IAES International Journal of Robotics and Automation (IJRA)
Vol. 15, No. 4, December 2026, pp. 1085~1093
ISSN: 2722-2586, DOI: 10.11591/ijra.v15i4.pp1085-1093

Contextual security system based on human detection, clothing attribute analysis, and temporal context using YOLOv8

Aris Budiyarto, Yuliadi Erdani, Mochammad Naufal
Department of Manufacturing Automation and Mechatronics Engineering, Bandung Manufacturing Polytechnic, Bandung, Indonesia

Article Info

Article history:

Received Jun 22, 2026
Revised Oct 8, 2026
Accepted Nov 10, 2026

Keywords:

Clothing attribute analysis
Context-aware security
Human detection
Temporal context
YOLOv8

ABSTRACT

Conventional surveillance systems primarily detect human presence and provide limited capability to interpret contextual security conditions. This study developed an automated contextual security system integrating YOLOv8n-based human detection, HSV-based clothing analysis, database-based identity verification, and temporal reasoning. A formalized rule-based mechanism mapped identity, clothing, and operational time into VALID, VIOLATION, GUEST, or INTRUDER categories, transmitted results to ThingsBoard via message queuing telemetry transport (MQTT), and selectively stored evidence for VIOLATION and INTRUDER events. The system was evaluated using 170 annotated samples from 27 controlled indoor simulation sessions involving three participants under normal and low-light conditions at distances of 1–3 m. Clothing classification achieved 27.06% accuracy and a 21.33% macro F1-score, identity verification achieved 100.00% across all metrics, temporal classification obtained 77.65% accuracy and a 77.58% macro F1-score, and contextual status achieved 55.29% accuracy and a 45.96% macro F1-score. The results identified HSV-based clothing analysis as the main performance bottleneck. The primary contribution was a traceable perception–context–decision–action mechanism that transforms heterogeneous observations into context-dependent security decisions, real-time monitoring, and automated evidence management rather than relying on a single visual attribute.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Mochammad Naufal
Department of Manufacturing Automation and Mechatronics Engineering, Bandung Manufacturing Polytechnic
Bandung, West Java, Indonesia
Email: madnaufal4@gmail.com

1. INTRODUCTION

Automated security monitoring requires both the detection of people and the interpretation of their presence under an operational policy. YOLO-based surveillance research addresses real-time object detection [1], integrated video monitoring [2], and intrusion detection for physical security [3]. Deep-learning surveillance in banks provides a further application context [4], while a survey of video anomaly detection examines methods for identifying unusual observations [5]. These contributions motivate visual monitoring, but a detected person alone does not specify whether an institutional clothing requirement or operating schedule has been satisfied.

Context provides a basis for converting observations into operational decisions. Research on context-aware Internet of Things (IoT) security examines the role of contextual information in security services [6]. Far-field safety assessment connects visual observations with contextual interpretation in construction monitoring [7], providing a relevant example of perception supporting automation. Adaptive

access control uses context to guide authorization in IoT environments [8], whereas behavior-graph analysis uses contextual semantics for anomaly detection [9]. These studies address different decision tasks. The specific problem considered here is how to map identity, clothing attributes, and operational time into explicit security states and corresponding software actions in a continuous monitoring workflow.

This study develops a rule-based supervisory monitoring system that integrates pre-trained YOLOv8n human detection, HSV-based clothing analysis, database-supported identity verification, and temporal context. Its contribution is an explicit mapping from these inputs to VALID, VIOLATION, GUEST, and INTRUDER states, followed by automatic telemetry publication and selective evidence archiving. The decision and archiving logic execute in the local application; ThingsBoard provides the monitoring interface. This arrangement automates the progression from camera observations to policy-based classification and evidence handling. It is relevant to monitoring access conditions around institutional workspaces and, potentially, manufacturing areas, although industrial deployment and robot control were not evaluated. The contribution is examined using component metrics, incremental ablation, and detection-stage timing on 170 annotated samples from controlled indoor sessions.

2. THEORETICAL BASIS AND PROPOSED FRAMEWORK

The detector references represent different technical contributions. YOLO9000 extends the early YOLO framework toward efficient detection across a broader object vocabulary [10], while YOLOv7 develops trainable bag-of-freebies methods for real-time detectors [11]. The Ultralytics implementation supplies the pre-trained YOLOv8n used in this study [12]. Subsequent work examines joint vehicle and pedestrian detection in complex scenes [13], pedestrian detection in dense scenes [14], and detection under foggy conditions [15]. Applications to autonomous-vehicle pedestrian and traffic-sign perception [16] and comparative pedestrian detection [17] further illustrate distinct uses of YOLOv8. These references motivate the perception stage; they do not establish the accuracy of the present contextual decision rules.

Pedestrian attribute recognition (PAR) concerns semantic properties of detected people. A PAR survey organizes the recognition problem and its methods [18], while the RAP dataset supplies attribute annotations together with viewpoint and occlusion information [19]. A surveillance-oriented review distinguishes frame-based and video-based approaches [20]. Attribute localization and correlation [21] and feature re-sampling for balanced recognition [22] address different modeling challenges. Here, fixed HSV thresholds provide a lightweight clothing-color proxy, rather than a trained PAR model. Reviews of the wider YOLO series [23], developments through YOLOv10 [24], and variants through YOLOv8 [25] provide architectural context, without implying that YOLOv8 is universally superior. The proposed framework places the selected detector and clothing proxy within an explicit policy-to-action workflow, whose performance is assessed separately from detector development.

3. METHOD

3.1. Research methodology

This study adopted a system development and experimental evaluation approach. The proposed framework integrates YOLOv8-based human detection, HSV-based clothing attribute analysis, identity verification, and temporal context through a rule-based decision mechanism. The system was evaluated under multiple surveillance scenarios involving different identities, clothing conditions, and operational times.

3.2. System architecture

Figure 1 shows six modules: image acquisition, human detection, identity verification, clothing analysis, contextual decision-making, and IoT monitoring. The webcam supplies frames to the local application, where YOLOv8n detections define the regions used for subsequent analysis. Identity, clothing, and time are evaluated against Table 1 to determine the security state. The application publishes telemetry via message queuing telemetry transport (MQTT) and selectively archives evidence for VIOLATION and INTRUDER states. ThingsBoard displays the transmitted results; the contextual decision and evidence-selection logic reside in the application.

3.3. Contextual decision process

The contextual decision process illustrated in Figure 2 begins with frame acquisition and preprocessing, followed by YOLOv8n-based human detection. The upper-body ROI of each detected person is analyzed using HSV segmentation to classify the clothing attribute as UNIFORM, FORMAL, or NON-FORMAL. This result is combined with identity verification (REGISTERED or UNKNOWN) and temporal context (WORKING or OUTSIDE WORKING HOURS).

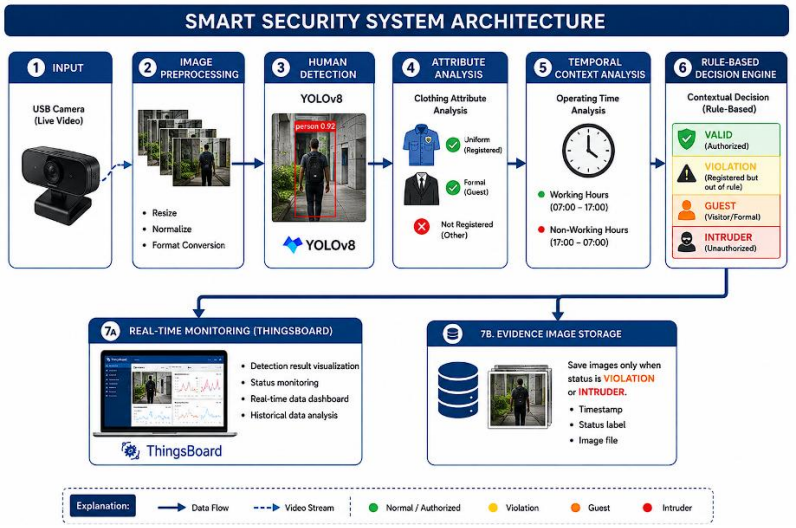


Figure 1. System architecture diagram

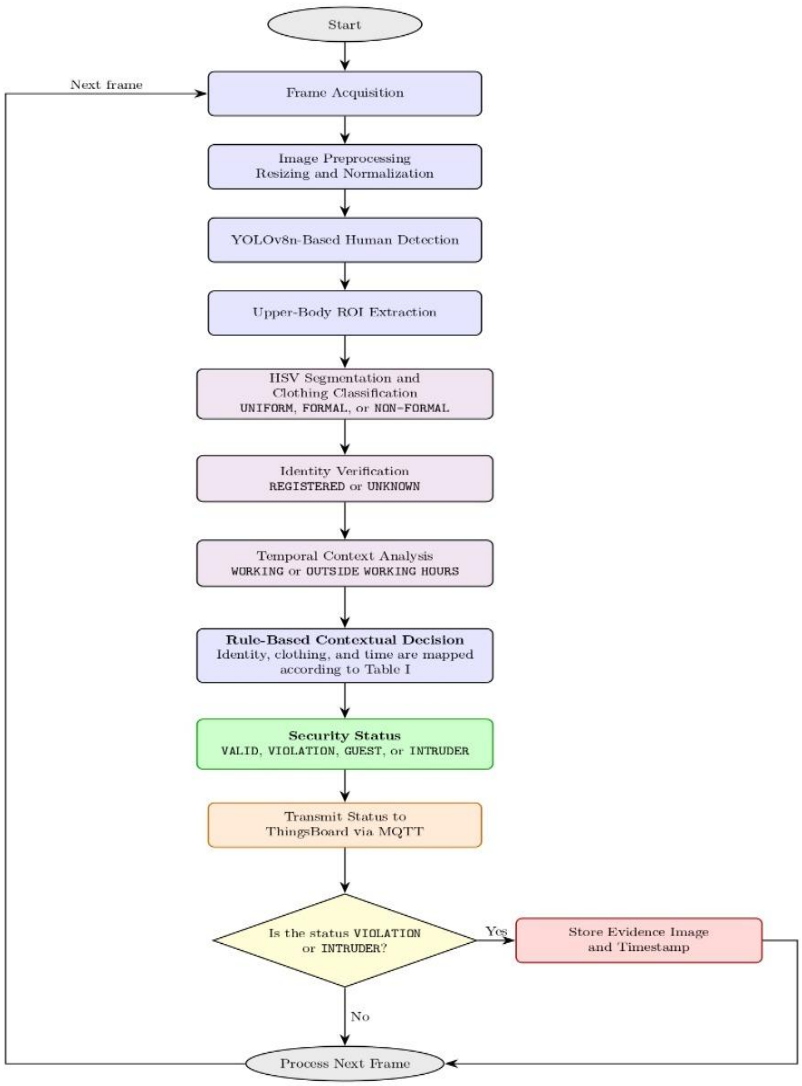


Figure 2. Flowchart diagram

The rule-based contextual decision mechanism is formally defined in Table 1, where each combination of identity status, clothing attribute, and temporal context is deterministically mapped to a corresponding security status. A registered person wearing a uniform is classified as VALID during working hours and as VIOLATION outside working hours. Registered persons wearing formal clothing are classified as VALID at any time, while registered persons wearing non-formal clothing at any time are classified as VIOLATION. Unknown persons are classified as GUEST when wearing formal clothing and as INTRUDER when wearing non-formal clothing, regardless of time. The resulting status is transmitted to ThingsBoard via MQTT. Only VIOLATION and INTRUDER events are stored with images and timestamps, after which the process returns to frame acquisition.

Table 1. Contextual security decision rules

Scenario	Identity	Clothing	Time	Expected Output	System Output
1	Registered	Uniform	Working Hours	VALID	VALID
2	Registered	Uniform	Outside Working Hours	VIOLATION	VIOLATION
3	Registered	Formal	Any	VALID	VALID
4	Registered	Non-formal	Any	VIOLATION	VIOLATION
5	Unknown	Formal	Any	GUEST	GUEST
6	Unknown Unknown	Non-formal	Any	INTRUDER	INTRUDER
7		Uniform	Any	INTRUDER	INTRUDER

3.4. Human detection and clothing attribute analysis

The system uses the pre-trained YOLOv8n model from Ultralytics [12] to detect people. The nano variant was chosen as a lightweight implementation for the available laptop; its measured inference time is reported in Section 4.3. Each bounding box defines the region for subsequent contextual feature extraction. No detector retraining or architectural modification was performed.

For each detected person, the system extracts an upper-body region of interest (ROI) to analyze clothing while reducing background interference. The ROI is converted into HSV space, which separates hue, saturation, and value for explicit color thresholding. This representation simplifies the clothing-color rules, but fixed thresholds remain sensitive to illumination, shadows, and overlapping colors.

The presence of predefined clothing attributes is determined by applying HSV thresholding to the extracted ROI. The proportion of pixels satisfying the predefined HSV range is calculated as (1).

$$R = \frac{Nc}{Nt} \times 100\% \quad (1)$$

where N_c represents the number of pixels satisfying the predefined HSV threshold and N_t denotes the total number of pixels within the upper-body ROI.

A clothing attribute is considered detected when the percentage of pixels satisfying the predefined HSV threshold exceeds 20% of the extracted ROI. Based on this analysis, the detected individual is categorized as wearing an institutional uniform, formal clothing, or non-formal clothing. The resulting clothing attribute is subsequently forwarded to the contextual decision module together with identity status and temporal information for security classification.

3.5. Dataset and evaluation protocol

The evaluation dataset was collected through 27 controlled indoor simulation sessions involving three individuals: one registered user and two unknown users. The simulations covered normal and low-light conditions at observation distances of approximately 1–3 m. Clothing conditions included UNIFORM, FORMAL, and NON-FORMAL, while temporal conditions included WORKING and OUTSIDE WORKING HOURS. These combinations were used to represent the REGISTERED and UNKNOWN identity states and the VALID, VIOLATION, GUEST, and INTRUDER security categories defined in Table 1.

A total of 392 surveillance images were initially collected. Because continuous monitoring produced consecutive and visually similar frames, representative-frame selection was applied to reduce temporal redundancy. This process produced 190 manually annotated frames containing ground-truth labels for lighting, clothing, identity, time, and expected contextual status. From these annotations, 170 samples with complete ground-truth labels, component predictions, and clothing-ratio values were retained for quantitative evaluation. The final set contained 10 UNIFORM, 38 FORMAL, and 122 NON-FORMAL samples, 137 REGISTERED and 33 UNKNOWN samples, and 87 WORKING and 83 OUTSIDE WORKING HOURS samples.

The dataset was used exclusively for system evaluation rather than model training. YOLOv8n remained pre-trained and the HSV thresholds and contextual rules were fixed before evaluation. For each sample, the predicted clothing, identity, temporal, and contextual-status labels were compared with their corresponding ground truth. Performance was measured using accuracy, macro precision, macro recall, and macro F1-score, with macro averaging used to give equal importance to each class. Object-detection mAP was not calculated because the evaluation dataset did not include bounding-box ground-truth annotations. Therefore, the quantitative evaluation focused on contextual classification performance and recorded YOLOv8n inference time.

4. RESULT AND DISCUSSION

4.1. System implementation

The proposed system was implemented as an integrated real-time surveillance application using Python, Ultralytics YOLOv8 for human detection, OpenCV for image processing, and Paho MQTT for communication with ThingsBoard. The experiments were conducted on an ASUS Vivobook 14X M1403QA equipped with an AMD Ryzen 5 5600H processor, 16 GB RAM, integrated AMD Radeon Graphics, and Windows 11 Home.

During operation, webcam frames are processed by YOLOv8n to detect persons, followed by upper-body ROI extraction and HSV-based clothing analysis. The clothing result is combined with identity and temporal information to generate a VALID, VIOLATION, GUEST, or INTRUDER status. Figure 3 shows the real-time detection interface and the resulting contextual security status, while Figure 4 presents the telemetry received by the ThingsBoard dashboard through MQTT. Only VIOLATION and INTRUDER images are archived with their timestamps, whereas VALID and GUEST events are displayed without image storage. The system was evaluated indoors under normal and low-light conditions at observation distances of approximately 1–3 m.



Figure 3. Webcam monitoring and YOLOv8 detection

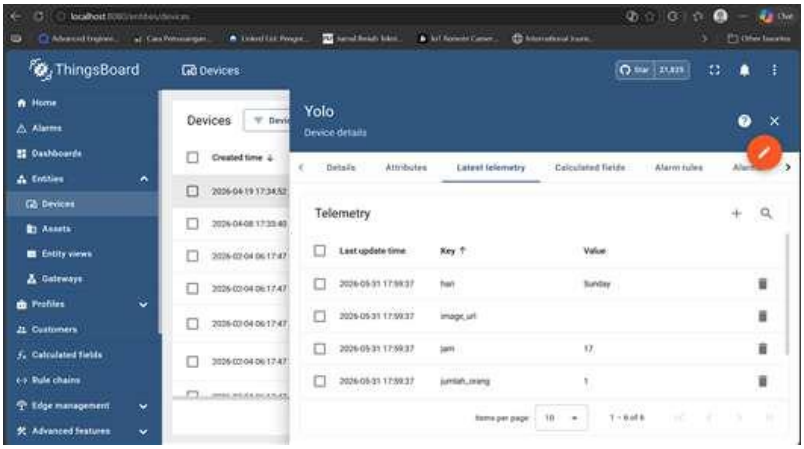


Figure 4. ThingsBoard dashboard

4.2. Detection and attribute analysis results

Figures 5 and 6 present representative detection results under normal and low-light conditions, respectively. Under normal lighting in Figure 5, the system detected an unknown individual and identified formal clothing, resulting in a GUEST classification according to the contextual decision rules. Under low-light conditions in Figure 6, person and identity detection remained operational, although the reduced illumination affected the visual appearance of the clothing ROI. These results demonstrate that YOLOv8-based human detection can operate under both lighting conditions, while HSV-based clothing analysis remains more sensitive to illumination changes.

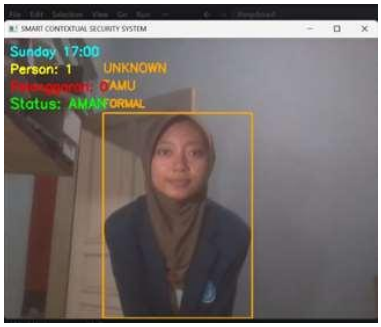


Figure 5. Human detection results under normal lighting



Figure 6. Human detection results under low lighting

4.3. Quantitative and contextual performance evaluation

Table 2 summarizes the performance of the clothing, identity, temporal, and contextual-status components using 170 evaluation samples. Accuracy, macro precision, macro recall, and macro F1-score were used to evaluate both overall correctness and performance across the respective classes.

Table 2. Overall classification performance on 170 evaluation samples

Component	Samples	Accuracy	Precision (Macro)	Recall (Macro)	F1-score (Macro)
Clothing	170	27.06%	27.88%	30.28%	21.33%
Identity	170	100.00%	100.00%	100.00%	100.00%
Time	170	77.65%	82.28%	78.63%	77.58%
Contextual Status	170	55.29%	48.50%	45.82%	45.96%

Clothing classification achieved 27.06% accuracy and a macro F1-score of 21.33%. This relatively low performance was caused by the sensitivity of fixed HSV thresholds to illumination, shadows, pose, camera distance, and upper-body ROI variations. The UNIFORM, FORMAL, and NON-FORMAL categories also contain overlapping colors, while the evaluated samples were distributed as 10 UNIFORM, 38 FORMAL, and 122 NON-FORMAL. Consequently, clothing was the most challenging component.

Identity verification achieved 100.00% across all metrics. The three-person simulation represented both required identity states through one registered and two unknown individuals. All 137 REGISTERED and 33 UNKNOWN samples were correctly classified, confirming consistent database-based identity verification throughout the experimental protocol. Temporal classification achieved 77.65% accuracy and a macro F1-score of 77.58%, indicating that most observations were correctly mapped to WORKING or OUTSIDE WORKING HOURS. The remaining differences were associated with the correspondence between the recorded timestamps, predefined operational schedule, and annotated temporal conditions. Contextual Status achieved 55.29% accuracy and a macro F1-score of 45.96% for the VALID, VIOLATION, GUEST, and INTRUDER classes. Because the final status is determined from the three contextual inputs, errors from clothing and temporal classification were propagated to the final decision, while the accurate identity component provided a stable contextual input.

Computational performance was evaluated using the recorded YOLOv8n inference times of 170 samples. The first frame required 3575.52 ms because it included model initialization and was excluded from the steady-state calculation. For the remaining samples, the average inference time was 86.54 ± 14.42 ms per frame, corresponding to an approximate detection-stage throughput of 11.56 FPS, while the 95th-percentile inference time was 106.81 ms. These results indicate that the YOLOv8n detection stage supports near-real-time processing under the evaluated hardware conditions.

The automation contribution is the executable connection between perception, policy evaluation, and software actions. Table 1 makes the mapping inspectable: formal clothing yields VALID for a registered person and GUEST for an unknown person, while time changes the decision for a registered person wearing a uniform. The application determines these states before MQTT transmission and uses them to select evidence for archiving. ThingsBoard therefore serves as a supervisory display within an automated monitoring workflow. This is related to context-aware safety assessment [7] and policy-based access decisions [8], with the present implementation focusing on identity, clothing, and time. Its potential use around manufacturing workspaces requires field validation; the experiments do not demonstrate robot actuation, physical access enforcement, or safety-rated machine control.

Table 3 reports a macro F1-score of 12.03% for clothing alone, 42.76% with identity added, and 45.96% with temporal context included. The respective gains of 30.73 and 3.20 percentage points support the usefulness of these contextual inputs on the evaluated samples. However, 55.29% final accuracy and 45.96% macro F1-score show that an executable automation workflow does not by itself ensure reliable security decisions. Clothing errors can trigger incorrect archiving or omit relevant evidence. The results support a supervised prototype, with broader validation required before operational reliance. The reported 11.56 FPS describes only the detection stage; end-to-end latency and reductions in operator workload were not measured.

Table 3. Ablation results of the contextual decision mechanism

Configuration	Contextual inputs	Accuracy	Macro F1-score
Visual-only baseline	Clothing	12.94%	12.03%
Identity-aware mechanism	Clothing + Identity	41.76%	42.76%
Full proposed mechanism	Clothing + Identity + Time	55.29%	45.96%

5. CONCLUSION

This study implemented an automated security-monitoring workflow that combines YOLOv8n detection, HSV clothing analysis, identity verification, and temporal rules. Its contribution is the explicit conversion of contextual observations into four security states and associated telemetry and evidence-archiving actions. On 170 samples, final classification achieved 55.29% accuracy and 45.96% macro F1-score. Incremental ablation increased macro F1-score from 12.03% with clothing alone to 42.76% with identity and 45.96% with time. Mean steady-state detector inference was 86.54 ms per frame, equivalent to approximately 11.56 FPS for that stage.

The prototype demonstrates software-level supervisory automation under controlled indoor conditions. HSV sensitivity, imbalanced classes, and the three-participant dataset limit reliability and generalization. Future work should improve clothing recognition, evaluate larger and more diverse datasets, and measure end-to-end latency before deployment in operational or manufacturing environments.

FUNDING INFORMATION

This research was supported by the university through internal academic research funding. The funding institution provided support for system development, experimental setup, and computational resources used in this study. No external commercial or industrial funding was involved in this work.

AUTHOR CONTRIBUTIONS STATEMENT

The authors contributed to all aspects of this research, including conceptualization, system design, implementation, experimentation, data analysis, and manuscript preparation. The authors also reviewed and approved the final version of the manuscript.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Aris Budiyarto	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Yuliadi Erdani	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mochammad Naufal	✓		✓		✓	✓	✓	✓	✓	✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest regarding the publication of this paper. The research was conducted independently without any commercial, financial, or personal relationships that could influence the results or interpretation of the study.

DATA AVAILABILITY

The data used in this study are not publicly available due to privacy and security considerations related to surveillance data. However, the experimental setup, system configuration, and implementation details are available from the corresponding author upon reasonable request.

REFERENCES

[1] S. Kukreti, R. Praveen, S. Bansal, H. Raju, N. Singh, and R. Begum, "Object detection in real-time surveillance using deep learning-based YOLO framework," in *2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*, Indore, India: IEEE, Feb. 2025, pp. 601–606, doi: 10.1109/ICCCIT62592.2025.10928084.

[2] N. Azatbekuly, A. Mukhanbet, and S. D. Bekele, "Development of an intelligent video surveillance system based on YOLO algorithm," in *2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)*, Astana, Kazakhstan: IEEE, May 2024, pp. 498–503, doi: 10.1109/SIST61555.2024.10629617.

[3] N. Chatterjee, A. V. Singh, and R. Agarwal, "You Only Look Once (YOLOv8) based intrusion detection system for physical security and surveillance," in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, IEEE, Mar. 2024, pp. 1–5, doi: 10.1109/ICRITO61523.2024.10522139.

[4] M. Zahrawi and K. Shaalan, "Improving video surveillance systems in banks using deep learning techniques," *Scientific Reports*, vol. 13, no. 1, p. 7911, May 2023, doi: 10.1038/s41598-023-35190-9.

[5] N. Choudhry, J. Abawajy, S. Huda, and I. Rao, "A comprehensive survey of machine learning methods for surveillance videos anomaly detection," *IEEE Access*, vol. 11, pp. 114680–114713, 2023, doi: 10.1109/ACCESS.2023.3321800.

[6] A. I. Alotaibi and A. Oracevic, "Context-aware security in the internet of things: what we know and where we are going," in *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, Doha, Qatar: IEEE, Oct. 2023, pp. 1–8, doi: 10.1109/ISNCC58260.2023.10323735.

[7] W.-C. Chern, J. Hyeon, T. V. Nguyen, V. K. Asari, and H. Kim, "Context-aware safety assessment system for far-field monitoring," *Automation in Construction*, vol. 149, p. 104779, May 2023, doi: 10.1016/j.autcon.2023.104779.

[8] R. Kalaria, A. S. M. Kayes, W. Rahayu, E. Pardede, and A. Salehi Shahraki, "Adaptive context-aware access control for IoT environments leveraging fog computing," *International Journal of Information Security*, vol. 23, no. 4, pp. 3089–3107, Aug. 2024, doi: 10.1007/s10207-024-00866-4.

[9] Q. Lin *et al.*, "Anomaly detection in smart IoT systems based on contextual semantics of behavior graphs," *IEEE Internet of Things Journal*, vol. 12, no. 10, pp. 13683–13696, May 2025, doi: 10.1109/JIOT.2024.3523998.

[10] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 7464–7475, doi: 10.1109/CVPR52729.2023.00721.

[12] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," GitHub Repository. Accessed: Jun 22, 2026. [Online]. Available: https://github.com/ultralytics/ultralytics

[13] S. Lei, H. Yi, and J. S. Sarmiento, "Synchronous end-to-end vehicle pedestrian detection algorithm based on improved YOLOv8 in complex scenarios," *Sensors*, vol. 24, no. 18, p. 6116, Sep. 2024, doi: 10.3390/s24186116.

[14] Y. Fang and H. Pang, "An improved pedestrian detection model based on YOLOv8 for dense scenes," *Symmetry*, vol. 16, no. 6, p. 716, Jun. 2024, doi: 10.3390/sym16060716.

[15] A. Esmaeil Abbasi, A. M. Mangini, and M. P. Fanti, "Object and pedestrian detection on road in foggy weather conditions by hyperparameterized YOLOv8 model," *Electronics*, vol. 13, no. 18, p. 3661, Sep. 2024, doi: 10.3390/electronics13183661.

[16] I. Logeswaran, H. Eissa, and R. Almadhoun, "Autonomous vehicle pedestrian detection & traffic sign recognition using YOLOv8," in *2024 29th International Conference on Automation and Computing (ICAC)*, IEEE, Aug. 2024, pp. 1–6, doi: 10.1109/ICAC61394.2024.10718752.

[17] S. G, S. E, and D. M, "YOLOv8 for pedestrian detection: a comparative study for pedestrian detection," in *2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, IEEE, Jul. 2024, pp. 1–7, doi: 10.1109/ICEEICT61591.2024.10718372.

[18] X. Wang *et al.*, "Pedestrian attribute recognition: A survey," *Pattern Recognition*, vol. 121, p. 108220, Jan. 2022, doi: 10.1016/j.patcog.2021.108220.

[19] D. Li, Z. Zhang, X. Chen, H. Ling, and K. Huang, "A richly annotated dataset for pedestrian attribute recognition," *arXiv preprint arXiv:1603.07054*, 2016, doi: 10.48550/arXiv.1603.07054.

[20] Y. Cao, W. Lu, J. Yu, Y. Zhou, and H. Hu, "Pedestrian attribute recognition in surveillance scenario: a survey and future perspectives on frame vs. video based methods," *Journal of Computer-Aided Design & Computer Graphics*, vol. 36, no. 3, pp. 336–356, Mar. 2024, doi: 10.3724/SP.J.1089.2024.2023-00362.

[21] D. Weng, Z. Tan, L. Fang, and G. Guo, "Exploring attribute localization and correlation for pedestrian attribute recognition," *Neurocomputing*, vol. 531, pp. 140–150, Apr. 2023, doi: 10.1016/j.neucom.2023.02.019.

[22] Y. Zhou, B. Li, H.-M. Hu, X. Zhang, D. Zhang, and H. Wang, "Heterogeneous feature re-sampling for balanced pedestrian attribute recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 2706–2722, Apr. 2025, doi: 10.1109/TPAMI.2025.3526930.

[23] R. Sapkota *et al.*, "YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series," *Artificial Intelligence Review*, vol. 58, no. 9, p. 274, Jun. 2025, doi: 10.1007/s10462-025-11253-3.

[24] C.-Y. Wang and H.-Y. M. Liao, "YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems," *APSIPA Transactions on Signal and Information Processing*, vol. 13, no. 1, pp. 1–38, Dec. 2024, doi: 10.1561/116.20240058.

[25] M. Hussain, "YOLOv1 to v8: Unveiling each variant–a comprehensive review of YOLO," *IEEE Access*, vol. 12, pp. 42816–42833, 2024, doi: 10.1109/ACCESS.2024.3378568.

BIOGRAPHIES OF AUTHORS



Aris Budiyarto is a lecturer and researcher at Politeknik Manufaktur Bandung (POLMAN Bandung), Indonesia. His research interests include digital systems, microprocessors, Internet of Things (IoT), embedded systems, cyber-physical systems, and industrial automation. He has been actively involved in applied research and engineering education, particularly in smart manufacturing and Industry 4.0 technologies. He currently serves as the Vice Director for Academic Affairs and Information Systems at POLMAN Bandung. He can be contacted at aris_b@polman-bandung.ac.id.



Yuliadi Erdani is a professor and faculty member at Politeknik Manufaktur Bandung (POLMAN Bandung), Indonesia. His research interests include intelligent manufacturing systems, cyber-physical systems, smart machine tools, digital twins, machine learning, computer vision, and industrial automation. He has actively contributed to research and innovation in Industry 4.0 technologies through numerous scientific publications and industrial projects. He currently serves as the Chair of the Academic Senate of POLMAN Bandung for the 2025–2029 term. He can be contacted at yul_erdani@yahoo.com.



Mochammad Naufal received the bachelor's degree in Computer Science from STMIK Mardira Indonesia, Indonesia. He is currently pursuing a master's degree in Cyber-Physical Systems at Politeknik Manufaktur Bandung (POLMAN Bandung). His research interests include computer vision, artificial intelligence, Internet of Things (IoT), cyber-physical systems, and intelligent security systems. He is the third author of this work and actively conducts research on contextual security systems based on human detection and attribute analysis using deep learning techniques. He can be contacted at madnaufal4@gmail.com.