Supplementary data

In response to reviewer concerns about the interpretation of performance metrics, Table S1 compiles all quantitative results cited in this review. Each entry specifies the reported value, its original source, and the experimental setting, whether laboratory, simulation, or real-world deployment. This helps clarify the conditions under which results were obtained and discourages unwarranted generalization.

Table S1. Reported performance values from cited studies and their experimental context

Study / Reference	Reported Performance Value(s)	Context (Lab Demo, Simulation, Real-world Deployment)	Notes / Clarifications
Huang et al. [43] – Spike camera "super vision system"	Microsecond temporal resolution; tracked objects at 30 m/s within 0.75 m, Mach 1 within 10 m, Mach 100 within 1 km	Lab Demo (controlled experiments with table-tennis balls and rotating fans)	Proof-of-concept only; real-world deployment at such speeds remains untested
Li et al. [104] – Bionic transistor for adaptive vision	Microsecond-level perception; response times 108 μs (scotopic), 268 μs (photopic); >98% recognition accuracy in dim/bright conditions	Lab Demo / Prototype	Prototype-level validation under controlled illumination
Penza et al. [128] – EnViSoRS robotic surgery	Surgical accuracy <5 mm; AR feature updates at 4 fps; high precision and recall	Lab Demo (phantom surgery)	Evaluated on phantoms and simulations, not yet in clinical deployment
Wei et al. [135] – YOLOv5 carton recognition in cold storage	mAP +0.7%; FPS +0.7%; fidelity +2.3%; response time –13.9%; positioning accuracy +7.1%	Lab Demo (experimental testbed)	Evaluated with PC + camera setup in controlled cold storage; values may differ in real-world scaling
Ghosal et al. [151] – Plant stress detection with DL	94.13% accuracy	Dataset / Controlled Imaging	Reported on controlled datasets; field performance may vary
Bai et al. [168] – Quantum SEQNN	87.56% accuracy on MNIST benchmark	Simulation	Quantum prototype simulation on benchmark dataset; exploratory stage
Li et al. [104], Dodda et al. [105] – Adaptive MV systems	>98% accuracy in image recognition across variable lighting	Lab Demo	Demonstrated on lab prototypes; field robustness remains open
Zhang et al. [141] – Construction site safety monitoring	97.4% increase in management efficiency	Simulation / Pilot test	Needs clarification of baseline and deployment environment

Building on this detailed inventory, Table S2 provides a domain-level synthesis, grouping findings by application area to highlight common tasks, prevailing methodologies, typical performance benchmarks, and emerging directions. This higher-level summary offers readers a streamlined overview that complements the granular evidence presented throughout the review.

Table S2. Domain-level synthesis of CV and MV applications. The table summarizes representative tasks, typical methodological approaches, performance ranges (as reported in the reviewed studies), and emerging trends.

Domain	Representative Tasks	Typical Methods	Reported Performance Ranges	Emerging Trends
Healthcare	Medical imaging analysis (tumor detection, surgical navigation, pathology classification)	CNNs, Transformers, hybrid DL models; edge–cloud integration	Accuracy ~85–97% depending on dataset/task; near-real-time feasible in optimized edge/cloud setups	Growing demand for explainability (XAI); federated and privacy-preserving learning
Manufacturing / Industrial	Defect detection, quality inspection, assembly verification, process monitoring	Classical MV (thresholding, template matching) combined with DL (CNNs, GANs); multimodal sensing	Detection accuracy ~90–99% in structured environments; inference speeds >30 FPS in industrial settings	Shift toward edge deployment; human-in-the-loop QA; integration with robotic systems.
Autonomous Vehicles & Robotics	Object detection, localization, SLAM, gesture, and action recognition	Deep CNNs, Transformers, reinforcement learning; SNNs for low-latency tasks; multimodal fusion (vision + lidar/radar)	Object detection mAP ~70–90%; latency <50 ms in optimized hardware; real-time operation at 30–60 FPS	Neuromorphic/edge processors for latency-critical perception; move toward standard benchmarks and reproducibility
Security & Surveillance	Person re-identification, anomaly detection, crowd monitoring, intrusion detection	CNNs, attention-based architectures, GNNs; emphasis on real-time tracking	Re-ID accuracy ~80–95%; anomaly detection AUROC ~0.85–0.95; real-time tracking at 25–40 FPS	Addressing bias/fairness; privacy-preserving vision; cross-domain generalization