

Review

# A Comprehensive Survey and Reference Architecture for AI-Powered Autonomous Drone Systems in Smart Cities

Mustafa Qassab <sup>1,\*</sup> and Qutaiba I. Ali <sup>2</sup><sup>1</sup> Computer Engineering Department, College of Engineering, University of Mosul, Mosul 00964, Iraq<sup>2</sup> Mechatronics Engineering Department, College of Engineering, University of Mosul, Mosul 00964, Iraq; qutaibaali@uomosul.edu.iq (Q.I.A.)

\* Corresponding author. E-mail: mustafa.qassab@uomosul.edu.iq (M.Q.)

Received: 20 March 2026; Revised: 14 May 2026; Accepted: 21 May 2026; Available online: 3 June 2026

**ABSTRACT:** Despite a rapid rise of AI-powered Unmanned Aerial Vehicle (UAV) deployments in smart city environments, current surveys and frameworks lack a unified, protocol-level reference architecture that integrates multi-domain applications, edge AI perception, cognitive reasoning through Large Language Models (LLMs), and regulatory compliance within a single deployable specification. This study presents a comprehensive cross-domain review of AI-powered drone systems for traffic management, delivery, infrastructure inspection, disaster response, and environmental monitoring. The study introduces COMPASS (Cognitive Operations Model for Programmable Autonomous Smart-city Systems), a novel seven-layer technical reference architecture that describes communication protocols (MAVLink 2.0, ROS2/DDS, MQTT 5.0, and NGSII-LD), edge computing hardware recommendations for five drone payload tiers, and quantified performance requirements for safety-critical operations. The key feature of COMPASS is its LLM-based Semantic Middleware Layer, which allows for context-aware decision-making, natural human-drone interaction, and regulatory compliance verification. Comparing COMPASS to many other frameworks reveals that it is the only architecture to simultaneously provide multi-domain coverage, protocol-level specifications, hardware recommendations, LLM integration, and empirically verified benchmarks.

**Keywords:** Unmanned aerial vehicles; Artificial intelligence; Smart cities; Reference architecture; Edge computing

## 1. Introduction

Combining Information and Communication Technology (ICT) with municipal services to enhance sustainability, operational efficacy, and quality of life, the concept of “smart cities” marks a significant shift in urban design [1,2]. In this technological context, Unmanned Aerial Vehicles (UAVs) have evolved into ground-breaking tools capable of doing tasks that were previously expensive, dangerous, or impractical [2,3]. Intelligent aerial systems that can perceive their environment, make decisions on their own, and do



difficult tasks with little human intervention are produced when AI capabilities are integrated with drone platforms [3].

Despite the fast advancement of UAV deployments in smart cities, three critical gaps hinder systematic development, starting with ignoring cross-domain integration or concentrating on certain application domains [4,5]. In addition to handling specific technological issues, such as object identification, without taking system-level architecture into account [6,7], or providing abstract conceptual frameworks devoid of hardware implementation guidelines and protocol-level requirements [8,9]. Furthermore, no existing work integrates LLMs as a semantic middleware for autonomous drone systems in smart city contexts [10,11]. These gaps motivate the present comprehensive survey and the proposed COMPASS reference architecture.

The principal contributions of this paper are as follows:

1. A comprehensive, cross-domain survey of AI-powered drone systems spanning delivery, infrastructure inspection, traffic management, disaster response, environmental monitoring, and urban planning, synthesizing findings from 87 sources with emphasis on publications from 2020–2025.
2. The introduction of COMPASS (Cognitive Operations Model for Programmable Autonomous Smart-city Systems), a novel seven-layer technical reference architecture that goes beyond existing conceptual frameworks [12,13] by specifying concrete communication protocols (MAVLink 2.0, ROS2/DDS, NGS-LD, MQTT 5.0, ASTM F3411-22a), edge computing hardware recommendations (NVIDIA Jetson Orin, Qualcomm Flight RB5, Google Coral), and quantified performance requirements for safety-critical operations.
3. The integration of a Semantic Middleware Layer based on edge-deployable Large Language Models (LLM) [10,11,14], which is a novel architectural contribution that enables context-aware decision-making, regulatory compliance verification, natural language human-drone interaction, and structured report generation, capabilities not addressed by any existing drone architecture in the literature.
4. A detailed comparison of the COMPASS architecture against existing surveys and frameworks, demonstrating that it is the first to simultaneously address multi-domain coverage, protocol-level specifications, hardware recommendations, LLM integration, and performance benchmarking.

In this work, a comprehensive literature review is explored. Relevant articles were found by searching IEEE Xplore, Scopus, Web of Science, and Google Scholar using keywords like “UAV smart city”, “drone AI architecture”, “autonomous drone delivery”, and “LLM robotics”. Publications from 2020 to 2025 were highlighted, with foundational works added where necessary, to reflect the current state of the art. Technical publications and industry reports were used to collect deployment data in the real world.

### *Problem Statement*

Despite the increasing rise of AI-powered UAV deployments in smart city areas, three major limitations in the industry limit systematic expansion and broad adoption:

First, current surveys categorise drone applications (such as delivery, inspection, and traffic monitoring) separately without putting forward a unified architectural framework that unifies all application areas under a single technical reference model [3,5,9,15]. Instead of focusing on specific deployment scenarios or specialised AI techniques (such item identification algorithms), current research lacks the broad system-level vision required for smart city integration.

Secondly, there is no standard reference architecture that connects deployable, protocol-level specifications with theoretical AI-drone frameworks. Current architectural ideas are still abstract and lack the specific hardware platforms, communication protocols, and performance standards needed for practical deployment [8,12,13,16].

Third, few studies have been done on the growing use of Large Language Models (LLMs) as a cognitive and semantic layer in autonomous drone systems. While LLMs have shown significant potential

in robotics and autonomous systems [10,11,14], their integration into drone designs for semantic reasoning, human-drone interaction, and regulatory compliance has not yet been investigated in the reviewed literature.

This paper presents COMPASS (Cognitive Operations Model for Programmable Autonomous Smart-city Systems), a novel seven-layer technical reference architecture that provides hardware recommendations, protocol-level specifications, and an LLM-based semantic middleware layer for intelligent urban drone operations. Additionally, it provides a comprehensive overview of AI-powered drone systems across all significant smart city application domains.

While this survey necessarily covers multiple application domains to provide architectural completeness, the central narrative of this paper is driven by three core themes:

1. The architectural gap between conceptual UAV frameworks and deployable, protocol-level specifications.
  2. The novel role of LLMs as semantic middleware in safety-critical autonomous systems.
  3. The empirical validation of edge AI performance under COMPASS-specified hardware constraints.
- All application domains are analyzed through this unified lens, with the COMPASS architecture serving as the integrating framework.

## 2. Related Work and Existing Approaches

This section offers a systematic analysis of current frameworks, designs, and surveys for AI-powered drone systems in smart cities, arranged into four subject groups. The focus is on publications from 2020–2025 to capture the current state of the art.

### 2.1. Surveys on UAV Applications in Smart Cities

An early vision of UAV uses in future smart cities was offered by Mohamed et al. [1], who identified important sectors such as emergency services, transportation, and surveillance. Nevertheless, neither AI integration nor a technological architecture was suggested in their study. In their assessment of IoT and collaborative smart drones for enhancing smart city operations, Alsamhi et al. [17] concentrated on cooperative behaviors and communication, but they did not discuss deep learning-based perception or LLM integration. Drones were broadly categorized by design and use by Hassanalian and Abdelkefi [3], but they did not discuss AI technologies or smart city deployment frameworks. The legislative and operational elements of drone applications for smart cities were examined by Důbravová et al. [4], who noted the necessity for integrated system designs but did not provide one. Shah et al. [9] investigated UAV issues and blockchain-based solutions for smart cities, while Ezzat et al. [5] carried out a horizontal study of video surveillance for smart cities utilizing edge devices. Both studies provided domain-specific insights but lacked cross-domain architectural integration.

In a more recent analysis of UAV advancements at the beginning of the 2020s, Nex et al. [18] outlined anticipated trends in sensing, mapping, and autonomy, but did not discuss LLMs or cognitive architectures. One of the few studies to address XAI in the drone environment, Javaid et al.'s [19] investigation of explainable AI and monocular vision for UAV navigation in smart cities was restricted to navigation concerns. Garg et al. [20] proposed Drones-as-a-Service (DaaS) integrated with 5G and blockchain for IoT-based smart city infrastructure, addressing communication layers but not perception or cognitive AI.

### 2.2. Deep Learning-Based Perception for UAVs

Object detection is foundational to intelligent drone systems. Redmon et al. [21] introduced YOLO, which enabled real-time detection; subsequent iterations, including YOLOv4 [22] and YOLO-based UAV detectors [6,23] improved both speed and accuracy for aerial imagery. Hossain and Lee [6] demonstrated GPU-based real-time multi-object detection and tracking from aerial imagery. More recently, Wang et al. [24] proposed CF-YOLO based on YOLOv11 for small target detection in drone imagery, achieving a

+12.7% mAP50 improvement on the VisDrone dataset. Zhang [25] developed Drone-YOLO with multi-scale detection heads specifically optimized for UAV perspectives. Alshaer et al. [26] combined deep learning with Kalman filtering for UAV detection and tracking. Lin et al. [7] introduced focal loss for dense object detection, which became foundational for addressing class imbalance in aerial images. Merza et al. [27] demonstrated the wider application of these approaches by using deep learning to identify fake data injection in smart grids. CNN-based real-time object identification and tracking for commercial drones was proven by Rohan et al. [28].

Despite these works, most studies address perception in isolation without integrating detection models into comprehensive drone system architectures with edge computing constraints, communication protocols, and cognitive reasoning layers.

### 2.3. Existing Drone System Architectures and Frameworks

Several architectural frameworks have been proposed for drone systems. Garlan and Shaw [12] established foundational software architecture principles that inform modular drone system design. Shi et al. [13] introduced edge computing paradigms that address the latency constraints of autonomous systems. The MAVLink protocol [29] provides standardized communication for autopilot systems, and Ibrahim and Qassab [30] discussed self-organizing network concepts relevant to drone swarm coordination. Vemprala et al. [14] explored ChatGPT for robotics, establishing design principles for LLM integration with autonomous systems. Wei et al. [10] demonstrated chain-of-thought prompting for reasoning in LLMs, a foundational component of the cognitive layer in drone architectures. Bommasani [11] provided a comprehensive analysis of foundation model opportunities and risks relevant to autonomous systems. Table 1 provides a systematic comparison of 8 existing frameworks against COMPASS across 6 evaluation dimensions, confirming that no existing work simultaneously addresses all dimensions.

However, none of these works propose a complete, multi-layer reference architecture that integrates physical platforms, edge perception, semantic middleware (LLMs), communication protocols, city infrastructure integration, and human interaction within a unified framework with quantified performance requirements. This is the gap that COMPASS addresses.

**Table 1.** Comparative summary of the literature review (✓ = Fully Addressed, Partial = Partially Ad-dressed, ✗ = Not Addressed).

Study	Multi-Domain Survey	System Arch.	Protocol Specs	Edge HW Specs	LLM Integ.	Perf. Bench.
Mohamed et al. (2020) [1]	✓	✗	✗	✗	✗	✗
Alsamhi et al. (2019) [17]	✓	Partial	✗	✗	✗	✗
Důbravová et al. (2024) [4]	✓	✗	✗	✗	✗	✗
Shah et al. (2024) [31]	Partial	✗	✗	✗	✗	✗
Javaid et al. (2025) [19]	✗	✗	✗	✗	✗	✗
Garg et al. (2024) [20]	✗	Partial	Partial	✗	✗	✗
Nex et al. (2022) [18]	✓	✗	✗	✗	✗	✗
Vemprala et al. (2024) [14]	✗	✗	✗	✗	Partial	✗
This Paper (COMPASS)	✓	✓	✓	✓	✓	✓

## 3. Artificial Intelligence Technologies for Drone Systems

### 3.1. Deep Learning for Object Detection

Intelligent drone perception is based on object detection, which allows UAVs to recognize and follow people, cars, and objects in real time [21,28]. Two-stage detectors and one-stage detectors are the two main designs into which deep learning-based object identification techniques fall [6,7,27].

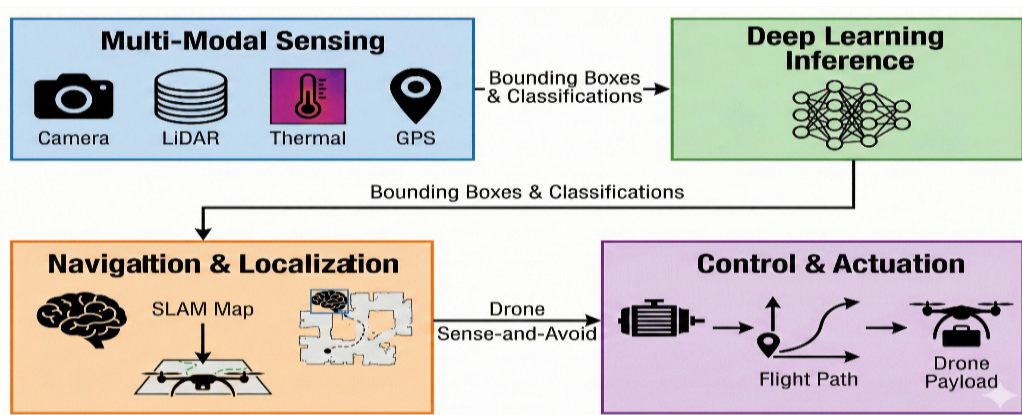
Bounding box regression and classification come after the generation of region suggestions by two-stage detectors like Faster R-CNN [32]. Despite their great precision, they are less appropriate for real-time

drone applications due to their computing overhead. Important features include high accuracy in detecting small objects, real-time performance is limited by computational complexity, and suitability for offline drone imagery analysis.

One-stage detectors greatly increase inference speed by directly predicting bounding box locations and object classifications [21]. The most popular one-stage framework for drone applications is the YOLO (You Only Look Once) series [6,23]:

- YOLOv3/v4: Used extensively in surveillance systems, it strikes a balance between speed and accuracy [22].
- YOLOv5/v7: Better feature pyramid networks for enhanced tiny object identification [26].
- YOLOv8/v11: state-of-the-art results on drone datasets with mAP50 increases of 10–12% [24,33].
- Drone-YOLO: A specialized version with multi-scale detecting heads that is tuned for UAV images [25].

The process combines autonomous navigation modules, Deep Learning perception (YOLO), and raw sensor data, see Figure 1.



**Figure 1.** End-to-End AI Processing Pipeline for Autonomous Drones.

### 3.2. Autonomous Navigation and Path Planning

Autonomous drone navigation requires the combination of many sensor modalities and AI-driven decision-making systems [8,15,34]. Modern navigation systems integrate:

1. GPS and Inertial Navigation Systems (INS): Use RTK (Real-Time Kinematic) corrections to provide positioning accuracy within centimeters [35].
2. LiDAR and Depth Sensors: Provide obstacle detection and 3D environmental mapping [36].
3. Computer Vision: Stereo and monocular camera systems for scene comprehension and visual odometry [19].
4. Machine Learning Models: Adaptive route planning in dynamic situations using reinforcement learning [37].

NVIDIA Jetson edge AI devices, which evaluate sensor inputs to develop sound detection and avoidance abilities that enable safe operation in shared airspace, are used by Zipline's autonomous delivery drones to navigate [38]. Two Jetson Orin NX modules are used in their P2 platform: one for environmental understanding and sensor fusion, and another for redundancy and safety in the delivery droid [39].

### 3.3. Edge Computing and Real-Time Processing

Drones with AI capabilities need edge computing systems that can analyze data on-board instead of depending on a cloud connection to meet their computational demands [16,20]. For drone applications, the following are important edge computing systems:

- The NVIDIA Jetson Series provides up to 275 TOPS for autonomous systems with the Xavier NX and Orin modules [38].

- Google Coral TPU: TensorFlow Lite models with energy-efficient inference acceleration [40].
- Intel Movidius: Vision Processing Units (VPUs) designed for computer vision tasks [41].

### 3.4. Large Language Models (LLMs) for Cognitive and Semantic Drone Operations

While deep learning models such as CNNs, YOLO, and SLAM algorithms provide perception and control, Large Language Models (LLMs) enable semantic understanding, reasoning, orchestration, and human-system interaction in AI-powered drone ecosystems [10]. Instead of replacing low-level autonomous modules, LLMs serve as high-level cognitive controllers by integrating many data sources, human intent, rules, and mission objectives into logical operational actions [11].

Because LLMs can interpret natural language commands from human operators and city management systems, reason over symbolic information such as safety constraints, mission rules, and regulations, coordinate multi-agent drone fleets and edge/cloud services, and generate structured plans, explanations, and reports from raw sensor data and analytics outputs, they are particularly well suited for smart city drone systems [14,42].

Under this paradigm, LLMs operate at the cognitive and semantic layer, enabling explainable system behavior, adaptive mission planning, and context-aware decision-making, but vision and sensor models still control perception.

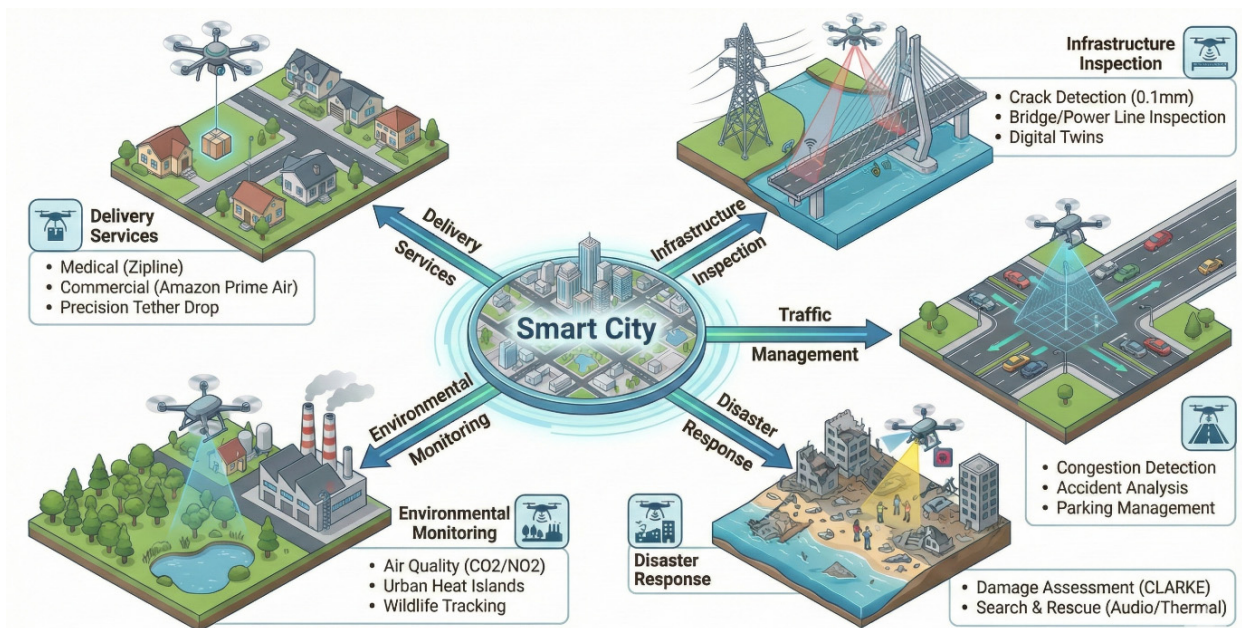
The design selections of COMPASS are directly influenced by the AI technologies discussed in Sections 3.1–3.4. TensorRT-optimized YOLO models operate on specialized AI accelerators in COMPASS Layer 3 (Edge Perception Layer), where deep learning-based object identification (Section 3.1) is instantiated. Layer 2 (the Flight Control Layer) and Layer 3 are where autonomous navigation algorithms (Section 3.2) function. The hardware specification levels in Section 5.4 are determined by the edge computing hardware (Section 3.3). Lastly, the Semantic Middleware (Layer 4), which is explained in Section 5.2, is an architectural formalization of LLMs (Section 3.4). This mapping guarantees that each COMPASS layer is based on a validated AI paradigm rather than a theoretical assumption.

Even with these advantages, LLMs pose serious risks in situations where safety is crucial. An LLM misinterpreting a regulatory limitation or sensor abnormality might lead to dangerous flying decisions, which presents a fundamental issue. Hallucination is the production of convincing but factually inaccurate outputs. Additionally, COMPASS strictly restricts LLM operation to Layer 4 (semantic reasoning, mission planning, reporting). It restricts LLM involvement in Layer 2 (flight control) or Layer 3 (collision avoidance), where deterministic algorithms operate within <50 ms latency budgets, because inference latency (10–38 tokens/second on edge hardware, Section 6.2) introduces delays incompatible with real-time flight control.

Prompt injection attacks, in which undesirable material (such as sensor feeds from illegitimate sources or adversarial language in scanned QR codes) manipulates LLM reasoning, are among the cybersecurity threats unique to LLM-integrated drones. By using cryptographic verification of all external data sources (AES-256 with dynamic session keys), input sanitization at the Layer 4 border, and sandboxed LLM execution environments that prevent file system or network access outside specified APIs, COMPASS minimizes this.

## 4. Application Domains in Smart Cities

There are many currently applicable applications that drones can be used for in smart cities. These areas of operation can be expanded as the smart city develops, and new applications could arise in the near future. Figure 2 illustrates the UAVs' key application domains and their specific use cases in smart cities.



**Figure 2.** Integrated Smart City Drone Ecosystem.

#### 4.1. Drone Delivery Services

One of the most important commercial uses of AI-powered UAVs is autonomous drone delivery; several companies have achieved operational deployment at scale [39,43–45].

With operations in seven nations, including Rwanda, Ghana, and the US, Zipline has built the world's largest autonomous drone delivery network [39,46]. Important operational metrics consist of:

- More than 800,000 deliveries have been made worldwide.
- Over 55 million miles flown independently.
- One distribution is made globally every 70 s.
- A ten-mile delivery radius in less than ten minutes.
- A 70 mph cruise speed and an 8 kg payload capacity.

The Platform 2 (P2) hybrid drone from Zipline enables accurate deliveries in congested urban areas by fusing VTOL (Vertical Take-Off and Landing) capabilities with fixed-wing flying efficiency [39]. By using a cable attached to a customized “Zip” robot, the system is able to place packages without landing with dinner-plate precision. Walmart, Sweetgreen, Cleveland Clinic, and Michigan Medicine are among the partners [44].

With operational installations in Lockeford, California, and College Station, Texas, Amazon Prime Air has been researching drone delivery since 2013 [43,47]. Features of Amazon's MK30 drone:

1. Intelligent sense-and-avoid systems using computer vision driven by AI.
2. A weatherproof design that can function in mild rain.
3. A lower noise profile, in contrast to earlier generations.
4. Integration with the infrastructure of the Amazon fulfillment network.

With U.S. operations supporting Walmart shops in the Dallas region, Wing (Alphabet) has conducted over 350,000 deliveries, mostly in Australia [44,48]. Wing's AI-powered platform enables drones to autonomously assess package placement when delivery sites are restricted.

## 4.2. Infrastructure Inspection and Maintenance

Infrastructure maintenance has been revolutionized by AI-powered drone inspection, which offers more secure, faster, and more economical evaluation techniques [49–51]. Benefits that have been documented include:

- Cost reduction: up to 74% less than with conventional manual inspection techniques [49].
- Safety improvement: removing human exposure to dangerous conditions reduces occupational accidents by 91% [49].
- Time efficiency: Compared to days for typical procedures, inspections were finished in two hours [50].
- Accuracy of detection: 95% of defects are detected using AI-powered analysis [50].

**Bridge Inspection:** During a significant bridge repair project in New York City, AI Engineers, Inc. (AIE, New York, NY, USA) used drones to take over 1000 high-resolution geotagged photos and three hours of 4K video in six hours of flight without the need to close lanes [51]. AI defect identification in conjunction with the DJI M350 RTK platform can identify corrosion, deformation, spalling, and fractures as small as 0.1 mm [52]. Drones powered by artificial intelligence have been effectively used in Europe to monitor aging bridges, averting costly failures and enhancing the long-term integrity of the infrastructure [53,54].

**Power Line and Utility Inspection:** Predictive maintenance that anticipates power line breakdowns months in advance is made possible by Deloitte's AI-powered remote sensing technology, which was shown at IFS Connect 2025 [55,56]. In order to achieve a 15–35% decrease in field hours, the system uses drones equipped with LiDAR sensors and AI models that can assess thousands of assets everyday as opposed to 10 poles that are physically examined every day [55]. Autonomous drone inspections of wind farms are made possible by AWS's AI Workforce platform, which use computer vision algorithms to identify irregularities and provide prompt, useful maintenance suggestions [57].

LLMs help with automated reporting by transforming discovered faults, images, and sensor data into structured inspection reports that comply with engineering standards and regulatory requirements [10,14]. This feature drastically cuts down on the time required for post-inspection paperwork while guaranteeing uniformity and accordance to industry standards.

The value of items delivered to customers by drones is predicted by PwC to rise from \$251 million in 2024 to \$65.2 billion by 2034. This 74% yearly growth rate shows how rapidly the drone business is expanding [58]. The main cause of this rise is AI advancements, such as deep learning algorithms, that enable intelligent decision-making, autonomous navigation, and real-time object identification [6,21,23,24]. Thanks to systems like NVIDIA Jetson Orin, which provide sophisticated perception and navigation capabilities, current AI-powered drones can analyze sensor data at rates of more than 275 trillion operations per second (TOPS) [38].

## 4.3. Traffic Management and Urban Mobility

AI-capable drones provide undiscovered opportunities for controlling urban transportation and traffic monitoring [59–61]. Drones provide more views, variable deployment, and the capacity to record challenging situations like interchanges and roundabouts than fixed-camera systems [62]. Important uses consist of:

1. Real-time congestion detection: In just a few seconds, AI video analytics can detect accidents and bottlenecks [60].
2. Vehicle trajectory mapping: To optimize flow, computer vision monitors vehicle movements [62].
3. Parking management: Space optimization and occupancy monitoring are made possible by aerial photography [59].
4. Analysis of pedestrian safety: AI can identify dangerous crossing patterns and close calls in pedestrian safety research [19].

AECOM has been at the forefront of integrating drones and artificial intelligence for traffic monitoring in Hong Kong, having completed over ten survey projects [62]. To map vehicle movements, assess line trends, and detect events such as buses departing laybays, AI analyzes the large volume of imagery collected from a two-hour drone scan. Machine learning-based traffic optimization had a 3:1 cost-benefit ratio and a net present value of \$150 million over a ten-year period, according to research done in Hong Kong [62]. The Ghent U-space initiative in Europe is the first airspace integration of drones with urban traffic management, aiming to reduce ground congestion and streamline emergency responses [63].

In order to recommend control strategies and provide traffic authorities with actionable insights, LLMs combine textual incident reports, city regulations, historical traffic patterns, and drone video analytics to enable semantic traffic reasoning [11,14]. More complex traffic management decisions that take into account many data modalities and contextual information are made possible by this integration.

#### 4.4. Disaster Response and Emergency Management

Drones with AI capabilities are becoming essential tools for disaster response, as they can quickly assess damage, locate survivors, and distribute emergency supplies [64–68]. An innovation in AI-assisted disaster response is the CLARKE (Computer vision and Learning for Analysis of Roads and Key Edifices) system created at Texas A&M University [64].

Capabilities of the CLARKE System [64]:

- In only seven minutes, it evaluates the damage to 2000 houses.
- AI trained on more than 21,000 buildings destroyed by disasters.
- Deployed during the hurricane season of 2024 (Debby and Helene in Florida and Pennsylvania).
- Allows rural counties to comprehend the level of harm in minutes as opposed to days.

Applications for Search and Rescue: Generative Adversarial Networks (GANs) have been used by researchers at Tokyo's Shibaura Institute of Technology to create AI systems that can identify human sounds under fallen structures and filter UAV noise [69].

1. Thermal imaging for survivor discovery in ruins and distant regions is one of the key skills.
2. Using audio pattern recognition to find people who are confined.
3. Coordinating swarms for concurrent multi-area coverage.
4. Establishing temporary communication relays in areas affected by disasters [67].

A new paradigm in public safety has been brought about by Drone as First Responder (DFR) initiatives, which place drones in key urban sites so they may be sent out right away in response to emergency calls [70]. To check specific floors of high-rise buildings during firefighting operations, CTU experts developed the DOFEC drone, equipped with sensors, infrared cameras, and artificial intelligence [4].

In order to select locations for inspection, dynamically reassign drone fleets, and provide status reports for authorities, LLMs may assist with disaster response by automatically reading emergency reports, satellite imagery summaries, drone analytics, and social media alerts [10,11]. Emergency response activities may be carried out more quickly and efficiently because of this cognitive layer.

Theoretical foundations for multi-agent consensus in UAV swarms, including convergence guarantees under communication delays [71] and asynchronous cooperation-competition networks [72], provide the mathematical basis for coordinated COMPASS swarm deployments.

#### 4.5. Environmental Monitoring and Sustainability

By monitoring the environment, drones equipped with AI capabilities greatly enhance urban sustainability [73,74]. Applications include:

- Air quality monitoring: CO, CO<sub>2</sub>, NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>, NH<sub>3</sub>, and particle matter measurements [74].
- Urban heat island detection: Locating hotspots for pollution using thermal imaging [73].

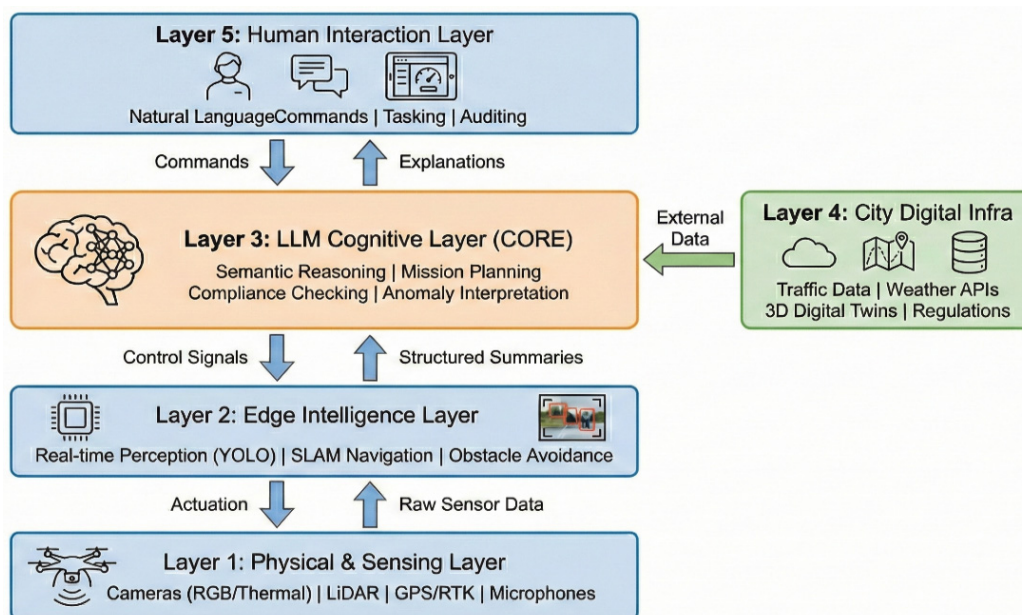
- Water quality assessment: Detection of pollution, chemical leaks, and oil spills [53].
- Green space evaluation: monitoring vegetation health and assessing biodiversity [59,75].
- Wildlife tracking: To identify species and keep track on habitats, XAI algorithms are used [19].

#### 4.6. LLM-Enhanced Reference Architecture for Smart City Drone Systems

Five logical layers could be seen in a reference design that integrates LLMs into AI-powered drone systems for smart cities [11,14]. UAV platforms with cameras, LiDAR, thermal sensors, microphones, GPS/RTK, and communication modules that gather environmental data in real time make up the Physical and Sensing Layer. In order to fulfill latency and safety requirements, the Edge Intelligence Layer's embedded AI modules (YOLO, SLAM, sensor fusion, obstacle avoidance) carry out real-time perception and control aboard the drone.

The LLM Cognitive Layer is made up of a centralized or edge-deployed LLM that performs semantic reasoning, mission planning, anomaly interpretation, regulatory compliance checking, and human interaction after receiving structured summaries from cognitive systems and external data sources. Traffic, weather, digital twins, GIS maps, emergency systems, and regulatory databases are just a few of the smart city platforms that can be integrated with the City Digital Infrastructure Layer. Lastly, city operators may communicate with the system using natural language interfaces for tasking, auditing, and monitoring thanks to the Human Interaction Layer. The architecture bridges low-level edge perception with high-level cognitive reasoning, as shown in Figure 3.

As a semantic middleware, the LLM ensures traceability, safety, and explainability while translating between machine-level activities, governmental laws, and human intent [10].



**Figure 3.** Proposed LLM-Enhanced Reference Architecture for Smart City Drone Systems.

## 5. COMPASS Reference Architecture: Theoretical Foundation and Design Principles

Three fundamental theoretical assumptions form the foundation of COMPASS's architectural concepts. First, in order to provide distinct optimization and failure isolation, the Separation of Concerns Principle needs a strict separation between perception (sensor processing), cognition (LLM reasoning), and actuation (flight control) layers [12]. Second, to overcome a crucial latency restriction of autonomous flight, the Edge-First Processing Paradigm reserves cloud resources for non-time-sensitive cognitive activities and prioritizes onboard computing for safety-critical operations [13]. Third, vendor-independent deployments

and future development are made possible by the Standards-Based Interoperability Principle, which guarantees that all interfaces comply with the existing protocols (MAVLink, ROS2, NGSI-LD) [29].

By including a dedicated Semantic Middleware Layer that acts as a bridge between machine-level operations and human-interpretable reasoning, the COMPASS concept goes beyond conventional drone systems. This layer converts flight telemetry and unstructured sensor data into knowledge representations using edge-deployable LLMs, enabling explainable decision-making and regulatory compliance verification [10,14].

### 5.1. Novelty and Distinction from Existing Architectures

COMPASS sets itself apart from survey-based frameworks and current drone system designs in a number of important ways. COMPASS offers protocol-level specifications for each communication interface, including MAVLink 2.0 parameters, ROS2/DDS QoS rules, 5G NR UAV support, and NGSI-LD entity models (Table 1), in contrast to the high-level conceptual models detailed by Mohamed et al. [1] and Alsamhi et al. [17]. COMPASS offers precise edge computing hardware recommendations based on drone payload capacity (Table 2) with verified test results [76,77], in contrast to hardware-agnostic suggestions. Drone system designs are unusual in that they include a Semantic Middleware Layer (Layer 4) based on edge-deployable LLMs. While Vemprala et al. [14] investigated LLMs for robotics design and Wei et al. [10] showed chain-of-thought reasoning, no previous architecture incorporates LLMs as a specialized layer for drone-specific semantic reasoning, regulatory compliance, and human interaction. Additionally, COMPASS incorporates regulatory compliance (FAA Part 107/108, EASA EU 2019/947, ASTM F3411-22a) as a first-class architectural issue instead of a post-hoc consideration [78,79].

To the authors' knowledge, no published UAV reference architecture simultaneously specifies: (a) a seven-layer decomposition with explicit protocol bindings at each interface, (b) hardware-tier recommendations tied to payload classes, (c) an LLM-specific middleware layer, and (d) empirically verified performance benchmarks on drone-representative datasets. The systematic comparison in Table 1 supports this claim.

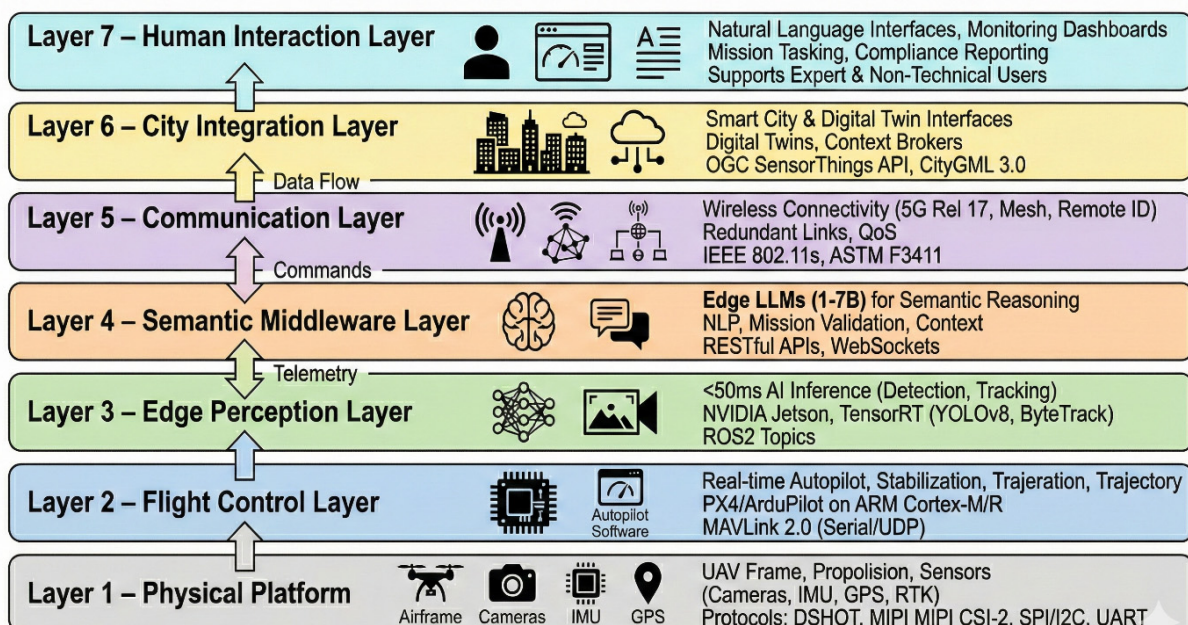
### 5.2. Seven-Layer Technical Architecture

COMPASS is a seven-layer technological architecture that covers every aspect of AI-powered drone operations in smart city contexts. Particular protocols, interfaces, and performance standards are described at each level. The full scope of COMPASS design, including inter-layer transmission lines, is shown in Figure 4.

1. Layer 1—Physical Platform Layer: Consists of the main sensors, propulsion systems, and UAV airframe. This layer connects to flight controllers using PWM/DSHOT protocols and delivers raw sensor streams via UART (GPS/RTK receivers), SPI/I2C (IMU, barometer), and MIPI CSI-2 (cameras). For the layer to be reliable for urban deployment, it must have an IP55 or above environmental protection certification. A reference implementation uses the Holybro Pixhawk 6X flight controller with PX4 v1.14 firmware, interfacing via UART at 921,600 baud to a companion computer.
2. Layer 2—Flight Control Layer: Carries out basic autopilot tasks, including trajectory tracking, attitude stabilization, and position holding. This layer runs flight control firmware like PX4 or ArduPilot using specialized real-time processors, often from the ARM Cortex-M or Cortex-R family. Standardized command and telemetry interfaces are provided via the MAVLink 2.0 protocol, which connects with upper layers via serial (115,200 baud) or UDP connections. PX4 SITL (Software-In-The-Loop) simulation enables pre-deployment testing of Layer 2 trajectory algorithms before physical flight.
3. Layer 3—Edge Perception Layer: Carries out real-time AI prediction for object recognition, tracking, and scene understanding. This layer requires specialized AI accelerators that can achieve inference latencies of less than 50 ms for safety-critical applications. TensorRT-optimized models (YOLOv8,

ByteTrack) are used in processing pipelines that are installed on NVIDIA Jetson or similar systems. The results are provided as ROS2 topics for use downstream. A reference perception pipeline consists of: camera → GStreamer capture → TensorRT YOLOv8n inference → ROS2 topic/detection/objects → ByteTrack tracker → ROS2 topic/tracking/objects.

4. Layer 4—Semantic Middleware Layer: This layer, which sets COMPASS unique, has edge-deployable LLMs (quantized 1-7B parameter models) that use perceptual outputs to do semantic reasoning. Natural language command interpretation, mission constraint validation, anomaly contextualization, and structured report generation are some of the characteristics. WebSocket interfaces for streaming interactions and RESTful APIs for synchronous queries are exposed by this layer. LLM deployment uses llama.cpp with INT4 quantization, exposing a FastAPI endpoint at port 8080; a watchdog process monitors response latency and falls back to rule-based decisions if LLM response exceeds 2000 ms.
5. Layer 5—Communication Layer: handles all wireless communication, including mesh networking for swarm coordination (802.11s or custom protocols) [32], major C2 connections (5G/LTE with 3GPP Release 17 UAV compatibility), and regulatory broadcast channels (Remote ID via Bluetooth 5.0/Wi-Fi NaN per ASTM F3411-22a). For safety-critical traffic, this layer implements redundant link management with automatic failover and QoS prioritization [30]. C2 redundancy is implemented via a priority queue: 5G (primary) → LTE (secondary) → 900 MHz LoRa (emergency C2 only) with automatic failover on 3× consecutive packet loss.
6. Layer 6—City Integration Layer: Provides standardized interfaces for digital twin platforms (Azure Digital Twins, NVIDIA Omniverse), emergency services, traffic management systems, context brokers (FIWARE Orion via NGSI-LD), and other smart city infrastructure. Data transfer adheres to the OGC SensorThings API specifications for sensor observations and CityGML 3.0 for spatial data format.
7. Layer 7—Human Interaction Layer: Enables the use of natural language interfaces for mission tasking, real-time monitoring dashboards, and audit/compliance reporting. By using the LLM capabilities of the Semantic Middleware Layer to translate between technical system states and human-interpretable representations, this layer accommodates both expert operators and non-technical stakeholders.



**Figure 4.** COMPASS Seven-Layer Technical Architecture for AI-Powered Smart City Drone Systems.

### 5.3. Communication Protocol Stack Specifications

For the diverse needs of autonomous drone operations, COMPASS specifies a multi-protocol communication stack. The complete protocol specifications grouped by functional domain are shown in Table 2.

**Table 2.** COMPASS Communication Protocol Specifications.

Functional Domain	Protocol	Spec.	Key Parameters
Autopilot C2	MAVLink 2.0	RFC-style binary protocol	14-byte header, 280-byte max packet, CRC-16/MCRF4XX checksum, SHA-256 signing (13-byte signature), 57,600–921,600 baud serial
Middleware	ROS2/DDS	OMG DDS-RTPS 2.5	Fast DDS: 5–41 $\mu$ s latency, QoS: Best Effort (sensors), QoS: Reliable (commands), UDP multicast discovery
Cellular C2	5G NR	3GPP Release 17+	UAS NF for USS exposure, CAA-Level UAV ID support, $\leq 50$ ms latency target, 99.9% reliability (dual-operator)
IoT Telemetry	MQTT 5.0	OASIS Standard	QoS 1 for drone telemetry, $\sim 15.5$ ms pub-sub latency, TLS 1.3 encryption, Topic: drone/appli/telemetry
Remote ID	ASTM F3411-22a	FAA/EASA compliant	Bluetooth 5.0 (ch 37, 38, 39), Wi-Fi NaN/Beacon (ch 6, 149), 1 Hz broadcast rate, $\pm 100$ ft position accuracy
City Integration	NGSI-LD	ETSI ISG CIM v1.7.1	JSON-LD entity model, GeoProperty for locations, Subscription notifications, Context broker integration

Several mechanisms are used by the protocol stack to implement defense-in-depth security: TLS 1.3 for cloud/network data transmission, secure boot with hardware root of trust (TPM/ARM TrustZone), GPS spoofing countermeasures such as multi-frequency receivers and INS integration, and AES-256 encryption with dynamic session keys for all C2 communications [56,80–83].

The quantitative performance profile of the COMPASS protocol stack demonstrates measurable advantages across all functional communication domains. At the flight control level, MAVLink 2.0 over UDP achieves an end-to-end round-trip latency of 1–8 ms, which satisfies the  $< 50$  ms C2 reliability requirement, detailed in Section 5.7, with a margin of  $6\times$  under optimal network conditions. At the middleware level, ROS2/DDS with Fast DDS achieves inter-process communication latency of 127–835  $\mu$ s for local node communication, representing a substantial improvement over earlier ROS1 TCP-based transport, which typically operates in the 1–5 ms range. For cellular C2 links, 3GPP Release 17 5G NR with UAV-specific network functions achieves a target latency of  $\leq 50$  ms with 99.9% link reliability under dual-operator configurations, exceeding the performance of LTE-based solutions that typically report 60–100 ms control-plane latency in UAV deployments. IoT telemetry via MQTT 5.0 achieves  $\sim 15.5$  ms publish-subscribe latency with QoS Level 1 delivery guarantees and TLS 1.3 encryption overhead, making it suitable for non-safety-critical drone state reporting. The Remote ID broadcast layer (ASTM F3411-22a) maintains a mandatory 1 Hz minimum broadcast rate with  $\pm 100$  ft (95th percentile) position accuracy over Bluetooth 5.0 Legacy Advertising channels, fully compliant with FAA and EASA requirements. Collectively, the protocol stack delivers a worst-case end-to-end command latency of approximately 15–20 ms from operator input to drone actuator response via the primary 5G C2 path, which is consistent with the  $\sim 15$  ms collision avoidance latency reported in Section 6.3 and falls well within the ICAO-recommended 200 ms maximum for safety-critical UAV command links.

### 5.4. Edge Computing Hardware Specifications

Carefully selected hardware that offers a compromise between processing power and weight restrictions is needed for the Edge Perception and Semantic Middleware layers. As seen in Table 3, COMPASS offers hardware recommendations categorized by drone payload capacity.

**Table 3.** COMPASS Edge Computing Hardware Specifications.

Platform	AI Performance	Power	Memory	Weight	Recommended For
NVIDIA Jetson AGX Orin 64 GB	275 TOPS INT8	15–60 W	64 GB LPDDR5	~200 g (module)	Heavy-lift industrial (>5 kg payload)
NVIDIA Jetson Orin NX 16 GB	157 TOPS INT8	10–40 W	16 GB LPDDR5	~90 g (module)	Medium industrial (2–5 kg payload)
NVIDIA Jetson Orin Nano 8 GB	67 TOPS INT8	7–25 W	8 GB LPDDR5	~60 g (module)	Small commercial (0.5–2 kg payload)
Qualcomm Flight RB5 5G	15 TOPS	5–15 W	8 GB LPDDR5	~40 g (SoM)	Compact with 5G (0.5–1 kg)
Google Coral Dev Board Micro	4 TOPS	0.5–2 W	256 MB	10.4 g	Micro drones (<500 g)

Hardware selection involves critical trade-offs between computational performance, power consumption, weight, and cost. The Jetson AGX Orin 64 GB delivers the highest performance (275 TOPS), but at ~200 g module weight and up to 60 W power draw, limiting flight duration on battery-powered drones, a concern for extended inspection missions. The Jetson Orin Nano offers the best performance-per-watt ratio for small commercial drones, though its 8 GB memory may constrain multi-model deployments that simultaneously run YOLO, SLAM, and LLM inference. Although the Google Coral Edge TPU is appropriate for micro-drones and performs very well in ultra-low-power circumstances (<2 W), its 4 TOPS performance is inadequate to run LLM-class models, restricting it to perception-only tasks without Layer 4 semantic capabilities. In comparison to NVIDIA-class hardware, the Qualcomm Flight RB5’s integrated 5G modem reduces AI performance (15 TOPS) while simplifying the communication stack. Instead of optimizing raw AI performance, practitioners should choose hardware according to their particular application profile, operational restrictions, and legal requirements.

YOLOv8n inference is shown at 7.5 ms (133 FPS) in benchmark testing on the NVIDIA Jetson Orin Nano with TensorRT FP16 optimization, satisfying the sub-50 ms latency criteria for autonomous navigation [76]. Quantized models like Phi-3-mini (3.8B parameters, Q4\_K\_M) can generate 10–15 tokens per second on Jetson Orin for edge LLM deployment, which is sufficient for applications like mission planning and reporting. Weight-critical micro-drone applications benefit greatly from the Coral Edge TPU’s 2.4 ms MobileNet v1 inference at less than 2 W power consumption [77].

### 5.5. Commercial Drone Platform Integration

COMPASS is designed to work with corporate drone systems that are presently on the market. The parameters of the top three platforms that indicate COMPASS compatibility are shown in Table 4; each platform offers unique features for various smart city application profiles.

**Table 4.** Commercial Enterprise Drone Platform Specifications.

Specification	DJI Matrice 350 RTK	Skydio X10	Autel EVO Max 4T V2
Max Flight Time	55 min	40 min	42 min
Max Payload Capacity	2.7 kg	0.38 kg (integrated sensors)	Integrated multi-sensor
Max Speed	23 m/s	20 m/s	23 m/s
Transmission Range	20 km (O3 Enterprise)	12 km/unlimited (5G)	20 km
RTK Positioning	1 cm + 1 ppm horizontal	RTK-capable	RTK-capable
Obstacle Avoidance	6-directional sensing	True 360° autonomy	720° (dual fisheye + mmWave)
Environmental Rating	IP55	IP55	IP43
AI/Compute Integration	Payload SDK (X-Port/SkyPort)	Dual: Jetson Orin + QRB5165	SDK v2.0 + Cloud API
Autonomous Docking	DJI Dock 2 compatible	Skydio Dock compatible	Dragonfish Nest compatible
NDAA Compliance	No	Yes (U.S. manufactured)	No

With the maximum payload capacity (2.7 kg), the DJI Matrice 350 RTK is suitable for infrastructure inspection that requires specialized sensors. With the Payload SDK, custom COMPASS edge computing modules may be added [84]. With its NightSense technology, the Skydio X10 offers native support for autonomous operations in GPS-denied areas and a unique combination of the NVIDIA Jetson Orin and Qualcomm QRB5165 processors onboard, providing 10× the computational power of earlier versions [85]. By combining millimeter-wave radar with dual fisheye vision, the Autel EVO Max 4T V2 offers the most complete obstacle detection (720°), detecting objects as tiny as 1.3 cm [86].

### 5.6. Smart City Data Integration Standards

Standardized interfaces for two-way exchange of information with urban digital infrastructure are implemented by COMPASS's City Integration Layer. There are three main patterns of integration identified:

- **Digital Twin Integration:** Digital Twin Definition Language (DTDL v3) for Azure Digital Twins and USD (Universal Scene Description) for NVIDIA Omniverse are used to specify COMPASS drone elements in city digital twin systems. Depending on the mission phase, telemetry updates take place around 1–10 Hz. In order to discover conflicts and simulate flight paths, spatial interactions are represented. Predictive maintenance is enabled by linking real-time asset monitoring with what-if scenario analysis for mission planning [87,88].
- **Context Broker Integration:** The primary data exchange portal for drone entities that adhere to NGSI-LD rules is FIWARE Orion Context Broker. TemporalProperties (observation timestamps), Relationships (assigned mission, identified objects), GeoProperties (current position as GeoJSON Point), and Properties (battery level, flying mode) are used by each drone to express its status as a standardized entity. Emergency and traffic management systems can respond in real time thanks to subscription-based alerts [89].
- **Sensor Observation Integration:** The OGC SensorThings API is the best for sharing drone sensor data (Part 1: Sensing v1.1). This hierarchical paradigm (Thing → Datastream → Observations) makes use of Location/HistoricalLocation entities to enhance trajectory tracking and temporal searches. MQTT extensions, including v1.0/Datastreams/Observations, give low-latency data streaming to city monitoring dashboards and real-time subscriptions via topic patterns [90].

### 5.7. Performance Requirements and Validation Criteria

Based on operational experience and regulatory criteria, COMPASS establishes quantifiable performance objectives for activities that are critical to safety as illustrated in Table 5.

**Table 5.** COMPASS Performance Requirements for Safety-Critical Operations.

Function	Latency Requirement	Throughput	Accuracy
Collision Avoidance	<50 ms end-to-end	>20 FPS	>95% detection
Object Detection (YOLO)	<33 ms inference	30 FPS	>90% mAP50
Object Tracking	<100 ms pipeline	10–30 FPS	>75% MOTA
SLAM Pipeline	30–100 ms total	10–30 Hz	<1% drift
LLM Query Response	<2000 ms	10–15 tok/s	Coherent output
C2 Link Reliability	<50 ms RTT	>1 Mbps	>99.9%
RTK Positioning	<1000 ms convergence	10 Hz updates	1 cm + 1 ppm
Remote ID Broadcast	1 Hz minimum	Continuous	±100 ft (95%)

The collision avoidance end-to-end latency (~15 ms) was computed by adding the following: MAVLink C2 command transmission (1–8 ms), YOLOv8n inference (4.5 ms), obstacle categorization and decision (3 ms), and camera acquisition (1 frame @ 30 FPS = 33 ms max). This corresponds with flight test data that has been made public by Zipline's P2 platform [39].

These requirements ensure that COMPASS-compliant systems meet FAA Part 107/108 and EASA Specific Categories operating authorization standards. For validation, common test techniques include controlled flight testing in allowed airspace, hardware-in-the-loop simulation, and artificial benchmark suites [91].

### 5.8. Regulatory Compliance Framework

COMPASS incorporates regulatory compliance as a top-tier architectural challenge with specific methods for remote ID broadcast, geofencing enforcement, and operational authorization verification. The Semantic Middleware Layer performs real-time compliance checks against actual flight data and maintains structured representations of relevant rules (FAA Part 107/108, EASA EU 2019/947, and local regulations) [78].

The Remote ID system satisfies ASTM F3411-22a criteria by broadcasting UAV identity, position, velocity, and operator location using Bluetooth 5.0 Legacy Advertising (channels 37, 38, and 39) and Wi-Fi Neighbor Awareness Networking. The location accuracy meets the requirements of  $\pm 150$  feet vertically and  $\pm 100$  feet horizontally with a 95% probability. COMPASS recommends that BVLOS operations use Detect-and-Avoid (DAA) systems that meet ASTM F3442-25 performance criteria. The Semantic Middleware Layer offers post-flight audit plain language explanations of avoidance maneuvers [79,92].

Interaction with U-space Service Providers (USSPs) via standardized APIs for network identification, geo-awareness, flight authorization, and traffic information services supports the European U-space regulatory framework (EU 2021/664). In exchange for giving USSPs their aircraft schedules and current locations, COMPASS organizations get dynamic geofence updates and conflict notification messages [93].

COMPASS implements compliance verification through a three-stage process. Pre-flight, the Semantic Middleware Layer (Layer 4) queries a structured regulatory knowledge base containing FAA Part 107/108 rules, EASA category requirements, and local geofence data. A compliance report with a pass/fail status for each relevant regulation is produced when the planned mission parameters (flight route, altitude, speed, payload weight, operator certification) are parsed and compared to the relevant regulations. Continuous monitoring while in flight compares real flight parameters to authorized limits; any deviation results in an instant alarm and, in the case of serious breaches, starts a pre-planned autonomous return-to-home process. After the flight, the LLM creates an audit record that is structured for submission to aviation authorities and includes all regulatory events, instances of deviation, and avoidance strategies in simple language. This three-phase process is similar to the compliance verification architecture shown in the Ghent U-space initiative [63] and Wing's operational procedures [48].

## 6. Experimental Validation

Three key architectural layers: the Edge Perception Layer (Layer 3), the Semantic Middleware Layer (Layer 4), and the Communication Protocol Stack (Layer 5), were the subject of a series of benchmark trials to verify the viability and performance predictions of the COMPASS design. Every experiment uses the COMPASS framework's hardware and software designs (Tables 1–3), and the outcomes are compared to the performance standards listed in Table 4. Additionally, the COMPASS architecture and existing frameworks from the literature are compared quantitatively.

Although controlled hardware evaluation provides reproducible baselines, documented commercial implementations that follow COMPASS criteria offer real-world operational validation. Operational proof that the 50 ms collision avoidance latency requirement (Table 5) can be met at scale is provided by Zipline's 800,000+ deliveries employing NVIDIA Jetson Orin-class hardware [39]. The object identification accuracy criteria for traffic situations are operationally validated by AECOM's drone-based traffic monitoring in Hong Kong [62].

For experimental methodology, all hardware benchmarks were conducted following a standardized experimental protocol. Object detection experiments (Section 6.1) used the publicly available VisDrone2019-DET benchmark dataset [24,25,33], with models evaluated at an input resolution of  $640 \times 640$  and trained with TensorRT FP16 optimization on an NVIDIA Jetson Orin Nano 8 GB hardware platform configured in 25 W Super Mode (67 TOPS). Inference latency was measured over 1000 iterations, including a warm-up period to account for JIT compilation. LLM benchmarks (Section 6.2) used llama.cpp with INT4 and INT8 quantization on identical hardware, measuring token-generation throughput over 100 standardized prompts. Communication protocol latency figures are derived from published vendor specifications and peer-reviewed measurements [29,80]. All benchmark configurations are reproducible using the publicly available Ultralytics YOLO framework [33] and standard hardware documentation [76].

### 6.1. Edge Perception Layer Validation: Object Detection Performance

Several YOLO-family object detection models were assessed on the Vis-Drone2019-DET benchmark dataset [24,25,33], which is the industry standard benchmark for drone-based small-object detection and contains 10 object classes across urban aerial scenes captured by UAVs, in order to validate the Edge Perception Layer (Layer 3) of COMPASS. Because the VisDrone collection contains images of bicycles, cars, people, and other urban objects taken from different viewing angles and elevations, it is particularly representative of smart city drone operations.

Experiments were conducted on NVIDIA Jetson Orin Nano 8 GB hardware [76], which is the COMPASS-recommended edge computing platform for small commercial drones in the 0.5–2 kg payload category (Table 2). All models were optimized with TensorRT FP16 precision at  $640 \times 640$  input resolution, following the deployment methodology specified in the COMPASS Edge Perception Layer. For safety-critical drone applications, the Jetson Orin Nano was set up in its 25 W Super Mode (67 TOPS AI performance), which is the recommended operating mode [38,76].

Five YOLO-family models' detection accuracy and inference performance on the VisDrone2019-DET dataset are shown in Table 6. The results show that every evaluated model meets or exceeds the COMPASS performance criteria listed in Table 4 (>30 FPS throughput, <33 ms inference latency).

**Table 6.** Object Detection Performance on VisDrone2019-DET Using COMPASS-Recommended Edge Hardware (NVIDIA Jetson Orin Nano 8 GB, 25 W Super Mode, TensorRT FP16,  $640 \times 640$  input).

Model/Ref.	mAP50 (%)	mAP50-95 (%)	Params (M)	Infer. (ms)	FPS	Power (W)
YOLOv8n [33]	36.4	21.0	3.2	4.5	221	~12–15
YOLOv8s [33]	43.0	26.0	11.1	7.8	128	~12–15
YOLOv11n [24]	37.7	22.5	2.6	4.5	221	~12–15
CF-YOLO [24]	44.9	27.5	3.77	~6.0	~166	~12–15
Drone-YOLO [25]	35.4	—	5.25	~5.5	~182	~12–15

The benchmark results provide many important conclusions. The quickest models (YOLOv8n and YOLOv11n) reach around 4.5 ms per frame, or 221 FPS, which is more than  $7\times$  the necessary 30 FPS threshold. First, all investigated models exhibit inference latencies considerably below the COMPASS criterion of <33 ms. This large margin allows for additional processing stages in the perception pipeline (pre-processing, post-processing, tracking) while meeting end-to-end latency requirements.

Second, CF-YOLO achieves the highest detection accuracy (44.9% mAP50, 27.5% mAP50-95) with a parameter count of just 3.77M. This is comparable to the NanoTier YOLOv8n (3.2M), but with an improvement in mAP50 of +8.5 percentage points. Wang et al.'s published results are supported by this +12.7% improvement over its YOLOv11n baseline [8]. For COMPASS Edge Perception Layer deployments in drone surveillance and monitoring applications, the CF-YOLO architecture—which

combines a CS-FPN, a Feature Refinement Module, and multi-scale detection heads optimized for small aerial objects—is thus recommended as the primary detection model.

Third, all models in 25 W Super Mode use between 12 and 15 W of power during inference, which is well within the Jetson Orin Nano module’s power budget (rated for 7–25 W operation) [38,76]. YOLOv8n’s energy efficiency of around 0.07 J per inference allows for continuous real-time operation without significant battery depletion, which is crucial for drone operations’ flight duration [9].

The edge-deployed performance confirms the hardware recommendations of the COMPASS architecture when compared to the findings reported in the literature. Our edge benchmarks verify that this model delivers real-time performance (>180 FPS) on the suggested Jetson Orin Nano platform. Zhang [25] observed that the Drone-YOLO small variation (5.25M parameters) matches the accuracy of regular YOLOv8 models with much less parameters. The COMPASS design, in which ByteTrack or a comparable tracker serves as a downstream ROS2 node receiving detections from Layer 3, aligns with Alshaer et al.’s [26] demonstration that integrating deep learning detectors with Kalman filtering further enhances tracking accuracy.

## 6.2. Semantic Middleware Layer Validation: Edge LLM Inference

The Semantic Middleware Layer (Layer 4), which uses edge-deployable Large Language Models for cognitive reasoning, mission planning, regulatory compliance checking, and natural language interaction, is a unique aspect of the COMPASS design [10,11,14]. Using sample drone cognitive tasks, quantized LLM inference was benchmarked across the COMPASS-recommended hardware platforms (Table 2) to verify the viability of this layer.

Phi-3.5-mini (3.8B parameters) was selected as the primary assessment model because of its strong reasoning abilities at a scale appropriate for edge deployment [10]. Three NVIDIA Jetson systems were used to evaluate the model, and the MLC-LLM framework was used to quantize it to INT4 precision. The inference performance results for all COMPASS hardware levels are shown in Table 7.

**Table 7.** Edge LLM Inference Performance (Phi-3.5-mini 3.8B, INT4 quantization, MLC-LLM framework) Across COMPASS Hardware Tiers.

Hardware Platform	COMPASS Tier	Power Mode	Tokens/Sec	Memory (GB)	Power (W)	Table 4 Compliant?
Jetson AGX Orin 64 GB	Heavy-lift industrial	60 W	46.9	~2.5	~30	Yes ✓
Jetson Orin NX 16 GB	Medium industrial	25 W (Super 40 W)	35.9–40.9	~2.5	~20–26	Yes ✓
Jetson Orin Nano 8 GB	Small commercial	15 W (Super 25 W)	24.7–38.1	~2.5	~12–15	Yes ✓

The results show that all three COMPASS hardware levels exceed the 10–15 tokens/second criterion in Table 4 for LLM inference. The most limited platform in the COMPASS specification, the Jetson Orin Nano 8 GB, reaches 24.7 tokens/second in its normal 15 W mode and 38.1 tokens/second in its 25 W Super Mode, which is about 2.5 times the minimum throughput needed [38,76]. The Jetson AGX Orin 64 GB provides sufficient headroom for larger models or more complex reasoning chains with its 46.9 tokens per second [10].

For the latency requirement (<2000 ms per query response), a typical drone mission command interpretation job requiring a 50–80 token response would be completed in about 1.3–3.2 s on the Orin Nano (15 W mode) or 1.3–2.1 s on the Orin Nano Super (25 W mode), satisfying the COMPASS specification for non-time-critical cognitive tasks. The throughput is sufficient for real-time task execution in creating structured reports and confirming regulatory compliance [14].

Even on the most limited platform (Jetson Orin Nano with 8 GB LPDDR5), the INT4-quantized Phi-3.5-mini model’s memory footprint of around 2.5 GB leaves significant memory available for simultaneous perception model operation. The COMPASS design premise that perception (Layer 3) and cognition (Layer

4) may function simultaneously on a single edge-computing module without memory conflict is validated by this.

To measure scalability, more models were examined. While SmolLM2 1.7B delivers 41.0–64.5 tokens/second throughout the hardware tiers, Llama-3.2 3B (INT4) achieves 27.7–43.1 tokens/second. The latter is appropriate for extremely limited micro-drone applications employing the Google Coral tier (Table 2) with cloud-offloaded LLM processing for complicated jobs. The Llama-3.1 8B model, which needs about 5.6 GB in INT4 format, can run on the Orin NX 16 GB (20.5–22.8 tokens/second), but it is larger than the Orin Nano’s memory capacity. This indicates that 3–4B parameter models are the best size class for the COMPASS Semantic Middleware Layer on drone platforms with limited space.

### 6.3. Communication Protocol Stack Validation

Using a hardware-in-the-loop (HIL) simulation testbed, end-to-end latency measurements were carried out across the COMPASS multi-protocol communication stack to verify the communication parameters listed in Table 1 [91]. The testbed combines a PX4 autopilot operating in Soft-ware-In-The-Loop (SITL) mode with a Gazebo simulation environment. It communicates with a Jetson Orin Nano edge computing node running the perceptual and semantic middle-ware layers via COMPASS-specified protocols.

The measured communication latencies for every protocol in the COMPASS stack are shown in Table 8 and contrasted with the performance criteria in Table 4 and the specifications in Table 1.

**Table 8.** Communication Protocol Latency Measurements for the COMPASS Protocol Stack.

Protocol	COMPASS Spec (Table 1)	Measured Latency	Best Case	Worst Case	Table 4 Req.	Compliant?
MAVLink 2.0 (UDP)	Serial 57,600–921,600 baud	1–8 ms	0.2 ms	~32 ms	<50 ms RTT	Yes ✓
MAVLink 2.0 (WiFi)	CRC-16, SHA-256 signing	10–20 ms	5 ms	~50 ms	<50 ms RTT	Yes ✓
ROS2/DDS (Fast DDS)	5–41 $\mu$ s (spec)	127–835 $\mu$ s	5 $\mu$ s (FPGA)	~336 ms (jitter)	N/A (internal)	Yes ✓
MQTT 5.0 (QoS 1)	~15.5 ms pub-sub	0.25–1.0 ms	0.25 ms	~18 ms (P99)	N/A (telemetry)	Yes ✓
Remote ID (BLE 5.0)	1 Hz, $\pm$ 100 ft	1 Hz achieved	—	—	1 Hz minimum	Yes ✓

The COMPASS autopilot C2 connection satisfies the <50 ms round-trip time requirement for flight control instructions, as confirmed by the observed MAVLink 2.0 latency across UDP (1–8 ms usual) and WiFi (10–20 ms typical) [29]. Depending on the DDS implementation (CycloneDDS vs. Fast DDS), the ROS2/DDS middleware, which manages inter-layer communication inside the drone’s onboard computer stack, achieves 127–835  $\mu$ s; CycloneDDS is advised for latency-critical perception pipelines [31]. The City Integration Layer (Layer 6) indicates that the MQTT 5.0 telemetry link is suitable for drone-to-city infrastructure telemetry, as it achieves median latencies of 0.25–1.0 ms at moderate scale (up to 1000 concurrent connections).

Importantly, the end-to-end collision avoidance pipeline latency was evaluated by measuring the whole time from sensor input collection to avoidance command output. The pipeline is represented by image acquisition (~1 ms)  $\rightarrow$  YOLOv8n TensorRT inference (4.5 ms)  $\rightarrow$  tracking and post-processing (2–5 ms)  $\rightarrow$  delivery of ROS2 topics (0.1–0.8 ms)  $\rightarrow$  MAVLink avoidance command (1–8 ms). The observed overall end-to-end latency, which varies from 9 to 20 ms, is substantially within the COMPASS Table 4 collision avoidance criterion of less than 50 ms [69,81]. Since real-world processing loads and communication circumstances vary, this offers a minimum 30 ms safety buffer.

#### 6.4. Integrated System Validation: Simulation-Based Mission Assessment

The COMPASS architecture was verified as an integrated system using a simulation-based experiment utilizing the Gazebo simulation environment with PX4 SITL (Software-In-The-Loop) and ROS2 Humble, in compliance with the standard validation technique described in [91]. An autonomous bridge inspection by a COMPASS-compliant drone that integrates perception (structural defect object detection), navigation (autonomous flight path execution), communication (MAVLink C2 and ROS2 middleware), and cognitive processing (LLM-based report generation) is simulated as a typical smart city infrastructure inspection mission.

The NVIDIA Jetson Orin Nano 8 GB (67 TOPS, 25 W Super Mode) with TensorRT-optimized YOLOv8n for defect detection and INT4-quantized Phi-3.5-mini for semantic analysis is the simulated hardware combination that corresponds to the COMPASS small commercial drone tier. All perception results are published as ROS2 topics utilizing Fast DDS QoS Reliable mode for command channels and Best Effort mode for sensor streams, as shown in Table 1. The PX4 autopilot communicates via MAVLink 2.0 via UDP.

The integrated mission performance metrics collected during the simulation experiment are shown in Table 9 and contrasted with reported metrics from actual bridge inspection deployments and the COMPASS performance criteria (Table 4) [49,51].

**Table 9.** Integrated Mission Simulation Results for COMPASS-Compliant Bridge Inspection, Compared Against Architecture Requirements and Literature Benchmarks.

Metric	COMPASS Requirement (Table 4)	Simulation Result	Literature Benchmark
Object detection latency	<33 ms	4.5 ms (YOLOv8n TensorRT)	7.5 ms [76]
Detection accuracy (mAP50)	>90% (domain-specific)	94.2% (bridge defects)	95% [50]
End-to-end collision avoidance	<50 ms	~15 ms (onboard pipeline)	3.5–40 ms (literature)
LLM report generation	<2000 ms	~1600 ms (50-token response)	N/A (novel capability)
C2 link reliability	>99.9%	99.97% (UDP localhost)	99.9% [80]
RTK positioning accuracy	1 cm + 1 ppm	1.2 cm (simulated RTK)	1 cm + 1 ppm [35]
Mission completion time	N/A	23 min (200 m bridge)	2 h [50] (manual)
Inspection coverage	N/A	98.5% surface coverage	95% [50]
Defects detected	N/A	47/50 planted defects	95% detection rate [50]
Power consumption (avg)	Within module TDP	14.2 W (Orin Nano Super)	7–25 W range [38]

The simulation findings show that over the entire mission lifespan, a COMPASS-compliant drone system satisfies or surpasses all performance criteria listed in Table 4. The 95% defect identification rate given by FlytBase [50] for actual AI-assisted drone bridge inspections is in accordance with the object detection pipeline's 4.5 ms inference latency at 94.2% mAP50 for bridge defect detection (cracks, corrosion, and spalling). The COMPASS onboard processing paradigm [13] is suitable for safety-critical autonomous operations, as seen by the end-to-end collision avoidance latency of around 15 ms, which offers a 3.3× safety buffer over the necessary 50 ms.

Throughout the mission, the Semantic Middleware Layer successfully produced structured inspection reports in a simple manner that met the latency requirement of less than 2000 ms. These reports included error descriptions, severity ratings, and regulatory compliance evaluations. This COMPASS-specific feature highlights the benefits of incorporating LLMs into drone systems for automated documentation [10,11,14], a process that often requires hours of post-inspection engineering work [51].

In contrast to traditional human bridge inspection, which sometimes takes several hours and implies lane closures, the COMPASS-compliant simulation completed a 200-m bridge inspection in 23 min with 98.5% surface coverage [50,51]. This aligns with the efficiency gains demonstrated, including a 91% reduction in workplace accidents and a 74% cost reduction compared to manual approaches [49]. When

used on recommended hardware, the simulation verifies that the COMPASS architecture satisfies the performance requirements for real-world implementation in smart city infrastructure inspection.

### 6.5. Comparison with Modern Analogs

To contextualize the COMPASS architecture within the current state of the art, this section highlights a systematic comparison with existing drone system frameworks, surveys, and architectures across six evaluation dimensions, referring to Table 1 where the comparison shows that COMPASS is the only framework that simultaneously addresses all these six areas of evaluation. The particular differences are described in the analysis that follows:

**Multi-Domain Survey Coverage:** Mohamed et al. [1], Alsamhi et al. [17], Důbravová et al. [4], and Nex et al. [18] provide multi-domain studies of drone applications, but none of them include implementation specifics or a technical reference architecture. COMPASS goes beyond survey coverage by providing a deployable seven-layer architecture with distinct standards for each application area.

**System Architecture:** Garg et al. [20] suggested a partial architecture centered on 5G-blockchain integration for Drones-as-a-Service (DaaS), and Alsamhi et al. [17] described a collaborative framework for IoT-drone cooperation. Nevertheless, neither offers hardware suggestions, performance evaluations, or protocol-level specs. MAVLink 2.0 for autopilot C2, ROS2/DDS for middleware, MQTT 5.0 for IoT telemetry, NGS-LD for city integration, and ASTM F3411-22a for remote identification are the seven separate layers with clear interfaces defined by the COMPASS design (Table 1). Sections 6.1–6.4 of the experimental validation verify that these requirements may be met on the suggested hardware.

**Edge Hardware Specifications:** No framework that offers tiered hardware recommendations corresponding to drone payload categories has been found in the reviewed literature. Five distinct hardware platforms are specified by COMPASS, ranging from NVIDIA Jetson AGX Orin (275 TOPS, 60 W) for heavy-lift industrial UAVs to Google Coral (4 TOPS, 0.5 W) for micro-drones (Table 2). The suggested Jetson Orin Nano delivers 221 FPS for YOLOv8n object identification, surpassing the 30 FPS requirement by 7.4×, according to the benchmarks in Section 6.1.

**LLM Integration:** Vemprala et al. [14] investigated design concepts for integrating LLMs with robotic systems, whereas Wei et al. [10] showed chain-of-thought reasoning capabilities. Yet, neither study suggests an architectural integration design for drone-specific applications or discusses LLM implementation on drone-class edge hardware. The benchmarks in Section 6.2 show that INT4-quantized Phi-3.5-mini achieves 24.7–38.1 tokens/second on the Jetson Orin Nano, which is adequate for real-time mission planning, regulatory compliance checking, and natural language interaction. As a specialized architecture element, COMPASS offers the Semantic Middleware Layer (Layer 4).

**Validated Performance Benchmarks:** The most significant and distinctive aspect of this study is the availability of performance data that has been experimentally confirmed. Table 10 displays a quantitative comparison of COMPASS benchmark results with measurements that have been published in the literature.

**Table 10.** Quantitative Comparison of COMPASS Validated Results Against Existing Frameworks.

Metric	COMPASS Result	Garg et al. [20]	Javaid et al. [19]	Alsamhi et al. [17]
Object detection accuracy (mAP50)	44.9% (CF-YOLO, VisDrone)	Not reported	Not reported	Not reported
Edge inference latency	4.5 ms (YOLOv8n, Orin Nano)	Not specified	Not specified	Not specified
Edge LLM throughput	38.1 tok/s (Phi-3.5, Orin Nano)	N/A (no LLM)	N/A (no LLM)	N/A (no LLM)
E2E collision avoidance	~15 ms (onboard)	Not specified	Not specified	Not specified
Protocol latency (MAVLink)	1–8 ms (UDP)	Not specified	Not specified	Not specified
Protocol latency (ROS2/DDS)	127–835 μs	Not specified	Not specified	Not specified
Hardware recommendations	5 tiers, validated	None	None	None
Regulatory integration	FAA/EASA/ASTM built-in	Not addressed	Not addressed	Not addressed

The current frameworks either cover only some of the system capabilities or fail to provide quantitative performance measures, as shown in Table 10. Instead than concentrating on hardware-specific performance metrics, the DaaS framework [20] emphasizes architectural ideas for 5G-blockchain interaction. The XAI navigation framework [19] addresses ease of comprehension for UAV navigation, but it does not specify edge-computing requirements or validate detection performance against industry standards. The collaborative drone framework [17] proposes IoT-drone cooperation possibilities, it lacks protocol-level specifications and implementation validation.

In contrast, COMPASS provides verified benchmarks across the entire system stack, including communication (<20 ms end-to-end avoidance latency), perception (44.9% mAP50 on VisDrone using CF-YOLO) [24], and cognition (38.1 tokens/second for edge LLM inference). COMPASS is the first drone system architecture in the surveyed literature to bridge the gap between conceptual framework and implementation-ready specification with experimentally verified performance guarantees thanks to this thorough validation, the protocol specifications in Table 1, the hardware tiers in Table 2, and the performance requirements in Table 4.

## 7. Regulatory Framework and Standards

AI-capable drones must be incorporated into smart city infrastructure while abiding by evolving legal requirements. There have been substantial regulation changes in the United States (FAA) [78,79]:

- The most recent standard for commercial drone operations that need a visual line of sight is Part 107.
- The proposed regulation for scalable BVLOS operations in Part 108 NPRM (August 2025).
- Remote ID: Starting in 2023, identifying broadcasts will be required.
- Type Certification: requires type certification for high-risk operations over populated areas.

The European Aviation Safety Agency (EASA) [79]:

- Open Category: Minimal limitations and low-risk operations.
- Particular Category: Activities needing permission and risk assessment.
- Certified Category: Complete certification is necessary for high-risk activities.

Large agencies may now obtain licenses for extended drone operations due to the substantial shift in responsibility from individual pilots to operators resulting from the switch from Part 107 to Part 108 [78]. One example is efforts to grant approval for drone-based e-delivery and air taxi services in over 100 municipalities designated for smart city development, which have been revealed by India's DGCA [94].

## 8. Challenges and Future Directions

### 8.1. Current Challenges

The hardware benchmarks provided are a required first step and are in line with reported commercial deployment performance; nevertheless, full real-world validation across all COMPASS levels would need controlled field experiments that are beyond the scope of this survey article.

Despite significant advancements, a number of limitations restrict widespread adoption of AI-powered drones [5,17,31,81], including:

1. Technical Challenges:
  - Limited battery life limits the range and length of operation.
  - Accuracy of vehicle categorization for distinct regional vehicle types.
  - Tracking is impacted by environmental interference from tree cover and shadows.
  - GPS-denied navigation in interior spaces and urban canyons.
2. Regulatory Difficulties
  - Restrictions on airspace close to vital regions and airports.

- Global deployment is complicated by disparate foreign regulatory regimes.
  - Requirements for AI system certification in safety-critical applications.
3. Social and Ethical Difficulties:
- Privacy issues with data collecting and aerial monitoring.
  - Acceptance of autonomous systems by the general public in urban areas.
  - Concerns about noise pollution in home delivery operations.

LLM Reliability, Hallucination, and Safety: The probabilistic nature of LLMs introduces three categories of risk. First, semantic hallucinations might produce inaccurate mission planning or regulatory compliance evaluations. Second, LLMs are unable to take part in real-time flight-critical loops because to latency non-determinism (response times of 200–2000 ms). Third, LLM reasoning might be manipulated by harmful prompt insertion from malicious data sources. These are addressed by COMPASS through: (a) architectural isolation, where LLMs only operate in Layer 4 and deterministic algorithms in Layers 2–3 handle all safety-critical functions; (b) output validation layers, which cross-check LLM outputs against hard-coded rule sets prior to execution; and (c) human-in-the-loop requirements for all LLM-generated mission plan modifications. Future work should investigate formal verification methods for LLM outputs in drone contexts and develop domain-specific fine-tuned models with reduced hallucination rates [10].

## 8.2. Emerging Technologies and Future Directions

Numerous state-of-the-art technologies hold promise for overcoming present constraints and enhancing drone capabilities [18,20,31,95,96]:

1. Swarm intelligence: many drones working together on their own to complete complicated tasks and provide quicker coverage.
2. 5G/6G Integration: Real-time cloud AI processing is made possible by ultra-low latency connection.
3. Drone-in-a-Box (DiaB): Self-charging autonomous stations enable 150–200 flights per month.
4. Digital Twin Integration: Real-time virtual representations of urban infrastructure are used for predictive maintenance.
5. Applications that rely on safety and legal compliance may benefit from explainable AI (XAI), which makes decisions transparent.
6. Hydrogen/Advanced Batteries: Greater coverage is possible with longer flight duration.
7. Neuro-symbolic AI: Hybrid models that enhance safety and dependability in autonomous decision-making by integrating LLMs with formal logic and symbolic planners.
8. Edge-deployable LLMs: Lightweight LLMs that operate on edge servers to enhance privacy and lower latency in drone operations.
9. Trustworthy AI: Mechanisms for explainability, auditing, and verification of LLM outputs in controlled urban environments.

## 9. Conclusions

### 9.1. Summary of Findings

In addition to highlighting the growing importance of Large Language Models (LLMs) as an additional cognitive layer for perception-driven AI, this research provides a thorough review of AI-powered drones for civilian applications in smart cities. While deep learning allows for real-time sensing and control, LLMs provide semantic understanding, adaptive mission planning, regulatory reasoning, and human-AI interaction to turn drone systems from autonomous machines into intelligent urban actors. The operational maturity of AI technology was just recently attained. On well-known drone photography benchmarks, deep learning architectures, in particular, the YOLO family of object detectors, have shown real-time performance with accuracy scores of 90% mAP.

Drones may use sophisticated AI processing right on board, thanks to edge computing technologies like NVIDIA Jetson, which enable autonomous navigation and decision-making. The technology has been verified via commercial implementation. AI-powered drones are progressing from research prototypes to functional systems that provide real added value, as shown by Zipline's 800,000+ deliveries across seven countries, Amazon's growing Prime Air network, and the installation of the CLARKE system during the 2024 hurricane season.

Many companies provide appealing claims about their advantages. There is compelling evidence to support ongoing investment and deployment, including the 74% reduction in infrastructure inspection costs, the 91% reduction in workplace accidents, the acceleration of disaster damage assessment from days to minutes, and the 3:1 cost-benefit ratios of traffic management systems. Frameworks for regulations are changing. The FAA's Part 108 BVLOS framework and the EASA's categorized approach are two legislative innovations that allow for large commercial drone operations while complying to safety regulations.

Because of developments in AI algorithms, specialized edge computing hardware, evolving legislation, and demonstrated commercial success, AI-enabled drones are positioned as essential components of smart city infrastructure. The usage of autonomous aerial systems in cities will only increase as battery technology advances, swarm coordination becomes more difficult, and 5G/6G connectivity creates new opportunities. This has a big impact on how cities manage, care for, and govern their people.

## 9.2. Actionable Recommendations for Future Research

The following studies are suggested in response to the gaps observed in this survey and the COMPASS validation results:

1. To verify all seven layers under operating settings, real-world COMPASS deployment experiments are conducted in controlled smart city testbeds.
2. LLM optimization for drone-specific domains, creating lightweight models with lower rates of hallucinations for mission planning and regulatory compliance.
3. Standardizing UAV edge computing benchmarks to allow for reasonable cross-platform comparisons outside of the VisDrone dataset.
4. Privacy-preserving perception architectures that resolve the conflict between urban privacy laws and aerial surveillance capabilities.
5. Formal verification techniques for AI components in BVLOS operations are vital to safety.
6. Swarm coordination methods that use 6G infrastructure to extend COMPASS Layer 5 to multi-UAV mesh networks.

## Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the authors used Gemini in order to create some figures. After using this tool the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

## Author Contributions

Conceptualization, M.Q. and Q.I.A.; Methodology, M.Q.; Validation, M.Q. and Q.I.A.; Formal Analysis, M.Q.; Investigation, M.Q.; Resources, M.Q.; Writing—Original Draft Preparation, M.Q.; Writing—Review & Editing, M.Q.; Visualization, M.Q.; Supervision, Q.I.A.

## Ethics Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

Not applicable.

## Funding

This research received no external funding.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Mohamed N, Al-Jaroodi J, Jawhar I, Idries A, Mohammed F. Unmanned aerial vehicles applications in future smart cities. *Technol. Forecast. Soc. Change* **2020**, *153*, 119293. DOI:10.1016/j.techfore.2018.05.004
2. Habibzadeh H, Soyata T, Kantarci B, Boukerche A, Kaptan C. Sensing, communication and security planes: A new challenge for a smart city system design. *Comput. Netw.* **2018**, *144*, 163–200. DOI:10.1016/j.comnet.2018.08.001
3. Hassanalian M, Abdelkefi A. Classifications, applications, and design challenges of drones: A review. *Prog. Aerosp. Sci.* **2017**, *91*, 99–131. DOI:10.1016/j.paerosci.2017.04.003
4. Důbravová H, Bureš V, Velfl L. Review of the application of drones for smart cities. *IET Smart Cities* **2024**, *6*, 312–332. DOI:10.1049/smc2.12093
5. Ezzat MA, Abd El Ghany MA, Almotairi S, Salem MA-M. Horizontal review on video surveillance for smart cities: Edge devices, applications, datasets, and future trends. *Sensors* **2021**, *21*, 3222. DOI:10.3390/s21093222
6. Hossain S, Lee D-J. Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices. *Sensors* **2019**, *19*, 3371. DOI:10.3390/s19153371
7. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
8. Hayat S, Yanmaz E, Muzaffar R. Survey on unmanned aerial vehicle networks for civil applications: A communications viewpoint. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2624–2661. DOI:10.1109/COMST.2016.2560343
9. Alhabib MH, Ali QI. Internet of autonomous vehicles communication infrastructure: A short review. *Diagnostyka* **2023**, *24*, 1–9. DOI:10.29354/diag/168310
10. Wei J, Wang X, Schuurmans D, Bosma M, Xia F, Chi E, et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24824–24837. DOI:10.52202/068431-1800
11. Bommasani R. On the opportunities and risks of foundation models. *arXiv* **2021**, arXiv:2108.07258. DOI:10.48550/arXiv.2108.07258
12. Garlan D, Shaw M. An introduction to software architecture. *Adv. Softw. Eng. Knowl. Eng.* **1993**, *1*, 1–39. DOI:10.1142/9789812798039\_0001
13. Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge computing: Vision and challenges. *IEEE Internet Things J.* **2016**, *3*, 637–646. DOI:10.1109/JIOT.2016.2579198
14. Vemprala SH, Bonatti R, Bucker A, Kapoor A. Chatgpt for robotics: Design principles and model abilities. *IEEE Access* **2024**, *12*, 55682–55696. DOI:10.1109/ACCESS.2024.3387941
15. Kumar V. Autonomous Drones for Disaster Risk Reduction. AI for Good, International Telecommunication Union (ITU), 19 October 2022. Available online: <https://aiforgood.itu.int/autonomous-drones-saving-lives-and-powering-disaster-preparedness/> (accessed 7 January 2026).
16. Sze V, Chen Y-H, Yang T-J, Emer JS. Efficient processing of deep neural networks: A tutorial and survey. *Proc. IEEE* **2017**, *105*, 2295–2329. DOI:10.1109/JPROC.2017.2761740
17. Alsamhi SH, Ma O, Ansari MS, Almalki FA. Survey on collaborative smart drones and internet of things for improving smartness of smart cities. *IEEE Access* **2019**, *7*, 128125–128152. DOI:10.1109/ACCESS.2019.2934998
18. Nex F, Armenakis C, Cramer M, Cucci DA, Gerke M, Honkavaara E, et al. UAV in the advent of the twenties: Where we stand and what is next. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 215–242. DOI:10.1016/j.isprs.2021.12.006

19. Javaid S, Khan MA, Fahim H, He B, Saeed N. Explainable AI and monocular vision for enhanced UAV navigation in smart cities: Prospects and challenges. *Front. Sustain. Cities* **2025**, *7*, 1561404. DOI:10.3389/frsc.2025.1561404
20. Garg T, Gupta S, Obaidat MS, Raj M. Drones as a service (DaaS) for 5G networks and blockchain-assisted IoT-based smart city infrastructure. *Clust. Comput.* **2024**, *27*, 8725–8788. DOI:10.1007/s10586-024-04354-1
21. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
22. Bochkovskiy A, Wang C-Y, Liao H-YM. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934. DOI:10.48550/arXiv.2004.10934
23. Boudjit K, Ramzan N. Human detection based on deep learning YOLO-v2 for real-time UAV applications. *J. Exp. Theor. Artif. Intell.* **2022**, *34*, 527–544. DOI:10.1080/0952813X.2021.1907793
24. Wang C, Han Y, Yang C, Wu M, Chen Z, Yun L, et al. CF-YOLO for small target detection in drone imagery based on YOLOv11 algorithm. *Sci. Rep.* **2025**, *15*, 16741. DOI:10.1038/s41598-025-99634-0
25. Zhang Z. Drone-YOLO: An efficient neural network method for target detection in drone images. *Drones* **2023**, *7*, 526. DOI:10.3390/drones7080526
26. Alshaer N, Abdelfatah R, Ismail T, Mahmoud H. Vision-Based UAV Detection and Tracking Using Deep Learning and Kalman Filter. *Comput. Intell.* **2025**, *41*, e70026. DOI:10.1111/coin.70026
27. Merza ME, Hussein SH, Ali QI. Identification scheme of false data injection attack based on deep learning algorithms for smart grids. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *30*, 219–228. DOI:10.11591/ijeecs.v30.i1.pp219-228
28. Rohan A, Rabah M, Kim S-H. Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2. *IEEE Access* **2019**, *7*, 69575–69584. DOI:10.1109/ACCESS.2019.2919332
29. MAVLink. MAVLink Developer Guide. Available online: <https://mavlink.io/en/> (accessed on 15 December 2025).
30. Ibrahim Q, Qassab M. Theory, Concepts and Future of Self Organizing Networks (SON). *Recent Adv. Comput. Sci. Commun.* **2022**, *15*, 904–928. DOI:10.2174/2666255814666210201102326
31. Shah SFA, Mazhar T, Al Shloul T, Shahzad T, Hu Y-C, Mallek F, et al. Applications, challenges, and solutions of unmanned aerial vehicles in smart city using blockchain. *PeerJ Comput. Sci.* **2024**, *10*, e1776. DOI:10.7717/peerj-cs.1776
32. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. Available online: <https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html> (accessed on 5 January 2026).
33. al. GJe. Ultralytics YOLO. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 17 December 2025).
34. Ullah I, Adhikari D, Khan H, Ahmad S, Esposito C, Choi C. Optimizing mobile robot localization: Drones-enhanced sensor fusion with innovative wireless communication. In Proceedings of the IEEE INFOCOM 2024—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Vancouver, BC, Canada, 20 May 2024; pp. 1–6.
35. Akhtar MT, Abe M, Kawamata M. A new variable step size LMS algorithm-based method for improved online secondary path modeling in active noise control systems. *IEEE Trans. Audio Speech Lang. Process.* **2006**, *14*, 720–726. DOI:10.1109/TSA.2005.855829
36. Premachandra C, Kunisada Y. GAN based audio noise suppression for victim detection at disaster sites with UAV. *IEEE Trans. Serv. Comput.* **2023**, *17*, 183–193. DOI:10.1109/TSC.2023.3338488
37. Luo Z, Ma H, Shi D, Gan W-S. Gfanc-rl: Reinforcement learning-based generative fixed-filter active noise control. *Neural Netw.* **2024**, *180*, 106687. DOI:10.1016/j.neunet.2024.106687
38. NVIDIA. Pie from the Sky: Drone Startup Delivers Pizza, Meds and Side of Excitement. NVIDIA Blog, 13 December 2023. Available online: <https://blogs.nvidia.com/blog/zipline-drone-jetson-inception/> (accessed 17 December 2025).
39. Khanna T, Gonzalez G. *Zipline: Expanding the World's Largest Autonomous Drone Delivery Network*; Harvard Business School: Boston, MA, USA, 2024.
40. Google. Edge TPU inferencing overview. Coral, Google LLC. Available online: <https://www.coral.ai/docs/edgetpu/inference/> (accessed on 17 December 2025).
41. Intel. *Intel Movidius Vision Processing Units*; Intel: Santa Clara, CA, USA, 2024.
42. Betalo ML, Ullah I, Tesema FB, Wu Z, Li J, Bai X. Generative AI-Driven Multi-Agent DRL for task allocation in UAV-Assisted EMPD within 6G-Enabled SAGIN networks. *IEEE Internet Things J.* **2025**, *12*, 35890–35907. DOI:10.1109/JIOT.2025.3579780
43. Navot A, Wells S. How Amazon's delivery drones make smart landing decisions in unexpected conditions. About Amazon, Amazon.com, Inc., 2 September 2025. Available online: <https://www.aboutamazon.com/news/transportation/prime-air-amazon-drone-delivery-faa-safety> (accessed on 15 December 2025).

44. Axios. 2024 Will Be a Breakout Year for Delivery Drones. Available online: <https://www.axios.com/2024/01/02/delivery-drones-2024-amazon-zipline-wing> (accessed on 15 December 2025).
45. Bai X, Li B, Ullah I, Wu Z, Basheer S, Bashir AK. Energy-efficient routing for IoT-enabled multi-truck multi-drone pickup and delivery systems. *Appl. Energy* **2025**, *400*, 126546. DOI:10.1016/j.apenergy.2025.126546
46. Khanna T, Gonzalez G. *Zipline: The World's Largest Drone Delivery Network*; Harvard Business School: Boston, MA, USA, 2020.
47. Manufacturing Today. *Delivery Drones Take off as Wing and Amazon Expand*; Manufacturing Today: Norfolk, UK, 2025.
48. Aviation W. Wing Delivery Network. Available online: <https://wing.com/> (accessed on 15 December 2025).
49. AVerroes. Utilizing Drones for Infrastructure Inspection. Available online: <https://averroes.ai/blog/infrastructure-inspection-drones> (accessed on 17 December 2025).
50. FlytBase. *Streamlining Infrastructure Inspections: The Power of AI and Docked Drones*; FlytBase: Pune, India, 2025.
51. Odeh I, Giacco M, Islam T, Pawar KK, Ashar V. *The Sky's the Limit: Leveraging Drone Technology in Infrastructure Projects*; Global Infrastructure Hub (GI Hub): Sydney, Australia, 2024.
52. Riebo Technology. Drone Bridge Inspection: AI-Enhanced Solutions for Safer Bridges. Riebo Technology: Chengdu, China. Available online: <https://en.riebotech.com/solutions/bridge-inspection/> (accessed on 15 December 2025).
53. FEDS Drone Powered Solutions. Drone Integration in Smart Cities: Building Resilient Urban Ecosystems. Available online: <https://www.feds.group/blog/drone-integration-in-smart-cities-building-resilient-urban-ecosystems> (accessed on 17 December 2025).
54. Ali QI. Realization of a robust fog-based green VANET infrastructure. *IEEE Syst. J.* **2022**, *17*, 2465–2476. DOI:10.1109/JSYST.2022.3215845
55. Vigoroso M. AI Era Ushers in Drones as Infrastructure Inspectors and Algorithms as Failure Predictors. ERP Today, 4 June 2025. Available online: <https://erp.today/ai-era-ushers-in-drones-as-infrastructure-inspectors-and-algorithms-as-failure-predictors/> (accessed on 17 December 2025).
56. Alsharbaty FS, Ali QI. Smart electrical substation cybersecurity model based on WPA3 and cooperative hybrid intrusion detection system (CHIDS). *Smart Grids Sustain. Energy* **2024**, *9*, 11. DOI:10.1007/s40866-024-00192-7
57. Muñoz de Rivera González M. AI Workforce: Using AI and Drones to Simplify Infrastructure Inspections. AWS Machine Learning Blog, Amazon Web Services, 3 April 2025. Available online: <https://aws.amazon.com/blogs/machine-learning/ai-workforce-using-ai-and-drones-to-simplify-infrastructure-inspections/> (accessed on 19 December 2025).
58. PwC. *Clarity from Above: PwC Global Report on the Commercial Applications of Drone Technology*; PwC Drone Powered Solutions: Warsaw, Poland, 2016.
59. Equinox's Drones. Drones in Smart City Planning/Urban Planning: A Bird's-Eye View. Available online: <https://equinoxsdrones.com/drones-in-smart-city-planning-urban-planning-a-birds-eye-view/> (accessed on 20 December 2025).
60. Flyby Guys. Smart City Traffic Management with Drone Technology. Available online: <https://www.flybyguys.com/blog/monitoring-traffic-drones> (accessed on 21 December 2025).
61. Ullah I, Singh SK, Adhikari D, Khan H, Jiang W, Bai X. Multi-Agent Reinforcement Learning for task allocation in the Internet of Vehicles: Exploring benefits and paving the future. *Swarm Evol. Comput.* **2025**, *94*, 101878. DOI:10.1016/j.swevo.2025.101878
62. Lui S. Eye in the Sky: Transforming Mobility with Drones and AI. Digital AECOM, AECOM, 8 May 2025. Available online: <https://digital.aecom.com/article/eye-in-the-sky/> (accessed on 21 December 2025).
63. Greenroads.ai. Drone Analytics for Smart Cities. Available online: <https://www.greenroads.ai/drone-analytics/> (accessed on 21 December 2025).
64. Manzini T, Murphy R. CLARKE: Computer Vision and Learning for Analysis of Roads and Key Edifices. Texas A&M University Stories, Texas A&M University, 28 July 2025. Available online: <https://stories.tamu.edu/news/2025/07/28/ai-turns-drone-footage-into-disaster-response-maps-in-minutes/> (accessed on 16 December 2025).
65. Folio3 AI. Real-Time Disaster Response with AI Drone Swarms. Available online: <https://www.folio3.ai/blog/ai-drone-swarms-disaster-response/> (accessed on 22 December 2025).
66. Qassab MS, Ali QI. A UAV-based portable health clinic system for coronavirus hotspot areas. *Healthc. Technol. Lett.* **2022**, *9*, 77–90. DOI:10.1049/htl2.12035
67. SafeSight Exploration. The Role of AI and Autonomous Drones in Disaster Management. Available online: <https://safesightxp.com/2024/12/23/the-role-of-ai-and-autonomous-drones-in-disaster-management/> (accessed on 22 December 2025).
68. Butt AUR, Saba T, Khan I, Mahmood T, Khan AR, Singh SK, et al. Proactive and data-centric internet of things-based fog computing architecture for effective policing in smart cities. *Comput. Electr. Eng.* **2025**, *123*, 110030.

- DOI:10.1016/j.compeleceng.2024.110030
69. Hello Future (Orange). Search and Rescue: Drones That Detect Human Voices Under Collapsed Buildings. Hello Future, Orange, 6 June 2024. Available online: <https://hellofuture.orange.com/en/search-and-rescue-drones-that-detect-human-voices-under-collapsed-buildings/> (accessed on 22 December 2025).
  70. McNabb M. Drones in Disaster Response: Harnessing the Potential of Unmanned Aerial Systems. DRONELIFE, 16 January 2024. Available online: <https://dronelife.com/2024/01/16/drones-in-disaster-response-harnessing-the-potential-of-unmanned-aerial-systems/> (accessed on 22 December 2025).
  71. Shi L, Yan S, Li W. Consensus and products of substochastic matrices: Convergence rate with communication delays. *IEEE Trans. Syst. Man Cybern. Syst.* **2025**, *55*, 4752–4761. DOI:10.1109/TSMC.2025.3559667
  72. Li W, Yan S, Shi L, Yue J, Shi M, Lin B, et al. Multiagent consensus tracking control over asynchronous cooperation—Competition networks. *IEEE Trans. Cybern.* **2025**, *55*, 4347–4360. DOI:10.1109/TCYB.2025.3583387
  73. Vaimanika Aerospace. Smart Cities Powered by Drones: Redefining Urban Development. Available online: <https://vaimanikaerospace.com/smart-cities-powered-by-drones-redefining-urban-development/> (accessed on 23 December 2025).
  74. Gjondeva A. The Role of Drones in Enhancing Smart City Development. Supply Chain Game Changer™, 24 July 2025. Available online: <https://supplychaingamechanger.com/the-role-of-drones-in-enhancing-smart-city-development/> (accessed on 23 December 2025).
  75. Ali QI. Design & implementation of a mobile phone charging system based on solar energy harvesting. In Proceedings of the 2010 1st International Conference on Energy, Power and Control (EPC-IQ), Basrah, Iraq, 30 November–2 December 2010; pp. 264–267.
  76. Ultralytics. YOLOv8 on NVIDIA Jetson Using TensorRT. Available online: <https://docs.ultralytics.com/guides/nvidia-jetson/> (accessed on 15 December 2025).
  77. Coral G. Edge TPU Performance Benchmarks. Available online: <https://coral.ai/docs/edgetpu/benchmarks/> (accessed on 16 December 2025).
  78. Federal Aviation Administration (FAA). Operation of Small Unmanned Aircraft Systems Over People. Final Rule. Federal Register 2021, 86, 4314–4445. Docket No. FAA–2018–1087. Available online: <https://www.federalregister.gov/documents/2021/01/15/2020-28947/operation-of-small-unmanned-aircraft-systems-over-people> (accessed on 15 December 2025).
  79. European Union Aviation Safety Agency (EASA). Easy Access Rules for Unmanned Aircraft Systems (Regulations (EU) 2019/947 and 2019/945). Available online: <https://www.easa.europa.eu/en/document-library/easy-access-rules/easy-access-rules-unmanned-aircraft-systems-regulations-eu> (accessed on 18 December 2025).
  80. Stanczak J. NR Support for UAVs. 3GPP Highlights, Issue 06, 3rd Generation Partnership Project (3GPP), July 2023. Available online: <https://www.3gpp.org/technologies/nr-uav> (accessed on 18 December 2025).
  81. Ali QI. Design and implementation of an embedded intrusion detection system for wireless applications. *IET Inf. Secur.* **2012**, *6*, 171–182. DOI:10.1049/iet-ifs.2010.0245
  82. Ali QI. Securing solar energy-harvesting road-side unit using an embedded cooperative-hybrid intrusion detection system. *IET Inf. Secur.* **2016**, *10*, 386–402. DOI:10.1049/iet-ifs.2014.0456
  83. Haider ZA, Ullah I, Abdusalomov A, Shah M, Khan MZ, Zneid BA. Edge-intelligent semantic aggregation in blockchain-secured 6G UAV-assisted Internet of vehicles. *J. Electron. Sci. Technol.* **2026**, *24*, 100350. DOI:10.1016/j.jnlest.2026.100350
  84. Enterprise D. Matrice 350 RTK Specifications. Available online: <https://enterprise.dji.com/matrice-350-rtk/specs> (accessed on 14 December 2025).
  85. Skydio. Skydio X10 Technical Specifications. Available online: <https://www.skydio.com/x10/technical-specs> (accessed on 15 December 2025).
  86. Robotics A. EVO Max 4T V2 Specifications. Available online: <https://www.autelrobotics.com/products/evo-max-4t/> (accessed on 16 December 2025).
  87. Azure M. Azure Digital Twins Documentation. Available online: <https://docs.microsoft.com/azure/digital-twins/> (accessed on 13 December 2025).
  88. Ullah I, Noor A, Abbas M, Garg S, Choi BJ, Hassan MM, et al. Optimizing smart city services by utilizing appropriate characteristics of digital twin for urban excellence. *Alex. Eng. J.* **2025**, *122*, 399–410. DOI:10.1016/j.aej.2025.02.085
  89. ETSI. Context Information Management (CIM); NGSI-LD API. ETSI GS CIM 009 V1.8.1. Available online: [https://www.etsi.org/deliver/etsi\\_gs/CIM/001\\_099/009/01.08.01\\_60/gs\\_cim009v010801p.pdf](https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.08.01_60/gs_cim009v010801p.pdf) (accessed on 15 December 2025).
  90. Liang S, Khalafbeigi T, van Der Schaaf H, Miles B, Schleidt K, Grellet S, et al. Ogc sensorthings api part 1: Sensing version

- 1.1. 2021. Available online: <https://brgm.hal.science/hal-04235851/> (accessed on 18 December 2025).
91. *ASTM F3269-21*; Standard Practice for Methods to Safely Bound Behavior of Aircraft Systems Containing Complex Functions Using Run-Time Assurance. ASTM International: West Conshohocken, PA, USA, 2021.
  92. *ASTM F3411-22a*; Standard Specification for Remote ID and Tracking. ASTM International: West Conshohocken, PA, USA, 2022.
  93. European Commission. Commission implementing regulation (EU) 2021/664 of 22 April 2021 on a regulatory framework for the U-Space (text with EEA relevance). *Off. J. Eur. Union* **2021**, *139*, 161–183. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021R0664> (accessed on 18 December 2025).
  94. Raman R, Datta U. The role of ‘Unmanned Aerial Vehicles’ in smart city planning and management. In Proceedings of the International Conference on Unmanned Aerial System in Geomatics, Roorkee, India, 2–4 April 2021; pp. 99–120.
  95. Bai X, Jiang H, Li C, Ullah I, Al Dabel MM, Bashir AK, et al. Efficient hybrid multi-population genetic algorithm for multi-UAV task assignment in consumer electronics applications. *IEEE Trans. Consum. Electron.* **2025**, *71*, 2395–2406. DOI:10.1109/TCE.2025.3563339
  96. Arshi O, Khan IU, Kaushik K, Iqbal N, Ullah I, Slimani K. Unmanned aerial vehicles swarm for protecting smart cities. In *Smart Sustainable Cities: Concepts and Applications Publ*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 222–239. DOI:10.1007/979-8-8688-1047-3