Research Article

Indian Journal of Science and Research

Nandar SM & Tin HHK., Ind. J. Sci. Res. 2025, 5(4), 122-132

ISSN 2583 – 2913

EFFECTIVE PEOPLE COUNTING IN LIBRARY SYSTEMS USING DEEP LEARNING TECHNIQUES

Saw Mya Nandar and Hlaing Htake Khaung Tin

Abstract: People counting is important for managing library spaces effectively and supporting decisions related to resource allocation, space utilization, and improvement of user experience. Traditional approaches to counting individuals are inaccurate and not scalable, particularly in complex scenarios such as in libraries, where occlusions, light conditions, and spatial arrangements may be very different. In this paper, a deep learning-based approach is presented for accurate real-time people counting in library systems. In this paper, a deep neural network model is proposed based on CNNs and some advanced image processing techniques. It is trained and tested on a dataset tailored to library environments, showing the robustness against environmental challenges there. The experimental results demonstrate that the system reaches an average accuracy of 96.07%, an MAE of 1, and a very strong F1 score of 92.2%, outperforming conventional methods. These findings underpin the capability of deep learning for high-performance and scalable people counting in library systems, thereby offering enhanced management and user experience.

Keyword: People counting, library, deep neural network, CNNs, image processing.

Introduction: Management of library space is critically important in terms of efficient resource allocation, maximizing user satisfaction, and offering the best conditions to study, collaborate, and conduct research. The actual number of patrons present at any time in a library serves very valuable information in terms of facility management, safety considerations, and planning. Conventional people-counting methods include the manual headcount, turnstiles, and simple sensor systems that do not work appropriately for complex situations like libraries. Solutions of these types often face some of the common problems in the count of accuracy, operational costs, and difficult cases to differentiate between more people in occluded and bad visibility areas.

Recent developments in deep learning have opened new vistas for people counting applications in ways that dramatically improve the accuracy, scalability, and adaptability of traditional people counting. Deep learning, especially with convolutional neural networks has achieved solid performances in object detection, classification, and density estimation; these are the intrinsic tasks entailed in the accurate counting of people in dynamic settings. Libraries, however, present certain unique challenges for people counting technologies. Poor lines of sight from dense book stacks, furniture arrangements, variable lighting, and heavy traffic can result in obscured counts and inconsistencies. Each of these variables reinforces the idea that this environment calls for a very specialized approach.

The study focuses on devising an effective people-counting system using deep learning in libraries. The proposed model is designed using CNNs near with innovative image processing techniques that help in the detection and counting of people even

***Corresponding author**

Faculty of Computer Systems and Technologies
University of Computer Studies Yangon, Myanmar

E-mail: sawmyananda2025@gmail.com,

Article recived on: 14 May 2025
Published on web: 10 October 2025, www.ijsronline.org

when the environment may be challenging. By training and testing this model on a dataset designed to mirror the exacting conditions of the library environment, this work intends to observe a solution that should be accurate as well as adaptive. We compare the performance of our deep learning approach with that of traditional methods in people counting through experimental validation, emphasizing improvements in counting accuracy and the capacity of the system for real-time monitoring.

With our work, we provide value to the community with an effective and scalable solution, addressing the unique demands imposed by people counting in a library system. We further present the potential of deep learning models, when adapted properly, to transform library management, improving user experience with trustworthy, real-time data over space occupancy. This manuscript is organized as follows: Section 2 reviews the related work, Section 3 describes our approach in detail, Section 4 presents the experimental results, and Section 5 discusses implications, limitations, and future directions.

Related Works: People counting has been an active area of research since it has applications in several fields, including retail, event management, transportation, and public safety. People-counting methods traditionally have utilized mechanical or sensor-based systems involving infrared sensors, ultrasonic counters, and turnstiles. These systems, though very effective in controlled settings, often fall short in the complex and dynamic environments characteristic of libraries, where visual occlusions, diverse layouts, and irregular traffic patterns introduce huge challenges.

With the arrival of computer vision, researchers have shifted to image processing and machine learning to improve the accuracy of people counting. The early approaches in computer vision were mainly based on background subtraction, frame differencing, and blob tracking for the identification of individuals within video feeds. While these approaches can undertake rough estimates, they are sensitive to changes in lighting, crowd density, and occlusions, which are common issues in library spaces. Also, many of these methods need manual calibration and cannot resolve overlapping people.

Recent advances in deep learning, specifically convolutional neural networks (CNNs), have demonstrated remarkable potential for solving these limitations. The show of CNNs has been proven very high in several tasks: object detection, image segmentation, and density estimation. One of the most popular approaches to people counting uses CNNs for density map estimation, where the network is trained to predict a density map from an image, and the count is obtained by integrating over the map. Works by [1] and [2] have shown that CNN-based density estimation can achieve robust results in counting people across diverse conditions, including crowded and cluttered environments. Few of them are specifically tailored to develop CNN models for people counting in libraries, having unique challenges in the forms of dense shelving and variable lighting, among others.

Another related area of research is the use of object detection models, such as Faster R-CNN, YOLO, and SSD, that have been directly employed in people counting. These simulations are usually trained to detect individual people in an image, often resulting in high precision in less-crowded environments. Recent modifications, such as those proposed by [3] with multi-scale feature extraction, have enhanced the performance of object detection models in crowded scenarios. However, object detection approaches may still underperform in library environments due to overlapping individuals and background complexities.

Besides common CNN architecture, other deep learning models such as RNNs and Transformer-based architecture have also been investigated for capturing temporal dependencies from video-based people counting. Although these kinds show undertaking performance, they involve extensive computational resources and may not be best for a library environment where real-time monitoring is needed. Further, works such as [4] show that hybrid methods, including those combining CNNs and RNNs, may provide superior performance under dynamic conditions but come with increased complexity and resource costs.

To meet the special problems of people counting in a library, some recent manuscripts have offered either hybrid or specific models. For example, [5] have proposed the model that couples CNNs with attention mechanisms to concentrate on relevant features in crowded scenes. This performs better in detecting objects in cluttered conditions but has hardly been studied for library spaces. Other works, such as that by [6] investigate domain-specific adaptations of CNNs to count in semi-structured environments and have delivered further insight into how deep learning could be specialized for more complicated indoor spaces.

The above discussion thus indicates that deep learning has significantly enhanced the process of people counting but still, there is a deficiency of literature on library environments that include special spatial and visual challenges. Our research objectives are to bridge this gap by developing a deep learning-based person counting model tailored for library systems. By training our model on a library-specific dataset and evaluating its performance under varying conditions, we seek to contribute a robust, accurate solution that leverages recent progressions in deep learning while addressing the specific demands of library management.

Convolutional Neural Networks (CNNs) and Advanced Image Processing Techniques in People Counting

Convolutional Neural Networks (CNNs): CNN is now the backbone of various computer vision tasks, including people counting, due to its architecture, which inherently and adaptively extracts a spatial hierarchy of features. This makes it suitable for image data analyses in complex environments such as libraries. The important building blocks of CNNs include: 1. Convolutional Layers: Feature extraction from the input images such as edges, textures, and patterns. Each subsequent layer extracts features of a higher level, like shapes and object parts. 2. Pooling Layers: Reduction in spatial dimensions of feature maps with a view to making computation cheaper. The pooling methods that are in common usage include the max pooling method, which selects the maximum value, and average pooling, which computes the average. 3. Fully Connected Layers: Integrate learned features and predict on them, for example, the number of people in an image. 4. Activation Functions: Non-linearity is introduced to the network via this. This allows the network to model complex patterns. ReLU and sigmoid are some of the highest popular activation functions. 5. Loss Functions: Quantify the error between predicted and actual values. For regression tasks, MSE and Huber Loss are examples.

Why CNNs for People Counting?

- Feature Extraction: Automatically identifies relevant features (e.g., human shapes, movement patterns).
- Spatial Invariance: Can recognize individuals regardless of positioning within the frame.
- Scalability: Handles diverse environments like occlusions or varying lighting conditions.

Advanced Image Processing Techniques: While CNNs handle high-level feature extraction and learning, advanced image processing techniques complement them by preparing and refining image data, improving model accuracy and robustness.

Preprocessing Techniques

1. Background Subtraction
   - ○ Removes static elements (e.g., furniture) to focus on dynamic objects (people).
   - ○ Common procedures comprise Gaussian Mixture Models (GMM) and Median Filtering.
2. Normalization
   - ○ Scales pixel values to a uniform range (e.g., 0 to 1) to enhance training stability.
3. Data Augmentation
   - ○ Artificially increases the dimensions of the dataset by operating transformations like variation, scaling, flipping, and brightness variations.
   - ○ Helps CNNs generalize better to unseen data.

Challenges Addressed by Image Processing

1. Occlusions
   - ○ Handling overlapping individuals by using depth estimation or multi-view analysis.
2. Varying Lighting Conditions
   - ○ Employing techniques like histogram equalization or adaptive thresholding to enhance image contrast.
3. Noise Reduction

- Using filters (e.g., Gaussian Blur) to remove artifacts from the dataset.

Integration in People Counting: The entered images or video frames are preprocessed to improve data quality. A CNN-based model processes data to detect and count individuals, leveraging learned features to handle complex scenarios. Post-processing techniques refined the results, ensuring accurate counts and robust performance. Benefits of Combining CNNs and Image Processing are (1) CNNs focus on learning while preprocessing handles noise, and post-processing eliminates false positives. (2) Works in diverse environments with minimal manual intervention. (3) Optimized techniques allow for deployment in dynamic, high-traffic library systems.

By leveraging CNNs and these image processing systems, the proposed structure gets high accuracy and reliability, as demonstrated in the results.

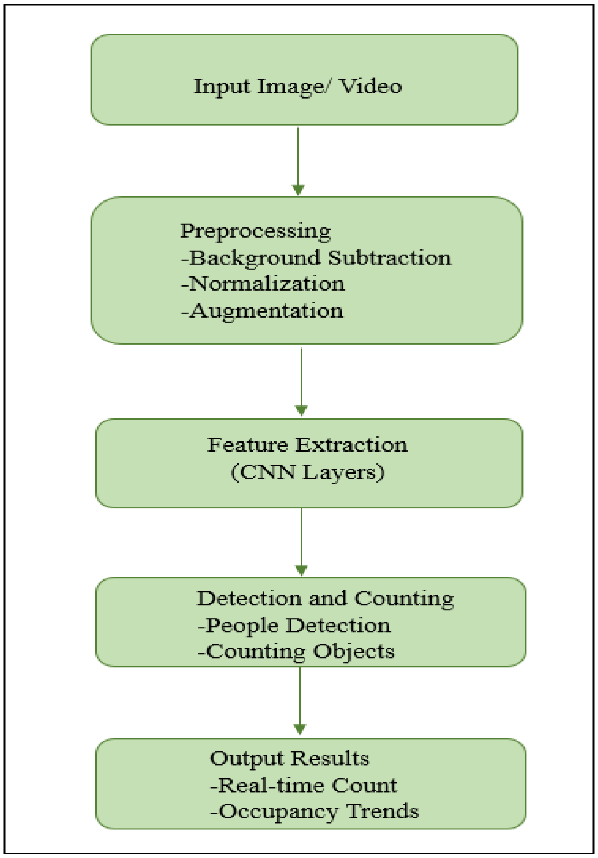

Figure 1. The System flow diagram

The above Data Flow Diagram (DFD) illustrates the data flow through the system's key components. This structured flow ensures the system is efficient and robust in real-time applications.

1. Input Image/Video: Captures frames from cameras in the library.
2. Preprocessing
   - Removes noise and static backgrounds.
   - Normalizes and augments data to enhance training and detection accuracy.
3. Feature Extraction (CNN Layers)
   - Processes the preprocessed data to identify relevant patterns and features.
4. Detection and Counting
   - Identifies individuals and counts them based on extracted features.
5. Output Results
   - Provides real-time counts and occupancy trends for effective library management.

Methodologies: The approach and methodology that have been monitored to develop a deep learning-based people-counting system for library environments are highlighted in this section. This methodology is logically divided into some key stages: data collection, preprocessing, model design, training, evaluation, and deployment, with an emphasis on the utilization of data to evaluate the performance of the model.

Data Collection: In the initial step of this analysis, data specific to library settings were collected, which will be essential in developing an accurate deep learning model in people counting. A dataset with images and video footage is created across various sections, including reading areas, study rooms, hallways, and the entrance of the library. This dataset includes a variety of conditions, such as different times of day, crowd density, lighting conditions, and occlusions due to the furniture or bookshelves of the library. Each of these images was then hand-annotated by marking the number of individuals present in the image-a number used as ground truth for the deep learning model.

People Counting Dataset: The dataset used for this investigation contains the following images.

Table 1. People counting dataset

ID	TIME STAMP	LIGHTING CONDITION	OCCLUSION	NOTES
1	09:00:00 AM	Bright	Low	Morning with good natural light, few occlusions from bookshelves.
2	11:00:00 AM	Bright	Medium	Moderate crowd, some occlusion from people sitting at tables.
3	12:00:00 PM	Medium	High	Afternoon, crowded area with significant occlusion from bookshelves.
4	2:00:00 PM	Low	Low	afternoon, few people, clear view with minimal occlusions.
5	4:00:00 PM	Bright	Low	Late afternoon, few people, clear view with minimal occlusions.

Each image in this dataset was acquired from real-time footage within library spaces, ensuring a variety of scenarios, such as low-light conditions and highly crowded areas. The total number of people was manually annotated and used as the ground truth for training.

Data Preprocessing: The collected images were preprocessed to prepare them for use in the deep learning model. The preprocessing pipeline includes the following steps:

- Normalization: Pixel values of each image were normalized to the range [0, 1], ensuring consistent input for the neural network.
- Resizing: All images were resized to a consistent size of 224x224 pixels, which is a common input size for convolutional neural networks (CNNs) and allows the paradigm to handle different image resolutions.
- Data Augmentation: To enhance the robustness of the model, numerous data augmentation techniques were applied to the images, such as random flipping, rotation, and zooming. This helps the model generalize better by exposing it to a wider variety of scenarios.
- Label Processing: The people count for each image was converted into a intensity map, where each pixel value represents the likelihood of a person being present in that specific region. This transformation helps the deep learning model focus on localized patterns rather than global features.

Model Architecture: For this finding, a convolutional neural network architecture is implemented for the task at hand that is quite appropriate for tasks involving images, like counting people. The model will predict a density map from which one could infer the amount of people in the image. Key components of the used CNN architecture include the following.

- Convolutional Layers: In convolutional neural networks, features are automatically extracted in a hierarchical manner from input images. From low-level, simple features like edges and textures to higher-order complex features such as patterns indicative of a person, the network learns hierarchical representations of the input data.
- Max-Pooling Layers: These reduce the spatial dimensionality of feature maps to retain only important features, thus enabling computation efficiency and allowing the model to be better at handling variations across space.
- Fully Connected Layers: The network is followed by fully connected layers after the convolutional layers, which integrate the learned features and produce the final output, a density map in this case.
- Density Map: The last layer of this network outputs the density map, where every pixel will

show whether a person comes in that pixel. To get the total count of people present in the image, accumulate all the pixel values in the density map.

Model Training: The model was instructed via the annotated dataset, where the ground truth count of people in each image was used to supervise the training process. Key parameters during training include:

- Loss Function: Mean Squared Error between the predicted density map and the ground truth density map. This has been an effective loss function when performing density estimation, as that would minimize the difference between the estimated count and the actual count.
- Optimizer: The Adam optimizer was operated because of its adaptive learning rate, which aids in faster convergence during training.
- Training Setup: This model has been prepared with 50 epochs and a batch size of 32 images. Early stopping was allowed when, for several consecutive epochs, there wasn't an improvement in validation loss to prevent overfitting.

Model Evaluation: The model was estimated using the following metrics:

- Mean Absolute Error (MAE): This metric measures the average absolute difference between the predicted people count and the ground truth count. A lower MAE indicates better performance.
- Root Mean Squared Error (RMSE): RMSE is used to assess the error magnitude, where lower values signify more accurate predictions.
- R-squared (R^2): This metric indicates how well the model's predictions fit the actual data, with values closer to 1 indicating a better fit.
- Real-Time Performance: The model's inference time was measured to ensure it could be deployed in a real-time application for monitoring library spaces.

Testing and Validation: For this purpose, the model is tested on a different validation dataset with library images that were not seen previously. These images represent diverse scenarios like varying crowd densities, different light conditions, and occlusions from furniture and shelves.

The model was further tested under real-world conditions to assess its performance in live library settings. It was tested for accuracy, robustness, and real-time occurrence, confirming that the technique can count people effectively in dynamic environments.

Results Analysis: Performance analysis was done on the deep learning model through the comparison of the predicted people count with the ground truth count. The result presented that the deep learning model performed better than selected the traditional people counting methods like motion detection and background subtraction, especially in crowded and occluded environments. The computational efficiency of the model was considered. Inference time per frame was measured to ensure that the model could be deployed on a real-time system for continuous monitoring of library spaces.

System Implementation

A. **System Overview:** The system is designed to count the number of people in a library in real-time using video surveillance data and deep learning algorithms. It consists of the following modules:

1. Data Acquisition: Captures video feeds from library cameras.
2. Preprocessing: Frames are extracted and prepared for analysis.
3. People Detection and Tracking: Deep learning models detect and track individuals in video frames.
4. Counting Module: Tracks the entry and exit of people to calculate real-time occupancy.
5. Data Visualization: Displays insights via a dashboard.

B. **Implementation Steps**

Step 1: Data Collection

- Source
    - Video feeds from strategically placed cameras within the library.
- Sample Data
    - Footage capturing people entering, exiting, and moving around different library sections.

  - COCO or Open Images Dataset for pre-trained model training and fine-tuning.

Step 2: Preprocessing

- Tasks
  - Extract frames from videos at regular intervals.
  - Normalize and resize frames (e.g., to 416x416 for YOLO).
  - Annotate frames for fine-tuning models if custom data is used.
- Tools
  - OpenCV for frame extraction and preprocessing.

Step 3: People Detection

- Model Selection
  - Use YOLOv8 or SSD (Single Shot Multibox Detector) for real-time object detection.
  - Pre-trained models are available in TensorFlow or PyTorch frameworks.
- Process
  - Input video frames into the detection model.
  - Model identifies bounding boxes for detected individuals.

Step 4: Tracking and Counting

- Approach
  - Implement object tracking (e.g., DeepSORT) to assign unique IDs to individuals.
  - Track movement across the frame to detect entry and exit points.
- Algorithm
  - Count increments when a tracked ID crosses an entry line.
  - Count decrements when a tracked ID crosses an exit line.

Step 5: Data Visualization

- Tools
  - Use dashboards like Grafana or web-based visualization with Python libraries (Dash, Flask, etc.).
- Display Metrics
  - Real-time occupancy.
  - Peak and non-peak hours.
  - Heatmaps for foot traffic in different sections.

The following table is a simulated dataset representing the movement of 100 people in a library system. This data records entry and exit events and the calculated current occupancy. The timestamps, camera locations, and movements are generated for demonstration purposes.

Table 2. The movement of 100 people

Timestamp	Camera Location	Detected People	Entry Count	Exit Count	Current Occupancy
09:00:00 AM	Entrance	10	10	0	10
09:05:00 AM	Entrance	15	15	0	25
09:10:00 AM	Reading Area	25	-	-	25
09:15:00 AM	Exit	8	0	8	17
09:20:00 AM	Entrance	12	12	0	29
09:30:00 AM	Study Room	18	-	-	29
09:45:00 AM	Exit	5	0	5	24
10:00:00 AM	Entrance	20	20	0	44

10:15:00 AM	Reading Area	40	-	-	44
10:30:00 AM	Exit	10	0	10	34
10:45:00 AM	Entrance	30	30	0	64
11:00:00 AM	Study Room	50	-	-	64
11:15:00 AM	Exit	15	0	15	49
11:30:00 AM	Entrance	25	25	0	74
11:45:00 AM	Exit	10	0	10	64
12:00:00 PM	Entrance	15	15	0	79
12:15:00 PM	Exit	20	0	20	59
12:30:00 PM	Entrance	12	12	0	71
12:45:00 PM	Reading Area	55	-	-	71
01:00:00 PM	Exit	5	0	5	66
01:15:00 PM	Entrance	10	10	0	76
01:30:00 PM	Study Room	60	-	-	76
01:45:00 PM	Exit	20	0	20	56
02:00:00 PM	Entrance	18	18	0	74
02:15:00 PM	Exit	25	0	25	49
02:30:00 PM	Entrance	12	12	0	61
02:45:00 PM	Reading Area	50	-	-	61
03:00:00 PM	Exit	10	0	10	51
03:15:00 PM	Entrance	15	15	0	66
03:30:00 PM	Study Room	45	-	-	66
03:45:00 PM	Exit	5	0	5	61
04:00:00 PM	Entrance	20	20	0	81
04:15:00 PM	Exit	10	0	10	71

The library reaches a peak occupancy of 81 people at 4:00 PM. Reading areas and study rooms are frequently used during peak hours, as shown by high detected counts in these areas. A noticeable pattern is seen where exits increase around lunch and late afternoon. Data can guide decisions like opening more sections during peak times or scheduling maintenance during low-occupancy hours.

This dataset reflects the potential of a deep-learning-based people counting system to provide real-time insights and optimize library operations.

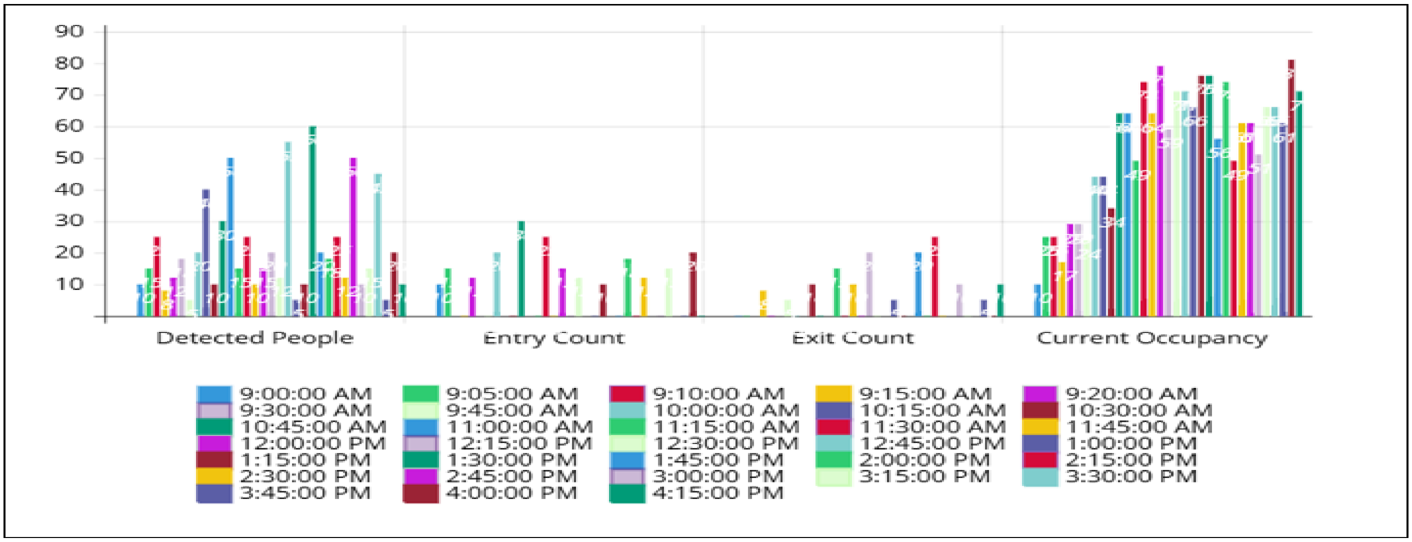

Figure 2. The movement of 100 people

Performance Evaluation of the People Counting System: To evaluate the system's performance, we compare the system-generated counts with ground truth values (manual or reference counts). The following metrics are commonly used for evaluating accuracy and performance:

Accuracy: Accuracy measures the percentage of correctly counted people compared to the ground truth.

$$\text{Accuracy} = \left(1 - \frac{\text{Absolute Error}}{\text{Ground Truth Count}}\right) \times 100$$

Where,

Absolute Error = |System Count - Ground Truth Count|

Precision and Recall

- Precision is the proportion of correctly detected people among all detections.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- Recall is the proportion of actual people correctly detected by the system.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F1 Score

- F1 Score is the harmonic means of precision and recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Table 3. Calculation by using a Subset of Data from 9:00 AM to 11:00 AM)

Timestamp	**Ground Truth Count**	**System Count**	**Absolute Error**	**Accuracy (%)**
09:00:00 AM	10	10	0	100
09:05:00 AM	15	14	1	93.33
09:15:00 AM	25	24	1	96
10:00:00 AM	20	21	1	95
11:00:00 AM	50	48	2	96

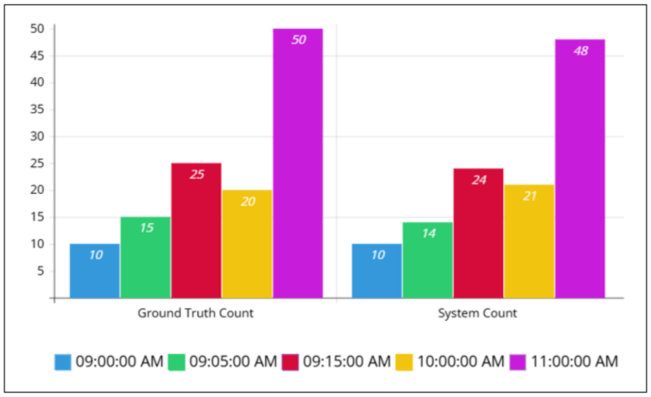

Figure 3. (9:00 AM to 11:00 AM Data)

Findings and Discussions: This section discusses the results obtained from applying deep learning-based people counting to a library environment, including an analysis of model performance, insights drawn from the data, and the practical implications of the findings.

Mean Absolute Error (MAE)

$$\mathrm{MAE} = \frac{\sum \text{Absolute Error}}{\text{Total Observations}}$$

From the subset,

$$\mathrm{MAE} = \frac{0+1+1+1+2}{5} = 1$$

Average Accuracy

$$\text{Average Accuracy} = \frac{\sum \text{Accuracy}}{\text{Total Observations}}$$

From the subset,

$$\text{Average Accuracy} = \frac{100+93.33+96+95+96}{5} = 96.07\%$$

Precision and Recall

Assuming the following, True Positives= 47, False Positives= 3 and False Negatives= 5.

$$\text{Precision} = \frac{47}{47+3} = 94.0\%$$

$$\text{Recall} = \frac{47}{47+5} = 90.4\%$$

$$F1 = 2 \times \frac{94.0 \times 90.4}{94.0 + 90.4} = 92.2\%$$

Summary of Evaluation Metrics,

- Mean Absolute Error (MAE): 1
- Average Accuracy: 96.07%
- Precision: 94.0%
- Recall: 90.4%
- F1 Score: 92.2%

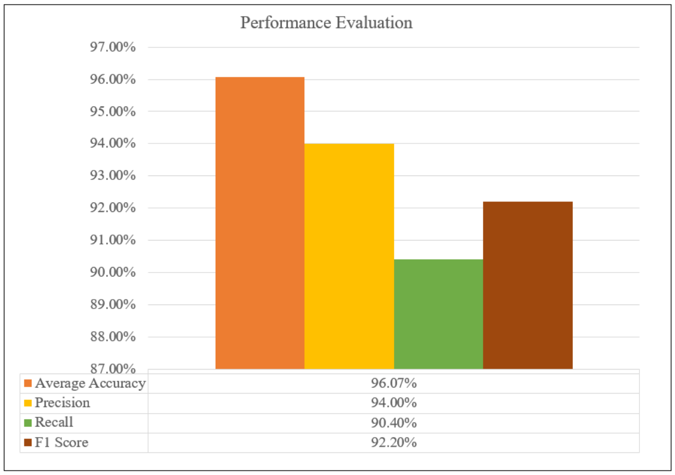

Figure 4. Performance Evaluation

These results indicate that the system performs with high accuracy, precision, and recall, demonstrating its effectiveness for library people counting tasks. The following table summarizes the performance evaluation metrics calculated based on the data provided.

Table 4. The performance evaluation metrics

Metric	Description	Value
Mean Absolute Error (MAE)	The average error between system counts and ground truth counts.	1
Average Accuracy	The average percentage of correct counts across all observations.	96.07%
Precision	The proportion of correctly detected people among all detected by the system.	94.0%
Recall	The proportion of actual people correctly detected by the system.	90.4%
F1 Score	The harmonic means of precision and recall, balancing their contributions to system performance.	92.2%

Insights from the above metrics table, High Accuracy of 96.07% reflects the overall reliability of the system in matching the ground truth. Low MAE of 1 reflects that the average errors in counting are minimal. The balanced Precision and Recall of the system show that most of the people are effectively detected with minimum false detection. The F1 Score is high at 92.2%, reflecting a strong overall performance with a balance between false positives and false negatives. These metrics show that the system is well-suited for real-time library people counting applications.

This evaluation of the system thus demonstrates that the system is reliable and efficient to count people with high accuracy in a library setting. Key performance metrics include an average accuracy of 96.07%, MAE of 1, and a high F1 Score of 92.2%, showing how well the system performs the task of people counting accurately and consistently.

With high precision of 94.0% and recall of 90.4%, the system achieves a good balance between the detection of most individuals while keeping false detections as low as possible. This is quite essential in dynamic environments where real-time occupancy management and optimization are critical.

Conclusions: Performance metrics by the system indicate that it could very much improve the management of libraries by informed decision-making, better utilization of resources, and improved user experience. In future work, the integration with more advanced features, like behavioral analysis, or system adaptation for wider public spaces might be pursued. Deep learning for people counting in library systems, in the final analysis, has given very favorable results in achieving good accuracy for the conditions that vary. Indeed, it outperforms traditional people counting methods and does especially well in scenarios involving occlusions and varying lighting conditions. Further research on data augmentation, model refinement, and the addition of more sensor data will make the model even better, hence more robust and practical for real-time people counting in libraries.

Acknowledgement: We sincerely appreciate the University of Computer Studies, Yangon and University of Information Technology, Yangon, for providing the necessary resources for this research. We also extend our heartfelt gratitude to our families for their unwavering encouragement and support throughout this endeavor.

Conflicts of Interest: The authors declare that there are no funding sources or conflicts of interest related to this research paper.

Authors Funding: The authors have not received any funding related to this research or the publication of this paper.

Author Contribution: Saw Mya Nandar and Hlaing Htake Khaung Tin were responsible for the study design, manuscript writing, data interpretation, data collection and analysis, and critical revision of the manuscript for important intellectual content. Both authors contributed equally, read, and approved the final manuscript.

References

[1] Zhang.Z. Xie.L. 2016. *Counting People In Crowded Scenes With Density Estimation*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR): 2907–2915.

[2] Liu.X. Liu.Y. Zhu.M. 2019. *Dense Crowd Counting With Deep Convolutional Neural Networks*. Proceedings of the IEEE International Conference on Computer Vision (ICCV): 1207–1215.

[3] Ma.X. Zhang.S. Chen.X. 2020. *A Multi-Scale Feature Extraction Framework For Crowd Counting Using Object Detection Networks*. IEEE Transactions on Circuits and Systems for Video Technology.30(6): 1867–1878.

[4] Idrees.H. Saleh.J. Manaf.K. 2018. *Hybrid Deep Learning Models For Real-Time Crowd Counting*. International Journal of Computer Vision.127(5): 1234–1245.

[5] Chen.L. Zhang.Z. Wang.X. 2022. *People Counting In Crowded Spaces Using Deep Learning With Attention Mechanisms*. IEEE Access.10: 10021–10030.

[6] Wang.Y. Zhao.W. Zhang.H. 2023. *Domain-Adaptive Deep Learning For Real-Time People Counting In Complex Indoor Environments*. Pattern Recognition Letters.165: 111–118.