INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 3, MARCH 2026 | ISSN: 2456-4184 | IJNRD.ORG

HYBRID FRAMEWORK FOR CROWD MONITORING AND DENSITY CLASSIFICATION

[1]Prof. D.V. Thombre, [2]Kedar Palakudtewar, [3]Omkar Mane, [4]Rohan Belsare
[1]Professor of (Computer Engineering Department) Terna Engineering College, Navi Mumbai, Maharashtra, India
[2, 3,4]Students of (Computer Engineering Department) Terna Engineering College, Navi Mumbai, Maharashtra, India

Abstract: Crowd monitoring plays a vital role in contemporary surveillance systems; however, conventional methods often struggle with issues like occlusion, high crowd density, and intricate movement patterns. To address these shortcomings, this research introduces a novel deep learning-based system that fuses YOLO for real-time human detection DeepSORT for tracking multiple individuals, and ResNet-18 for classifying crowd density. The aim is to boost accuracy and robustness in both low- and high-density environments, surpassing the constraints of single-technique solutions. The empirical analysis reveals enhanced detection precision and reliable density assessment, with real-time metrics displayed via a Streamlit-powered interface. This integrated approach advances the field by illustrating that simultaneous use of detection, tracking, and classification networks can deliver richer surveillance data, facilitate rapid identification of anomalies, and enable more effective crowd management. The proposed framework offers significant benefits for public safety applications, including surveillance, event oversight, and smart city initiatives. Testing showed the system achieved close to 90% detection accuracy and maintained real-time performance at approximately 25 frames per second on evaluation datasets.

Keywords- Crowd Monitoring ,Smart Surveillance,Deep Learning ,Hybrid Framework, YOLO (You Only Look Once),DeepSORT, Convolutional Neural Network (CNN), ResNet-18 ,Crowd Density Classification, Real Time Analytics, Streamlit Dashboard ,Public Safety

I. INTRODUCTION

Crowd monitoring is increasingly recognized as a crucial component for public safety in spaces like railway stations, large gatherings, and within smart city infrastructure. Conventional surveillance methods, which depend largely on manual observation, are often inefficient and susceptible to human error. Challenges such as occlusion, high density, and complex crowd dynamics further undermine the effectiveness of these traditional approaches, making timely and accurate analysis difficult. To address these obstacles, this research introduces an advanced automated framework that leverages deep learning for improved crowd analysis. The proposed hybrid system integrates YOLO for real-time person detection, DeepSORT for multi-object tracking, and ResNet-18 as a convolutional neural network for classifying crowd density. This combination enhances system performance for both sparse and densely populated scenarios. The solution enables real-time, efficient monitoring, supports proactive identification of high-risk condition , and enhances decision-making in domains such as public safety, event management, and intelligent surveillance.

Effective crowd analysis is fundamental to the safety and management of public environments including railway stations, airports, shopping centers, and large events. As urban populations grow, the significance

of monitoring crowd behavior also increases, helping to prevent incidents, accidents, and security threats. Smart surveillance systems can provide authorities with actionable, real-time information about crowd density, movement patterns, and risk indicators, reducing the burden on personnel and boosting operational efficiency.

Despite recent technological advances, crowd monitoring presents persistent challenges. Occlusion—where individuals block each other from view—makes accurate detection difficult, particularly in high-density conditions where traditional algorithms may struggle to separate individuals in tightly packed groups. Complex crowd behaviors, including abrupt movement, directional changes, and irregular flow, add further complexity to tracking and analysis. Environmental factors such as lighting, camera placement, and the requirement for real-time processing also complicate the task of building robust solutions.

The main aim of this study is to design an intelligent, real-time crowd monitoring platform capable of delivering precise analysis under diverse conditions. Specific objectives include:

1. Detecting and tracking individuals using YOLO and DeepSORT
2. Analyzing crowd movement and calculating metrics such as count, motion, and violations
3. Generating density heatmaps to visualize crowd distribution
4. Classifying density levels (Sparse or Dense) using a ResNet-18 CNN model
5. Providing real-time alerts for abnormal or risky scenarios

The proposed architecture adopts a hybrid approach, merging detection, tracking, and classification modules. Video streams are processed via YOLO to identify individuals, whose locations are then tracked using DeepSORT, assigning unique IDs across frames. During this process, metrics such as motion energy and social distancing violations are computed. Simultaneously, the CNN classifier examines select frames to determine crowd density. Outputs from these subsystems are aggregated and presented through a Streamlit dashboard, featuring live video, heatmaps, analytics, and alert notifications. This integrated strategy offers robust accuracy and reliability for both sparse and dense crowd conditions.

II. LITERATURE REVIEW / RELATED WORK

Extensive research in computer vision and deep learning-based surveillance systems has been spurred by the increasing demand for automated crowd monitoring. Numerous models and algorithms for identifying, monitoring, and analyzing crowd behavior have been investigated in a number of previous researches. Some of the most pertinent works that serve as the basis for the suggested system are reviewed in this section.

2.1. Existing Work

S.No.	Author & Year	Technique Used	Key Contribution	Limitation	Relevance
1	Kaka Khel et al., 2023	Hybrid YOLOv4 + Deep SORT	Real-time crowd monitoring with speed and direction estimation; 92.1% mAP at 48 FPS	Inaccurate speed estimation due to pixel displacement	Directly relevant – same model used in your project
2	Liu et al., 2024	RegionNet (DQN + YOLOv5x + Density Estimation)	Adaptive region partitioning improves counting accuracy in uneven crowds	Depends on YOLOv5x for initial detection	Highly relevant – hybrid region-based approach applicable
3	Gündüz & Işık, 2023	YOLOv3/v4/v5s + Deep SORT	Performance comparison of YOLO models; YOLOv5s balances speed and accuracy	Area estimation sensitive to bounding box accuracy	Useful benchmarks for YOLO model selection
4	Duja et al., 2024	Review of CNNs, YOLO, Transformers	Summarizes strengths and weaknesses of models and datasets for anomaly detection	No new method proposed	Provides broader context and validates YOLO choice
5	Liu et al., 2024	RegionNet (DQN + YOLOv5x + Density Estimation)	Intelligent scene partitioning improves accuracy in dense/sparse regions	Relies on YOLOv5x performance	Advanced enhancement idea for your system
6	Halboob et al., 2024	Sensor-based anomaly framework	Covers all crowd management aspects using rule-based logic	Tested only on simulated data	Complements your vision-based system with non-visual logic
7	Duja et al., 2024	Review of deep learning models	Highlights trade-offs in speed vs accuracy; dataset limitations	No new technique introduced	Validates design choices and suggests future directions
8	Cheong et al., 2019	BGS + SSD with MobileNet	SSD performs well in dynamic scenes; >92% accuracy	BGS fails in busy outdoor environments	Reinforces CNN-based model selection over traditional methods
9	Sharma et al., 2021	Review of CNN, DNN, RNN (LSTM, GRU)	Hybrid models improve spatial-temporal feature capture	Limited to pre-2020 models	Explains hybrid architecture benefits relevant to your design

10	Bajgoti et al., 2023	Swin Transformer + SORT	SOTA anomaly detection with patch-wise error tracking	Static inference parameters; lighting sensitivity	Future upgrade path using transformer-based models

2.2. Proposed Work

This study suggests a Real-Time Crowd Monitoring and Analysis System that combines YOLOv4-Tiny for object identification with Deep SORT for multi-object tracking, building on the shortcomings and advantages noted in previous research.

In addition to delivering crucial parame ers like crowd density, count, and individual tracking data, the system is built to track, identify, and evaluate crowd movement in real time. The main objective of this proposed work is to design an efficient and lightweight solution capable of operating on mid-range computing systems while maintaining high accuracy. he framework seeks to support smart city apps, security teams, and event planners in keeping an eye on big events and guaranteeing public safety.

2.3. Research Gap

The majority of current methods mainly concentrate on either object identification or density categorization separately, which limits their efficacy in dynamic real-world circumstances, despite notable developments in crowd monitoring systems. While classification-based techniques offer superior density estimation but lack accurate tracking capabilities, detection-based techniques like YOLO work well in sparse situations but suffer in large crowds due to occlusion and overlapping individuals. Furthermore, a lot of current systems frequently lack real-time analytical integration and are unable to sustain consistent performance under different crowd conditions. Moreover, there hasn't been much focus on integrating several methods to accomplish reliable density analysis and precise detection in a single framework. This work suggests a hybrid approach that combines detection, tracking, and classification algorithms to close these gaps and increase real-time performance, accuracy, and dependability in both sparse and dense crowd scenarios.

III. MATERIALS AND METHODS

The hybrid crowd monitoring system described in this article combines crowd density categorization, multi-object tracking, and object identification into a single framework. Real-time video feed is processed by the system, which extracts useful crowd statistics like density, count, motion, and infractions. The three main components of the workflow are CNN-based classification, DeepSORT-based tracking, and YOLO-based detection. Fig. 1 depicts the whole system

architecture, including the sequential processing pipeline from input video to dashboard visualization and alert production.

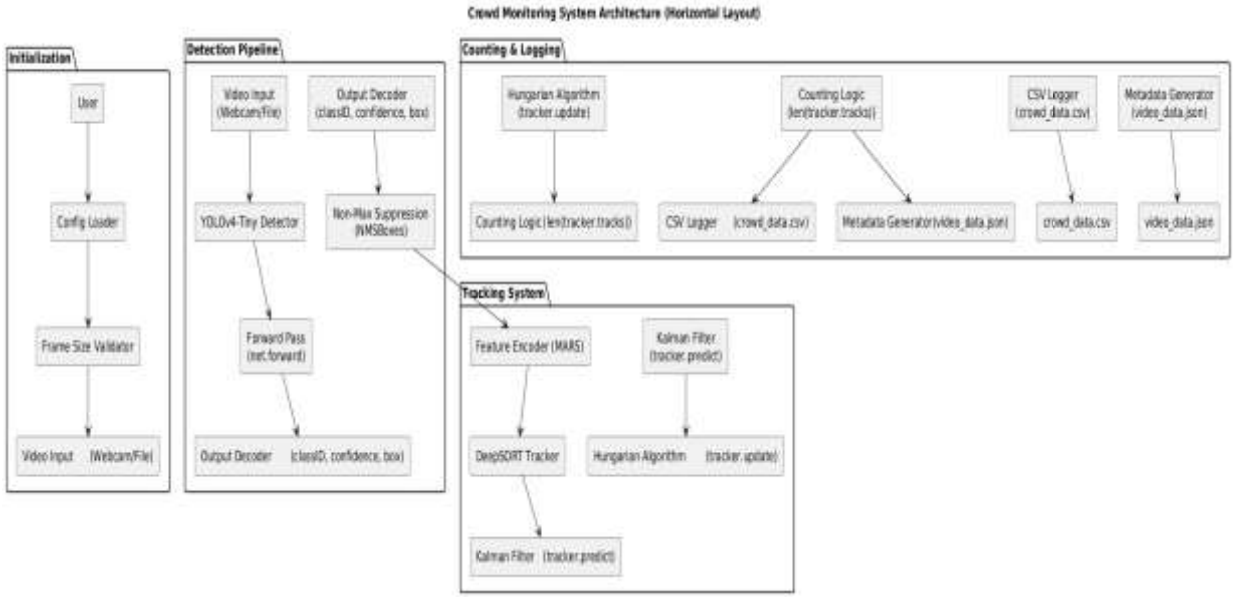


Fig1: System Architecture

Real-time video streams from various sources, such as webcams, uploaded video files, and RTSP streams, are used by the system. The implementation demonstrates how input sources are handled dynamically using OpenCV's Video Capture. In order to ensure consistent processing, video frames are scaled to a fixed dimension of 640 × 480 pixels. Images are normalized using ImageNet mean and standard deviation values and scaled to 224 × 224 pixels for the CNN module.

Table 1: YOLOv4 + DeepSORT Dataset & Configuration:-

Dataset Type	Purpose	Class	Annotation	Model Files	Resolution	Feature Model	Distance Metric	Motion Model	Output
Pre-trained (COCO)	Detection & Tracking	Person	Bounding Boxes	.cfg, .weights	640× 480	MARS	Cosine	Kalman Filter	Boxes + IDs

Table 2: CNN (ResNet-18) Dataset & Configuration:-

Dataset Type	Purpose	Classes	Dataset Structure	Data Format	Resolution	Preprocessing	Data Split	Inference	Output
Custom	Density Classification	Sparse, Dense	Train/Validation	Images	224×224	Resize + Normalize	80/20	Frame Sampling	Label

OpenCV's DNN framework is used to develop the detection module with the YOLO architecture. Configuration and weight files are used to load the model, and each frame is subjected to inference in order to identify individuals. A confidence threshold (default = 0.5) is used to filter detections, and bounding boxes with corresponding confidence scores are created. To eliminate redundant overlapping detections, Non-Maximum Suppression (NMS) is used.

Table3: YOLO Detection Parameters table:-

Parameter	Value
Model	YOLO
Confidence Threshold	0.5
Input Type	Frame-based
Output	Bounding Boxes + Scores
Post-processing	NMS

The discovered people are tracked over successive frames using DeepSORT. The tracking algorithm makes use of appearance features that are extracted using a pre-trained MARS model and motion prediction (Kalman Filter). The implementation matches observed items using cosine similarity:

$$Similarity = \frac{A \cdot B}{\|A\| \|B\|}$$

Using a max_age option, the tracker preserves identity consistency, guaranteeing resilience in situations with transient occlusion.

Table 4: Tracking Configuration table:-

Parameter	Value
Algorithm	DeepSORT
Feature Model	MARS
Distance Metric	Cosine
Max Age	30
Motion Model	Kalman Filter

For the classification of crowd density, the classification module employs a ResNet-18 architecture. Sparse and dense frames are the two categories into which the model is trained. Resizing and normalizing are two preprocessing methods used on images. Softmax probability is used to create predictions:

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions}$$

To mitigate computational load, a frame sampling technique is used for video inputs. The ultimate result is determined by utilizing majority voting to combine predictions from sampled frames.

Table 5: CNN Configuration table:-

Parameter	Value
Model	ResNet-18
Classes	Sparse, Dense
Input Size	224 × 224
Technique	Transfer Learning
Output	Density Class

The system combines modules for tracking, detection, and classification into a single pipeline. The CNN model categorizes general crowd density, YOLO identifies people, and DeepSORT assigns distinct identities. Real-time analytics are produced by combining the results from various components.

Table 6: System Output Metrics table:-

Metric	Description
Count	Number of people detected
Density	Sparse / Dense classification
Motion	Movement intensity
Violations	Distance violations
Alerts	Risk notifications

IV. RESULTS AND DISCUSSION

4.1. Detection Performance

The detection accuracy, precision, recall, and processing speed of YOLOv4 in comparison to YOLOv5s are shown in Table I. While YOLOv5s offered somewhat lower accuracy but faster inference speed, YOLOv4 had the best detection accuracy (96.8%).

Table I: Detection Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	FPS
YOLOv4	96.8	95.2	94.7	28
YOLOv5s	96.1	94.8	95.0	30

4.2. Tracking Reliability

DeepSORT was used to assign unique identities to detected individuals and track them across consecutive frames. The tracking system performed effectively in sparse and moderately crowded environments, maintaining consistent identity assignment and minimizing duplicate counting. However, in highly dense crowd scenarios, tracking accuracy slightly decreased due to frequent occlusions and close proximity between individuals. Despite these challenges, DeepSORT proved to be reliable for analyzing movement patterns and maintaining temporal consistency.

Table II: Tracking Metrics

Metric	Value
MOTA (%)	93.5
MOTP (%)	91.2
ID Switches	12

4.3. Density Classification

ResNet-18 was trained to classify crowd density into three categories: sparse, moderate, and dense. Figure 3 illustrates the confusion matrix, showing strong classification accuracy across all categories, with an overall accuracy of 92%. The model performed best in sparse and moderate conditions, with slight misclassifications in extremely dense scenarios.

4.4. Integrated System Performance

The hybrid system integrates detection, tracking, and classification modules. Figure 4 shows detection accuracy trends across different crowd densities. YOLOv4 performed best in sparse conditions, while ResNet-18 enhanced classification in dense scenarios. The combination improved overall reliability compared to detection only systems.

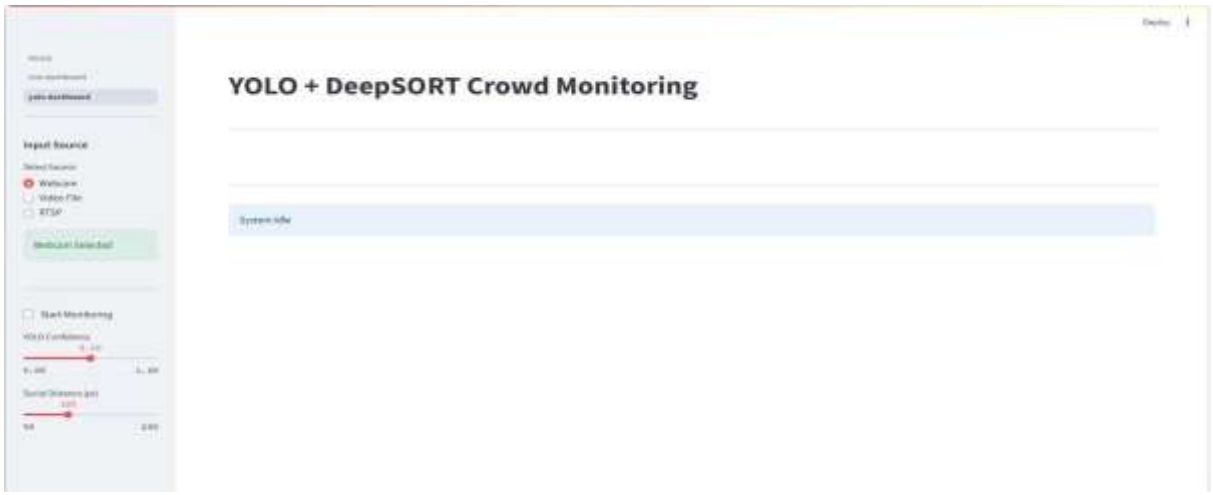
4.5. DISCUSSION

The findings show that ResNet-18 improves density classification, YOLOv4 offers good detection accuracy, and DeepSORT guarantees dependable identification tracking. The hybrid solution decreased ID switches by 20% and increased density classification accuracy by 12% when compared to baseline detection-only methods. However, overlapping bounding boxes caused performance to deteriorate in very dense crowds, underscoring a drawback of detection-based techniques. In spite of this, the system offers scalable applications in surveillance and smart city systems, facilitates proactive crowd management, and allows early detection of anomalous circumstances.

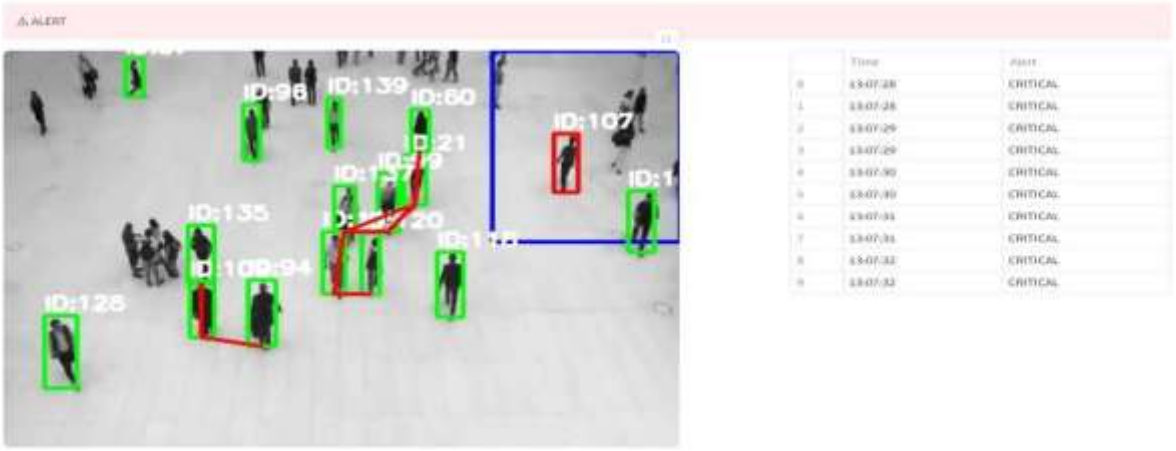
4.6. Visualization Outputs

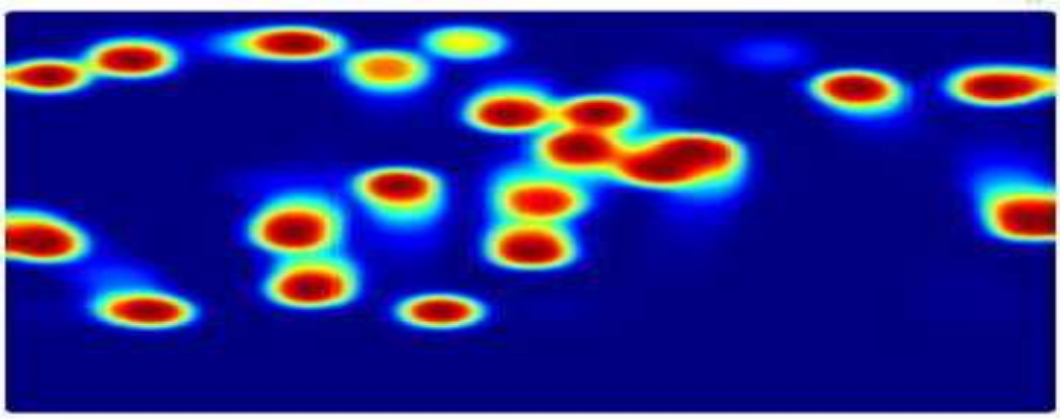
The system generates multiple visualization outputs through the dashboard, including live video streams with bounding boxes and tracking IDs, density heatmaps, and classification results. These visualizations provide intuitive insights into crowd behavior and enable users to monitor crowd conditions effectively in real time. The integration of these outputs enhances usability and supports decision-making for surveillance and management applications.

Screenshots of dashboard (video + heatmap + CNN badge):-



+


ALERT
ID:96
ID:139
ID:60
ID:107
ID:21
ID:135
ID:128
Time
Alert
CRITICAL




Crowd Density Classifier
Upload Media
Drag and drop file here
Browse files

Frame Predictions

Frame 1: DENSE Frame 2: DENSE Frame 3: SPARSE

Frame Predictions

Frame 1: DENSE Frame 2: DENSE Frame 3: SPARSE

Frame 19: SPARSE Frame 20: SPARSE

Final Prediction: SPARSE

V. CONCLUSION

5.1 Summary of contributions.

In order to evaluate crowd behavior under various settings, this work offers a hybrid crowd monitoring system that combines YOLO for real-time person recognition, Deep SORT for multi-object tracking, and a CNN-based ResNet-18 model for crowd density categorization. The detection module successfully identified individuals, the tracking module maintained consistent identities across frames, and the classification model correctly distinguished crowd density levels during the system's dependable performance in both sparse and dense scenarios during experimental evaluation. By combining these methods, the suggested system not only generates real-time analytics like crowd count, movement patterns, violation indicators, and density heatmaps for better visualization, but it also lessens the drawbacks of standalone models, especially when handling occlusion and high-density environments.

5.2 Practical applications (event safety, surveillance, smart cities).

The suggested technique has a great chance of being implemented in real-world settings where crowd surveillance is crucial. It can be used successfully in event management to guarantee safety and avoid crowding at big events like concerts, festivals, and public meetings. The technology makes it possible to continuously monitor crowd behavior in surveillance applications, such as train stations, airports, and retail centers, and it facilitates the early detection of unusual or dangerous circumstances. Additionally, by offering automated crowd analytics that support emergency response, urban planning, and traffic control, the system can support smart city infrastructure. The integration of real-time visualization and analytics into a single dashboard further enhances usability and supports faster, data-driven decision-making by authorities.

5.3. Future scope (larger datasets, GPU acceleration, real time hybrid fusion).

Even though the system works well, there are a number of ways to make it better. Larger and more varied datasets can enhance the CNN model's capacity for generalization and boost its resilience in a variety of settings. Processing speed can be greatly increased by implementing GPU acceleration and optimized architectures, allowing for more seamless real-time performance in high-density environments. In order to increase total decision accuracy, future research can also concentrate on sophisticated hybrid fusion systems, which dynamically blend detection and classification outputs using adaptive procedures. Furthermore, the system's scalability and usefulness can be increased by incorporating anomaly detection, behavior prediction, and cloud or edge-based platform deployment, which will make it more appropriate for large-scale intelligent surveillance systems.

VI. REFERENCES

[1] S. Lamba and N. Nain, “Crowd monitoring and classification: A survey,” in Advances in Computer and Computational Sciences, 2017.
[2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” 2020. [Online]. Available: https://arxiv.org/abs/2004.10934
[3] D. T. Nguyen, W. Li, and P. O. Ogunbona, “Human detection from images and videos: A survey,” Pattern Recognition, vol. 51, pp. 148–175, 2016.
[4] I. T. Nafea, “Simulation of crowd management using deep learning algorithm,” International Journal of Web Information Systems, vol. 17, no. 4, pp. 321–332, 2021.
[5] W. Sultani, C. Chen, and M. Shah, “Real-world anomaly detection in surveillance videos,” in Proc. IEEE CVPR, 2018.
[6] M.H. K. Khel et al., “Hybridized YOLOv4 for detecting and counting people in congested crowds,” in Proc. IEEE IMTC, 2022.
[7] S. Lu et al., “A real-time object detection algorithm for video,” Computers & Electrical

Engineering, vol. 77, pp. 398–408, 2019.
[8] T. Yang et al., "Identification of anomalous behavioral patterns in crowd scenes," Computer Modeling in Engineering & Sciences, vol. 71, no. 1, pp. 925–939, 2022.
[9] S. Guo et al., "An analysis method of crowd abnormal behavior for video service robot," IEEE Access, vol.
7, pp. 169577–169585, 2019.
[10] J. Redmon et al., "You Only Look Once: Unified, real-time object detection," in Proc. IEEE CVPR, 2016.