Object Detection in Surveillance - One-Page Reference

What detection is, how it surfaces into the VMS, its honest accuracy, and the false-alarm math. Representative 2026 figures - all move with scene, lighting, angle, and tuning.

A. Detection vs classification

- Image classification labels a whole frame with one tag ('this image contains a car') - useless alone for surveillance.
- Object detection draws a box + class for every object (person, vehicle, bag) and gives its location - the foundation analytic.
- Detection makes a scene actionable: not 'a person is present' but 'a person is inside this zone at these coordinates'.
- Every richer analytic - tracking, behavioral rules, search-by-event - is built on object detection.

B. Accuracy is a range, never 100%

- Precision = share of alerts that are real; recall = share of real objects found; they trade off as you move the sensitivity dial.
- Strict COCO mAP@0.5:0.95 for real-time detectors ~38-55% (YOLOv12-N ~40.6, -X ~55.2) - low only because the benchmark is hard.
- In a fixed, tuned scene with few classes, operational precision/recall run much higher (~85-96% mAP@0.5 on scene footage).
- Ask a vendor for precision AND recall on YOUR footage - especially for small or distant objects. Never accept a single '99%'.

C. False alarms: motion vs object detection

- Motion detection fires on any pixel change - shadows, foliage, animals, weather, headlights - and runs 90%+ false alarms.
- Object detection filters on class (person/vehicle), removing ~80-95% of nuisance alarms.
- But volume still bites: 1% false-positive x 100,000 candidates/day = 1,000 false alarms/day - tune the survivors down.
- Camera placement and scene control set accuracy as much as the model - demand a short on-site pilot before committing.

D. Where it runs and how it surfaces

- Detection is light enough for the camera's own NPU (~1-6 TOPS runs a quantized detector at 15-30 FPS) - the edge tier.
- On the edge it keeps latency low, cuts bandwidth (a tiny event, not full video), and keeps raw video on-device.
- It surfaces into the VMS via ONVIF Profile M: generic object classification + metadata for vehicle, plate, face, body.
- ONVIF conformance is a baseline - a vendor's special attribute may still need that maker's SDK.

E. The privacy line: a class is not an identity

- Detecting 'person' or 'vehicle' tells you what, not who - generally NOT biometric data under GDPR Art. 9. Low privacy weight.
- Face recognition and license-plate recognition DO identify - biometric/personal-data, a legal gate (GDPR, EU AI Act, BIPA).
- Plain detection gains weight once it feeds cross-camera tracking (a movement trail without a name) - medium weight.
- Engineering guidance, not legal advice; confirm specifics with qualified counsel. The full law is the Privacy & Compliance block.

How detection becomes a searchable event: ONVIF Profile M standardizes generic object classification and metadata (vehicle, plate, face, body) plus event interfaces, carried over the stream, the ONVIF event service, or MQTT - a conformant analytic can be a camera, a server, or a cloud service. The detection MODEL (YOLO lineage, transformer/open-vocabulary detectors) is engineered in the AI for Video Engineering section; this is the surveillance APPLICATION. Sources: ONVIF Profile M (spec v1.1); IEC 62676 (esp. Part 6, video content analytics performance grading); GDPR Reg. (EU) 2016/679 Art. 9; YOLOv12 (arXiv 2502.12524); COCO/Roboflow mAP; surveillance fine-tuning literature; edge-AI chip analyses.