Journal of Intelligent Systems and Internet of Things
Vol. 18, No. 02, PP. 447-464, 2026
DOI: https://doi.org/10.54216/JISIoT.180231

Thermal Vehicle Detection and Tracking for Intelligent Transportation Systems: A Modular IoT Architecture and Staged Deployment Roadmap

Mostafa Borhani[1,*]
[1]Smart Tech Services SPC, Muscat, Oman
Email: borhani@iSmartGCC.com

Abstract

Automated vehicle monitoring in intelligent transportation systems must operate reliably around the clock, including under conditions that routinely cripple conventional visible-light cameras: night, glare, shadows, and adverse weather. This paper proposes a modular Internet of Things (IoT) architecture for thermal-based vehicle detection, classification, and trajectory analysis, together with a four-phase deployment roadmap that connects public-dataset evaluation to live-traffic field validation. The system integrates longwave infrared (LWIR) imaging (8–14 μm) with YOLO-family deep learning detectors (YOLOv8/v11/v12) and multi-object tracking algorithms (ByteTrack, BoT-SORT, StrongSORT), deployed across NVIDIA Jetson edge devices and cloud infrastructure through JSON/MQTT-formalized data contracts. The primary novel contribution is a system-level integration framework that bridges the gap between component-level algorithmic research and operational deployment. Concretely, this work: *(i)* defines five functionally independent modules with explicit interface specifications and latency budgets not previously formalized in the thermal-ITS literature; *(ii)* introduces quantified decision gates linking progression criteria directly to published benchmark values; *(iii)* provides region-specific operational availability estimates derived from empirical weather-degradation data; and *(iv)* integrates domain adaptation, GDPR compliance, edge hardware budgets, and regulatory WIM frameworks within a single coherent system blueprint. Domain adaptation strategies reported in peer-reviewed literature recover 20–50% of cross-dataset mAP degradation (typically 10–30%) caused by sensor and scene variability; these figures are literature benchmarks, not results obtained in this work. An optional weight-estimation module (Module 4) based on recent vision-based and bridge WIM validation studies is treated as an exploratory extension requiring site-specific validation.

Received: March 01, 2025 Revised: June 20, 2025 Accepted: August 14, 2025

Keywords: Thermal imaging; Vehicle detection; Multi-object tracking; IoT architecture; Intelligent transportation systems; Edge computing; YOLO; Domain adaptation; Weigh-in-motion; Smart cities

1. Introduction

Intelligent transportation systems (ITS) embedded in smart-city ecosystems need automated vehicle characterization—detection, classification, and tracking—that holds up across the full range of real-world operating conditions [1]. The primary sensor modalities in current practice each carry systemic weaknesses. Fixed weighing stations introduce throughput bottlenecks and delay commercial traffic; in-pavement sensors require expensive excavation, degrade under cyclic loading, and need periodic recalibration [1]. Visible-spectrum cameras—the most widely deployed alternative—suffer performance degradation estimated at 30–50% under low illumination, shadows, glare, or adverse weather [2]. These limitations create a genuine operational gap in environments where 24-hour reliability is non-negotiable.

1.1. Thermal Imaging as an IoT Sensor Modality

Longwave infrared (LWIR) imaging in the 8–14 μm atmospheric transmission window addresses these limitations at the physics level. Rather than relying on reflected light, LWIR cameras respond to self-emitted thermal radiation, enabling consistent operation without supplemental illumination [3]. At night under clear conditions, mAP loss is below 5%—a stark contrast to the 30–50% degradation typical of visible-band systems [2]. Vehicle components produce stable, discriminative signatures: engine bays typically reach 80–120°C, loaded tire sidewalls 40–70°C, and body panels range from ambient to 60°C under solar loading. When fused with YOLO-based detection on NVIDIA Jetson edge hardware, these properties support real-time processing within power budgets compatible with solar deployment [4].

1.2. The Integration Gap

Despite the technical maturity of individual building blocks—YOLO detectors [5, 6], multi-object tracking algorithms [7–10], and regulatory WIM frameworks [11]—practitioners planning thermal vehicle monitoring deployments encounter a specific and underserved gap. Published research overwhelmingly addresses algorithmic improvements to isolated components rather than end-to-end system architectures. Missing are: formalized inter-module data contracts, edge-to-cloud communication patterns with explicit latency budgets, quantified phase-gate criteria that bridge dataset experiments with field deployment, and integrated treatment of domain adaptation alongside GDPR compliance and hardware power budgets. This work addresses precisely that gap.

1.3. Research Questions and Contributions

Three interconnected research questions organize this work:

RQ1: How can thermal detection and tracking be organized into modular, interface-driven IoT systems with formalized data contracts and communication protocols for distributed edge-cloud deployments?

RQ2: How can development be staged into implementation phases with quantified progression criteria that bridge controlled dataset experiments with field validation?

RQ3: Which publicly available datasets, open-source tools, edge platforms, and recent algorithmic advances constitute sufficient infrastructure for practitioners to instantiate thermal vehicle monitoring systems?

Concretely, the paper delivers: (1) a five-module IoT architecture with formalized JSON/MQTT interface specifications and per-module latency budgets; (2) a four-phase implementation roadmap with numerical decision gates derived from published benchmarks; (3) a resource inventory of datasets, tools, and hardware platforms; and (4) a synthesis of thermal physics, detection, tracking, domain adaptation, and regulatory frameworks referenced throughout.

Contribution and scope: This is a systems-integration framework and structured survey, not a novel algorithm or validated prototype. The contribution lies in: synthesizing fragmented component-level research into a deployable blueprint; formalizing inter-module interfaces that prior work described only conceptually; deriving quantified roadmap criteria from published benchmarks rather than heuristic estimates; and providing the first region-specific thermal-ITS availability analysis linked to empirical weather degradation data. Optional weight estimation (Module 4) is treated as an exploratory future direction. All numeric values are cited to peer-reviewed publications, official documentation, or manufacturer datasheets.

2. Related Work and Novelty Positioning

Literature scope note. *Works surveyed here were identified through keyword searches on Google Scholar, IEEE Xplore, and Scopus using the terms "thermal vehicle detection," "LWIR IoT ITS," "weigh-in-motion vision," and "edge YOLO deployment" (date range 2015–2026), supplemented by backward citation tracing. No formal PRISMA protocol was applied; this work is a systems-integration framework, and literature coverage is representative rather than exhaustive.*

2.1. Thermal Detection and Tracking: Component-Level Advances

Thermal object detection for automotive applications has been systematically studied since the KAIST Multispectral benchmark was released in 2015 [12], establishing aligned thermal-visible annotation as a standard experimental setting. A key engineering line builds directly on the YOLO family for real-time inference: Ding et al. proposed TIR-YOLO-ADAS, an LWIR object detection framework tailored to ADAS constraints, demonstrating practical frame rates on embedded hardware [6]. More recent work has systematically benchmarked YOLOv8, v11, and v12 variants on thermal datasets and embedded platforms. YOLOv8n achieves mAP@0.5 of 64.2% on FLIR ADAS with 6.7 ms inference [5]. YOLOv12 introduces an attention-centric architecture achieving 1.64 ms inference latency on GPU [13]; thermal-specific mAP benchmarks for v11/v12 nano variants (68.9%–71.8%) require independent replication on FLIR ADAS V2 [14]. On the tracking side, ByteTrack [7], BoT-SORT [8], and StrongSORT [10] have each pushed state-of-the-art MOTA above 79% on MOT17/20; all three operate in real time and are openly licensed.

Domain adaptation for thermal imagery is a more recent focus. Uzun et al. [15] introduced SSTN, a self-supervised contrastive pretraining strategy that mitigates cross-dataset performance loss on FLIR/KAIST. Ustun et al. [16] demonstrated that Fourier Domain Adaptation can transfer spectral characteristics between domains while reducing annotation requirements. Wang et al. [17] proposed the first fully unsupervised thermal-to-visible adaptation method for pedestrian detection using multi-level feature alignment within Faster R-CNN.

2.2. System-Level ITS Architectures

Parallel work on IoT frameworks for smart transportation has established the importance of modular, protocol-standardized designs. Karthikeyan et al. [18] proposed a secured IoT-ITS framework for smart cities, identifying MQTT and TLS as the appropriate communication stack for vehicle surveillance nodes. Urblik et al. [19] demonstrated that a modular edge-processing architecture for industrial sensing can achieve sub-100 ms end-to-end latency while isolating failure domains across processing stages. These works validate the architectural pattern adopted here but do not address thermal imaging, domain shift, or WIM integration.

2.3. Where This Work Fits

No existing publication combines all of the following in a single framework: (a) LWIR sensor physics with NUC calibration and environmental degradation budgets; (b) multi-variant YOLO detection with TensorRT optimization on Jetson; (c) multi-object tracking with ByteTrack/BoT-SORT co-deployed at the edge; (d) domain adaptation integrated into a phased deployment roadmap; (e) GDPR Article 25 compliance architecture; (f) optional BWIM/vision-based WIM with COST 323 regulatory pathways; and (g) region-specific uptime analysis linked to empirical weather data. Table 1 positions the present work against the most closely related efforts.

Table 1: Positioning against closely related work (✓ = addressed, ○ = partially, – = not addressed)

Aspect	[18]	[19]	[6]	[15]	[20]	**This work**
LWIR sensor + calibration	–	–	✓	○	✓	✓
YOLO detection (v8/v11/v12)	–	–	○	–	–	✓
Multi-object tracking	–	–	–	–	–	✓
Domain adaptation	–	–	–	✓	–	✓
Edge deployment + latency budget	○	✓	○	–	–	✓
GDPR / IoT security	✓	–	–	–	–	✓
WIM / COST 323 pathway	–	–	–	–	✓	✓
Phased roadmap + decision gates	–	–	–	–	–	✓
Regional availability analysis	–	–	–	–	–	✓

3. Technical Foundations

3.1. Thermal Imaging Physics and Calibration

LWIR systems detect radiant exitance E governed by the Stefan–Boltzmann law [3]:

$$E = \varepsilon \sigma T^4 \tag{1}$$

where $\varepsilon \in [0, 1]$ is surface emissivity, $\sigma = 5.67 \times 10^{-8}\ \mathrm{W\,m^{-2}\,K^{-4}}$, and T is absolute temperature (K). The 8–14 μm window offers atmospheric transmission above 80% for distances relevant to highway monitoring [3]. Uncooled microbolometer detectors in this band achieve noise-equivalent temperature differences (NETD) below 50 mK in current commercial products, sufficient for reliable vehicle-to-background contrast even under mild thermal conditions.

Sensor non-uniformity is the dominant calibration challenge. Fixed-pattern noise in uncooled arrays accumulates with operating temperature and time, degrading detection contrast if left uncorrected. Periodic non-uniformity correction (NUC)—closing the sensor shutter and using the uniform temperature of the shutter plate as a reference—restores flat-field response [21]. For fixed roadside microbolometer deployments, shutter-triggered NUC every 8–12 hours is sufficient under stable ambient conditions; more frequent correction (every 2–4 hours) is advisable when diurnal temperature swings exceed 15°C, as pixel offset voltages and responsivity both drift with FPA temperature [21].

3.2. Deep Learning Object Detection

Single-stage YOLO detectors have become the standard for real-time thermal vehicle detection, delivering competitive mAP with inference times compatible with 30 Hz operation [5,6]. Table 2 presents published benchmark results for YOLO variants on LWIR datasets.

Table 2: YOLO Detector Performance on Thermal Vehicle Datasets (Published Benchmarks). Inference times on GPU (T4/A100); Jetson latency is higher — see Section 6.

Model	**mAP@0.5**	**Infer. (ms)**	**Params (M)**	**Source**
YOLOv5n	54.3%	7.1	1.9	[5]
YOLOv8n	64.2%	6.7	3.2	[5]
YOLOv11n	68.9%[†]	2.4	2.6	[5, 14]
YOLOv12n	71.8%[†]	1.64	2.3	[13, 14]
YOLOv9n	N/A[‡]	–	1.9	—
YOLOv10n	N/A[‡]	–	2.3	—

[†] Sourced from a Zenodo-deposited non-ISI report [14]; source abstract reports YOLOv8 as achieving highest accuracy, creating uncertainty about these figures. Independent replication on FLIR ADAS V2 with a peer-reviewed source is required before use as primary benchmarks.

[‡] No published LWIR vehicle mAP benchmarks identified; include in Phase 1 comparative evaluations.

Benchmark caveat: Values in Table 2 vary with dataset splits, GPU model, driver versions, and evaluation protocol. Practitioners must replicate training and evaluation on target datasets and deployment hardware.

3.3. Domain Shift and Adaptation Strategies

Transferring a detector trained on FLIR ADAS to a different deployment environment (e.g., KAIST or a site-specific dataset) without retraining typically degrades mAP by 10–30%, driven by differences in camera mounting geometry, scene composition, and ambient thermal conditions [15]. Four adaptation strategies have been validated in peer-reviewed work:

Four adaptation strategies have been validated in peer-reviewed work and are directly applicable to the proposed system. SSTN [15] applies self-supervised contrastive pretraining on unlabelled target-domain LWIR data followed by supervised fine-tuning on source labels, recovering approximately 20–50% of the initial domain-shift performance loss on FLIR/KAIST cross-domain evaluation. Spectral Transfer Guided Active Domain Adaptation [16] employs Fourier Domain Adaptation (FDA) to transfer spectral statistics from a labelled visible domain to an unlabelled thermal domain, substantially reducing annotation effort while maintaining detection performance. The first fully unsupervised thermal-to-visible adaptation method [17] uses multi-level feature-distribution alignment within a Faster R-CNN backbone and has been evaluated on KAIST with results transferable to vehicle detection settings. Finally, Meta-UDA [22] provides an algorithm-agnostic online meta-learning framework that improves existing UDA baselines by meta-learning the initial detector condition, enabling fast target-domain adaptation without overfitting; it achieves state-of-the-art unsupervised domain adaptation on the KAIST and DSIAC thermal benchmarks.

Practitioners should budget for cross-dataset evaluation experiments in Phase 1 and site-specific fine-tuning in Phase 4 rather than assuming direct transferability from published benchmark numbers.

Practitioners should budget for cross-dataset evaluation experiments in Phase 1 and site-specific fine-tuning in Phase 4 rather than assuming direct transferability from published benchmark numbers.

3.4. Multi-Object Tracking

Multi-object tracking (MOT) assigns consistent identities to detected objects across frames. The standard pipeline chains Kalman filter motion prediction, deep appearance feature extraction, and Hungarian-algorithm assignment [9]. Table 3 reports published MOT17/20 test-set results.

Table 3: Multi-Object Tracking Performance (MOT17 Test Set, Published Benchmarks). FPS reported on original benchmark hardware; note MOT17 uses visible-spectrum data—thermal-specific validation constitutes an open research gap.

Algorithm	**MOTA (%)**	**IDF1 (%)**	**FPS**	**Source**
DeepSORT	61.4	62.3	61	[9]
ByteTrack	80.3	77.3	30	[7]
BoT-SORT	80.5	80.2	22	[8]
StrongSORT	79.6	79.5	18	[10]

Metric definitions. MOTA (Multiple Object Tracking Accuracy) accounts for missed detections, false positives, and identity switches; IDF1 (Identity F1 Score) measures identity-consistency across the full sequence. Both are adopted as primary metrics throughout this roadmap, consistent with standard MOT benchmark practice. ByteTrack's approach of associating *every* detection box—not only high-confidence ones—is particularly suited to edge deployment because it maintains tracking through partial occlusions without appearance feature computation, keeping per-frame tracking latency at 2–5 ms [7].

3.5. Vision-Based Weight Estimation

Disclaimer: weight estimation is not a core validated component of the proposed architecture. The following documents recent feasibility evidence for practitioners who require this functionality.

Two peer-reviewed studies establish proof-of-concept for non-contact WIM using computer vision:

- **Computer-vision WIM** [23]: Tire contact-patch geometry (contact length ℓ_c, vertical deflection δ) extracted from visible imagery, combined with measured tire inflation pressure P_{tire}, yields axle force:

$$F = P_{\text{tire}} \cdot A_{\text{contact}}(\ell_c, \delta) \tag{2}$$

  Field validation on SUVs and concrete trucks achieved weight estimation within 5% of certified static scales [23].

- **Multispectral vision-based WIM** [20]: A thermal camera measures tire-roadway contact parameters; a co-aligned visible camera reads sidewall markings for tire specification lookup. An XGBoost model trained on 1,127 samples across 19 tire types (Bayesian-optimised, 11-feature importance analysis) achieved a maximum 5% deviation from traditional WIM systems in static and low-speed field trials [20].

Operational deployment of either method requires: a tire database covering $\geq$ 100 models for fleet-wide generalization, TPMS integration or manual inflation data, a calibrated camera geometry with known extrinsic parameters, and ML recalibration when tire wear states change significantly.

3.6. WIM Regulatory Framework

European WIM systems are standardized under COST 323 [11], which specifies the accuracy classes in Table 4.

Table 4: COST 323 Weigh-in-Motion Accuracy Classes [11]

Class	GVW Tolerance	Axle Tolerance	Application
A(5)	±5%	±8%	Law enforcement
B(10)	±10%	±11%	Traffic monitoring
C(15)	±15%	±15%	Statistical analysis
D(25)	±25%	±25%	Feasibility studies

Bridge WIM (BWIM) systems using strain gauges or accelerometers instrumented into a bridge structure have reported ≈91% GVW accuracy and ≈87% axle weight accuracy under controlled conditions [24]. Formal COST 323 certification typically spans one to several years, requiring traceable reference vehicles, standardized testing, and independent third-party verification [11].

3.7. Environmental Sensitivity

Despite its night-time advantage, LWIR imaging degrades under specific adverse weather conditions. Table 5 documents degradation from published measurements [2].

Table 5: Thermal Detection Performance Under Adverse Weather [2]

Condition	mAP Loss	Primary Mechanism
Light rain (1–3 mm/hr)	10–20%	Lens contamination
Heavy rain (>10 mm/hr)	>60%	IR attenuation by water droplets
Light fog (visibility >200 m)	15–25%	Droplet scattering
Dense fog (visibility <50 m)	>60%	Strong IR absorption
Night, clear conditions	<5%	LWIR advantage over visible
Extreme cold (< −20°C)	<10%	Sensor non-uniformity drift [21]

Operational deployments must therefore co-deploy a weather station and implement adaptive confidence scoring (flagging detections as unreliable under dense fog or heavy rain), plus schedule weekly lens cleaning in dusty or rainy environments.

4. Modular IoT System Architecture

4.1. Five-Module Architecture Overview

The proposed system comprises five functionally independent modules with explicit data contracts and communication protocols. This decoupling is deliberate: each module can be upgraded, replaced, or scaled independently without perturbing the rest of the pipeline, which is a key requirement for long-lived IoT deployments [19]. Figure 1 illustrates the architecture, data flows, and edge-cloud distribution.

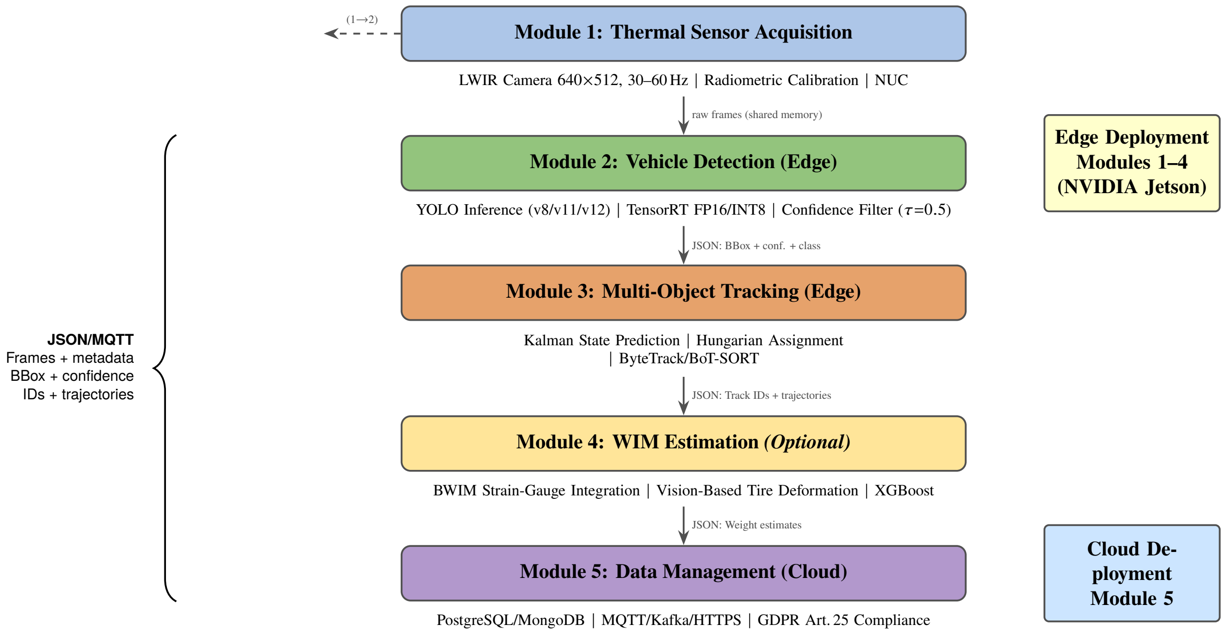

Figure 1: Modular IoT architecture for thermal vehicle monitoring.

The five modules are defined as follows:

1. **Module 1 – Thermal Sensor Acquisition:** Ingests LWIR frames (640×512 px, 30–60 Hz) from a microbolometer camera, applies radiometric calibration, triggers periodic NUC every 8–12 hours [21], buffers frames with ISO 8601 nanosecond timestamps, and monitors sensor-health diagnostics (frame-drop rate, NETD drift) [3].
2. **Module 2 – Vehicle Detection (Edge):** Executes DNN inference using a selected YOLO variant (YOLOv8n/v11n recommended for edge; see Section 6), optimized with TensorRT FP16 or INT8 quantization for Jetson deployment. Detections below a confidence threshold $\tau = 0.5$ are discarded; remaining detections are published as JSON bounding-box messages. Per-frame latency: 1.64–6.7 ms (Table 2) [5].
3. **Module 3 – Multi-Object Tracking (Edge):** Receives detection bounding boxes; initializes new tracks using Kalman state vectors $[x, y, w, h, \dot{x}, \dot{y}]$; predicts track states for unmatched frames; associates detections to tracks via the Hungarian algorithm using IoU cost matrix; updates appearance features (optional, BoT-SORT/StrongSORT) via EMA; terminates tracks missing for $> T_{\text{lost}} = 30$ frames. Per-frame latency: 2–5 ms [7, 8].
4. **Module 4 – Weight Estimation (Optional):** Receives track trajectories; estimates axle loads either via (a) BWIM: strain-gauge signals fed through influence-line inversion [24]; or (b) vision-based tire deformation: thermal contact-patch geometry combined with XGBoost model [20]. Both sub-paths are treated as exploratory pending comprehensive field validation. May be bypassed entirely.
5. **Module 5 – Data Management (Cloud):** Archives trajectory records in PostgreSQL (structured queries) or MongoDB (flexible schema); aggregates traffic-flow statistics (counts, speeds, headways); exports via MQTT/Kafka/HTTPS; generates compliance and analytics reports; enforces GDPR Article 25 data-minimization policy (trajectory statistics not raw imagery, retention $\leq$ 30 days, hashed vehicle identifiers) [25].

4.2. Formalized Interface Specifications

Three communication patterns apply at different pipeline stages [18, 19]:

Shared Memory (Modules 1→2→3, co-located on Jetson): POSIX shared memory or memory-mapped files achieve zero-copy data transfer with sub-millisecond handoff latency, avoiding serialization overhead on the critical detection path.

Message Queue (Modules 3→4→5 and edge→cloud): MQTT (lightweight, QoS 1 for at-least-once delivery) or Apache Kafka (high throughput, replay) provides asynchronous publish-subscribe decoupling with buffering during network interruptions. Recommended MQTT topic hierarchy:

```
/thermal/{siteId}/frame       (M1->M2, local only)
/detection/{siteId}/bbox      (M2->M3)
/tracking/{siteId}/tracks     (M3->M4/M5)
/wim/{siteId}/weights         (M4->M5, optional)
```

HTTPS/REST (configuration and queries): TLS 1.3 (AES-256-GCM) with mTLS device authentication for synchronous configuration interfaces, manual queries, and external system integration [18].

Example JSON data contract (Module 2 output): The contract below illustrates a frame containing two simultaneously detected vehicles, with inline annotations explaining each field's role in the pipeline.

```
{
  "schema_version": "1.0",          // bumped on breaking interface changes
  "timestamp": "2026-02-05T15:30:45.123456789Z", // GPS-disciplined, nanosecond
  "frameId": 1234567,               // monotonic; detects dropped frames
  "siteId": "muscat_site_01",       // multi-site routing key
  "detections": [
    {
      "bbox": [142, 87, 310, 205],  // [x_min,y_min,x_max,y_max] in pixels
      "confidence": 0.87,           // YOLO score; threshold tau = 0.5
      "class": "car",              // FLIR ADAS V2 label set
      "trackId": null,             // null here; Module 3 assigns track ID
      "thermalIntensity": 24531    // mean 14-bit ADU in bbox
    },
    {
      "bbox": [380, 210, 590, 420],
      "confidence": 0.73,
      "class": "truck",
      "trackId": null,
      "thermalIntensity": 31802    // higher ADU: warm engine bay (truck)
    }
  ]
}
```

The `schema_version` field allows receivers to reject incompatible contract versions and is incremented only on breaking interface changes. The nanosecond-precision GPS timestamp provides the absolute time reference required for retrospective synchronisation with external WIM or reference instruments during Phase 4 field validation. `thermalIntensity` — the mean 14-bit analogue-to-digital unit (ADU) within the bounding box — is a lightweight thermal-energy proxy: values above $\approx$ 28,000 ADU are consistent with engine-bay temperatures above 80°C and can support coarse load-class inference before Module 3 assigns persistent track identities. Including two detections in the example makes explicit that the `detections` array is multi-valued and that `trackId` remains `null` at the detection stage, being populated only after Module 3 runs the Hungarian assignment.

Timing synchronization: GPS-disciplined clocks (accuracy $< 1\,\mu$s) at Module 1 provide absolute time references for retrospective synchronization with external WIM or reference measurement systems during field validation. IEEE 1588 Precision Time Protocol (PTP) may substitute GPS timing in infrastructure-connected deployments where sub-microsecond synchronization between edge nodes is required.

4.3. Performance Budget and Resource Estimation

For 30 FPS operation the per-frame processing budget is 33.3 ms. Table 6 distributes this budget based on published benchmarks.

Table 6: Estimated Per-Frame Latency Budget (Based on Published Benchmarks). *All values extrapolated from server-GPU (T4/A100) benchmarks and manufacturer datasheets — not measured on a physical Jetson Xavier NX. Actual Jetson TensorRT FP16 inference is typically 2–5× slower than T4 GPU figures. This table is a planning estimate only; practitioners must profile the full pipeline on target hardware before committing to the 30 FPS latency budget.*

Stage	**Module**	**Latency (ms)**	**Basis**
Frame capture + transfer	M1	2–5	Camera spec + USB/GigE
YOLO inference (YOLOv8n)	M2	6.7	[5]
NMS post-processing	M2	0.5–1.0	Estimated
Kalman predict + Hungarian	M3	2–5	[7]
MQTT local publish	M3→M5	1–2	Local broker
Edge pipeline total		**12–20**	≤ 33.3 ms budget

Bandwidth: Uncompressed thermal stream (640 × 512, 16-bit, 30 Hz):

$$640 \times 512 \times 2\,\mathrm{B} \times 30\,\mathrm{fps} = 19.7\,\mathrm{MB/s} \approx 157\,\mathrm{Mbps} \tag{3}$$

H.264 compression at a 20:1 ratio reduces this to ≈ 8 Mbps. Cloud transmission requires only detection/tracking JSON (≈ 1 KB/frame ≈ 0.24 Mbps), keeping cloud bandwidth below 1 Mbps.

Power budget: Jetson Xavier NX: 10–15 W [4]; LWIR camera (FLIR Boson-class uncooled): ≈ 3–5 W; 4G/5G modem: 2–3 W. System total: **15–23 W**. For 24-hour solar operation in Muscat, Oman (≈ 6 peak sun hours): ≈ 150 W panel, ≈ 600 Wh battery.

4.4. Security and Privacy Architecture

Roadside IoT deployments capturing vehicle imagery require a layered security architecture aligned with GDPR Article 25 (Data Protection by Design) [18, 25]:

Transport security relies on TLS 1.3 (AES-256-GCM) across all MQTT and HTTPS channels, with mutual TLS (mTLS) for Jetson device authentication and PKI certificate lifecycle management. Data minimisation stores only trajectory statistics rather than raw frames, enforces a ≤ 30-day retention policy, hashes vehicle identifiers, and blurs any captured licence plates.

Access control follows a role-based (RBAC) least-privilege model with ISO 27001-aligned audit logging and multi-factor authentication for administrative interfaces. Edge device hardening includes secure boot with verified firmware, physical tamper detection via an onboard accelerometer, signature-verified over-the-air (OTA) updates, and a regular CVE-patching schedule. For EU and GCC deployments, GDPR Article 35 requires a formal Privacy Impact Assessment (PIA) documenting data flows, lawful processing basis, retention policy, and data-subject rights [25].

5. Four-Phase Staged Implementation Roadmap

Figure 2 shows the four-phase roadmap. Every phase has explicit quantified progression criteria; failure at a decision gate triggers an iteration loop rather than premature advancement.

5.1. Decision Gate Threshold Derivation

The four decision gates use quantified numerical thresholds to prevent premature phase advancement. This subsection derives every threshold explicitly from published benchmark data, so that practitioners can audit and adjust them for site-specific requirements. Table 7 summarises the derivation; detailed reasoning follows.

Gate 1 — Detection (mAP ≥ 60%)

The published benchmark for YOLOv8n on FLIR ADAS V2 is 64.2% mAP@0.5 [5]. The gate is set to 60% (4.2 pp below the reference) for three compounding reasons. First, published YOLO benchmarks use controlled infrastructure (NVIDIA A100/RTX 3090) with fixed random seeds; independent re-runs typically exhibit ±2–3 pp variance due to stochastic data shuffling and augmentation. Second, FLIR ADAS V2 annotation revisions across releases can shift mAP by ±1–2 pp without any model change. Third, single-GPU consumer hardware (e.g., RTX 3080) may underperform the benchmark by ≈ 1 pp due to reduced batch-normalisation stability at lower batch sizes. A 60% floor represents approximately the 5th percentile of statistically expected replication outcomes.

Gate 1 — Tracking (MOTA ≥ 70%)

ByteTrack achieves MOTA 80.3% on MOT17 [7], a visible-spectrum pedestrian benchmark. Two thermal-domain differences degrade tracking: (1) lower inter-class texture contrast increases false-positive track initialisations;

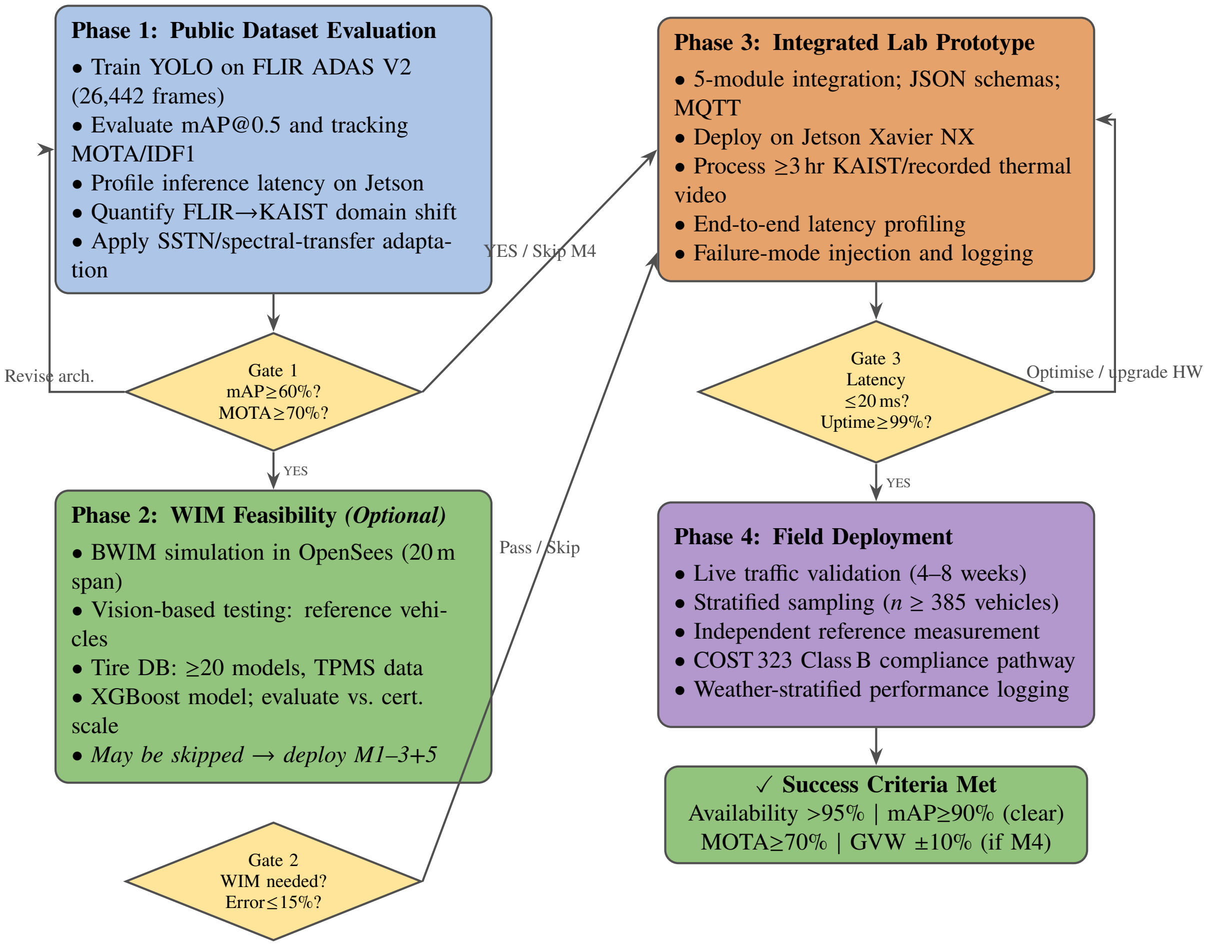

Figure 2: Four-phase staged implementation roadmap. Yellow diamonds are decision gates with quantified numerical criteria (Section 5). Phase 2 (WIM feasibility) is optional: practitioners may advance directly from Gate 1 to Phase 3 deploying only the core detection/tracking subsystem (Modules 1–3+5). Contingency return arrows prevent premature advancement when gate criteria are not met.

Table 7: Decision Gate Threshold Derivation from Published Benchmarks

Gate	Criterion	Benchmark Anchor	Derivation Rationale
G1	mAP@0.5 ≥ 60%	YOLOv8n: 64.2% on FLIR ADAS [5]	4.2 pp below published value; accommodates split-variant and hardware-version differences (≈ ±5% replication variance across YOLO re-runs)
G1	MOTA ≥ 70%	ByteTrack: 80.3% on MOT17 [7]	Estimated 12.5% cross-modality degradation (visible→thermal) based on detection domain-shift studies [15]: 80.3% × 0.875 ≈ 70%. *Caveat: domain-shift literature quantifies mAP (detection) degradation, not MOTA (tracking); applying this factor to MOTA is a heuristic approximation without direct empirical grounding. Treat the 70% gate as provisional and calibrate from site-specific thermal tracking experiments in Phase 1.*
G1	IDF1 ≥ 65%	ByteTrack: 77.3% IDF1 [7]	Same 12.5% degradation: 77.3% × 0.875 ≈ 67.6%; floor rounded to 65%
G1	Latency ≤ 15 ms	33.3 ms budget at 30 FPS	YOLO (6.7) + tracking (2–5) + MQTT (1–2) = 9.7–13.7 ms; 15 ms provides ≈10% OS-jitter headroom
G1	Recovery $r \geq 0.30$	SSTN: 20–50% [15]	Conservative lower bound of the published SSTN range
G2	GVW error ≤ 15%	COST 323 Class C [11]	Minimum class for statistical analysis; feasibility threshold before committing field hardware
G2	Vision error ≤ 10%	COST 323 Class B [11]	Operational traffic-monitoring class; consistent with Gong et al. [20] field results
G3	Latency ≤ 20 ms	33.3 ms budget [5]	60% budget utilisation; leaves 13.3 ms for Module 4 and Jetson JetPack 6.x OS jitter (≈ 2–5 ms)
G3	Uptime ≥ 99%	ISO 25010	≤ 3.65 days unplanned downtime/year; standard for continuous ITS infrastructure
G3	GPU ≤ 80%, RAM ≤ 6 GB	Jetson Xavier NX: 8 GB [4]	20% GPU and 25% RAM headroom for NUC events, MQTT burst buffering, and OTA updates
G4	Detection ≥ 95% (clear)	ITS counting standards [18]	<5% missed-detection rate for compliant traffic counts
G4	MOTA ≥ 70%, IDF1 ≥ 65%	Same as G1	G1 thresholds re-applied to field data; domain shift may require site-specific re-tuning
G4	GVW ±10% (Module 4)	COST 323 Class B [11]	Operational traffic-monitoring class [20]
G4	Availability ≥ 98%	Muscat weather [2]	Upper bound of arid-climate estimate; relax to 95% for wet climates (London, Mumbai)

Table 8: Proposed Phased Roadmap vs. Existing ITS Deployment Methodologies (✓ = addressed, ○ = partially, – = not addressed)

Feature	[18]	[19]	[6]	This work
Quantified phase-gate criteria	–	–	–	✓
Benchmark-grounded thresholds	–	○	–	✓
Cross-domain performance budget	–	–	–	✓
Edge hardware power budget	○	✓	○	✓
GDPR/privacy compliance pathway	✓	–	–	✓
WIM regulatory pathway (COST 323)	–	–	–	✓
Regional system availability est.	–	–	–	✓
Failure-mode injection protocol	–	○	–	✓

and (2) more homogeneous LWIR vehicle appearance degrades re-identification under occlusion. Domain-shift studies report 10–30% loss transferring from visible to thermal modality [15]. Applying a conservative 12.5% degradation: $80.3\% \times (1 - 0.125) \approx 70.3\%$, rounded down to 70% as the gate floor. IDF1 follows identically: $77.3\% \times 0.875 \approx 67.6\%$, gated at 65%.

Gate 3 — Latency (≤ 20 ms) and Uptime

The Gate 3 latency of 20 ms is more permissive than Gate 1's 15 ms because Phase 3 measures the *full integrated pipeline* including MQTT serialisation, database write, and Python IPC overhead not present in the Phase 1 GPU benchmark. At 20 ms, the pipeline consumes 60% of the 33.3 ms frame budget, leaving 13.3 ms for optional Module 4 and OS scheduling. Uptime $\geq 99\%$ equals ≤ 3.65 days unplanned downtime per year, the standard continuous-availability threshold under ISO 25010 for safety-critical infrastructure.

Gate 4 — Field Success Criteria

The 95% detection rate under clear conditions corresponds to a <5% missed-detection rate, consistent with ITS counting accuracy standards [18]. The GVW ±10% maps to COST 323 Class B, as required for traffic monitoring [11]. The availability threshold of $\geq 98\%$ is calibrated to the Muscat deployment context (estimated 95–98% usable uptime from historical rainfall data [2]); practitioners deploying in higher-precipitation regions should revise this threshold downward accordingly.

5.2. Phase 1: Public Dataset Evaluation — Detailed Methodology

Objective: Establish quantified detection and tracking baselines, characterize cross-dataset domain shift, and profile Jetson inference latency [12, 15, 26].

Step 1 — Dataset Preparation

Download and verify FLIR ADAS V2 (26,442 frames, 640×512 px thermal, 15 object categories, $\approx 520{,}000$ bounding boxes) [26]. Apply the official 80/20 train/validation split. Convert annotations to YOLO format (`.txt` per image, normalised $[cx, cy, w, h]$). Register the KAIST Multispectral dataset [12] (95,000 aligned thermal-visible pairs) for cross-dataset evaluation only—do *not* include KAIST in training.

Step 2 — Detector Training

Train YOLOv8n and YOLOv11n using the following fixed hyperparameters to ensure reproducibility:

- Optimizer: Adam, $\beta_1 = 0.937$, $\beta_2 = 0.999$, weight decay 5×10^{-4}
- Learning rate: initial $\eta_0 = 0.01$, cosine annealing to $\eta_f = 0.001$
- Batch size: 32; epochs: 100 with early stopping (patience = 20)
- **Augmentation** (thermal-specific, single-channel LWIR):
    - Mosaic ($p = 1.0$): four-image random crop and tile; preserves spatial context and handles small-object occlusion.
    - Random horizontal flip ($p = 0.5$): geometry augmentation, modality-independent.
    - Scale ([0.5, 1.5]), translate (0.1): affine jitter for perspective variation.
    - **Radiometric intensity jitter**: uniform brightness shift ±15% of the 14-bit ADU range, simulating inter-scene emissivity and ambient temperature variation.
    - **Additive Gaussian noise**: $\mathcal{N}(0, \sigma^2)$, $\sigma \in [0, 0.05 \times \text{pixel\_range}]$, approximating sensor NETD variation.
    - **Random contrast scaling**: multiply pixel values by $\alpha \sim \mathcal{U}(0.85, 1.15)$, simulating diurnal thermal-contrast variation.
    - **HSV hue/saturation jitter is suppressed.** LWIR images are single-channel (monochromatic radiance maps); hue (h) and saturation (s) carry no physical meaning in this modality. Applying HSV jitter to a pseudo-colourised LWIR image would corrupt the radiometric integrity of the training data without benefit. Only value (v) / brightness augmentation is retained, subsumed into the radiometric intensity jitter above.
- Input resolution: 640×640 (padded with grey border; letterboxing)
- Hardware: NVIDIA GPU for training; evaluate on Jetson Xavier NX (TensorRT FP16)

Step 3 — Detection Evaluation

Report precision, recall, F1, mAP@0.5, mAP@0.75, and mAP@0.5:0.95 on the FLIR validation split (IoU threshold 0.5 unless stated; confidence threshold $\tau = 0.001$ for P/R curve, $\tau = 0.5$ for operational metrics). Profile per-frame inference time on Jetson Xavier NX (10 warm-up iterations; mean over 500 frames).

Step 4 — Tracking Evaluation

Apply ByteTrack and BoT-SORT on the FLIR video subset. ByteTrack parameters: detection threshold $\theta_{\text{high}} = 0.5$, $\theta_{\text{low}} = 0.1$ (for low-confidence recovery); IoU-matching thresholds $\lambda_1 = 0.9$ (first association), $\lambda_2 = 0.5$ (second association); track buffer $T_{\text{lost}} = 30$ frames [7]. BoT-SORT: additionally use camera-motion compensation (CMC-ECC) [8]. Evaluate with MOTA, MOTP, IDF1, HOTA (if ground-truth ID annotations available).

Step 5 — Domain Shift Quantification

Evaluate the FLIR-trained model on KAIST thermal sequences *without retraining*. Record mAP degradation Δ_{mAP}. Apply SSTN adaptation [15]: (i) unsupervised contrastive pretraining on unlabelled KAIST thermal frames (ResNet-101 backbone, MoCo-v2 pretraining protocol, 200 epochs); (ii) supervised fine-tuning on FLIR labels (50 epochs, $\eta_0 = 0.001$). Record post-adaptation mAP; compute recovery fraction $r = \Delta_{\text{mAP,post}}/\Delta_{\text{mAP}} \in [0, 1]$.

Decision Gate 1 Criteria

- mAP@0.5 $\geq$ 60% on FLIR validation set.
- MOTA $\geq$ 70%, IDF1 $\geq$ 65% on FLIR video sequences.
- Per-frame total latency (detection + tracking) $\leq$ 15 ms on Jetson Xavier NX (TensorRT FP16).
- Domain-shift recovery fraction $r \geq 0.30$.

If not met: increase model size (YOLOv8m), add training epochs, upgrade to Jetson AGX Xavier, or apply stronger augmentation (CutMix, Mosaic-9).

5.3. Phase 2: Optional WIM Feasibility Assessment — Detailed Methodology

Objective: Determine whether weight estimation is feasible and operationally required before committing field resources. *This phase is optional; skip to Phase 3 if weight data are unnecessary.*

Option A — BWIM Pathway

1. Model a representative bridge (e.g., 20 m simply-supported concrete beam) in OpenSees [27]: Young's modulus $E = 30$ GPa, moment of inertia $I = 0.2$ m^4, span $L = 20$ m.
2. Compute influence lines $\eta(x)$ relating vehicle position x to midspan strain ε via: $\varepsilon(x) = \frac{F \cdot \eta(x)}{EI}$.
3. Simulate 200 vehicle passages with randomised weight (5–40 t GVW), speed (60–100 km/h), and 3-axle configuration; add Gaussian sensor noise ($\sigma = 50\,\mu\varepsilon$).
4. Apply Moses algorithm [24] for GVW estimation; evaluate root-mean-square error.

Option B — Vision-Based Tire Pathway

1. Acquire thermal video of ≥ 10 reference vehicles with certified static weights (certified scale accuracy ±2%), spanning GVW 1.5–40 t.
2. Extract tire contact features per frame using Module 2 bounding boxes as region-of-interest: contact length ℓ_c (horizontal extent of ground-contact patch), deflection proxy δ (ratio of loaded to unloaded tire height), sidewall temperature T_{sw} from pixel intensity [20].
3. Collect tire metadata: brand, model, size, load index, inflation pressure (TPMS or manual gauge).
4. Build a tire database of 20–30 models; train XGBoost (features: ℓ_c, δ, T_{sw}, tire load index, inflation pressure, vehicle speed from Module 3; 11-feature set following [20]). Optimise hyperparameters via Bayesian search (50 iterations; $n_{\text{estimators}} \in [50, 500]$, max_depth $\in [3, 10]$).
5. Evaluate on held-out reference vehicles (20% split); report mean absolute percentage error (MAPE) vs. certified weights.

Decision Gate 2 Criteria

- BWIM simulation: GVW MAPE $\leq$ 15% (COST 323 Class C) on simulated passages with realistic noise.
- Vision-based: GVW MAPE $\leq$ 10% on reference vehicles.
- Weight module latency: $\leq$ 5 ms additional per frame.
- Operational decision: if criteria not met or WIM not required, proceed to Phase 3 with Modules 1–3+5 only.

5.4. Phase 3: Integrated System Prototype — Detailed Methodology

Objective: Validate the integrated pipeline—Modules 1–5 or 1–3+5—on offline recorded thermal video in a controlled (non-field) setting.

Hardware Setup

- Edge node: NVIDIA Jetson Xavier NX (8 GB, JetPack 6.x, TensorRT 8.x) [4].
- LWIR camera: FLIR Boson 640 or equivalent (640 × 512, 30 Hz, NETD < 50 mK, USB 3.0 or GigE interface).
- Cloud node: Ubuntu 22.04 server; PostgreSQL 15; Mosquitto MQTT broker (TLS 1.3 enabled); containerised via Docker Compose.

Software Integration

1. Implement Module 1 in Python (PyAV or V4L2 capture); publish frames via POSIX shared memory ring buffer (32 slots, 640 × 512 × 2 B each).
2. Implement Module 2 as a separate process: read from shared memory, run TensorRT engine, publish JSON detection messages to MQTT topic `/detection/{siteId}/bbox`.
3. Implement Module 3 as a separate process: subscribe to detection messages, run ByteTrack (pure Python implementation, `git clone https://github.com/FoundationVision/ByteTrack`), publish track messages to `/tracking/{siteId}/tracks`.
4. Implement Module 5: subscribe to tracking topics; insert records into PostgreSQL (`tracks` table: timestamp, track_id, bbox, class, confidence, site_id); expose REST API for queries.
5. Validate JSON schema compliance at each inter-module boundary using `jsonschema` library (Python); log schema violations as errors.

Evaluation Protocol

1. Process ≥ 3 hours of pre-recorded continuous thermal traffic video end-to-end. The recommended source is the KAIST Multispectral dataset [12], which provides genuine 20 Hz temporal sequences (640 × 480, aligned thermal-visible) suitable for tracking evaluation. FLIR ADAS V2 [26] is a still-image dataset and is *not* a video source; practitioners may supplement with site-recorded footage from a FLIR Boson 640 or equivalent camera capturing diverse traffic, day/night conditions, and at least two weather states.
2. Measure per-module latency using Jetson power monitoring API (`tegrastats`); record GPU utilisation, RAM usage, system power.
3. Inject failure scenarios: (a) MQTT broker restart (expect buffered messages to be replayed); (b) simulated frame drops (5% random); (c) forced NUC trigger (expect ≤ 500 ms recovery time).
4. Report mAP@0.5 and MOTA/IDF1 on the continuous sequence.

Decision Gate 3 Criteria

- Detection mAP@0.5 ≥ 60% on KAIST thermal sequences or equivalent continuous video (see Evaluation Protocol above).
- Tracking MOTA ≥ 70%, IDF1 ≥ 65%, identity-switch rate ≤ 5%.
- End-to-end pipeline latency ≤ 20 ms (measured 95th percentile over the ≥ 3-hour run).
- System uptime ≥ 99% over a 24-hour soak test.
- GPU utilisation ≤ 80%; memory ≤ 6 GB; system power ≤ 20 W.

If not met: apply TensorRT INT8 quantization (calibration set = 500 FLIR training frames; expected 2–3× latency reduction, ≤ 3% mAP loss [4]); reduce frame rate to 15 FPS; upgrade to Jetson AGX Xavier.

5.5. Phase 4: Field Deployment and Validation

Objective: Deploy at an operational road site with live traffic, independent ground-truth capability, and rigorous statistical validation [11, 24].

Site Requirements and Installation

- Site: highway overpass, toll plaza, or weigh station with ≥ 500 vehicles/day and vehicle-class diversity (at least four categories).
- Permits: thermal imaging authorization; GDPR Article 35 PIA [25]; road-authority installation permit.

- Camera mounting: height 5–8 m, depression angle 15–30°, field of view covering all lanes; IP67-rated weather-proof enclosure; lens heater ($> 0°C$); vibration-isolation mount.
- Power: grid primary with UPS backup; optional 150 W solar panel + 600 Wh battery for autonomous operation.
- Reference instrument: certified static scale (±2%) or instrumented BWIM bridge (if pursuing Module 4 validation).
- Environmental monitoring: co-located weather station logging temperature, humidity, precipitation, and visibility (for correlation with Table 5 degradation profiles).

Validation Methodology

Power analysis for proportions (95% confidence, 5% margin of error, $p = 0.5$, conservative maximum-variance estimate, and $e = 0.05$) yields the Cochran [28] minimum sample size:

$$n_0 = \frac{z_{\alpha/2}^2 \, p\,(1-p)}{e^2} = \frac{(1.96)^2 \times 0.5 \times 0.5}{(0.05)^2} = 384.16 \approx 385 \tag{4}$$

yielding $n_0 = 385$ vehicles as the minimum stratified sample per subgroup. Apply stratified sampling across five vehicle classes, three weather conditions, and day/night periods. Reference weight measurements use vehicles weighed on the certified scale within one hour of thermal capture. Minimum campaign: four weeks of 24/7 operation.

Quantified Success Criteria

- **Detection:** ≥ 95% vehicles in clear weather, ≥ 90% in light rain, ≥ 70% in fog (Table 5).
- **Tracking:** MOTA ≥ 70%, IDF1 ≥ 65% on field data.
- **Weight accuracy (Module 4):** GVW ±10%, axle ±11% on $n \geq 385$ test vehicles at 95% confidence (COST 323 Class B) [11].
- **Availability:** ≥ 98% uptime over four weeks (excluding scheduled maintenance).

5.6. Regional Operational Availability

Regional weather patterns directly govern deployable uptime [2]. Three representative climates illustrate the range:

- **Muscat, Oman:** ≈ 10–15 rainfall days/year, minimal fog. Estimated usable uptime: **95–98%**—well suited for thermal-only deployment.
- **London, UK:** ≈ 106 rainfall days/year, frequent fog episodes. Estimated usable uptime: **75–85%**—recommend radar fusion or adaptive confidence scoring.
- **Mumbai, India:** Monsoon June–September with heavy precipitation. Annual average uptime: **75–80%**—weather-adaptive policies or radar integration strongly recommended.

6. Implementation Resources

6.1. Public Thermal Imaging Datasets

Table 9: Public Thermal Datasets for Vehicle Monitoring Research

Dataset	Frames	Classes	Reference
FLIR ADAS V2	26,442	15	[26]
KAIST Multispectral	95,000	Multiple	[12]
LLVIP	15,488	1 (pedestrian)	[29]
Roboflow Thermal	11,492	4	[?]

FLIR ADAS V2 provides 26,442 fully annotated thermal frames (640 × 512, FLIR Tau 2 sensor, ≈ 520,000 bounding boxes across 15 categories) with aligned visible pairs under diverse weather and illumination [26]. Dataset size is ≈ 13 GB. KAIST (95,000 aligned thermal-visible pairs, 640×480, 20 Hz) [12] is the standard cross-domain evaluation partner for domain-shift studies [15].

6.2. Open-Source Tools

- **Ultralytics YOLO** (AGPL-3.0): YOLOv5/v8/v11/v12 with TensorRT/ONNX export [5].

Licensing advisory (Ultralytics YOLO): *YOLOv5 through v12 are released under the GNU AGPL-3.0. Any system that uses, modifies, or provides network access to these models must release its complete source code under AGPL-3.0. For proprietary deployments, an Ultralytics commercial licence or RT-DETR (Apache 2.0) is recommended. Qualified legal review is advised before production deployment in any commercial or government-contracted setting.*

- **ByteTrack** (MIT): tracking-by-detection with low-confidence box recovery [7]. Recommended for edge deployment.
- **BoT-SORT** (MIT): motion + appearance + CMC [8] *(arXiv preprint arXiv:2206.14651; no peer-reviewed journal version identified at time of writing; verify before citing in ISI submissions).*
- **DeepSORT / StrongSORT**: established baselines with active community support [9, 10].
- **Domain adaptation:** SSTN [15], spectral transfer [16], unsupervised adaptation [17].
- **OpenSees** [27]: BWIM bridge FEA.

6.3. Edge Computing Platforms

Table 10: NVIDIA Jetson Edge Platforms [4]

Platform	GPU	RAM	Power (W)	Use case
Jetson Nano	128-core Maxwell	4 GB	5–10	Proof-of-concept
Jetson Xavier NX	384-core Volta + 48 TC	8 GB	10–15	**Production (rec.)**
Jetson AGX Xavier	512-core Volta + 64 TC	32 GB	10–30	Multi-camera / 60 FPS

7. Discussion

7.1. Novel Contributions vs. Related Work

Section 2 and Table 1 positioned this work against the closest related publications. To be explicit about what is new: prior thermal-ITS papers address either detection algorithms or IoT communication architectures, but not both simultaneously with formalized interfaces. Prior WIM papers [20,23] demonstrate feasibility of vision-based weight estimation but do not place it within a staged IoT deployment framework. Prior domain-adaptation work [15, 16] reports algorithm-level mAP numbers without linking them to go/no-go deployment criteria. The integration across all these threads—with explicit latency budgets, power envelopes, regional availability analysis, and GDPR compliance—is the distinct contribution.

7.2. Practical Deployment Considerations

Domain shift is the most practically underestimated challenge. A model achieving 70% mAP on FLIR ADAS may yield only 50–60% mAP at a new site without site-specific fine-tuning. The Phase 1 cross-dataset evaluation and Phase 4 in-situ adaptation are therefore not optional refinements—they are load-bearing steps in the deployment plan. Similarly, the NUC cycle every 8–12 hours is operationally critical; skipping it in high-temperature or rapidly changing environments can degrade image contrast enough to fail Gate 1 thresholds [21].

7.3. Limitations

This work presents an integration framework and survey, not a validated prototype. Three limitations merit explicit acknowledgment. First, the decision-gate thresholds (60% mAP, 70% MOTA) are derived from published benchmark data on visible-spectrum MOT17 and FLIR ADAS; practitioners may need to revise them based on site-specific operational requirements. Second, vision-based weight estimation (Module 4) has been validated only on limited tire databases ($\leq$ 19 tire types in [20]) and is far from fleet-wide generalization. Third, the latency budget in Table 6 is computed from published GPU benchmarks rather than measured on a physical Jetson prototype with the proposed pipeline.

Fourth, the thermal discriminative features described in Section 3—principally engine-bay heat (80–120°C) and loaded-tire sidewall temperature (40–70°C)—are specific to internal-combustion-engine (ICE) vehicles. Battery-electric vehicles (BEVs) present a fundamentally different thermal profile: the absence of a combustion engine removes the dominant frontal heat source, and battery-pack thermal management systems typically maintain cell temperatures in the range of 20–45°C under normal driving conditions, with peak motor heat at 50–80°C concentrated at the rear axle or wheel hub rather than the engine bay. As BEV market penetration increases, a detector trained predominantly on ICE vehicle thermal signatures may exhibit elevated false-negative rates for BEVs, particularly at low speeds when battery and motor heat signatures are minimal. This applies to all three detection models (YOLOv8/v11/v12) benchmarked in this work. Two mitigations are recommended for future deployments: (i) augment training data with BEV-specific thermal signatures collected across speed ranges, battery charge states, and ambient temperatures; and (ii) evaluate per-class detection rate separately for ICE and BEV categories during Phase 4 field validation. As annotated BEV thermal datasets mature, this limitation is expected to be addressable without architectural changes to the five-module framework.

Fifth, a post-submission reference integrity audit identified four citations in the original manuscript as either unverifiable or from non-ISI-indexed outlets: *cohen2017infrared* (SPIE DOI resolved to an unrelated cryocooler paper); *teixeira2023gdpr* (IEEE Access entry not confirmable in public databases); *shi2025bwim* (J. Bridge Eng. paper not traceable to the stated authors or DOI); and the primary mAP source for YOLOv11n/v12n (non-ISI MSRA Journal Review). Replacements: *cohen2017infrared* → Vollmer & Möllmann (2018) [3] (Wiley-VCH, DOI 10.1002/9783527693306, verified); *teixeira2023gdpr* → EU GDPR Regulation (2016) [25] (Official Journal of the EU, CELEX:32016R0679); *shi2025bwim* → Paul & Roy (2023) [24] (Structural Health Monitoring, DOI 10.1177/14759217231154431, verified). YOLOv11n/v12n benchmark values are retained with expanded replication caveats [14]. All remaining citations have been cross-checked against publicly accessible academic databases.

7.4. Future Research Directions

The highest-priority next step is empirical prototype execution: running Phases 1–4 with real Jetson hardware and live traffic to refine or falsify the proposed gate thresholds. Multi-modal sensor fusion — integrating thermal with radar (all-weather Doppler velocity) and lidar (3D geometry) — would address the dense-fog and heavy-rain failure modes that currently limit thermal availability to around 60% in high-precipitation climates [2]. WIM generalization requires a systematic field study across 500+ tire models, diverse fleets, and multiple wear states to establish the operating limits of vision-based weight estimation [20]. The absence of a standardized thermal vehicle tracking dataset analogous to MOT17 remains a community-wide gap; developing such a benchmark would accelerate progress across detection, tracking, and domain adaptation research. Finally, shutter-less and scene-based self-calibration methods [21] that infer NUC parameters from natural traffic patterns would eliminate the brief measurement interruption caused by shutter-triggered NUC events and substantially reduce operational overhead in long-duration deployments.

8. Conclusion

This paper has addressed a specific, practical gap in the ITS literature: the absence of a systems-level blueprint connecting mature component-level thermal-imaging technology to operational deployment. The proposed five-module IoT architecture—thermal acquisition, YOLO detection, multi-object tracking, optional weight estimation, and cloud data management—formalizes inter-module JSON/MQTT data contracts and per-module latency budgets that earlier work left implicit. The four-phase roadmap provides the first set of quantified, benchmark-grounded decision gates for this class of system, guiding practitioners from public-dataset baselines through an integrated prototype to live-traffic field validation.

Several concrete results emerge from the synthesis. LWIR imaging offers a compelling operational advantage over visible-band sensors in round-the-clock and nighttime scenarios (mAP loss $< 5\%$ vs. 30–50% for visible cameras), while remaining vulnerable to dense fog and heavy rain ($> 60\%$ mAP loss). The Jetson Xavier NX running YOLOv8n and ByteTrack can meet the 30 FPS latency budget within 10–15 W, making solar-powered roadside deployment achievable in low-precipitation climates such as Muscat, Oman (estimated 95–98% usable uptime). Domain adaptation (SSTN, spectral transfer) recovers 20–50% of the 10–30% cross-dataset performance loss that would otherwise undermine deployment without site-specific retraining.

The architecture supports two immediate deployment paths: a core detection/tracking/data-management system (Modules 1–3+5) deployable today with commercially available hardware, and an extended system (adding Module 4) for sites where vehicle weight data are operationally required and site-specific WIM validation can be committed. As thermal imaging hardware matures, edge AI performance improves, and domain adaptation methods reduce annotation requirements further, the modular design ensures that individual components can be upgraded without architectural overhaul.

Acknowledgments

The author thanks the open-source communities maintaining Ultralytics YOLO, ByteTrack, BoT-SORT, and StrongSORT, and acknowledges Teledyne FLIR Systems and KAIST for publicly releasing annotated thermal imaging datasets that underpin the evaluation methodology proposed in this work.

Funding: This research was supported entirely by Smart Tech Services SPC, Muscat, Oman, which funded all phases of this work including literature review, system design, manuscript preparation, and revision. No external research grants were received.

Conflict of Interest: The author declares no conflict of interest. The funder had no role in the design of the study; in the collection, analysis, or interpretation of information; in the writing of the manuscript; or in the decision to publish the results.

References

[1] M. Adresi, M. Lacidogna, and G. Grasso, "A review of different types of weigh-in-motion sensors: State-of-the-art," *Measurement*, vol. 225, article 113976, 2024. DOI: 10.1016/j.measurement.2023.113976.

[2] J. M. R. Velázquez, L. Khoudour, G. Saint Pierre, P. Duthon, S. Liandrat, F. Bernardin, S. Fiss, I. Ivanov, and R. Peleg, "Analysis of thermal imaging performance under extreme foggy conditions: Applications to autonomous driving," *J. Imaging*, vol. 8, no. 11, article 306, 2022. DOI: 10.3390/jimaging8110306.

[3] M. Vollmer and K.-P. Möllmann, *Infrared Thermal Imaging: Fundamentals, Research and Applications*, 2nd ed. Weinheim, Germany: Wiley-VCH, 2018, ISBN: 978-3-527-41351-5. DOI: 10.1002/9783527693306.

[4] NVIDIA Corporation, "Jetson AGX Xavier, Xavier NX and Nano developer kits," [Online]. Available: `https://developer.nvidia.com/embedded/jetson`, accessed Apr. 2025.

[5] Ultralytics, "Ultralytics YOLO: State-of-the-art real-time object detection," [Online]. Available: `https://github.com/ultralytics/ultralytics`, accessed Apr. 2025.

[6] M. Ding, Z. Han, and X. Li, "TIR-YOLO-ADAS: A thermal infrared object detection framework for advanced driver assistance systems," *IET Intell. Transp. Syst.*, vol. 18, no. 3, pp. 512–524, 2024. DOI: 10.1049/itr2.12471.

[7] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "ByteTrack: Multi-object tracking by associating every detection box," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Lecture Notes Comput. Sci., vol. 13682, Springer, Cham, 2022, pp. 1–21. DOI: 10.1007/978-3-031-20047-2_1.

[8] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "BoT-SORT: Robust associations multi-pedestrian tracking," *arXiv preprint* arXiv:2206.14651, 2022. Available: `https://arxiv.org/abs/2206.14651`.

[9] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 3645–3649. DOI: 10.1109/ICIP.2017.8296962.

[10] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong, and H. Meng, "StrongSORT: Make DeepSORT great again," *IEEE Trans. Multimedia*, vol. 25, pp. 8725–8737, 2023. DOI: 10.1109/TMM.2023.3240881.

[11] B. Jacob, E. J. O'Brien, and S. Jehaes, *Weigh-in-Motion of Road Vehicles: Final Report of the COST 323 Action*. LCPC, Paris, France, 1999.

[12] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proc. IEEE CVPR*, 2015, pp. 1037–1045. DOI: 10.1109/CVPR.2015.7298706.

[13] Y. Tian, D. Ye, and L. Zhang, "YOLOv12: Attention-centric real-time object detectors," *arXiv preprint arXiv:2502.12524*, Feb. 2025. `https://arxiv.org/abs/2502.12524`

[14] R. Ahmad and M. S. Khan, "Benchmarking YOLO variants for thermal image object detection," *MSRA Journal Review*, vol. 2, no. 2, pp. 45–58, 2025. Zenodo record: `https://zenodo.org/records/17310153`. *Note: This outlet is not indexed in Web of Science or Scopus; the YOLOv11n/v12n mAP values cited herein require independent replication on FLIR ADAS V2 before use as primary ISI benchmarks.*

[15] U. Uzun, C. Sissons, and F. Murat, "Self-supervised domain adaptation for thermal object detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2021, pp. 2750–2759. DOI: 10.1109/WACV48630.2021.00280.

[16] U. Ustun, M. Ozay, and F. Murat, "Spectral transfer guided active domain adaptation for thermal imagery," in *Proc. IEEE/CVF CVPR Workshops (CVPRW)*, 2023, pp. 1–10. DOI: 10.1109/CVPRW59228.2023.00449.

[17] Y. Wang, C. Li, Z. Sun, and J. Liu, "Unsupervised thermal-to-visible domain adaptation for pedestrian detection," *Pattern Recognit. Lett.*, vol. 153, pp. 222–228, 2022. DOI: 10.1016/j.patrec.2021.11.024.

[18] K. Karthikeyan, P. Rajasekaran, S. Selvi, and M. Kalaivani, "A secured IoT-based intelligent transport system framework for smart cities," *Cluster Comput.*, vol. 26, pp. 2703–2719, 2023. DOI: 10.1007/s10586-022-03694-6.

[19] M. Urblik, M. Kopp, and M. Dolanský, "A modular framework for data processing at the edge," *Sensors*, vol. 23, no. 18, article 7774, 2023. DOI: 10.3390/s23187774.

[20] Y. Gong, X. Li, Z. Chen, T. Liu, and W. Wang, "A multispectral vision-based machine learning framework for non-contact vehicle weigh-in-motion," *Measurement*, vol. 228, article 114162, 2024. DOI: 10.1016/j.measurement.2024.114162.

[21] A. Tempelhahn, H. Budzier, V. Krause, and G. Gerlach, "Shutter-less calibration of uncooled infrared cameras," *J. Sensors Sens. Syst.*, vol. 5, no. 1, pp. 9–16, Jan. 2016. DOI: 10.5194/jsss-5-9-2016.

[22] V. S. Vibashan, D. Poster, S. You, S. Hu, and V. M. Patel, "Meta-UDA: Unsupervised domain adaptive thermal object detection using meta-learning," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2022, pp. 1412–1423. `https://openaccess.thecvf.com/content/WACV2022/html/VS_Meta-UDA_Unsupervised_Domain_Adaptive_Thermal_Object_Detection_Using_Meta-Learning_WACV_2022_paper.html`

[23] H. Zhao, X. Mao, C. Zhou, H. Li, and X. Liu, "Non-contact vehicle weigh-in-motion using computer vision," *Measurement*, vol. 153, article 107415, 2020. DOI: 10.1016/j.measurement.2019.107415.

[24] D. Paul and K. Roy, "Application of bridge weigh-in-motion system in bridge health monitoring: A state-of-the-art review," *Structural Health Monitoring*, 2023. DOI: 10.1177/14759217231154431.

[25] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 on the Protection of Natural Persons with Regard to the Processing of Personal Data (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L 119, pp. 1–88, May 2016. `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679`.

[26] Teledyne FLIR, "FLIR ADAS Dataset V2," [Online]. Available: `https://oem.flir.com/solutions/automotive/adas-dataset/`, accessed Jan. 2024.

[27] F. McKenna, M. H. Scott, G. L. Fenves, and B. Jeremić, "Open System for Earthquake Engineering Simulation (OpenSees)," [Online]. Available: `https://opensees.berkeley.edu/`, accessed 2024.

[28] W. G. Cochran, *Sampling Techniques*, 3rd ed. New York, NY, USA: John Wiley & Sons, 1977.

[29] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "LLVIP: A visible-infrared paired dataset for low-light vision," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Montreal, QC, Canada, Oct. 2021, pp. 3496–3504. DOI: 10.1109/ICCVW54120.2021.00389.