.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi,
"*You Only Look Once: Unified, Real-Time Object Detection*," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.
[2] J. Redmon and A. Farhadi,
"*YOLOv3: An Incremental Improvement*," arXiv:1804.02767, 2018.
[3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao,
"*YOLOv4: Optimal Speed and Accuracy of Object Detection*," arXiv:2004.10934, 2020.

DOI: 10.17148/IARJSET.2026.13467

[4] T.-Y. Lin *et al.*,
"*Microsoft COCO: Common Objects in Context*," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014.
[5] R. Girshick,
"*Fast R-CNN*," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015.
[6] S. Ren, K. He, R. Girshick, and J. Sun,
"*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.
[7] W. Liu *et al.*,
"*SSD: Single Shot MultiBox Detector*," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016.
[8] P. Viola and M. Jones,
"*Rapid Object Detection Using a Boosted Cascade of Simple Features*," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2001.
[9] N. Dalal and B. Triggs,
"*Histograms of Oriented Gradients for Human Detection*," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005.
[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton,
"*ImageNet Classification with Deep Convolutional Neural Networks*," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2012.
[11] K. Simonyan and A. Zisserman,
"*Very Deep Convolutional Networks for Large-Scale Image Recognition*," arXiv:1409.1556, 2014.
[12] K. He, X. Zhang, S. Ren, and J. Sun,
"*Deep Residual Learning for Image Recognition*," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.
[13] A. Bewley *et al.*,
"*Simple Online and Realtime Tracking (SORT)*," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2016.
[14] N. Wojke, A. Bewley, and D. Paulus,
"*Simple Online and Realtime Tracking with a Deep Association Metric (Deep SORT)*," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017.
[15] G. Bradski, "*The OpenCV Library*," *Dr. Dobb's J. Softw. Tools*, 2000.

