

Intelligent Surveillance System Using Deep Learning to Reduce Shoplifting in Minimarkets in Santiago de Surco, Lima, Peru

Yosep Alexeis Solorzano Aguero and Jose Karim Candela Rengifo

Professional School of Systems Engineering, Peruvian University of Applied Sciences (UPC), Lima, Peru

Keywords: Deep Learning, YOLO, YOLOv8, Convolutional Neural Networks (CNN), Computer Vision, Smart Surveillance, Shoplifting, Shoplifting Detection, Theft, Theft Detection.

Abstract: This article presents an intelligent video surveillance system for theft detection in minimarkets located in Santiago de Surco, Lima. The proposed solution integrates computer vision techniques with deep learning models such as Convolutional Neural Networks (CNN) and You Only Look Once (YOLO), implemented using PyTorch. The system analyzes customer movements in real time to detect suspicious behavior patterns, including torso twists and concealment attempts. Trained on a dataset of over 2700 real and simulated images, the model achieved an accuracy of 82%, outperforming traditional surveillance systems by more than 30%. The solution includes a web interface developed with FastAPI (Fast Application Programming Interface, a high-performance Python framework for building APIs) and Angular, enabling remote monitoring. Practically, the system can reduce economic losses by up to 15%, offering a scalable and cost-effective alternative for improving security in small commercial environments.

1 INTRODUCTION

The increase in thefts in minimarkets located in Santiago de Surco, Lima, highlights significant shortcomings in current security systems. In 2022, 142 cases were officially reported, placing this district among the most affected by property crimes (Asociación de Bodegueros del Perú, 2022). These incidents, often carried out by offenders known as “tenderos”, individuals who disguise themselves as regular customers, have resulted in income losses of up to 15% and a 20% decline in customer traffic. Current surveillance systems are mostly reactive, heavily dependent on favorable visual conditions, and struggle to detect concealed behaviors. Moreover, their adoption is hindered by both technical and financial constraints.

In response to this challenge, the present study proposes an intelligent video-surveillance system based on Deep Learning (DL), specifically tailored for small businesses. The core of the system leverages Convolutional Neural Networks (CNNs), capable of analyzing visual features in video frames to identify suspicious actions. The model operates in real time without human intervention, integrating an automatic alert mechanism that enables timely responses to

abnormal activities, thereby contributing to loss prevention and enhanced commercial security.

This research is particularly relevant in the current context of rising theft rates in economically vulnerable areas such as Lima’s minimarkets, further exacerbated by the post-COVID-19 crisis. In 2022, for instance, 30% of these businesses reported a 45% increase in robberies (Asociación de Bodegueros del Perú, 2022). Under such conditions, an autonomous, efficient, and affordable DL-based solution offers a viable alternative to strengthen surveillance in resource-constrained retail environments.

Detecting theft in real time is inherently challenging due to the speed and subtlety with which offenders operate. Traditional methods—such as human surveillance or Closed-Circuit Television (CCTV)—face critical limitations, including operator fatigue, low-light sensitivity, and blind spots (Kakadiya et al., 2019). In contrast, deep learning models can process large volumes of video streams, identify complex behavioral patterns, and continually improve detection accuracy, making them more effective tools in small commercial settings (Zhang et al., 2020).

Nevertheless, many existing DL-based systems are designed for large-scale environments and require costly infrastructure. For example, the framework

proposed by Zhang et al. (2020) achieves an accuracy of 83% (dimensionless) but is impractical for small businesses due to its complex architecture and high computational requirements. Similarly, Horng and Huang (2022) designed a system dependent on multiple cameras, which significantly increases implementation costs. In contrast, the solution proposed in this study is operationally simple, cost-effective, and scalable. It requires fewer cameras, adapts to different store layouts, and is deployable on modest hardware resources, making it an accessible option for local minimarkets.

The contributions of this research are threefold. First, it introduces an efficient Deep Learning (DL) model based on You Only Look Once (YOLO), capable of detecting theft without human intervention across diverse commercial contexts (Ultralytics, n.d.). Second, it presents the design of a complete intelligent-surveillance system that integrates real-time monitoring with automated alert generation, reducing reliance on human operators. Third, it provides empirical evidence demonstrating the model's accuracy and practical impact, supporting future research in object detection and anomaly detection, and highlighting its potential to reduce theft in low-income retail environments.

This article is organized as follows: Section 2 reviews the state of the art in computer vision and deep learning techniques for object detection and anomalous behavior recognition. Section 3 details the methodology of the proposed system, including its technical components, model architecture (YOLOv8 and Convolutional Neural Networks, CNNs), and dataset split (80% training / 15% validation / 5% testing). Section 4 describes the experimental setup, evaluation metrics—precision (%), recall (%), and Mean Average Precision (mAP). For clarity, mAP50 and mAP50–95 are reported as dimensionless metrics that combine both classification and localization performance. Section 5 presents the discussion, conclusions, limitations, and recommendations for future research in intelligent video-surveillance.

2 RELATED WORKS

Several studies have explored the application of Deep Learning (DL) techniques to enhance video-surveillance systems, particularly in contexts where the automatic detection of suspicious behavior is critical for theft prevention. The following five studies provide relevant support for the development of the proposed system.

Kim et al. (2021). The authors presented a system that uses Three-Dimensional Convolutional Neural Networks (3D-CNNs) to detect shoplifting in convenience stores from surveillance footage. The network was trained on datasets collected from both real and simulated environments, achieving 85% accuracy (dimensionless) in detecting individual actions and 98.9% accuracy (dimensionless) in predicting criminal intent. However, the architecture relies on 3D convolutional layers that process spatial and temporal information simultaneously, which requires substantial computational infrastructure and high memory consumption, limiting deployment on small-scale retail hardware.

De Paula et al. (2022). This study introduced CamNuvem, a dataset designed for theft-detection model training in commercial environments. Built from real videos sourced from social media and weakly labeled to indicate the presence or absence of theft, CamNuvem provides an important benchmark for anomaly detection. The authors evaluated Robust Temporal Feature Magnitude (RTFM), Weakly-Supervised Anomaly Localization (WSAL), and Real-Time Anomaly Detection System (RADS) models. Reported results show 78–88% accuracy (dimensionless). Nevertheless, accuracy declined significantly when analyzing videos containing specific theft events, highlighting the challenge of temporal localization. This limitation illustrates the need for sequence-based approaches (e.g., Long Short-Term Memory, LSTM) that explicitly capture temporal dependencies, although such models increase latency and computational cost.

Han et al. (2024). An indoor surveillance system was developed combining YOLOv8 with DeepSORT (Simple Online and Realtime Tracking with Deep Features). YOLOv8 performs single-frame object detection, while DeepSORT assigns consistent IDs across frames, enabling real-time person tracking even in occluded areas. The system achieved 93.56% accuracy (dimensionless) and operated at 22 Frames per Second (FPS). Although performance metrics are strong, the authors note that large-scale deployment is constrained by high computational resource requirements.

Gawande et al. (2023). The proposed architecture integrates Mask R-CNN (two-stage instance segmentation) with YOLOv5 (single-stage detection) to enhance recognition under poor visual conditions, such as occlusions and low resolution. Tested in academic environments, the system achieved 87.41% accuracy (dimensionless), outperforming baselines such as RetinaNet and Region-based Fully Convolutional Networks (R-FCN). However,

validation was limited to controlled settings, and the lack of deployment in real commercial environments (supermarkets or minimarkets) reduces its external validity.

Finally, Santos et al. (2024). The authors developed a system for automatic weapon detection using Faster R-CNN and YOLO models. The system considered contextual factors such as object size and lighting conditions, reporting 85.44% accuracy (dimensionless) for firearms and 46.68% accuracy (dimensionless) for knives. The study highlights that incorporating contextual variables such as body posture and hand movement improves robustness. While the focus is weapon detection rather than theft, the emphasis on context-aware modeling is highly relevant for shoplifting detection.

3 METHODOLOGIES

3.1 Preliminary Concepts

This surveillance system is designed to reduce theft in small businesses such as minimarkets by leveraging Artificial Intelligence (AI). It integrates computer vision, Convolutional Neural Networks (CNNs), and a high-speed object detection model known as You Only Look Once (YOLO). The prototype is implemented in Python using frameworks such as PyTorch, which enables real-time identification of suspicious activities without requiring constant human oversight.

3.1.1 Convolutional Neural Networks (CNN)

CNNs are a class of Deep Neural Networks (DNNs) specialized in image and video analysis. They extract essential visual features—such as object shapes, contours, and textures—that are critical for accurate scene interpretation. In this system, CNNs process video frames captured by surveillance cameras to identify abnormal behaviors, including concealment gestures. The networks learn to autonomously detect body positions and hand movements associated with shoplifting (Cao et al., 2021).

3.1.2 Real-Time Detection with YOLO

YOLO is a single-stage object detector that divides the image into a grid and predicts bounding boxes and class probabilities in a single pass. This design enables simultaneous multi-object detection with very low latency. YOLO is particularly effective for tracking customer–product interactions in confined or

complex retail environments because it maintains real-time inference rates (Frames per Second, FPS) even on modest hardware (Redmon et al., 2016).



Figure 1: the system identifies customers upon entry and tracks their interaction with products.

3.1.3 Smart Surveillance

Traditional video surveillance is transformed into an intelligent system through the integration of deep learning models. These models can interpret live visual data and generate alerts for suspicious behavior in real time, eliminating the need for constant human monitoring (Krizhevsky et al., 2012; Valera et al., 2005).

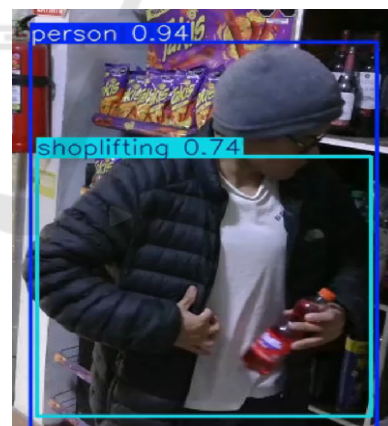


Figure 2: The system detects hand movements associated with product concealment and generates an automatic alert.

3.1.4 Learning by Demonstration

According to Nguyen et al. (2021) and Paszke et al. (2019), the system is trained on a dataset that includes both typical customer behavior and shoplifting attempts. This approach allows the neural networks to learn to distinguish between normal actions and those that represent risk patterns.

3.2 Method

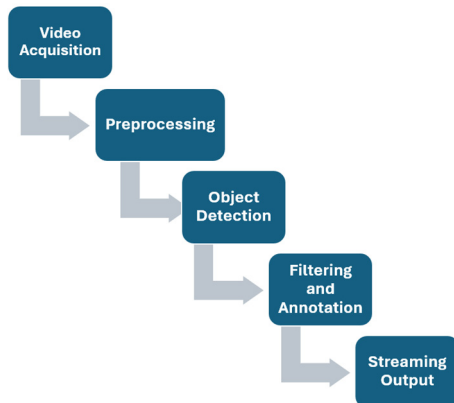


Figure 3: Shoplifting detection process is illustrated step by step.

This section describes the principal contributions of the proposed intelligent surveillance system. The core innovation lies in the integration of computer vision and Deep Learning (DL) to deliver a comprehensive solution capable of detecting and classifying suspicious behavior across multiple video streams in real time. The system leverages YOLO-based architectures to perform high-speed inference on each captured frame, ensuring efficient operation even in resource-constrained environments.

Figure 4 illustrates the overall processing workflow, where independent video streams from multiple cameras are analyzed in parallel through the following stages:

- **Video Acquisition.** The system captures real-time video streams from one or more cameras connected to the local processing unit. Each device produces a continuous, synchronized sequence of digital frames, ensuring no latency or frame misalignment across different camera angles in the minimarket environment.
- **Preprocessing.** All frames are automatically resized to 640×640 pixels, normalized, and adjusted for color and brightness. This preprocessing step guarantees that the YOLO detector can process inputs consistently, mitigating distortions from lighting changes or variable resolutions.
- **Object Detection.** YOLO analyzes each frame in a single pass, dividing the image into a grid to predict bounding boxes, categories (e.g., person, theft attempt), coordinates, and confidence scores. This single-stage architecture yields robust, real-time

surveillance, outperforming conventional multi-stage pipelines.

- **Filtering and Annotation.** Predictions are refined using Non-Maximum Suppression (NMS), which removes redundant, overlapping bounding boxes while preserving the most confident detections. Detected objects are annotated with bounding boxes, class labels, and accuracy percentages (%), directly overlaid on the video frames.
- **Streaming Output.** Annotated frames are encoded in JPEG format and transmitted through dedicated WebSocket channels (one per camera). A centralized web interface developed with FastAPI (Fast Application Programming Interface) and Angular displays real-time streams. The interface includes connection status and failure notifications, enabling secure, multi-camera, and remote monitoring. Next, we show the process of the YOLO algorithm below:

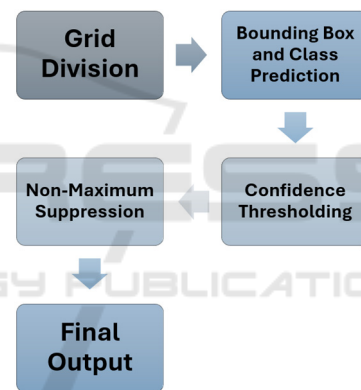


Figure 4: Object detection process using YOLO is illustrated.

YOLO is a real-time object detection algorithm that processes the entire image in a single evaluation cycle, unlike traditional models that analyze image regions independently. Its detection pipeline consists of the following steps:

- **Grid Division:** The input image is partitioned into a grid (e.g., 13×13 or 19×19), where each cell is responsible for detecting objects whose center falls within its boundaries. This approach supports distributed and localized detection across the frame.
- **Bounding Box and Class Prediction:** Each grid cell predicts multiple bounding boxes. For each box, the algorithm outputs the object's spatial coordinates (x, y, width, height), a confidence score, and a class label.

- **Confidence Thresholding:** A Minimum Confidence Threshold (Typically 0.5) Is Applied to Discard Low-Confidence Predictions. Only Bounding Boxes with High Reliability Are Retained.
- **non-Maximum Suppression (NMS):** to Prevent Duplicate Detections of the Same Object, NMS Filters Overlapping Bounding Boxes, Keeping Only the One with the Highest Confidence Score for Each Object Class.
- **Final Output:** the Algorithm Produces a Refined List of Detected Objects, Each with Bounding Box Coordinates, Class Labels, and Confidence Scores Ready for Visualization and System Response.

3.2.1 Learning by Demonstration

the System Was Trained Using an Annotated Dataset Consisting of 2,782 Images, Captured from both Real Minimarket Scenarios and Simulated Shoplifting Environments. the Dataset Included Examples of Normal Customer Behavior as Well as Theft Attempts, Ensuring Class Diversity for Model Learning.

to Guarantee Robust Evaluation and Prevent Data Leakage, the Dataset Was Divided into Three Subsets:

- **Training Set (80%).** Used to Fit the Model Parameters by Minimizing the Loss Function During Iterative Updates.
- **Validation Set (15%).** Employed to Fine-Tune Hyperparameters, Monitor Learning Curves, and Mitigate Overfitting.
- **Test Set (5%).** Reserved Exclusively for Final Performance Measurement, Ensuring Unbiased Assessment of Generalization Capacity.

This Partition Was Conducted with Stratification by Class Labels (Normal vs. Suspicious) to Maintain Balance, and all Random Splits Were Generated with Fixed Random Seeds for Reproducibility. Reported Evaluation Metrics, Precision (%), Recall (%), and Average Precision (Map, Dimensionless), Were Computed Exclusively on the Held-out Test Set.

4 EXPERIMENTS

This Section Describes the Environment and Resources Used to Validate the Proposed Intelligent Video-Surveillance System. It Covers the Development Setup, Cloud Infrastructure, Dataset Sources, and Supporting Tools Employed During Experimentation.

4.1 Experimental Protocol

4.1.1 Development Environment

All experiments were conducted on a local workstation with the following specifications:

- **Central Processing Unit (CPU):** Intel® Core™ i7-9750HF.
- **Graphics Processing Unit (GPU):** NVIDIA® GeForce® GTX 1650 (4 GB memory).
- **Random Access Memory (RAM):** 8 GB DDR4.
- **Operating System (OS):** Windows 11 x64.
- **Frameworks/Libraries:** PyTorch 2.0, OpenCV 4.7, YOLOv8 (Ultralytics implementation).
- **Programming Languages:** Python 3.10 for back-end development and Angular Material v19 for the web interface.
- **Complementary Tools:** TensorBoard (for visualization of training metrics, loss curves, and convergence plots).
- **Database:** MongoDB for storage and retrieval of annotated data and logs.

4.1.2 Additional Infrastructure

To support large-scale training and experimentation, the system also leveraged Google Colab Pro (cloud platform), providing access to high-performance GPUs for accelerated training.

4.1.3 Code Repository

For reproducibility, the complete source code and trained models are publicly available at: <https://github.com/YOSS201/DeepEyes.git>

This repository includes training scripts, configuration files, and annotation formats, enabling replication of results and facilitating future improvements.

4.1.4 Dataset Used

The system was trained on a proprietary dataset comprising 2,782 images, manually captured and annotated in both real-world minimarket environments and simulated shoplifting scenarios. Two labels were defined: “Person” as normal customer behavior and “Shoplifting” as suspicious or theft-related actions.

Images were collected under varying conditions (lighting, camera angles, crowd density) to improve generalization.

4.1.5 External Sources

In addition to the proprietary dataset, complementary resources were used to enhance annotation quality and balance class representation:

- Roboflow: Provided supplementary annotated images and labeling assistance.
- Label Studio: Used for annotation management, data cleaning, and curation.

4.2 Results

This subsection details the results obtained from the experiments.

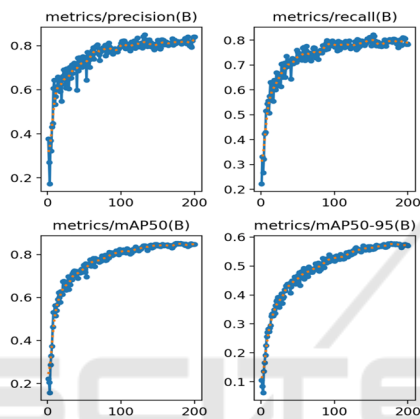


Figure 5: The precision, recall, Mean Average Precision (mAP50), and mAP50–95 metrics (y-axis) for each training epochs (x-axis) for the theft detection model presented.

4.2.1 Validation

The system was evaluated using standard deep learning performance metrics:

- Precision (%). Indicates the proportion of correct detections among all predicted suspicious activities. The system achieved 82%, confirming that most alerts correspond to true shoplifting behaviors (low false positives).
- Recall (%). Represents the proportion of actual theft events correctly detected. The system reached 80%, showing strong detection capacity while leaving room for improvement in minimizing false negatives.
- mAP50 (Mean Average Precision at $\text{IoU} \geq 50\%$, dimensionless). Combines classification and localization performance. The system achieved 84%, demonstrating reliable detection and localization when bounding boxes overlap ground truth by at least 50%.
- mAP50–95 (Mean Average Precision across IoU thresholds from 50% to 95%,

dimensionless). Evaluates performance under stricter localization thresholds. Our system reached 57%, reflecting moderate robustness under challenging conditions such as occlusion, low light, and camera variability.

Below are the labeled images in the dataset versus the images predicted by the model.

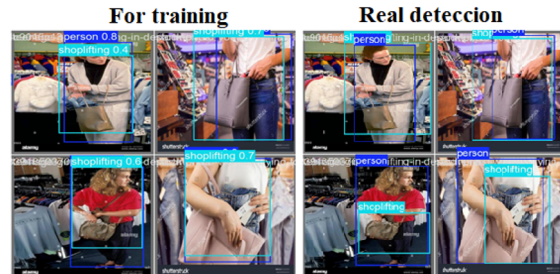


Figure 6: Labeled images from the dataset used for training the model are presented on the left, and images processed by the model showing detection results after training are shown on the right.

4.2.2 Conclusion of the Trained Model

The trained detector demonstrated **precision of 82%** and **recall of 80%**, confirming strong detection performance in real and simulated minimarket environments. The **mAP50 of 84%** validates reliable localization, while the **mAP50–95 of 57%** highlights the system's ability to handle more complex scenes, albeit with performance degradation under extreme conditions.

The system correctly identified the following suspicious behaviors:

- Prolonged presence in sensitive areas
- Sudden movements or leaning towards shelves
- Concealment of items within clothing

These results suggest that the proposed architecture is well-suited for **real-time deployment in small commercial environments**, balancing accuracy and computational efficiency.

4.3 Comparative Evaluation

4.3.1 Comparison with Traditional Systems

To measure effectiveness, the proposed system was compared against conventional Closed-Circuit Television (CCTV) surveillance commonly deployed in small businesses. Traditional systems rely on human monitoring, which introduces limitations in real-time detection, accuracy, and event analysis.

Table 1: Comparison of Traditional CCTV Systems vs. Proposed YOLO-based System.

Criterion Evaluated	Traditional System (CCTV)	Proposed System (YOLO + Deep Learning)
Real-time detection	No (requires constant human monitoring)	Yes (automatic and immediate detection)
Detection Precision	50–60% (variable due to visual fatigue)	82% (low false positive detection)
Recall rate	Very low	80% (detection of 3 out of 4 thefts)
Generation of automatic alerts	Not available	Available for suspicious actions
Stability and remote viewing	Limited	Multi-platform and cloud access

4.3.2 Comparison Between Yolo Versions

To examine improvements across YOLO versions, three architectures were evaluated during training: YOLOv5, YOLOv8, and YOLOv11. Metrics are reported as precision (%), recall (%), and mAP (dimensionless).

Table 2: Comparison of metrics for YOLO versions for training the proposed model.

Version	Metric			
	Precision (%)	Recall (%)	mAP50 (%)	AP50-95 (%)
YOLOv5	0.81256	0.80784	0.83403	0.49208
YOLOv8	0.82451	0.80422	0.84456	0.57787
YOLOv11	0.82752	0.78377	0.815	0.55488

5 DISCUSSIONS

Prior research highlights the potential of deep learning for theft detection but also reveals practical limitations. Kim et al. (2021) achieved 98.9% accuracy with a 3D-Convolutional Neural Network (3D-CNN), though its computational cost prevents small-scale deployment. De Paula et al. (2022) introduced the CamNuvem dataset, but their models addressed theft only at a binary level, without recognizing specific suspicious actions. Han and Feng et al. (2024) combined YOLOv8 with DeepSORT to track individuals, focusing mainly on

crowd dynamics rather than theft behavior. Similarly, Gawande et al. (2023) improved detection under occlusion and low resolution, but only in academic scenarios. Santos et al. (2024) developed a Faster R-CNN and YOLO system for weapon detection, targeting object-specific threats instead of behavioral patterns.

In contrast, our approach integrates YOLOv8 with Convolutional Neural Networks (CNNs) to detect fine-grained theft-related behaviors, such as product concealment, in real time and with modest resource requirements. Controlled experiments validated its performance with precision = 82%, recall = 80%, and mAP50 = 84% (dimensionless), showing reliable detection in realistic minimarket conditions.

Beyond technical accuracy, the system demonstrates clear economic relevance, with the potential to reduce $\approx 15\%$ in financial losses, equivalent to S/.12,000 annually for a typical minimarket. By automating alerts and reducing dependence on human monitoring, it enhances both operational security and customer trust.

In summary, this research advances theft detection by offering a cost-effective, scalable, and behavior-focused solution, addressing gaps left by previous deep learning approaches and adapting effectively to resource-constrained retail environments.

6 CONCLUSIONS

The results obtained demonstrate that the proposed system, based on YOLOv8, CNN and learning by demonstration significantly outperforms traditional video surveillance methods. Conventional approaches typically rely on continuous human supervision, making them vulnerable to errors caused by fatigue or distraction. In contrast, our system operates autonomously, identifying suspicious behaviors and generating real-time alerts.

The model achieved a precision of 82%, a recall of 80%, and a Mean Average Precision at 50% IoU (mAP50) of 84%, reflecting a strong balance between accuracy and detection sensitivity. These metrics confirm the system’s ability to detect theft-related behaviors in real-world convenience store environments with high reliability

In practical terms, the system has the potential to reduce economic losses by up to 15%, based on data from local commercial associations and our experimental results. For a typical minimarket with annual revenues of S/.80,000, this translates into potential savings of approximately S/.12,000 per

year, demonstrating the economic value of deploying intelligent surveillance in small-scale businesses.

The integration of computer vision and deep learning represents a robust, scalable, and cost-effective solution to enhance security in vulnerable commercial settings, particularly where resources are limited.

Future work will explore the integration of Internet of Things (IoT) components, such as shelf pressure sensors or RFID systems, to provide multi-source behavioral analysis and contextual awareness. Additionally, the adoption of edge computing architectures (e.g., NVIDIA Jetson or Raspberry Pi) is proposed to enable faster, on-device processing and improve system performance in environments with limited connectivity.

REFERENCES

- Asociación de Bodegueros del Perú. (2022). Statistical report about losses caused by thefts in minimarkets. <https://surl.li/cznuvu>
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields. <https://arxiv.org/abs/1812.08008>
- De Paula, D. D., Salvadeo, D. H. P., & De Araujo, D. M. N. (2022). CamNuvm: A robbery dataset for video anomaly detection. *Sensors*, 22(24), 10016. <https://doi.org/10.3390/s222410016>
- Gawande, U., Hajari, K., & Golhar, Y. (2023). Real-time deep learning approach for pedestrian detection and suspicious activity recognition. *Procedia Computer Science*, 218, 2438–2447. <https://doi.org/10.1016/j.procs.2023.01.219>
- Han, L., Feng, H., Liu, G., Zhang, A., & Han, T. (2024). A real-time intelligent monitoring method for indoor evacuation distribution based on deep learning and spatial division. *Journal of Building Engineering*, 92, 109764. <https://doi.org/10.1016/j.jobbe.2024.109764>
- Horng, S., & Huang, P. (2022). Building unmanned store identification systems using YOLOv4 and Siamese network. *Applied Sciences*, 12(8), 3826. <https://doi.org/10.3390/app12083826>
- Kakadiya, R., Lemos, R., Mangalan, S., Pillai, M., & Nikam, S. (2019). AI based automatic robbery/theft detection using smart surveillance in banks. 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA). <https://doi.org/10.1109/ICECA.2019.8822186>
- Kim, S., Hwang, S., & Hong, S. H. (2021). Identifying shoplifting behaviors and inferring behavioral intention based on human action detection and sequence analysis. *Advanced Engineering Informatics*, 50, 101399. <https://doi.org/10.1016/j.aei.2021.101399>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 25, 1097–1105. <https://dx.doi.org/10.1145/3065386>
- Nguyen, H. H., Ta, T. N., Nguyen, N. C., Bui, V. T., Pham, H. M., & Nguyen, D. M. (2021). YOLO based real-time human detection for smart video surveillance at the edge. In *IEEE Eighth International Conference on Communications and Electronics (ICCE)*. <https://doi.org/10.1109/ICCE48956.2021.9352144>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 32. <https://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Policía Nacional del Perú. (2024). Police statistical bulletin I quarter 2024. <https://www.policia.gob.pe/estadisticopnp/documentos/boletin-2024/Boletin%201%20Trimestre%202024.pdf>
- 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 (CVPR)*, <https://doi.org/779-788>. 10.1109/CVPR.2016.91
- Santos, T., Oliveira, H., & Cunha, A. (2024). Systematic review on weapon detection in surveillance footage through deep learning. *Computer Science Review*, 51, 100612. <https://doi.org/10.1016/j.cosrev.2023.100612>
- Ultralytics. (n.d.). YOLOv8 documentation. <https://docs.ultralytics.com/>
- Valera, M., & Velastin, S. A. (2005). Intelligent distributed surveillance systems: A review. *IEEE Proceedings – Vision, Image and Signal Processing*, 152(2), 192–204. <https://doi.org/10.5220/0001936803140319>
- Wang, H., Wang, C., & Zhang, J. (2020). Human behavior recognition in surveillance video based on 3D skeleton information. *Sensors*, 20(3), 1–15. <https://doi.org/10.3390/s23115024>
- Zhang, Y., Jin, S., Wu, Y., Zhao, T., Yan, Y., Li, Z., & Li, Y. (2020). A new intelligent supermarket security system. *Neural Network World*, 30(2), 113–131. <https://doi.org/10.14311/nnw.2020.30.009>