

Multi-Camera Subject Tracking Through Deep Metric Embeddings and YOLO-Based Detection: A Layered System Architecture

¹ Serif Oyindamola Oyesiji, ² Toussida Fatah Tanguy Minoungou, ³ Kingsley Chinazaekpere Ndupu, ⁴ Chukwudera Obumneke Anunagba.

¹ Harrisburg University of Science and Technology, Harrisburg, PA; oyesijiserif@gmail.com

² Independent Researcher, Dallas, TX, USA; tanguyminoungou@gmail.com

³ Kennesaw State University, Georgia, USA; kingsleynd34@gmail.com

⁴ Barclays, USA; chiokemuo@gmail.com

Corresponding Author; oyesijiserif@gmail.com

Abstract: Tracking individual subjects across networks of non-overlapping cameras is a foundational capability for traffic analytics, retail and facility understanding, sports analysis, and safety monitoring, yet it remains substantially harder than either of its component problems. Single-camera multi-object tracking and appearance-based re-identification are each mature; their composition, maintaining consistent identities as subjects leave one view and appear in another under changed pose, illumination, and viewpoint, multiplies their failure modes and adds association problems neither addresses alone. This paper develops a research concept for a multi-camera subject tracking system built from three specified layers: per-camera detection and short-term tracking using YOLO-family detectors with motion-and-appearance trackers; a metric embedding layer trained with re-identification objectives to produce identity-discriminative appearance vectors under cross-view variation; and a cross-camera association layer that matches tracklets across views using appearance similarity gated by spatio-temporal feasibility derived from camera topology, managed through a gallery of identity prototypes with principled creation, update, and retirement policies. The concept specifies the embedding training regime, the association algorithm as a constrained assignment over tracklet pairs, and a real-time implementation architecture with approximate nearest neighbor indexing for gallery search at network scale. A phased evaluation plan measures per-camera tracking, embedding quality, and end-to-end cross-camera identity consistency with modern higher-order metrics on public multi-camera benchmarks, with prespecified ablations isolating each layer's contribution. Applications, failure modes, and the privacy and governance constraints that subject tracking obligates are analyzed. The concept's claim is architectural: cross-camera identity is best engineered as a composition of separately measurable layers with explicit contracts, not as a monolithic learned system.

1. Introduction

The components of multi-camera tracking have matured separately. Object detection provides reliable per-frame localization at real-time rates, with the YOLO lineage the standard choice where latency budgets bind (Redmon et al., 2016; Redmon & Farhadi, 2018; Bochkovskiy et al., 2020; Ge et al., 2021; Wang et al., 2023) within a broader detector landscape surveyed comprehensively elsewhere (Zou et al., 2023; Ren et al., 2015; Lin et al., 2017), a primitive whose applied deployments in corridor inspection, encroachment detection, airport wildlife hazard monitoring, and medical image screening show it travelling well beyond benchmark conditions (Dagodzo et al., 2022; Dagodzo & Ahiaekwe Patrick, 2022; Adeyelu & Dagodzo, 2024; Ejofodomi et al., 2014). Single-camera multi-object tracking converts detections into trajectories, from minimal motion-based association (Bewley et al., 2016; Kalman, 1960; Kuhn, 1955) through appearance-augmented designs (Wojke et al., 2017) to detection-centric and joint formulations that define the current state of practice (Bergmann et al., 2019; Zhang et al., 2021; Zhang et al., 2022), evaluated on standard benchmark series (Milan et al., 2016; Dendorfer et al., 2020) under metrics whose evolution itself reflects the field's maturity (Bernardin & Stiefelhagen, 2008; Ristani et al., 2016; Luiten et al., 2021). Person re-identification supplies the appearance machinery: metric embeddings trained so that images of the same identity lie close across views, built on contrastive and triplet formulations (Chopra et al., 2005; Hadsell et al., 2006; Schroff et al., 2015; Hermans et al., 2017), standardized datasets (Zheng et al., 2015), architecture and training practice (Luo et al., 2019; Zhou et al., 2019; He et al., 2016), and consolidated surveys (Zheng et al., 2016; Ye et al., 2021), with adjacent biometric work documenting the spoofing and demographic-bias pressures that any deployed appearance-matching system inherits (Adeniyi et al., 2025).

Their composition is where deployed systems struggle. Cross-camera tracking must decide when a tracklet ending in one view and a tracklet beginning in another belong to the same subject, a decision that appearance alone cannot make reliably at network scale, where galleries grow, distractors accumulate, and viewpoint change is adversarial by construction, and that spatio-temporal reasoning alone cannot make either, since topology constrains but rarely determines identity. The research systems that established the joint problem demonstrated both the value and the difficulty of the composition, learning association features across cameras (Ristani & Tomasi, 2018), formulating tracking with re-identification in unified combinatorial objectives (Tang et al., 2017), and

building the multi-camera benchmarks that make progress measurable (Ristani et al., 2016; Chavdarova et al., 2018). What remains under-specified in the literature is the systems form of the problem: an architecture whose layers have explicit contracts, whose gallery is managed rather than assumed, whose association is gated by calibrated feasibility, and whose evaluation isolates where identity is actually lost. Comparable layered specifications in adjacent applied-AI domains, where model behaviour, data handling, and release gating are held to separate and separately auditable contracts, suggest that the discipline generalizes (Nwakamma et al., 2024a; Oyesiji et al., 2024; Ominyi et al., 2026b).

This paper develops that specification as a research concept. Section 2 reviews background. Section 3 defines the problem and requirements. Sections 4 through 7 specify the architecture, the embedding layer, the association layer, and the implementation. Section 8 sets out the evaluation plan, Section 9 applications, Section 10 limitations and governance, and Section 11 concludes.

2. Background and Related Work

Single-camera tracking establishes the tracklet substrate. Tracking-by-detection dominates practice: detections are associated frame to frame using motion prediction, with the linear-Gaussian filter the standard state estimator (Kalman, 1960) and optimal assignment solving the per-frame matching (Kuhn, 1955), the combination that minimal trackers showed sufficient at high frame rates (Bewley et al., 2016). Appearance integration extended association across occlusions by matching learned descriptors alongside motion (Wojke et al., 2017); regression-based designs showed detectors themselves can carry association (Bergmann et al., 2019); joint detection-embedding models fold re-identification features into the detection network for efficiency (Zhang et al., 2021); and association over the full detection set, including low-confidence boxes, recovers identities that thresholding discards (Zhang et al., 2022). Benchmark practice fixed evaluation: the MOT challenge series standardizes sequences and protocols (Milan et al., 2016; Dendorfer et al., 2020), and metric development moved from event-counting scores (Bernardin & Stiefelhagen, 2008) through identity-preservation measures introduced with multi-camera evaluation (Ristani et al., 2016) to higher-order metrics that decompose detection and association quality (Luiten et al., 2021), the last two being the natural currency for cross-camera claims.

Re-identification supplies the cross-view appearance model. The problem is defined by its invariance demands, identity constancy under viewpoint, pose, illumination, and background change, and the field's answer is deep metric learning: siamese and contrastive origins (Chopra et al., 2005; Hadsell et al., 2006), triplet objectives with hard mining that shaped modern training (Schroff et al., 2015; Hermans et al., 2017), strong conventional baselines documenting the training details that matter (Luo et al., 2019), and architectures designed for identity-scale discrimination (Zhou et al., 2019) over standard backbones (He et al., 2016). Datasets and surveys consolidate the field (Zheng et al., 2015; Zheng et al., 2016; Ye et al., 2021), while adjacent work on parameter-efficient adaptation and context-aware data curation frames the domain-transfer problem in the same terms this concept adopts (Oyesiji et al., 2023; Oyesiji et al., 2024), and two of its standing findings bind this concept's design: embedding quality degrades across domains, so deployment-domain adaptation is a requirement rather than an option, and gallery composition governs practical accuracy as strongly as embedding quality, motivating the managed-gallery layer of Section 4.

Multi-camera tracking research defines the composed problem. Network-scale systems formulated cross-camera association as correlation clustering over tracklets with learned features (Ristani & Tomasi, 2018), lifted combinatorial formulations coupled within-camera and across-camera decisions with re-identification terms (Tang et al., 2017), and dedicated benchmarks supplied ground truth across camera sets, both large non-overlapping networks (Ristani et al., 2016) and dense overlapping rigs (Chavdarova et al., 2018). Geometric infrastructure underpins the spatio-temporal side: calibration renders image evidence into common coordinates where overlap exists (Zhang, 2000; Hartley & Zisserman, 2004), and site models of the kind maintained in geospatial asset management and building information modelling supply the same class of structure at facility scale (Dagodzo & Ahiaeke Patrick, 2021a; Dagodzo & Ahiaeke Patrick, 2021b; Uduokhai et al., 2023a); where views do not overlap, topology, which exits feed which entries at what transit-time distributions, becomes the feasibility structure the association layer of Section 6 formalizes. Finally, scale pushes the retrieval problem into the approximate regime: gallery search at network size runs on approximate nearest neighbor indexes whose accuracy-throughput trade-offs are established infrastructure (Johnson et al., 2021; Malkov & Yashunin, 2020), with model and index compression for constrained deployment targets an active adjacent concern (Oyesiji et al., 2025; Nwakamma et al., 2025).

3. Problem Definition and Requirements

The system's function is stated as follows: given synchronized video from a network of calibrated or topology-mapped cameras, maintain a set of global identities such that each detected subject is assigned a global identity consistent across all views and times at which the subject appears, with per-assignment confidence, under latency compatible with the consuming application, live monitoring at seconds-scale delay or forensic analysis offline. Inputs are per-camera streams; outputs are global tracklets, per-camera trajectory segments labeled with global identities, together with the association evidence, appearance similarity and spatio-temporal feasibility, behind each cross-camera link.

Requirements follow from the failure structure. Identity precision dominates identity recall for most consuming applications: a merged identity, two subjects fused into one, corrupts downstream analysis silently, while a fragmented identity, one

subject split into two, is visible and repairable, so association operates at high-precision operating points with explicit abstention, unresolved tracklets remain unlinked rather than forced. Contract separation is required between layers: per-camera tracking, embedding quality, and cross-camera association are separately measurable, and the evaluation plan holds each to its own metrics so that end-to-end failures localize. Scale behavior must be specified: gallery size grows with subjects and time, and both association quality and latency must degrade predictably, not catastrophically, as it does, the same predictability discipline that governs energy, reliability, and latency trade-offs in large distributed sensing deployments (Asiedu & Quainoo, 2024; Asiedu & Quainoo, 2023b). And governance is a requirement, not a feature: subject tracking is surveillance infrastructure by capability, so purpose limitation, retention bounds, and access control are architectural obligations stated in Section 10 (Mbonu et al., 2018; Ominyi & Anichukwueze, 2022a; Annan, 2022).

4. System Architecture

The architecture composes three layers with a managed gallery as shared state. The per-camera layer runs a YOLO-family detector selected for the deployment's latency budget (Redmon & Farhadi, 2018; Bochkovskiy et al., 2020; Ge et al., 2021) feeding a tracking-by-detection tracker in the modern configuration, motion prediction with full-detection-set association and appearance assistance across short occlusions (Zhang et al., 2022; Wojke et al., 2017; Bewley et al., 2016). Its output contract is tracklets: temporally contiguous single-camera trajectories with per-frame boxes, confidences, and sampled appearance crops, terminated conservatively, since the cross-camera layer can join what the per-camera layer splits, but joins made wrongly within a camera propagate outward.

The embedding layer converts sampled crops into identity vectors under the training regime of Section 5, aggregating per-tracklet, quality-weighted pooling over samples with blur, occlusion, and resolution screening, into one or a few prototype vectors per tracklet, the representation on which all cross-camera reasoning operates. The association layer, specified in Section 6, links tracklets across cameras into global identities using appearance similarity gated by spatio-temporal feasibility, writing its decisions into the gallery: the set of active global identities, each carrying its prototype vectors, its camera-time itinerary, and its confidence history. Gallery management is specified as policy, not left implicit: identities are created from unlinked tracklets, updated by exponentially weighted prototype refresh on confident links, demoted to dormant after absence horizons, and retired at retention bounds, with all transitions logged, because in deployed systems the gallery is where scale, drift, and governance all concentrate (Ye et al., 2021; Johnson et al., 2021), and because durable identity state is a governed data asset subject to sensitivity classification, retention control, and regulatory traceability like any other (Aliliele et al., 2024a; Aliliele et al., 2025b; Mbonu et al., 2022a).

5. Embedding Design and Training

The embedding is trained as a re-identification metric under the field's established regime: a backbone of deployment-appropriate capacity (He et al., 2016; Zhou et al., 2019) trained with a combined objective, classification over training identities plus triplet loss with batch-hard mining (Hermans et al., 2017; Schroff et al., 2015), under the practice details that the strong-baseline literature shows decisive, sampling identities per batch, appropriate augmentation, and normalization choices (Luo et al., 2019). Training data composition follows the deployment gap finding: public re-identification corpora (Zheng et al., 2015) provide the base, and a deployment-domain adaptation phase fine-tunes on tracklet-derived pairs from the target network, using within-tracklet samples as positives and temporally co-occurring cross-tracklet samples as guaranteed negatives, the supervision the system's own per-camera layer generates for free, curated under the context-aware selection discipline that fine-tuning pipelines elsewhere have found decisive (Oyesiji et al., 2024; Oyesiji et al., 2023).

The embedding contract is stated for the association layer's benefit: similarity scores are calibrated on held-out deployment data into match probabilities, so that gating and assignment in Section 6 operate on probabilities rather than raw distances; embedding versioning is explicit, since prototype vectors in the gallery are comparable only within a version, and re-embedding on version change is a managed migration; and evaluation of the layer in isolation uses standard retrieval protocols, cumulative match characteristics and mean average precision on held-out identities (Zheng et al., 2015; Ye et al., 2021), so that embedding regressions are caught before they surface as association failures, an automated regression-control and release-gating obligation rather than a manual one (Mbonu et al., 2021b; Borah et al., 2022; Nwakamma et al., 2024a).

6. Cross-Camera Association

Association is formulated as constrained assignment over tracklet pairs within a sliding decision window. Candidate generation applies spatio-temporal gating: for non-overlapping views, the camera topology, exit zones, entry zones, and empirical transit-time distributions between them, defines which cross-camera tracklet pairs are feasible and weights them by transit likelihood; for overlapping views, calibrated geometry tests co-location directly (Zhang, 2000; Hartley & Zisserman, 2004). Topology and transit distributions are themselves estimated from data, bootstrapped from confident appearance matches and refined continuously,

so the feasibility structure adapts to the site rather than being hand-fixed, the bootstrapped-then-continuously-refined pattern that operational predictive models follow in other monitoring domains (Mayo et al., 2021; Tonoyan et al., 2024b).

Scoring combines calibrated appearance similarity between tracklet prototypes with the transit likelihood into a link probability per feasible pair. Decision-making solves the assignment at high-precision thresholds, minimum-cost matching within the window (Kuhn, 1955), with clustering formulations available where windows are long and transitive constraints bind, in the lineage of correlation-clustering association (Ristani & Tomasi, 2018; Tang et al., 2017). Three behaviors are specified beyond the optimizer. Abstention: pairs between threshold bounds remain unlinked and are revisited as evidence accrues, honoring the precision-dominance requirement. Late binding: links are provisional within the decision window and harden at its trailing edge, allowing short-horizon revision when a better candidate arrives. And evidence retention: every link stores its appearance and transit scores, making downstream audit and threshold retuning possible on the deployed record, the operational property that distinguishes an accountable system from an oracle, and the precondition for the algorithmic accountability that automated decision systems are increasingly expected to demonstrate (Mbonu et al., 2019; Ominyi & Anichukwueze, 2024; Annan, 2024).

7. Implementation Architecture

The implementation maps the layers onto network infrastructure. Per-camera detection and tracking run at the edge, on per-camera or per-cluster compute, emitting compact tracklet records, boxes, prototypes, and quality metadata, rather than video, an allocation that scales bandwidth with subjects rather than pixels and keeps raw footage local by default, sized against the link budgets and throughput degradation that real wireless deployments exhibit (Quainoo et al., 2025a; Asiedu & Quainoo, 2023b). On-device execution under quantization, pruning, and distillation, carried on portable cloud-to-edge runtimes, is the enabling substrate for that allocation (Nwakamma et al., 2025; Ojukwu et al., 2025). The embedding layer runs where its inputs are, at the edge for prototype extraction, centrally for re-embedding migrations. The association layer and gallery run centrally, on clustered accelerator infrastructure whose availability is itself a maintained engineering property (Oyeleke et al., 2026): prototype search over the gallery uses approximate nearest neighbor indexing with the graph-based structures whose recall-latency behavior is well characterized (Malkov & Yashunin, 2020; Johnson et al., 2021), candidate sets are pre-filtered by the spatio-temporal gate before any distance computation, and the decision window advances on a fixed cadence so that association latency is bounded and measurable. State durability follows the gallery policies of Section 4, with the audit log of link evidence persisted alongside identities under the log-analytics and incident-response practices applied to other regulated telemetry (Mbonu et al., 2019; Aliliele et al., 2025b), and with every identity query mediated by enforced identity and access management (Mbonu et al., 2020; Mbonu et al., 2022b; Ahmed et al., 2021). The architecture's load parameters, cameras, concurrent subjects, gallery size, and window length, are the axes along which the evaluation plan's scale phase measures degradation, because the concept's scale claim is predictability, not unboundedness.

8. Evaluation Plan

Evaluation proceeds in three phases with layer isolation as the organizing principle. Phase one validates layers separately. Per-camera tracking is evaluated on standard sequences with higher-order and identity metrics (Luiten et al., 2021; Ristani et al., 2016; Bernardin & Stiefelhagen, 2008; Milan et al., 2016; Dendorfer et al., 2020), establishing the tracklet substrate's quality independent of anything cross-camera. The embedding is evaluated on held-out identities under retrieval protocols, with the deployment-adaptation phase's contribution measured as the delta over the public-corpus baseline (Zheng et al., 2015; Ye et al., 2021; Luo et al., 2019). Gating infrastructure is validated directly: learned transit-time distributions are scored against held-out confirmed transitions.

Phase two evaluates the composed system on multi-camera benchmarks with cross-camera ground truth, non-overlapping network data and dense overlapping rigs (Ristani et al., 2016; Chavdarova et al., 2018), reporting identity-preservation and higher-order metrics computed globally across cameras (Luiten et al., 2021), with the operating point swept to expose the precision-recall structure of association and with abstention rates reported alongside, since unlinked-but-correct is this design's intended failure mode. Prespecified ablations isolate contributions: appearance-only versus gated association quantifies the spatio-temporal layer; single-prototype versus quality-weighted aggregation quantifies the pooling design; and per-camera tracker variants quantify upstream sensitivity, the decomposition that tells future work where identity is actually lost.

Phase three measures systems behavior: end-to-end association latency against gallery size and camera count, index recall-latency operating points under the deployment's candidate distributions (Malkov & Yashunin, 2020; Johnson et al., 2021), gallery dynamics under long-duration operation, growth, dormancy, retirement, and identity-merge incident rates under stress conditions, crowded scenes and appearance-similar cohorts, the conditions under which precision claims earn their keep. Instrumentation for this phase follows established continuous-monitoring and performance-dashboard practice, so that degradation is observed rather than inferred (Ogbole et al., 2021; Tonoyan et al., 2024b; Basnet et al., 2023). Acceptance is defined per requirement: layer metrics at or above phase-one baselines, composed identity precision at the specified operating point, and scale degradation within the predicted envelopes, with acceptance treated as a release gate rather than a report (Nwakamma et al., 2024a; Dosunmu & Ogundele, 2024a).

9. Anticipated Applications

The capability serves analytics wherever subject-level continuity across views carries the value. Facility and retail analytics consume global trajectories for flow, dwell, and conversion understanding, complementing the geospatial and predictive methods already used for siting, market penetration, and lifecycle analysis (Tonoyan et al., 2022a; Seyi-Lande et al., 2020; Oziri et al., 2020); transport hubs consume them for crowd management and journey-time estimation, within transport infrastructures whose sensing layers carry their own security obligations (Jimoh et al., 2023) and alongside evacuation and emergency-preparedness planning in high-density facilities (Obriki et al., 2022b); sports analysis consumes them for player tracking across broadcast and tactical cameras, a setting where the overlapping-rig configuration and dense benchmarks apply directly (Chavdarova et al., 2018); and safety applications consume them for lost-person search and incident reconstruction, the forensic mode where late binding and evidence retention matter most, adjacent to continuous hazard monitoring and predictive safety analytics in industrial and construction settings (Obriki et al., 2023; Obriki & Arumosoye, 2023; Obogo et al., 2026) and to the alerting and public-communication layer that incidents ultimately trigger (Odejide, 2024). Industrial and logistics environments add a further consumer, since autonomous mobile robot fleets need the same subject continuity for safe coexistence and fleet coordination (Akanbi & Sunday, 2024; Akanbi & Sunday, 2025; Akanbi et al., 2025), and airport and shared-airspace environments raise the same continuity question for non-human subjects under regulatory oversight (Adeyelu, 2023; Adeyelu & Dagodzo, 2024). Research applications include behavioral studies at ecological scale, and the architecture's audit-log design doubles as an instrument for studying association itself, since every deployment accumulates labeled hard cases at its own operating point (Ristani & Tomasi, 2018; Ye et al., 2021). Where population-level rather than individual-level output suffices, aggregate flow statistics feed the same dashboard and surveillance-reporting infrastructures used in public health informatics, which is also the setting where the case for de-identified aggregation is strongest (Oparah et al., 2023; Anioke & Atima, 2023c). Related identity-linked operational systems, from biometric attendance to RFID asset tracking, indicate both the demand for continuity and the narrower, consent-anchored designs that often satisfy it (Ussher-Eke et al., 2025; Yusuf et al., 2023). Across applications, the deployment posture follows Section 10: the concept is specified for consented, purpose-limited settings, and its evaluation claims are scoped to those settings.

10. Limitations, Failure Modes, and Governance

Technical limits are structural. Appearance is not identity: clothing change defeats any appearance embedding, uniformed cohorts compress inter-identity distances, and cross-view extremes, top-down to eye-level, stretch invariance beyond what training distributions cover, which is why abstention and spatio-temporal gating are load-bearing rather than decorative (Ye et al., 2021; Zheng et al., 2016). Appearance matching is also attackable and unevenly accurate across cohorts, the two failure modes the biometric literature treats as primary (Adeniyi et al., 2025), and adversarial manipulation of a perception pipeline embedded in operational infrastructure is a modelled threat rather than a hypothetical one (Adebayo et al., 2022; Adebayo et al., 2023). Topology feasibility weakens in open spaces with unconstrained paths, degrading the gate toward permissiveness precisely where appearance is least reliable, crowds. Error propagation crosses layers: per-camera identity switches contaminate prototypes, and contaminated prototypes corrupt galleries, which the conservative-termination contract and quality-weighted pooling mitigate but do not eliminate (Zhang et al., 2022; Luiten et al., 2021). Domain drift is permanent: seasonal appearance, camera changes, and re-lamping shift distributions under the embedding, making the deployment-adaptation loop and embedding versioning standing maintenance obligations rather than one-off tasks (Oyesiji et al., 2023; Mbonu et al., 2021b).

Governance is the concept's hard boundary. Multi-camera subject tracking is, by capability, surveillance infrastructure, and the difference between an analytics system and an instrument of harm lies in deployment constraints: explicit purpose limitation with capability disabled beyond it, retention bounds enforced by the gallery's retirement machinery, access control and audit on every identity query, and deployment only where a lawful basis and appropriate notice exist. Each of these has an established literature the concept borrows from rather than reinvents: harmonized data-protection compliance across divergent regimes and privacy by design as a structuring commitment (Ominyi & Anichukwueze, 2022a; Ominyi, 2021; Annan, 2022); anti-discrimination and accountability duties attaching to automated decisions (Annan, 2025a; Ominyi & Anichukwueze, 2024; Ominyi et al., 2024b; Ominyi et al., 2026b); data protection impact assessment and sensitivity classification for the identity store itself (Mbonu et al., 2022a; Mbonu et al., 2018; Aliliele et al., 2024a); privacy-preserving handling patterns drawn from comparably sensitive domains (Afrihyia et al., 2024; Orise & Niniola, 2023; Orise & Niniola, 2025); and the access-governance, insider-risk, and human-factors controls without which an access policy is only nominal (Mbonu et al., 2020; Mbonu et al., 2022b; Onche et al., 2024; Onche et al., 2026). The system's own security belongs to the same posture, since a compromised tracking network is a surveillance capability handed to whoever compromised it: zero-trust segmentation, encrypted and attribute-scoped access to identity records, misconfiguration and API exposure monitoring, secure development practice, and continuous control validation all apply (Ahmed et al., 2021; Adebayo et al., 2025; Shittu & Nabil, 2023; Aliliele et al., 2023a; Aliliele et al., 2023b; Odejobi et al., 2025; Dosunmu & Ogundele, 2024a). Two design choices support these constraints structurally, appearance prototypes rather than raw imagery as the circulating representation, and edge-local video with only tracklet metadata leaving the camera, but structure does not substitute for policy, and the concept declines the fiction that it could. Evaluation claims in this paper are correspondingly scoped: performance

is defined on consented benchmarks and deployments satisfying the governance posture, and no claim is made or implied for adversarial identification of unwilling subjects.

11. Conclusion

This paper has developed a research concept for multi-camera subject tracking as a composition of specified layers: YOLO-based per-camera detection and tracking emitting conservative tracklets; a calibrated re-identification embedding with deployment-domain adaptation and versioned prototypes; and a cross-camera association layer that gates appearance by learned spatio-temporal feasibility, decides at high-precision operating points with explicit abstention, and manages a governed gallery with auditable link evidence. The evaluation plan operationalizes the architecture's central discipline, contract separation, measuring each layer in isolation before composing them on multi-camera benchmarks under identity-preservation and higher-order metrics, with ablations that localize where identity is lost and a systems phase that renders scale behavior predictable.

The concept's claim is that cross-camera identity is an engineering composition, not a monolith: the mature components, detection, tracking, metric embedding, retrieval indexing, become a trustworthy capability only when their interfaces carry calibrated probabilities, their shared state is managed policy, and their failure modes are designed for rather than discovered. Executed, the plan yields both a deployable architecture for the analytics settings that legitimately need subject continuity and a measurement template that makes multi-camera claims comparable across works; stated plainly, it also yields capability that obligates governance, and the concept treats that obligation as part of the specification rather than an afterthought, in line with the view that governance functions are institutional standards for auditable and defensible deployment rather than documentation exercises (Ominyi et al., 2026b; Annan, 2024).

References

- Adebayo, A., Adegbite, M. P., & Ahmed, M. O. (2022). Adversarial machine learning in critical infrastructure: A conceptual framework for threat modeling AI enabled OT systems. *World Journal of Innovation and Modern Technology*, 6(1), 184–234.
- Adebayo, A., Adegbite, M. P., & Ahmed, M. O. (2023). AI augmented threat detection in industrial control systems: A systematic review of machine learning approaches for ICS anomaly detection. *International Journal of Engineering and Modern Technology*, 9(3), 287–340.
- Adebayo, A., Anunagba, C. O., & Ozowara, D. E. (2025). A review of zero trust security models and cost effectiveness in healthcare infrastructure. *International Journal of Advanced Multidisciplinary Research and Studies*, 5(6), 2269–2283.
- Adeniyi, A. I., Odejebi, O., & Taiwo, T. (2025). Countermeasures against bias and spoofing in modern facial recognition systems. *World Journal of Advanced Research and Reviews*, 25(1), 1914–1930.
- Adeyelu, O. O. (2023). A risk-tiered oversight model for unmanned aircraft operations in shared controlled airspace over Nigerian airports. *International Journal of Scientific Research in Civil Engineering*, 7(4), 147–184.
- Adeyelu, O. O., & Dagodzo, D. (2024). Advances in artificial intelligence and data-driven wildlife hazard monitoring and incident reduction at international airports in West Africa. *International Journal of Scientific Research in Science and Technology*, 11(5), 872–910.
- Afrihiya, E., Ojukwu, P. U., & Akinse, S. G. (2024). Privacy-preserving health data governance models: A comparative review of blockchain and cryptographic strategies in U.S. and developing healthcare systems. *International Journal of Advanced Multidisciplinary Research and Studies*, 4(6), 3071–3086.
- Ahmed, M. O., Adegbite, M. P., & Adebayo, A. (2021). Zero trust architecture for operational technology in North American critical infrastructure: A framework for implementation and resilience optimization. *International Journal of Engineering and Modern Technology*, 7(1), 66–113.
- Akanbi, O., & Sunday, E. A. (2024). An integrated path-planning and slotting optimization model for AMR-enabled high-density warehousing. *International Journal of Advanced Multidisciplinary Research and Studies*, 4(6), 3226–3243.
- Akanbi, O., & Sunday, E. A. (2025). A systematic review of AI-driven autonomous mobile robots (AMR) in scalable micro-fulfillment centers. *International Journal of Advanced Multidisciplinary Research and Studies*, 5(6), 2363–2379.
- Akanbi, O., Ganiu, O. S., & Sunday, E. A. (2025). A multi-agent AI framework for swarm intelligence in autonomous mobile robot (AMR) fleet coordination. *International Journal of Scientific Research in Humanities and Social Sciences*, 2(1), 81–109.
- Aliliele, C., Mbonu, I. S., & Iwuanyanwu, U. (2023a). A conceptual framework for continuous cloud misconfiguration monitoring and enterprise risk mitigation strategies. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 9(10), 373–394.
- Aliliele, C., Mbonu, I. S., & Iwuanyanwu, U. (2023b). A review of API governance and risk prioritization frameworks in modern financial institutions. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 9(10), 395–433.
- Aliliele, C., Mbonu, I. S., & Iwuanyanwu, U. (2024a). A conceptual framework for enterprise data sensitivity classification and regulatory traceability mechanisms. *International Journal of Advanced Multidisciplinary Research and Studies*, 4(6), 3103–3124.
- Aliliele, C., Mbonu, I. S., Uzoka, E., & Iwuanyanwu, U. (2025b). Advances in data lakehouse governance architectures for enterprise data loss prevention and compliance assurance. *Shodhshauryam, International Scientific Refereed Research Journal*, 8(4), 171–213.
- Anioke, S. C., & Atima, M. E. (2023c). Public health informatics frameworks for protecting vulnerable populations through data-driven policy enforcement. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2564–2579.

- Annan, A. O. (2022). Data privacy governance models for cross-border digital platforms. *International Journal of Multidisciplinary Research and Growth Evaluation*, 3(6), 949–964.
- Annan, A. O. (2024). Algorithmic accountability and trade secret protection in artificial intelligence. *Shodhshauryam, International Scientific Refereed Research Journal*, 7(5), 315–347.
- Annan, A. O. (2025a). Automated decision-making and anti-discrimination compliance under U.S. law. *International Journal of Advanced Multidisciplinary Research and Studies*, 5(2), 2522–2540.
- Asiedu, W., & Quainoo, R. (2023b). Toward maintenance free wireless infrastructure: Simulating performance and reliability in large scale intermittently powered IoT networks. *Gyanshauryam, International Scientific Refereed Research Journal*, 6(1), 489–510.
- Asiedu, W., & Quainoo, R. (2024). Rethinking energy, reliability, and latency trade-offs in green communication for next generation IoT. *International Journal of Multidisciplinary Research and Growth Evaluation*, 5(6), 1987–1994.
- Basnet, A., Oghenemaiga, E., & Anene, U. N. (2023). Real time analytics and monitoring systems for livestream media performance using Google platforms. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10(1).
- Bergmann, P., Meinhardt, T., & Leal-Taixé, L. (2019). Tracking without bells and whistles. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 941–951.
- Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. *EURASIP Journal on Image and Video Processing*, 2008, 246309.
- Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. *Proceedings of the IEEE International Conference on Image Processing*, 3464–3468.
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Borah, S., Aliliele, K. C., Rakshit, S., & Vajjhala, N. R. (2022). Applications of artificial intelligence in software testing. In P. K. Mallick et al. (Eds.), *Cognitive informatics and soft computing (Lecture Notes in Networks and Systems, Vol. 375, pp. 727-736)*. Springer.
- Chavdarova, T., Baqué, P., Bouquet, S., et al. (2018). WILDTRACK: A multi-camera HD dataset for dense unscripted pedestrian detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5030–5039.
- Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 539–546.
- Dagodzo, D., & Ahiaeké Patrick, M. C. (2021a). A review of GIS applications in utility asset management and infrastructure planning. *Iconic Research and Engineering Journals*, 5(3), 468–492.
- Dagodzo, D., & Ahiaeké Patrick, M. C. (2021b). An integrated framework for UAV, LiDAR, and GIS in infrastructure corridor management. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 7(5), 497–524.
- Dagodzo, D., & Ahiaeké Patrick, M. C. (2022). A review of right-of-way encroachment detection methods using geospatial technologies. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 8(1), 638–667.
- Dagodzo, D., Ahiaeké Patrick, M. C., & Aliliele, C. (2022). AI and deep learning for vegetation classification in power corridor management: A review. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 8(1), 668–698.
- Dendorfer, P., Rezatofighi, H., Milan, A., et al. (2020). MOT20: A benchmark for multi object tracking in crowded scenes. *arXiv preprint arXiv:2003.09003*.
- Dosunmu, A. A., & Ogundele, P. O. (2024a). Breach and attack simulation frameworks for continuous validation of enterprise security controls. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10(3), 1100–1119.
- Ejofodomi, O. A., Gideon, E. N., Oladipo, G. O., & Oshomah, E. R. (2014). Automated detection of architectural detection in mammograms using template matching. *International Journal of Biomedical Science and Engineering*, 2(1), 1–6.
- Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. *arXiv preprint arXiv:2107.08430*.
- Hadsell, R., Chopra, S., & LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1735–1742.
- Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
- Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*.
- Jimoh, H. O., Abolle-Okoyeagu, C. J., Ahmed, M. O., & Lawal, N. O. (2023). Advancing security in IoT-driven critical infrastructure: A focus on smart transportation system. *American Journal of Engineering Research*, 12(12), 33–46.
- Johnson, J., Douze, M., & Jégou, H. (2021). Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3), 535–547.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45.
- Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2), 83–97.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988.

- Luiten, J., Osep, A., Dendorfer, P., et al. (2021). HOTA: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, 129(2), 548–578.
- Luo, H., Gu, Y., Liao, X., Lai, S., & Jiang, W. (2019). Bag of tricks and a strong baseline for deep person re-identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 1487–1495.
- Malkov, Y. A., & Yashunin, D. A. (2020). Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4), 824–836.
- Mayo, W., Ogbale, J. I., Okoruwa, P. O., & Babatope, O. M. (2021). Designing an AI-predictive maintenance model for e-commerce systems using machine learning and cloud analytics. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 7(5), 416–440.
- Mbonu, I. S., Aliliele, C., Iwuanyanwu, U., & Oluoha, O. M. (2018). A conceptual framework for legal and ethical risk modeling in enterprise data protection governance systems. *Iconic Research and Engineering Journals*, 2(2), 207–226.
- Mbonu, I. S., Aliliele, C., Iwuanyanwu, U., & Uzoka, E. (2020). A review of identity and access management integration strategies in hybrid and multi cloud environments. *International Journal of Multidisciplinary Research and Growth Evaluation*, 1(5), 795–810.
- Mbonu, I. S., Aliliele, C., Iwuanyanwu, U., & Uzoka, E. (2021b). Advances in artificial intelligence techniques for secure software testing and automated regression control mechanisms. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 7(5), 468–496.
- Mbonu, I. S., Iwuanyanwu, U., Aliliele, C., & Uzoka, E. (2022a). A review of data protection impact assessment models in multi cloud financial infrastructure systems. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 8(1), 589–623.
- Mbonu, I. S., Iwuanyanwu, U., Aliliele, C., & Uzoka, E. (2022b). Advances in cloud identity and access governance optimization in large scale AWS enterprise environments. *Shodhshauryam, International Scientific Refereed Research Journal*, 5(3), 403–438.
- Mbonu, I. S., Iwuanyanwu, U., Uzoka, E., & Oluoha, O. M. (2019). Advances in enterprise log analytics and automated incident response architectures using Python and SIEM platforms. *Iconic Research and Engineering Journals*, 3(2), 1000–1019.
- Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. *arXiv preprint arXiv:1603.00831*.
- Nwakamma, S., Ojukwu, J., & Oyesiji, S. O. (2024a). LLM-as-a-judge for automated model governance: A review of methods, biases, and release-gating practices. *World Journal of Innovation and Modern Technology*, 8(6), 185–216.
- Nwakamma, S., Ojukwu, J., & Oyesiji, S. O. (2025). On-device multimodal small language models for privacy-preserving edge intelligence through quantization, pruning, distillation, and energy-efficient inference. *World Journal of Innovation and Modern Technology*, 9(12), 271–302.
- Obogo, S. F., Ozobu, C. O., Garba, B. M. P., & Adio, S. A. (2026). Predictive safety analytics model for early detection of high-risk construction activities. *Global Journal of Engineering and Technology Review*, 2(3), 92–109.
- Obriki, O. D., & Arumosoye, O. M. (2023). Conceptual framework for proactive hazard identification using digital safety data streams. *Gyanshauryam, International Scientific Refereed Research Journal*, 6(3), 457–481.
- Obriki, O. D., Arumosoye, O. M., & Obogo, S. F. (2023). Advances in continuous hazard monitoring systems for workplace safety. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2742–2759.
- Obriki, O. D., Obogo, S. F., & Arumosoye, O. M. (2022b). Review of emergency preparedness and evacuation systems in high density commercial buildings. *Gyanshauryam, International Scientific Refereed Research Journal*, 5(6), 411–438.
- Odejide, G. S. (2024). A trust-centered conceptual model for AI-driven multilingual emergency communication. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10(6), 2781–2811.
- Odejobi, O. D., Okonkwo, C. S., Ahiaeké Patrick, M. C., Okeke, O. T., & Mayo, W. (2025). AI-augmented secure software engineering: Leveraging deep learning for autonomous threat detection and mitigation. *International Journal of Engineering and Modern Technology*, 11(12), 101–121.
- Ogbale, J. I., Okoruwa, P. O., Babatope, O. M., & Oyewole, T. (2021). Developing an integrated data visualization model for continuous business performance monitoring and optimization. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 7(5), 441–467.
- Ojukwu, J., Oyesiji, S. O., & Nwakamma, S. (2025). WebAssembly component model and WASI for portable cloud-to-edge software: Language interoperability, security, performance, and container comparisons. *International Journal of Computer Science and Mathematical Theory*, 11(12), 232–263.
- Ominyi, M. (2021). Toward a conceptual framework for embedding privacy by design in corporate restructuring and transaction structuring. *Journal of Accounting and Financial Management*, 7(5), 152–173.
- Ominyi, M., & Anichukwueze, C. C. (2022a). A conceptual framework for harmonizing data protection compliance across GDPR, NDPR, and United States privacy regimes. *International Journal of Social Sciences and Management Research*, 8(5), 148–174.
- Ominyi, M., & Anichukwueze, C. C. (2024). Conceptualizing responsible AI governance structures for automated decision systems in regulated industries. *International Journal of Social Sciences and Management Research*, 10(11), 403–430.
- Ominyi, M., Anichukwueze, C. C., & Uzougbo, N. S. (2024b). A systematic review of organizational readiness for the EU Artificial Intelligence Act in multinational enterprises. *World Journal of Innovation and Modern Technology*, 8(6), 185–203.
- Ominyi, M., Anichukwueze, C. C., & Uzougbo, N. S. (2026b). Conceptualizing enterprise AI governance functions as institutional standards for auditable and defensible risk management. *Journal of Accounting and Financial Management*, 12(7), 326–356.

- Onche, V. O., Adegbite, M. P., & Ogbonna, C. S. (2026). Human factors in information security: A capability framework for administrative professionals in the digital workplace. *World Journal of Innovation and Modern Technology*, 10(5), 154–179.
- Onche, V. O., Ogbonna, C. S., & Adegbite, M. P. (2024). Effectiveness of cybersecurity awareness training on insider-threat behavior among university administrators: An integrative academic review. *International Journal of Computer Science and Mathematical Theory*, 10(2), 117–144.
- Oparah, O. S., Ezeh, F. E., Olatunji, G. I., & Ajayi, O. O. (2023). Framework for designing national real-time disease surveillance dashboards for public health stakeholders. *Shodhshauryam, International Scientific Refereed Research Journal*, 6(1), 208–227.
- Orise, C., & Niniola, S. (2023). Privacy, safeguarding, and ethical data practices in healthcare EdTech: A conceptual model for compliance and trust. *Shodhshauryam, International Scientific Refereed Research Journal*, 6(1), 507–529.
- Orise, C., & Niniola, S. (2025). Data governance in digital health learning systems: A systematic review of implementation frameworks and best practices. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 11(5), 520–538.
- Oyeleke, A. V., Eze, F. C., & Asiedu, W. (2026). Reliability-centered maintenance strategies for minimizing downtime and maximizing performance in high-density GPU cluster environments. *International Journal of Engineering Technology Research & Management*, 10(7), 18–38.
- Oyesiji, S. O., Nwakamma, S., & Ojukwu, J. (2023). Parameter-efficient multilingual adaptation of on-device language models: A survey. *International Journal of Engineering and Modern Technology*, 9(3), 287–320.
- Oyesiji, S. O., Nwakamma, S., & Ojukwu, J. (2024). A conceptual framework for context-aware data curation in fine-tuning pipelines. *International Journal of Engineering and Modern Technology*, 10(11), 197–228.
- Oyesiji, S. O., Nwakamma, S., & Ojukwu, J. (2025). Compressing recommender and ranking models for edge deployment: A survey. *International Journal of Engineering and Modern Technology*, 11(12), 205–237.
- Oziri, S. T., Arowogbadamu, A. A. G., & Seyi-Lande, O. B. (2020). Predictive analytics applications in reducing customer churn and enhancing lifecycle value in telecommunications markets. *International Journal of Multidisciplinary Futuristic Development*, 1(2), 40–49.
- Quainoo, R., Ogundapo, O., & Asiedu, W. A. (2025a). Predicting throughput degradation in Wi-Fi 6 networks under varying channel conditions: A physical layer performance model. *International Journal of Engineering and Modern Technology*, 11(12), 205–264.
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28, 91–99.
- Ristani, E., & Tomasi, C. (2018). Features for multi-target multi-camera tracking and re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6036–6046.
- Ristani, E., Solera, F., Zou, R., Cucchiara, R., & Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. *Proceedings of the European Conference on Computer Vision Workshops*, 17–35.
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 815–823.
- Seyi-Lande, O. B., Arowogbadamu, A. A. G., & Oziri, S. T. (2020). Geomarketing analytics for driving strategic retail expansion and improving market penetration in telecommunications. *International Journal of Multidisciplinary Futuristic Development*, 1(2), 50–60.
- Shittu, H., & Nabil, M. (2023). Smart supply chain management with attribute-based encryption access control. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 198–204). IEEE.
- Tang, S., Andriluka, M., Andres, B., & Schiele, B. (2017). Multiple people tracking by lifted multicut and person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3539–3548.
- Tonoyan, A., Dada, O., & Ayivi-Donkor, S. S. (2022a). Advances in geospatial analysis: Predictive analytics methods for store location selection and market entry return on investment. *International Journal of Marketing and Communication Studies*, 6(2), 78–105.
- Tonoyan, A., Dada, O., & Ayivi-Donkor, S. S. (2024b). Real-time KPI tracking systems: A review of automated performance monitoring and data-driven decision making. *World Journal of Innovation and Modern Technology*, 8(6), 247–281.
- Uduokhai, D. O., Giloid, S., Nwafor, M. I., & Adio, S. A. (2023a). Evaluating the role of building information modeling in enhancing project performance in Nigeria. *International Journal of Advanced Multidisciplinary Research and Studies*, 3(6), 2154–2161.
- Ussher-Eke, D., Okoh, O. F., & Ijiga, O. M. (2025). The role of biometric and IoT-based attendance systems in streamlining HR administrative functions, enhancing workforce accountability, and reducing labor inefficiencies. *International Journal for Multidisciplinary Research*, 7(4).
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7464–7475.
- Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *Proceedings of the IEEE International Conference on Image Processing*, 3645–3649.
- Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. H. (2021). Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6), 2872–2893.

- Yusuf, M., Akintunde, A., Shittu, H., & Quadri, A. (2023). Radio frequency identification (RFID) based library management system. *International Journal of Technology and Systems*, 8(2), 21–35.
- Zhang, Y., Sun, P., Jiang, Y., et al. (2022). ByteTrack: Multi-object tracking by associating every detection box. *Proceedings of the European Conference on Computer Vision*, 1–21.
- Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2021). FairMOT: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 129(11), 3069–3087.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.
- Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: A benchmark. *Proceedings of the IEEE International Conference on Computer Vision*, 1116–1124.
- Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*.
- Zhou, K., Yang, Y., Cavallaro, A., & Xiang, T. (2019). Omni-scale feature learning for person re-identification. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3702–3712.
- Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3), 257–276.