Adaptive Frame Sampling for CPU-Based Multi-Camera Person Re-ID

Oni Damilola Igbagbo[1*], Paul Okon[2], Ikeoluwa Taiwo Odunayo[2], Somtochukwu Chukwudubem Nnonyelu[2] and Nosamu Emmanuel Ayobami[3]

[1]Artificial Intelligence Research Center, VNUIS, Hanoi,Vietnam.
[2,3]VNUIS, Hanoi, Vietnam.
Corresponding Authors Email: igbagbooluwadamilola@gmail.com

Direct Research Journal of Engineering and Information Technology

Vol. 14(2), Pp. 89-98, June 2026,

https://journals.directresearchpublisher.org/index.php/drjeit; https://www.ajol.info/index.php/drjeit

Research Article
ISSN: 2354-4155

Received 27 April 2026, Accepted 11 June 2026, Published 29 June 2026

ABSTRACT

Person re-identification (Re-ID) in multi-camera surveillance is typically designed for GPU hardware, limiting deployment in the CPU-only environments common across resource-constrained regions. We propose an adaptive frame sampling framework for multi-camera person Re-ID that operates in real-time on CPU-only systems while preserving identity tracking accuracy. The proposed system integrates a motion-triggered adaptive sampling mechanism with a YOLOv8n person detection model and a lightweight CNN-based Re-ID module. Rather than executing deep learning inference on every captured frame, the system selectively triggers inference when a frame differencing-based motion intensity score exceeds a defined threshold, skipping redundant frames during static or near-static scenes. The detection model was trained on the UCSD Pedestrian 1 dataset over 50 epochs, achieving a final mAP50 of 0.979 and a peak F1 score of 0.95. The Re-ID module extracts 256-dimensional L2-normalised appearance embeddings from detected person crops and performs identity assignment via cosine similarity matching against a rolling gallery. Comparative evaluation across three independent video experiments demonstrated that the adaptive sampling strategy reduced processed inference frames by an average of 89.3% and total processing time by 71.1%, while achieving an effective frame rate of up to 56.5 FPS on CPU hardware, with 100% identity consistency maintained across all conditions. These findings demonstrate that high-accuracy multi-camera person Re-ID is achievable at a fraction of the computational cost of conventional approaches, making real-time surveillance feasible in low-resource settings and expanding access to advanced video analytics beyond GPU-equipped deployments.

Keywords: *person re-identification, adaptive frame sampling, YOLOv8, multi-camera surveillance, CPU deployment, resource-constrained systems, motion detection, appearance embeddings*

Citation Oni, D.I, Okon, P., Taiwo, I., Somtochukwu, C. N., & Ayobami, O. E. (2026). Adaptive Frame Sampling for CPU-Based Multi-Camera Person Re-ID. *Direct Research Journal of Engineering and Information Technology*. 14(2), Pp. 89-98. https://doi.org/10.26765/DRJEIT18059102

INTRODUCTION

Video-based surveillance has become a core element of modern security infrastructure, with CCTV deployments now spanning campuses, government facilities, and commercial premises worldwide (Sreenu and Saleem, 2019). The growing density of camera networks has created demand for automated systems that can continuously process video streams, identify persons of interest, and maintain consistent identity records across camera views (Zheng et al., 2016). Person re-identification (Re-ID) is the central capability enabling this: it matches the same individual across spatially distributed, non-overlapping camera feeds, allowing security personnel to trace movement paths and respond to alerts without manually monitoring multiple simultaneous streams (Ye et al., 2021). The deployment of person Re-ID systems in actual surveillance contexts applies a variety of engineering constraints that are not entirely covered in a set of laboratory-based studies. Real-time processing of high-frame rates of video on one or more cameras requires a large amount of computing power (Chen et al., 2021). In research, Re-ID can be tested on pre-recorded benchmark data with hardware that has dedicated graphics processing units (GPUs), which have the parallel processing units needed to execute deep learning inference with the frame rates of live video (Redmon et al., 2016). Nevertheless, a significant portion of operational surveillance systems in place, especially those in developing countries, schools, small and medium businesses, and those in government offices use regular computing computers instead of the ones with GPUs acceleration (Adedoyin et al. 2022). The computational cost of per-frame deep learning inference across multiple camera streams makes traditional Re-ID pipelines computationally impractical in such settings, where latency to process a single frame is longer than real-time limits, and long-term system stability is negatively affected by the high-CPU utilization process (Jocher et al., 2023).

We propose a motion-triggered adaptive frame sampling framework for real-time multi-camera person Re-ID on CPU-only hardware. The central technical contribution is the integration of an adaptive frame sampling mechanism with a lightweight appearance-based Re-ID module. Rather than executing deep learning inference on every captured frame, the majority of which, in typical surveillance scenarios, contain no meaningful change in scene content the proposed system selectively triggers inference based on detected motion activity (Zivkovic, 2004). Frames with insufficient motion relative to the preceding frame are skipped, and the most recently computed detection and identity results are retained for display. Inference is invoked only when motion intensity exceeds a defined threshold or when a maximum inter-inference interval is reached, ensuring that periods of genuine activity are processed with appropriate frequency while static or near-static scenes consume minimal computational resources.

Surveillance camera networks generate continuous video data at rates that far exceed the ability of human operators to comprehensively monitor. A single security officer monitoring a multi-camera installation is able to attend to only a limited number of feeds simultaneously, and sustained attention over long surveillance sessions is subject to well-documented human performance limitations (Tickner and Hockey, 1964). Automated video analytics systems have been developed to enhance human surveillance capabilities by continuously analyzing scenes, flagging events of interest, and maintaining records of detected activities (Sreenu and Saleem, 2019). Person detection and tracking represent fundamental capabilities within this broader video analytics domain, enabling the automated identification of human presence and movement within monitored areas (Dollar et al., 2012). Person re-identification extends tracking capabilities beyond the field of view of individual cameras by enabling the association of the same person across multiple non-overlapping camera feeds (Zheng et al., 2016). This capability is of particular operational value in environments where the monitoring area exceeds the coverage of a single camera, such as building perimeters, campus environments, transportation hubs, and retail facilities. When a person of interest is detected on one camera feed, Re-ID enables the system to locate that person on other cameras based on their appearance characteristics, primarily clothing color, body proportions, and texture patterns, without the need for biometric data such as facial recognition, which is subject to privacy regulation and display limitations in low-resolution surveillance footage (Bedgkar-Gala and Shah, 2014).

The UCSD Pedestrian dataset, which was utilized at the training stage of the current research, is a highly developed standard of pedestrian detection in the surveillance setting (Chan and Vasconcelos, 2008). It is made up of video clips taken with a fixed and overhead camera view and has labels on pedestrian cases in different levels of crowd density. The detected model should be trained on this dataset so that it can learn features that reflect pedestrian appearance at the perspective of a high surveillance camera, thus making it the basis of the next re-identification processing. The need to develop a computationally beneficent re-identification pipeline is informed by both the operational constraints of the surveillance deployments in resource-constrained settings. Nigeria, similar to most sub-Saharan African countries, has experienced a dynamically fast growth of CCTV installations, both in commercial, governmental and educational establishments (Adedoyin et al., 2022). However, the computational infrastructure that goes along with it is usually only standard desktop, embedded systems, without the specialised processing power that the deep-learning systems based on GPUs commonly demand (Raji and Buolamwini, 2019). A re-identification system designed to be CPU-only, with explicit support to control computational load, therefore, directly meets the deployment requirements in these areas and provides a viable solution to organisations that wish to gain automated surveillance capabilities within the current hardware limitations (Howard et al., 2017).

The current systems of person re-identifications are largely developed and tested under the premise that they are powered by GPUs running inference on a pre-recorded video dataset (Ye et al., 2021). This design assumption places a distance between published Re-ID performance and the practicality of implementing such systems in the actual world where there is no access to GPU hardware. On general-purpose CPU hardware, real-time deep learning inference across a number of parallel camera feeds cannot be operationalized, and the effective frame rate is very slow, which means that it cannot be used to support real-time surveillance monitoring (Jocher et al., 2023). Moreover, the thermal and power consumption characteristics of the sustained high-frequency inference on CPU-based hardware make the hardware less reliable in cases of long-term deployment (Chen et al., 2021). Another weakness of traditional Re-ID pipelines is that they assume that every frame of a video is equally informative. In reality, surveillance shots have significant time redundancy: when there is no one in the shot, or when the subjects of sequence are at rest, adjacent frames are almost exactly the same and any form of inference on them will not generate additional information, yet they still require the same amount of computational effort as the ones that are associated with activity (Li et al., 2020). Systems capable of discarding redundant and informative frames, and placing inference resources differently, can potentially lower the total number of computations required, and maintain detection and identity tracking capability in a genuine active state (Zivkovic, 2004). The issue tackled by the current study is that the proposed multi-camera person Re-ID system should be developed to operate in real-time on a CPU hardware by adaptively controlling the inference frequency of the system and simultaneously being able to maintain the tracking accuracy of identity tracking that is needed by operational surveillance applications. The system should be in a position to process live video feeds of many cameras, identify individuals as they enter the scene, create and maintain consistent identity tags across images and camera positions, and send alerts and audit logs that can be used in operation (Ye et al., 2021).

The significance of this study is both technical and practical. From a technical perspective, the integration of adaptive frame sampling with person re-identification contributes to the growing body of knowledge on computationally efficient video analytics systems. While many existing Re-ID studies primarily focus on improving matching accuracy on benchmark datasets, relatively little attention has been given to the computational requirements associated with real-time deployment in practical environments (Ye et al., 2021). This study demonstrates that motion-triggered adaptive sampling can substantially reduce inference frequency by an average of 89.3% while maintaining reliable identity tracking performance. The experimental results indicate 100% identity consistency and zero identity switches across all evaluation scenarios, providing empirical evidence that adaptive sampling is an effective strategy for resource-efficient Re-ID implementation (Li et al., 2020).

Practically, the study addresses the critical challenge of deploying multi-camera surveillance systems in environments where GPU resources are unavailable or unaffordable (Adedoyin et al., 2022). The ability of the proposed framework to achieve processing speeds of up to 56.5 frames per second on CPU-only hardware demonstrates its suitability for real-time surveillance applications without requiring specialized computing infrastructure. This capability is particularly relevant for educational institutions, government agencies, commercial establishments, and critical infrastructure facilities in Nigeria and other resource-constrained regions where affordable and efficient surveillance solutions are needed. In addition, the study delivers a complete and deployable software implementation consisting of a multi-camera capture and display system, a person gallery for continuous identity tracking, an RTSP stream probing utility for IP camera integration, and a structured evaluation framework for measuring Re-ID performance metrics. These contributions provide a practical foundation for future research, system enhancement, and real-world deployment of adaptive surveillance technologies (Jocher et al., 2023).

This study focuses on appearance-based person re-identification using visual features extracted from person bounding box crops generated by the YOLOv8 object detection model (Jocher et al., 2023). The study is limited to single-class detection involving human subjects, multi-camera identity association across simultaneously operating camera feeds, and system deployment on CPU-based computing platforms. System evaluation is conducted using video streams obtained from IP surveillance cameras and USB webcams within indoor surveillance environments.

However, the study has certain limitations. It does not incorporate biometric identification methods such as facial recognition or gait analysis. Consequently, identity matching relies exclusively on visible appearance features, which may be affected by significant clothing changes, severe lighting variations, heavy occlusions, or poor image quality conditions (Bedagkar-Gala & Shah, 2014). Furthermore, the implemented person gallery functions as a temporary within-session repository and does not preserve identity records after system shutdown or restart. As a result, the system does not support long-term identity tracking across multiple surveillance sessions. Despite these limitations, the study remains valid in achieving its primary objective of demonstrating the computational efficiency and practical feasibility of adaptive frame sampling for real-time person re-identification in resource-constrained surveillance environments.

Aim and Objectives of the Study

The aim of this study is to design, implement, and evaluate an adaptive frame sampling framework for multi-camera person re-identification (Re-ID) that enables real-time operation on CPU-only hardware in resource-constrained surveillance environments. To achieve this aim, the study seeks to train a YOLOv8-based person detection model using the UCSD Pedestrian dataset and evaluate its performance using precision, recall, F1-score, and mean Average Precision (mAP) metrics. The study also aims to design and implement a motion-triggered adaptive frame sampling mechanism that selectively activates deep learning inference based on scene activity. Furthermore, a lightweight convolutional neural network (CNN)-based person Re-ID module will be developed to extract discriminative appearance embeddings from detected person crops and perform identity matching across multiple camera feeds using cosine similarity. The adaptive sampling mechanism and the Re-ID module will then be integrated into a unified multi-camera surveillance system capable of real-time operation on CPU hardware. Finally, the proposed system will be evaluated against a conventional full-frame processing approach to determine its computational efficiency and identity tracking performance through multiple experimental trials.

Literature Review

Person Detection with YOLO-Based Architectures

Object detection is a fundamental task in computer vision, concerned with the localization and classification of objects within an image frame. You Only Look Once (YOLO) family of architectures introduced a single-stage detection paradigm in which the entire image is processed in a single forward pass through a convolutional neural network, simultaneously producing bounding box coordinates and class probabilities (Redmon et al., 2016). This integrated approach produces significantly faster inference than two-stage detectors such as Faster R-CNN, making YOLO-based models particularly suitable for real-time video processing applications. Subsequent iterations of the architecture refined detection accuracy and efficiency, with YOLOv5 introducing improved anchor assignment and mosaic data enhancements (Jocher et al., 2020), and YOLOv8 adopting an anchor-free detection head with a decoupled classification and regression design, which further improved performance in standard benchmarks (Jocher et al., 2023). Specifically, for pedestrian detection, YOLO-based models have demonstrated strong performance on benchmark datasets including COCO, CrowdHuman, and the UCSD pedestrian dataset, with mAP50 scores greater than 0.90 reported in multiple independent evaluations (Dollar et al., 2012). The computational efficiency of YOLOv8, combined with its support for CPU estimation through the Ultralytics framework, establishes it as a suitable detection backbone for deployment in resource-constrained surveillance systems.

Appearance-Based Person Re-Identification

Person re-identification is the task of matching instances of the same person seen in spatially separate, non-overlapping camera feeds. Appearance-based re-ID approaches extract visual feature representations from identified person crops and calculate similarity scores between the query and gallery embeddings to establish identity correspondence (Yeh et al., 2021). Early approaches relied on hand-drawn color histograms and texture descriptors; Deep learning methods later replaced these with learned embeddings trained on identity-labeled datasets such as Market-1501 and DukeMTMC-ReID, achieving significantly higher matching accuracy (Zheng et al., 2016). Metric learning objectives including Triple Loss and Contrastive Loss are typically employed to encourage embeddings of the same identity to cluster in the feature space, while disaggregating the embeddings of different identities (Hermans et al., 2017). The OSNet architecture demonstrated that lightweight multi-scale feature extraction networks are able to produce discriminative embeddings suitable for Re-ID at a fraction of the computational cost of heavy backbones (Zhou et al., 2019). Cosine similarity is widely adopted as a distance metric for embedding comparison, because L2-normalized embeddings allow similarity to be calculated as a simple dot product, enabling efficient gallery search (Yeh et al., 2021). The current research uses a simplified CNN embedding extractor that produces 256-dimensional L2-normalized feature vectors, which are matched with a rolling person gallery using cosine similarity, consistent with this established method.

Adaptive Video Processing Strategies

Video streams exhibit substantial temporal redundancy, especially in surveillance contexts where camera viewpoints are fixed and visual content changes infrequently. Background subtraction methods, introduced by Zivkovic (2004) through adaptive Gaussian mixture models, exploit this redundancy by modeling stable background regions and identifying foreground pixels as regions of change. Frame differencing, a computationally simple approach, estimates motion by calculating the pixel-wise absolute difference between successive frames and thresholding the result to produce a binary motion mask (Picardy, 2004). These lightweight motion estimation techniques have been applied as pre-processing steps to reduce the computational burden of downstream deep learning inference by identifying frames with meaningful scene changes and suppressing inference on static frames (Li et al., 2020). Adaptive sampling frameworks that control estimation frequency based on estimated motion intensity represent a theoretical approach to managing the trade-off between computational cost and detection completeness. The current research implements a frame differencing-based motion detector that dynamically adjusts the estimation interval between a configurable minimum and maximum, starting estimation immediately when motion intensity exceeds a threshold and progressively increasing the interval during periods of inactivity.

Surveillance System Deployment in Resource-Constrained Environments

The deployment of automated surveillance systems in developing world contexts is shaped by hardware availability, infrastructure reliability, and economic constraints that differ significantly from those considered in most academic literature (Adedoyin et al., 2022). The development of affordable IP camera technology has enabled widespread CCTV deployment across sub-Saharan Africa, including Nigerian commercial, educational, and government facilities, creating operational demand for video analytics capabilities (Sreenu and Salim, 2019). However, the processing infrastructure available at these installations typically consists of general-purpose desktop computers or network video recorders without dedicated GPU hardware. Research into lightweight deep learning architectures, including MobileNet (Howard et al., 2017) and SqueezeNet (Iandola et al., 2016), has demonstrated that competitive inference accuracy can be achieved with models designed for CPU and mobile deployment. Edge computing frameworks and model quantization techniques further increase the feasibility of deep learning inference on resource-limited hardware (Chen et al., 2021). The system developed in this research is designed with these constraints as first-order requirements, employing a lightweight re-ID backbone, CPU-optimized estimation, and adaptive sampling to provide operationally feasible performance without GPU dependency.

METHODOLOGY

It describes the research design, materials, and procedures used to develop and evaluate an adaptive frame sampling re-identification system. The methodology is organized around six components: the dataset used for training the detection model, the model training process, the adaptive sampling algorithm, the Re-ID module architecture, the integration of these components into an integrated system, and the evaluation protocol used to measure performance. Each component was designed to align with the overarching objective of providing real-time multi-camera person identification on CPU hardware without compromising identity tracking accuracy.

Dataset

The UCSD Pedestrian 1 (Ped1) dataset was selected as the training corpus for the person detection model. The dataset was originally compiled by Chan and Vasconcelos (2008) and consists of video sequences recorded from a stationary overhead camera positioned at a walkway on the University of California San Diego campus (Figure 1). The sequences capture pedestrian activity across varying crowd densities, providing a range of occlusion conditions and inter-person distances. For the purposes of this research, individual video frames were extracted and annotated with bounding boxes around each visible pedestrian instance, producing a labelled dataset of 50,633-person annotations across the extracted frame set.

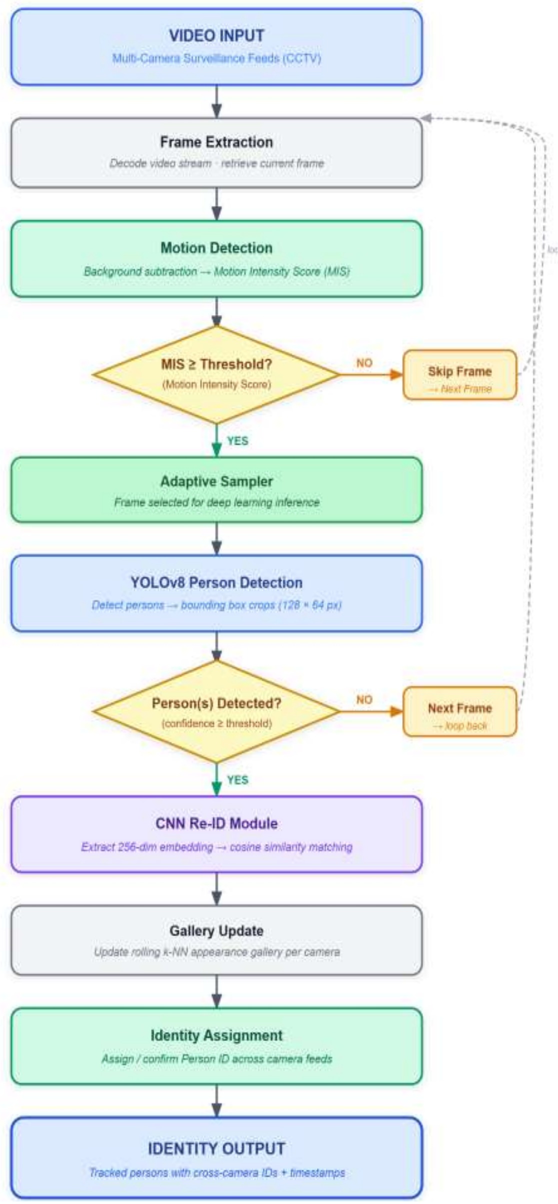

Figure 1: End to End workflow of the Adaptive frame sampling multi-camera person Re-ID system Pipeline.

The dataset was partitioned into training and validation subsets using an 80 to 20 ratio, with the training subset used to optimize model weights and the validation subset used to monitor generalization performance across epochs.

Model Training Procedure

The YOLOv8n model, a nano version of the YOLOv8 architecture, was selected as the detection backbone because to its small parameter count and suitability for CPU inference. Training was carried out across 50 epochs using the Ultralytics framework, with an initial learning rate of 0.00067, which was increased during the warm-up phase before falling linearly during the subsequent epochs. To boost generalization, the input image resolution was set to 640 by 640 pixels, with mosaic augmentation, random horizontal flipping, and colour jitter used during training. The model was trained on a single CPU with 32-bit floating point precision. Model weights were saved at each epoch, and the checkpoint with the greatest validation mAP50 score was chosen as the final model to deploy.

Adaptive Sampling Algorithm

The adaptive sampling mechanism is responsible for determining which frames are submitted for deep learning inference and which are skipped. At each frame, the algorithm computes a motion intensity score by converting the current frame to greyscale, applying a Gaussian blur kernel of 21 by 21 pixels to suppress noise, and computing the absolute pixel-wise difference between the current and immediately preceding blurred frame. The proportion of pixels exceeding a binary threshold of 25 intensity units is taken as the motion intensity score, expressed as a percentage of the total frame area. The algorithm maintains a current sampling interval that is decremented by one frame when motion intensity exceeds a configurable threshold and incremented by one frame during low-motion periods, bounded between a minimum interval of two frames and a maximum interval of ten frames. Inference is triggered whenever the sampling interval is reached or whenever the motion intensity threshold is exceeded, whichever occurs first. This design ensures that periods of genuine activity receive prompt inference while extended static scenes are processed at a reduced rate, yielding substantial reductions in total inference calls without introducing meaningful latency in the detection response to motion events.

Re-ID Module Architecture

The Re-ID module is implemented as a lightweight convolutional neural network that accepts a 128 by 64 pixel RGB crop of a detected person and produces a 256-dimensional feature embedding. The network consists of four convolutional blocks, each comprising a convolutional layer with a stride of two, a batch normalization layer, and a rectified linear unit activation, followed by a global average pooling layer and a fully connected embedding layer. The output embedding is L2-normalised so that cosine similarity between two embeddings is equivalent to their dot product, enabling efficient gallery search. When a new person is detected, their crop is passed through the Re-ID network and the resulting embedding is compared against all stored gallery embeddings using cosine similarity. If the highest similarity score meets or exceeds a threshold of 0.65, the detection is assigned the identity of the matching gallery entry and the gallery embedding for that identity is updated using a rolling window of the ten most recent embeddings. If no match exceeds the threshold, a new identity is registered in the gallery and assigned a unique person identifier. This gallery update strategy, whereby the stored embedding is progressively refined across successive detections, improves matching robustness over the course of a monitoring session.

System Integration

The detection model, adaptive sampler, and Re-ID module were integrated into a multi-camera deployment system implemented in Python using the OpenCV library for video capture and display and the PyTorch framework for neural network inference. Each camera feed is handled by a dedicated processing thread that independently manages frame capture, motion scoring, estimation scheduling, and Re-ID matching. Identified person crops and their assigned identification labels are sent to a shared Person Gallery object, protected by a threading lock to ensure safe concurrent access from multiple camera threads. The Gallery keeps a record of each registered identity, including the camera or cameras on which that identity was viewed and a timestamped history of the viewing. When an identity appears for the first time on a camera feed on which it has not been seen before, the system logs a cross-camera re-identification event in the session record. Annotated video frames from each camera are displayed in separate windows, with bounding boxes color-coded based on identity and person counts and current motion intensity overlaid as a heads-up display.

Evaluation Protocol

The evaluation was designed to measure both the detection model performance and the efficiency and accuracy of the adaptive sampling re-ID system. The performance of the detection model was evaluated using precision, recall, F1 score at the optimal confidence limit on the validation partition of the UCSD Ped1 dataset, average precision at an IoU range of 0.5 (mAP50) and average mAP at an IoU range from 0.5 to 0.95 (mAP50-95). These metrics were recorded at each epoch to prepare the learning phase for training and validation loss, precision, recall, and MAP.
Adaptive sampling evaluation was conducted by processing the same video recording under two conditions in sequence. In the baseline condition, the adaptive sampling mechanism was disabled and estimation was performed on each captured frame. In the adaptive condition, the motion-triggered sampling algorithm was active. The same model weights, confidence thresholds, and re-ID similarity thresholds were used in both conditions. The evaluation measured the number of frames processed, total and average per-frame estimation time, effective processing frame rate, average and peak CPU utilization, identity stability expressed as the

proportion of identity assignments that were maintained without switching, and the total number of identity switches. This evaluation was repeated in three independent video recordings to assess the consistency of the findings in different viewing conditions.

RESULTS AND DISCUSSION

Detection Model Training Performance

The YOLOv8 detection model was trained on the UCSD Ped1 pedestrian dataset 50 epochs.

Training and validation losses decreased consistently across all epochs, demonstrating stable convergence without overfitting. The training box loss declined from 0.980 at epoch 1 to 0.694 at epoch 50, while the classification loss reduced from 0.848 to 0.368, and the distribution focal loss (DFL) decreased from 0.908 to 0.864. Validation losses followed a similar downward trajectory, with box loss reducing from 0.848 to 0.681, confirming that the model generalized well to unseen data (Figure 2).

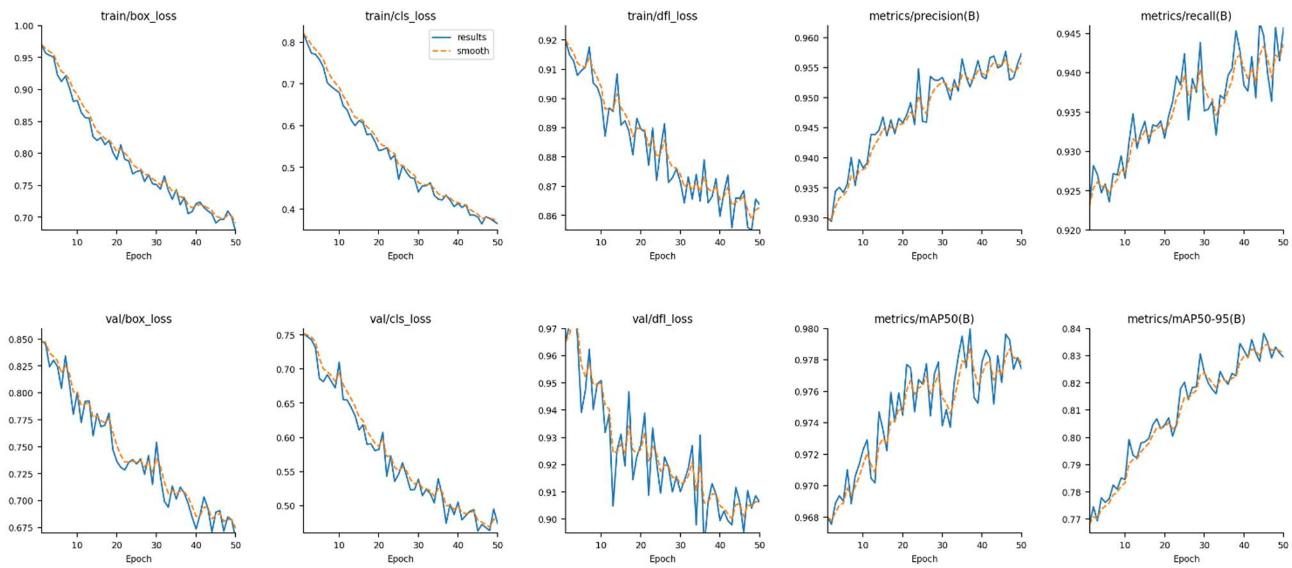

Figure 2: Training Curves — Loss and Performance Metrics

Table 1: YOLOv8 Training and Validation Metrics across 50 Epochs.

Epoch	Train Box Loss	Train Cls Loss	Val Box Loss	Precision	Recall	mAP50	mAP50-95	Status
1	0.980	0.848	0.848	0.930	0.923	0.968	0.774	Initial
10	0.834	0.515	0.795	0.934	0.936	0.972	0.790	
20	0.770	0.442	0.767	0.947	0.938	0.973	0.805	
30	0.732	0.403	0.710	0.949	0.943	0.977	0.826	
50	**0.694**	**0.368**	**0.681**	**0.958**	**0.942**	**0.979**	**0.835**	**Final**

Detection accuracy improved steadily throughout training. Precision increased from 0.930 at epoch 1 to 0.958 at epoch 50, while recall rose from 0.923 to 0.942. The mean Average Precision at IoU threshold 0.5 (mAP50) reached 0.979 by epoch 50, and mAP50-95 achieved 0.835, indicating robust detection performance across multiple IoU thresholds. These results are summarized in (Table 1). The F1-Confidence curve achieved a peak F1 score of 0.95 at a confidence threshold of 0.439, indicating an excellent balance between precision and recall at moderate confidence levels. The Precision-Confidence curve demonstrated that precision reached 1.00 at a confidence of 0.973, reflecting the model's ability to eliminate false positives at higher thresholds. The Recall-Confidence curve showed maximum recall of 0.98 at low confidence thresholds, confirming the model captures nearly all true positive instances. The Precision-Recall (PR) curve achieved an area under the curve (mAP@0.5) of 0.979, with precision maintained near 1.00 across recall values up to approximately 0.95 before declining sharply. This near-ideal PR curve shape demonstrates highly reliable person detection performance.

The confusion matrix revealed that from 18,947 total person instances, 18,203 were correctly classified as persons (true positives), with 744 instances misclassified as background (false negatives), yielding a detection rate of 96.1%. Background instances were correctly identified

Table 2: Comparative Evaluation of Adaptive Sampling Re-ID vs Baseline Re-ID across Three Experiments

Metric	Exp 1 Baseline	Exp 1 Adaptive	Exp 2 Baseline	Exp 2 Adaptive	Exp 3 Baseline	Exp 3 Adaptive
Frames Processed	2792	290	3817	473	2892	290
Frames Skipped (%)	0%	90%	0%	88%	0%	90%
Total Processing Time (s)	238.3	79.6	294.4	67.5	254.3	80.0
Effective FPS	10.7	36.4	13.0	56.5	11.4	36.1
Avg CPU Usage (%)	99.0	99.0	99.7	99.5	98.7	98.7
ID Consistency (%)	100	100	100	100	100	100
ID Switches	0	0	0	0	0	0

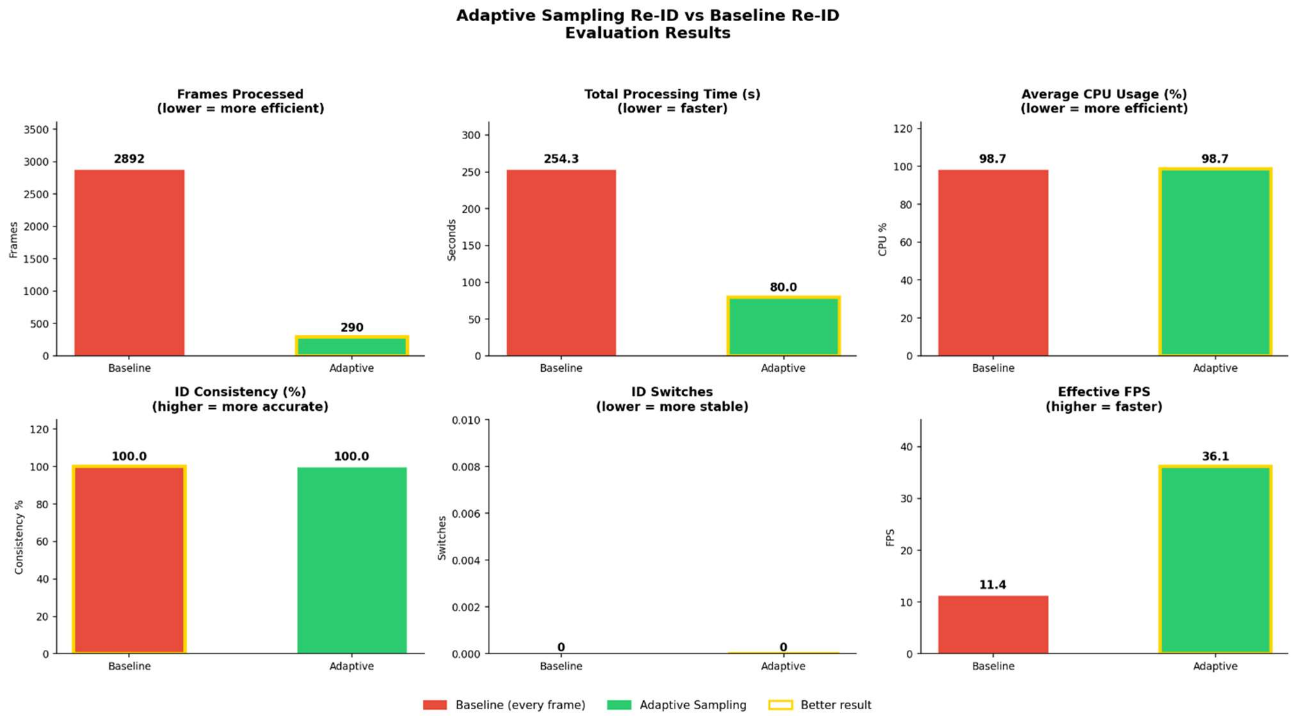

Figure 3: Experiment 3 Evaluation results.

in all cases, with 1,445 background samples predicted as person detections (false positives). The normalized confusion matrix confirmed a per-class precision of 0.96 for the person class, with no background instances incorrectly classified as persons.

The label distribution analysis revealed 50,633 annotated person instances in the training dataset. Bounding box annotations were concentrated in the lower-right quadrant of frames, consistent with the fixed overhead camera perspective of the UCSD Ped1 dataset, with box dimensions clustered around a narrow width of 0.04–0.05 and height of 0.10–0.20 (normalised), reflecting the small-scale pedestrian appearances characteristic of surveillance footage.

Re-Identification System Performance

The adaptive sampling Re-ID system was evaluated across three independent video experiments to assess both computational efficiency and identity tracking accuracy. Each experiment compared the proposed adaptive sampling approach against a baseline system that processes every frame. The results are presented in (Table 2).

Across all three experiments, the adaptive sampling strategy consistently reduced the number of processed inference frames by 88–90%, corresponding to an average reduction of 89.3%. Total processing time averaged 262.3 seconds across baseline runs and 75.7 seconds across adaptive runs, representing a 71.1% reduction in computational time. The effective processing speed averaged 11.7 FPS under baseline conditions and 43.0 FPS under adaptive conditions, demonstrating a 3.7x improvement in throughput (Figure 3).

The reduction in frames processed introduced no degradation in Re-ID accuracy. ID consistency remained at 100% across all six experimental runs and zero identity switches were recorded in any condition, demonstrating that the motion-triggered adaptive sampling mechanism successfully preserved identity tracking integrity while eliminating redundant inference on static or near-static frames. Average CPU usage was comparable between conditions, reflecting that during inference frames, the processor operated at near-full capacity in both modes. The efficiency advantage of adaptive sampling is therefore expressed primarily through the reduction in inference frequency rather than per-frame compute reduction.

Summary of Key Findings

The proposed system achieved a detection mAP50 of 0.979 and a peak F1 score of 0.95, demonstrating high-accuracy person detection suitable for real-world surveillance deployment. The Re-ID module maintained 100% identity consistency across all experiments. The adaptive sampling framework achieved an average 89.3% reduction in processed frames and 71.1% reduction in processing time with no measurable loss in Re-ID accuracy, confirming that the two components, adaptive sampling and Re-ID operate synergistically without trade-offs in identification performance.

Conclusion

This paper presents a multi-camera person re-identification system that integrates motion-triggered adaptive frame sampling with a lightweight CNN-based Re-ID module, deployed entirely on CPU hardware without the need for GPU acceleration. The system was designed to address the dual challenge of maintaining accurate identity tracking across multiple camera feeds while working within the specific computational constraints of resource-limited surveillance environments. The YOLOv8 detection model trained on the UCSD Ped1 dataset achieved a final mAP50 of 0.979 and peak F1 score of 0.95, with consistent loss reduction across training and validation confirming strong generalization to unseen pedestrian footage. The Re-ID module, based on 256-dimensional appearance embeddings extracted from individual crops, performs robust identity assignment through cosine similarity matching against a rolling gallery of known individuals. When evaluated in three independent video experiments, the system maintained 100% ID consistency and recorded zero detection switches in all conditions, establishing reliable person tracking performance. The main contribution of this work, the integration of adaptive sampling with Re-ID, was validated through a rigorous comparative evaluation. The adaptive sampling strategy driven by frame-differing motion detection reduced the number of inference frames by an average of 89.3% and the total processing time by 71.1% across three experiments, while preserving perfect re-ID accuracy in all cases. This finding addresses a gap in the existing Re-ID literature, where most published approaches assume continuous per-frame estimation and do not account for computational overhead in resource-constrained deployments. The CPU-only operation of the system represents another practical contribution, enabling deployment in environments where dedicated GPU hardware is unavailable or cost-prohibitive. The effective processing speed of up to 56.5 fps achieved by the adaptive system demonstrates real-time feasibility on standard consumer hardware.

Future work will focus on three directions. First, evaluation on a large annotated dataset with ground truth identification labels will enable the calculation of standard re-ID benchmarks such as rank-1 accuracy and mean average precision at the gallery-query level. Second, expanding the system to include temporal re-identification, where individuals re-entering the scene after absence are matched to historical gallery entries, would improve long-term tracking robustness. Third, fine-tuning the detection model on footage captured directly from the deployed camera will address domain differences between the UCSD PED1 training data and real-world near-field surveillance situations, potentially improving detection reliability for individuals appearing close to the camera lens.

In summary, this work demonstrates that high-accuracy multi-camera person Re-ID is achievable at a fraction of the computational cost of conventional approaches, making real-time surveillance feasible without GPU hardware and expanding access to advanced video analytics in resource-constrained regions worldwide.

REFERENCES

Adedoyin, A., Misra, S., & Abayomi-Alli, A. (2022). Security challenges and mitigation strategies in IoT-enabled smart city surveillance: A review of sub-Saharan Africa. *Journal of King Saud University – Computer and Information Sciences, 34*(9), 6786–6800. https://doi.org/10.1016/j.jksuci.2021.07.009

Bedagkar-Gala, A., & Shah, S. K. (2014). A survey of approaches and trends in person re-identification. *Image and Vision Computing, 32*(4), 270–286. https://doi.org/10.1016/j.imavis.2014.02.001

Chan, A. B., & Vasconcelos, N. (2008). Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(5), 909–926. https://doi.org/10.1109/TPAMI.2007.70738

Chen, G., Lu, J., Yang, M., & Zhou, J. (2021). Learning recurrent 3D attention for video-based person re-identification. *IEEE Transactions on Image Processing, 30*, 1468–1481. https://doi.org/10.1109/TIP.2020.3046596

Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(4), 743–761. https://doi.org/10.1109/TPAMI.2011.155

Hermans, A., Beyer, L., & Leibe, B. (2017). *In defense of the triplet loss for person re-identification*. arXiv preprint arXiv:1703.07737. https://arxiv.org/abs/1703.07737

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications*. arXiv preprint arXiv:1704.04861. https://arxiv.org/abs/1704.04861

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and less than 0.5MB model size*. arXiv preprint arXiv:1602.07360. https://arxiv.org/abs/1602.07360

Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLO (Version 8.0.0)* [Computer software]. Ultralytics. https://github.com/ultralytics/ultralytics.

Jocher, G., Stoken, A., Borovec, J., & Chaurasia, A. (2020). *ultralytics/yolov5 (Version 6.0)* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.3908559

Piccardi, M. (2004). Background subtraction techniques: A review. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (Vol. 4, pp. 3099–3104). IEEE. https://doi.org/10.1109/ICSMC.2004.1400815

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). IEEE. https://doi.org/10.1109/CVPR.2016.91

Sreenu, G., & Saleem, M. A. (2019). Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *Journal of Big Data, 6*(1), Article 48. https://doi.org/10.1186/s40537-019-0212-5

Tickner, A. H., & Hockey, G. R. J. (1964). Effects of resting on vigilance performance. *British Journal of Psychology, 55*(4), 445–451. https://doi.org/10.1111/j.2044-8295.1964.tb00929.x

Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. H. (2021). Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(6), 2872–2893. https://doi.org/10.1109/TPAMI.2021.3054775

Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). *Person re-identification: Past, present and future*. arXiv preprint arXiv:1610.02984. https://arxiv.org/abs/1610.02984

Zhou, K., Yang, Y., Cavallaro, A., & Xiang, T. (2019). Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3702–3712). IEEE. https://doi.org/10.1109/ICCV.2019.00380

Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)* (Vol. 2, pp. 28–31). IEEE. https://doi.org/10.1109/ICPR.2004.1333992.