

Edge AI for Smart, Automated Awareness — No Cloud Required

 Project Overview

One-line takeaway: A Raspberry Pi 5 paired with a Hailo-8L AI accelerator performs real-time person detection entirely at the edge — no cloud, no lag — and fires instant Telegram notifications with delayed snapshots the moment a person is detected.

Purpose: Demonstrate how compact, low-power edge hardware can handle demanding AI inference workloads that traditionally required cloud servers or desktop GPUs. The Hailo-8L delivers up to 13 TOPS (Tera Operations Per Second) of AI compute in a small M.2 module, enabling real-time neural network inference at a fraction of the power draw of a GPU. The system runs an optimized person detection model, monitors a live camera feed, and autonomously sends Telegram alerts with 0.25-second delayed snapshots for full-body image clarity — all without leaving the device.

 Key Capabilities Real-Time Detection

The Hailo-8L delivers up to 13 TOPS of neural network inference, enabling live person detection at high frame rates with minimal latency — far exceeding what the Pi's CPU alone could achieve.

 Instant Telegram Alerts

A Telegram bot fires a notification the moment a person is detected in frame, reaching any device with Telegram — phone, tablet, or desktop — in real time.

 Delayed Full-Body Snapshot

A 0.25-second delay between detection and capture ensures the person has fully entered the frame, producing complete, clear images instead of partial or cut-off detections.

 Unattended Autonomous Operation

Runs continuously without manual intervention. The system boots, initializes the camera pipeline, and begins monitoring automatically — ideal for entryways and unmanned spaces.

 Private — No Cloud Dependency

All inference, image processing, and logic runs locally on the Pi. No data is sent to external servers — only outbound Telegram notifications leave the device.

 Event Logging

Detection events are logged locally with timestamps, providing a lightweight audit trail for later review, analytics, and system health monitoring.

How It Works

① Raspberry Pi 5 + Hailo-8L Accelerator

The Raspberry Pi 5 acts as the host, managing the camera, running application logic, and sending notifications. The Hailo-8L AI accelerator — connected via the Pi 5's PCIe interface — handles all neural network inference, delivering detection speeds impossible on CPU alone. Together they form an efficient, fully self-contained edge AI node.



② AI Detection Pipeline

A Python application manages the full pipeline: GStreamer handles the live camera stream and feeds frames to the Hailo runtime. The Hailo-8L runs an edge-optimized person detection model, returning bounding boxes and confidence scores per frame. OpenCV handles snapshot capture. The pipeline filters by confidence threshold to minimize false positives.



③ Notification & Logging System

When a detection crosses the confidence threshold, the Telegram Bot API fires an instant alert message to a configured channel or user. After a 0.25-second delay to allow the subject to fully enter frame, a snapshot is captured and sent as a follow-up Telegram message. All events are logged locally with timestamps for auditing and review.

Technologies Used

HARDWARE

- Raspberry Pi 5
- Hailo-8L AI Accelerator (M.2)
- USB / CSI Camera
- MicroSD Card
- HDMI Display (optional)

SOFTWARE

- Raspberry Pi OS
- Python 3
- OpenCV (image processing)
- GStreamer (video pipeline)
- Telegram Bot API
- Hailo Runtime & SDK

AI MODEL

- Edge-optimized detection model
- Person bounding box tracking
- Confidence score filtering
- Real-time frame processing
- Motion-aware event logic

Potential Use Cases

Entry Monitoring

Detect and log when people enter doorways, lobbies, or restricted areas without a dedicated security system.

Package Delivery Alerts

Notify when a delivery person arrives at a door — with a snapshot — even when no one is home.

Campus & Office Spaces

Monitor lab, office, or study room occupancy autonomously, enabling automated notifications for space management.

Edge AI Research

A practical platform for exploring real-time neural network inference, computer vision pipelines, and edge deployment.