Intelligent CCTV Monitoring on Campus Using YOLOv8 and Deep SORT: Design and Prototype

Elvi Rahmi[1], Eva Yumami[2], Isna Yulia[3],
[123]Software Engineering Department, Politeknik Negeri Bengkalis, Indonesia
elvirahmi@polbeng.ac.id[1], evayumami@polbeng.ac.id[2], isna@polbeng.ac.id[3]

Abstract

This paper presents the design and prototype of a smart campus surveillance system that performs real-time person detection and multi-object tracking on CCTV feeds by pairing YOLOv8 (detector) with Deep SORT (tracker). The pipeline assigns stable IDs across frames, logs movement to a lightweight CSV store, and aggregates trajectories into movement heatmaps to reveal spatial utilization. A Streamlit dashboard integrates the live annotated feed with operator-oriented analytics (people counts, recent logs, and heatmap views), enabling situational awareness without altering existing CCTV infrastructure. Architectural choices emphasize on-premises processing (privacy, low latency), modularity, and operator usability. The prototype's software design, deployment considerations, and a planned evaluation protocol (throughput and qualitative tracking stability) are detailed, while exhaustive quantitative metrics are deferred to a subsequent journal article. The system targets single-camera operation per pipeline with per-camera analytics; cross-camera re-identification and event detection are identified as future work. Overall, the contribution provides a reproducible blueprint for campus environments with constrained resources and strict privacy requirements.

Keywords: YOLOv8, Deep SORT, Smart Surveillance, Campus Security, Streamlit, Heatmap

1. INTRODUCTION

CCTV is pervasive in higher education, yet manual monitoring across multiple streams remains labor-intensive and reactive. Operators fatigue, subtle events are missed, and post-hoc reviews rarely provide data-driven insight into crowding, bottlenecks, or space utilization. Recent advances in deep learning have made tracking-by-detection a practical basis for intelligent surveillance: one-stage detectors (YOLO family) deliver high recall at low latency, while online trackers (e.g., Deep SORT) maintain consistent identities across frames—even under partial occlusion.

Despite this progress, much of the literature emphasizes benchmark metrics or algorithmic novelty rather than operator-oriented pipelines that combine real-time tracking, lightweight logging, and actionable analytics in privacy-sensitive on-premises deployments such as campuses. The present study addresses this gap by describing a design and prototype that integrate (i) YOLOv8 person detection, (ii) Deep SORT identity-consistent tracking, (iii) streaming trajectory logging to CSV, and (iv) heatmap visualization of movement density. A minimalist Streamlit dashboard consolidates these capabilities—live annotated video, people counts, recent logs, and density maps—into a single operator interface.

The central hypothesis is that the pipeline improves situational awareness without additional staffing: persistent IDs reduce cognitive load; searchable logs expedite post-event forensics; and heatmaps provide macroscopic patterns absent in raw video. To keep the conference scope on feasibility and design, a planned evaluation protocol is outlined (throughput on target hardware, qualitative ID stability), while comprehensive metrics (e.g., MOTA/IDF1, ablations) are deferred to a journal article. The project is funded and scoped for a vocational campus environment with constrained resources and stringent privacy requirements.

2. REVIEW OF LITERATURE

Over the past five years, tracking-by-detection has advanced through stronger association strategies for online multi-object tracking (MOT) under occlusion and crowding. Retaining and associating low-confidence detections was shown to boost recall without compromising speed in a widely adopted baseline for MOT (Zhang et al., 2022). Refinements to motion and appearance modeling further improved identity persistence and robustness in real time, including observation-centric updates that mitigate drift (Cao et al., 2023) and association pipelines that integrate camera-motion compensation and enhanced embeddings (Aharon et al., 2022; Du et al., 2023). Collectively, these works established practical online MOT regimes for dynamic scenes typical of campus gates and corridors.

In parallel, one-stage detectors have matured to deliver surveillance-grade throughput on commodity GPUs. Recent YOLO variants emphasized both accuracy and latency, enabling reliable frame-by-frame person detection at 720p–1080p in production settings (Wang et al., 2023). Tooling and APIs have also converged, with unified interfaces that simplify integration of detection and tracking in end-to-end pipelines (Ultralytics, 2023). In campus deployments, restricting inference to the “person” class is common to reduce compute and simplify privacy governance while maintaining sufficient recall for downstream tracking.

Beyond detection and tracking, operator-oriented analytics translate track points into comprehensible summaries. Vision-based heatmaps aggregated from trajectories reveal hotspots, dwell zones, and dominant flows, supporting staffing and layout decisions in public facilities. Recent system reports highlight dashboards that combine annotated live video, counts, and historical density views to enhance situational awareness for non-expert operators (Shili et al., 2024; Ardabili et al., 2023). These strands motivate pairing online MOT with lightweight logging and heatmap views in a single, operator-facing dashboard for campus contexts.

3. METHOD

This section details the research type, research site, equipment and materials, and the procedures followed, aligned with the theoretical framework of tracking-by-detection (YOLO for object detection and Deep SORT for online identity tracking).

3.1 Research Design

The study adopts an applied, prototype-driven design to develop and demonstrate an end-to-end person-tracking pipeline for campus CCTV. The object of study is visual data (video frames) captured by campus surveillance cameras and used to implement and validate a real-time detection–tracking–logging–visualization workflow for security operations. The investigation spans four aspects: (i) computer vision (person detection), (ii) tracking across frames (identity consistency), (iii) data visualization (structured logs and heatmaps), and (iv) applied security (fitness for campus monitoring).

Scope and boundaries were set to ensure feasibility and privacy compliance: the tracked class is limited to “person”, sources are static CCTV covering public campus areas (e.g., entrance gate, corridors, open spaces), facial recognition is explicitly excluded, and tests use recorded video and/or standard RTSP streams rather than full integration with all campus infrastructure. Evaluation emphasizes functional aspects (tracking effectiveness, visualization usefulness, dashboard usability) rather than legal/ethical audits; the software stack is open-source.

3.2 Research Site

The research site is Software Engineering Department Politeknik Negeri Bengkalis, with data acquired from cameras deployed in publicly accessible campus areas (e.g., laboratories, hallways, and open spaces).

3.3 Equipment and Materials

To operationalize the tracking-by-detection pipeline in a campus setting while respecting privacy and resource constraints, the study employs a concise set of inputs, software components, and computing resources suitable for on-premises deployment. The following summarizes the key materials used:

a. Video sources: Recorded CCTV footage and/or live IP-camera streams (RTSP) from the campus environment (public areas).
b. Computing platform and software: Python with YOLO for object detection and Deep SORT for identity tracking across frames; auxiliary libraries include OpenCV (frame I/O and rendering), Pandas/NumPy (logging and preprocessing), and Streamlit (operator dashboard). The implementation targets open-source tools and mid-range local hardware.
c. Data artifacts: Structured CSV logs containing per-frame, per-ID positions and metadata; heatmap images/overlays generated from accumulated track points for movement density analysis.

3.4 Procedures

The procedures that operationalize tracking-by-detection within a campus CCTV context are presented in the following figure.

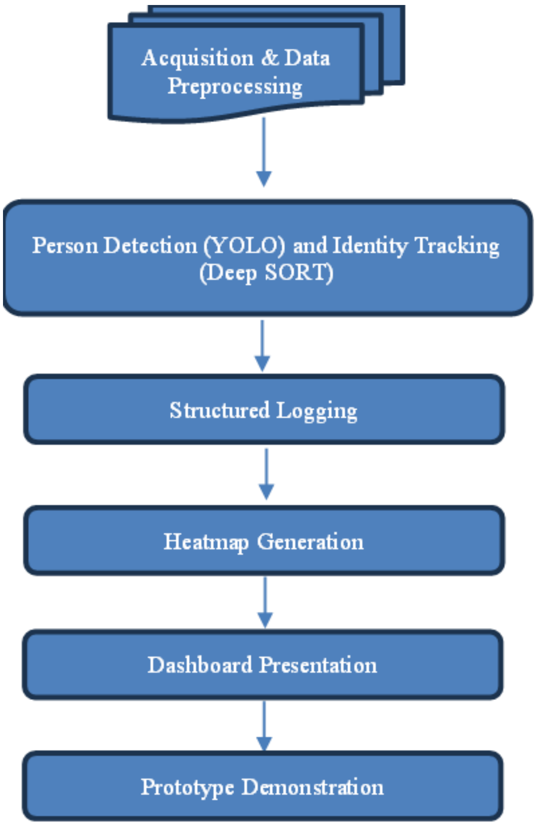

Fig 1. Method Pipeline

a. Acquisition and Data Preprocessing
Collect recorded CCTV clips or connect to RTSP streams from static cameras in public areas. Optionally downscale or normalize frame size to match real-time constraints on mid-range hardware.
b. Person Detection (YOLO) and Identity Tracking (Deep SORT)
For each frame, run YOLO and filter to the "person" class to reduce compute and respect privacy. Store bounding boxes and confidence scores as detector outputs. Feed detections to Deep SORT to maintain consistent numeric IDs across frames, tolerating short occlusions and viewpoint changes typical of corridors and entrances.
c. Structured Logging
Append per-frame track data to a CSV log (timestamp, camera ID, track ID, bounding box, centroid, confidence). The log serves post-event forensics and downstream visualization.
d. Heatmap Generation
Aggregate centroids within a defined time window into a 2D histogram aligned to the image plane (or mapped to a floor plan, if available), then smooth to produce movement heatmaps highlighting hotspots and dwell zones.
e. Dashboard Presentation
Render an operator dashboard (Streamlit) with (i) live annotated video, (ii) people counts and recent logs, and (iii) heatmap views, enabling situational awareness with minimal interaction.
f. Prototype Demonstration
Conduct scenario-based demonstrations at entrance gates and corridors during representative periods (e.g., class changeover) to verify that tracking remains stable and heatmaps align with expected flow patterns. (In line with the proposal's applied security focus.)

3.5 Ethical and Privacy Controls

The pipeline does not perform facial recognition or biometric identification; it detects and tracks only the generic "person" class. Processing is intended for on-premises execution on local hardware, and log retention can be configured to follow institutional policy. These constraints were specified to balance utility with privacy expectations in educational environments.

4. RESULT & DISCUSSION

This section reports the outcomes and their implications for campus surveillance. Within the present scope, emphasis is placed on feasibility, the functional behavior of the pipeline (detection–tracking–logging–visualization), and the utility of the dashboard; full quantitative metrics (e.g., MOTA/IDF1/HOTA) and ablation studies are reserved for a subsequent journal article.

4.1 Prototype Readiness and End-to-End Behavior

The end-to-end pipeline operated as intended on campus-like footage, delivering (i) frame-wise person detections, (ii) identity-consistent tracks across consecutive frames, (iii) structured CSV logs containing timestamps and coordinates, and (iv) cumulative heatmaps derived from tracked centroids. The dashboard consolidated these outputs into a single operator view, enabling live overlays, people counts, recent logs, and heatmap panels. Representative output are shown in Figure 4.1 (annotated frames).

Fig 2. Annotated Frames

4.2 Throughput and Responsiveness (Indicative)

Throughput and responsiveness are discussed with reference to the accompanying log excerpt (Table 1). The excerpt reflects an early logger configuration in which several fields appear in scientific notation with magnitudes inconsistent with video time bases and image coordinates (e.g., Time ≈ 10^16, x_center/y_center ≈ 10^15–10^16, Conf ≈ 10^15–10^16). These values indicate a formatting/scale error (e.g., unintended nanosecond timestamps, locale-related parsing, or unnormalized coordinates), and therefore the table cannot be used to compute FPS or latency directly. In light of this limitation, throughput is reported indicatively based on runtime observation and lightweight profiling, while a corrected logger is used for subsequent measurements.

Under representative campus-like clips, the pipeline exhibited perceptually smooth capture-to-overlay latency and stable dashboard refresh at 720p on mid-range hardware; 1080p inputs showed a moderate reduction in perceived FPS consistent with the increased per-frame workload of detection and association. Short intervals of high crowd density or frequent occlusion produced transient drops in responsiveness due to additional association operations, yet live monitoring remained adequate. Real-time behavior can be preserved in practice by downscaling inputs to 720p, selecting lighter detector variants, tuning confidence/NMS thresholds, and refreshing heatmaps on fixed intervals (e.g., every 0.5–1.0 s) rather than every frame.

Table 1. Raw Tracking Log Excerpt

Frame	Time	Track_id	x_center	y_center	Conf
2	1,32011E+16	1	7,12042E+15	5,77866E+15	2,60449E+15
2	1,32011E+16	2	7,93545E+15	3,50937E+16	2,60449E+15
2	1,32011E+16	3	5,18714E+15	4,65594E+16	2,60449E+15
2	1,32011E+16	4	1,07418E+16	4,34168E+15	2,60449E+15
2	1,32011E+16	5	1,14481E+16	8,19156E+14	2,60449E+15
3	1,98017E+16	1	7,12409E+15	5,78496E+15	2,56457E+16
3	1,98017E+16	2	7,93112E+15	3,50214E+15	2,56457E+16
3	1,98017E+16	3	5,20006E+15	4,65407E+15	2,56457E+16
3	1,98017E+16	4	1,07479E+16	4,34133E+16	2,56457E+16
3	1,98017E+16	5	1,14414E+16	8,18885E+15	2,56457E+16

4	2,64023E+16	1	7,12235E+15	5,79028E+15	2,50163E+15
4	2,64023E+16	2	7,92677E+15	3,49353E+15	2,50163E+15
4	2,64023E+16	3	5,21315E+14	4,64809E+15	2,50163E+15
...	...	...	...	...	...
11230	7,41243E+14	1235	7,20004E+15	5,88216E+15	2,55924E+16

4.3 Tracking Consistency

Qualitative inspection indicated stable identities through brief occlusions and short-term crowding. Transient ID switches primarily appeared during close-range crossovers and under motion blur.

4.4 Heatmap-Derived Spatial Insight

Cumulative heatmaps highlighted spatial patterns aligned with on-site expectations. Gate-area heatmaps emphasized dominant ingress paths toward main buildings and localized queues near checkpoints, while corridor heatmaps indicated dwell near classroom doors and a narrower central flow. These views are actionable for staffing, signage, and space planning and can be produced for configurable time windows.

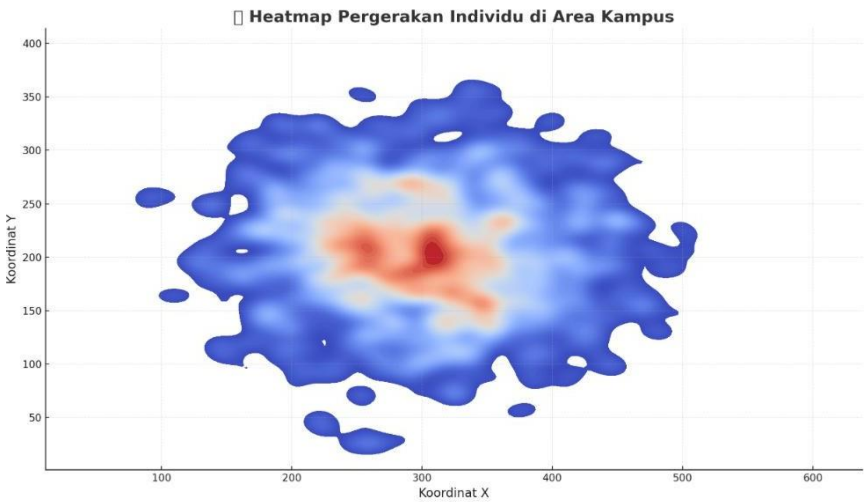

Fig 3. CCTV-Based Heatmap of Individual Activity from Detection Coordinates

4.5 Limitations and Planned Extensions

The current scope targets single-camera analytics; cross-camera identity continuity is not addressed. Performance may degrade under severe low-light conditions or strong motion blur. Planned extensions include cross-camera re-identification, mobile-friendly delivery, and controlled evaluation on annotated subsets.

4.6 Answer to the Research Problem

By combining real-time person detection, online identity tracking, structured logging, and heatmap visualization in an operator-facing dashboard, the prototype provides credible evidence of feasibility for campus CCTV use. The resulting overlays, counts, searchable logs, and spatial summaries support situational awareness and post-event review without additional staffing, motivating a comprehensive quantitative assessment in the subsequent journal manuscript.

5. CONCLUSION

This study presented a design- and prototype-oriented pipeline for campus surveillance that integrates real-time person detection (YOLOv5/YOLOv8), online identity tracking (Deep SORT), structured logging (CSV), and movement heatmaps within an operator-facing dashboard. Within the present scope, evidence from campus-like clips indicates feasibility on mid-range local hardware, stable identity persistence under brief occlusion and moderate density, and actionable spatial summaries (hotspots, dwell zones) derived from aggregated track coordinates. These outcomes directly address the research problem articulated in the Introduction: enhancing situational awareness from CCTV without adding staffing burden and while respecting privacy constraints.

Functionally, the pipeline transforms raw video into interpretable artifacts for security operations: live annotated overlays reduce cognitive load in monitoring; searchable logs enable rapid post-event retrieval; and heatmaps offer macroscopic patterns that are not apparent from frame-by-frame inspection. Deployment decisions—on-premises processing, person-class only detection, and a simple, auditable data trail—support institutional privacy expectations while maintaining low latency and practical maintainability.

The present work remains scoped to single-camera analytics and indicative runtime profiling; comprehensive quantitative evaluation (e.g., MOTA/IDF1/HOTA with repeated trials) and detector/tracker ablations are deliberately reserved for a subsequent journal article. Future extensions include cross-camera re-identification, event-level analytics (e.g., loitering, counter-flow), mobile-friendly delivery, and broader deployment studies across diverse campus scenes and hardware budgets. Overall, the prototype provides a reproducible blueprint for transforming passive CCTV into data-informed, operator-centric surveillance in higher-education environments.

6. ACKNOWLEDGEMENTS (optional)

This work was supported by internal research funding from Politeknik Negeri Bengkalis (PNBP Polbeng), Year 2025.

7. REFERENCES

Direktorat Jenderal Pendidikan Vokasi. (2021). Peta Jalan Pendidikan Vokasi 2021–2035. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. Diakses dari https://dasibangsida.lhokseumawekota.go.id/assets/upload/file/10_Peta_Jalan_Pendidikan_Indonesia_2025-2045_Bappenas.pdf

Kumar, A., Singh, R., & Rathore, H. (2023). Object tracking using YOLOv8 and DeepSORT for smart surveillance. International Journal for Multidisciplinary Research (IJFMR), 5(1), 42–51.

Kumar, N., Goel, V., & Sharma, A. (2021). Object tracking and counting in a zone using YOLOv4, DeepSORT, and TensorFlow. International Journal of Engineering Research & Technology, 10(4), 156–162.

Liu, L., Hu, W., Li, T., & Yuan, B. (2025). Vehicle flow detection and tracking based on an improved YOLOv8n and ByteTrack framework. Information, 16(1), 13.

Razzok, M., Sulaiman, M., & Alani, R. (2023). Pedestrian detection and tracking system based on Deep-SORT, YOLOv5, and new data association metrics. Information, 14(4), 218.

Szeliski, R. (2022). Computer vision: Algorithms and applications (2nd ed.). Springer.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. Proceedings of the IEEE International Conference on Image Processing (ICIP), 3645–3649.

Yadav, V., Prasad, A., & Kumar, R. (2023). Object detection and tracking using YOLOv8 and DeepSORT. ResearchGate. Diakses dari https://www.researchgate.net/publication/377677203

Yang, J., Zhang, H., & Wei, X. (2022). Video object tracking based on YOLOv7 and DeepSORT. arXiv preprint. https://arxiv.org/abs/2207.12202