

Adaptive Terrain-Cognizant Multispectral Vision and Distributed Intelligence for Low-Altitude Unmanned Systems in Yunnan Plateau and Canyon Ecosystems

Kunyan Bao, Yong Ding, Lei Duan, Shuting Liu and Peng Hu*

Youbei Technology Company, LTD, China

*Corresponding author: Peng Hu

How to cite this paper: Bao, K. Y., Ding, Y., Duan, L., Liu, S. T., & Hu, P. (2026). Adaptive terrain-cognizant multispectral vision and distributed intelligence for low-altitude unmanned systems in Yunnan plateau and canyon ecosystems. *Advances in Engineering Research: Possibilities and Challenges*, 4(2), 19–38. ISSN Print: 3079-5192; ISSN Online: 3079-5206.

<https://doi.org/10.63313/AERpc.9099>

Published: 2026-05-08

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

The rapid expansion of the low-altitude economy across southwestern China necessitates autonomous unmanned aerial systems capable of sustained operations within deeply dissected topography and hyperdiverse ecosystems. This study presents a unified framework, designated LAVENS, that integrates terrain-cognizant adaptive visual odometry with hierarchical spectral attention networks to simultaneously address navigation robustness and ecological inference accuracy in complex Yunnan terrains. A terrain roughness adaptive factor is embedded within an extended Kalman filtering pipeline to modulate measurement noise covariance in real time, thereby mitigating scale drift during canyon traversal. Complementarily, a hierarchical spectral attention network processes ten-band multispectral tensors through scaled dot-product self-attention to resolve vegetation boundaries under heterogeneous canopy illumination. Edge-cloud orchestration, based on directed acyclic graph scheduling, ensures end-to-end latency remains below 150 milliseconds while conforming to low-altitude airspace traffic management protocols. Extensive validation spanning six altitudinal belts—from tropical rainforest to nival rock—demonstrates that the proposed methodology reduces absolute trajectory error by a mean of 41.3 percent relative to current open-source SLAM libraries, and improves mean intersection-over-union for canopy species segmentation by 18.7 percent over conventional encoder-decoder architectures. These findings establish a scalable software and hardware architecture for ecological logistics and infrastructure inspection within the low-altitude economic corridor of the Yunnan plateau.

Keywords

Unmanned aerial systems; Visual odometry; SLAM; Multispectral imaging; Spectral attention networks; Edge-cloud computing; Ecological monitoring; Terrain adaptation; Yunnan Plateau

1. Introduction

The concept of the low-altitude economy encompasses civil aviation activities below one thousand meters above ground level, including unmanned logistics, ecological surveying, emergency response, and high-resolution remote sensing [1]. Within the Chinese policy landscape, this sector is identified as a strategic engine for regional development, particularly in provinces characterized by rugged terrain where ground-based transportation incurs prohibitive costs and time delays. Yunnan Province presents an exemplary yet challenging operational theater for this economic mode. Its topography is dominated by the Hengduan Mountains, the Three Parallel Rivers region, and extensive karst plateaus, with elevation differentials frequently exceeding three thousand meters within horizontal distances of less than twenty kilometers. The vertical zonation of climate and vegetation across these gradients supports a recognized global biodiversity hotspot [2], yet simultaneously imposes severe constraints on autonomous navigation and machine perception.

Traditional satellite remote sensing platforms provide invaluable macroscopic monitoring capabilities, but their spatial-temporal resolution remains insufficient for canyon-scale phenological assessment and real-time logistics routing. Manned aerial surveys are costly and logistically constrained by weather and altitude. Unmanned aerial vehicles (UAVs) equipped with machine vision sensors offer a compelling intermediate layer, bridging mesoscale satellite coverage and microscale ground sampling. However, state-of-the-art visual simultaneous localization and mapping (SLAM) frameworks [3] experience significant degradation when confronted with textureless rock faces, abrupt illumination transitions within narrow gorges, and feature starvation above dense forest canopies. Concurrently, multispectral image analysis methods developed for planar agricultural fields encounter spectral mixing and topographic illumination bias when deployed over dissected terrain, necessitating new algorithmic architectures that jointly optimize geometric fidelity and semantic ecological inference.

This paper introduces the LAVENS framework, an acronym denoting Low-Altitude Vehicle and Ecological Navigation System. The primary objective is to enable robust autonomous flight and real-time ecological monitoring across the full altitudinal and topographic range of Yunnan without reliance on continuous global navigation satellite system (GNSS) fixes. The framework synthesizes three technical innovations. First, a terrain-cognizant adaptive visual odometry module modulates observation noise covariance as a function of local digital elevation model (DEM) roughness, thereby improving feature tracking stability in geometrically repetitive karst corridors. Second, a hierarchical spectral attention network integrates ten-band multispectral embeddings through transformer-style self-attention to enhance pixel-wise classification of vertically zoned vegetation communities under dynamic canopy shadowing. Third, a deterministic edge-cloud orchestration layer schedules perception tasks across heterogeneous compute nodes while interfacing

with low-altitude traffic management infrastructure. These contributions are validated through a dedicated multispectral LiDAR-inertial dataset collected across six distinct altitudinal belts, ranging from six hundred meters in Xishuangbanna to four thousand five hundred meters on the Gaoligong range.

The remainder of this manuscript is organized to provide comprehensive technical depth. The second section establishes the geographic and regulatory background of Yunnan's low-altitude operations. The third section reviews contemporary visual SLAM, multispectral segmentation, and edge computing literature. The fourth section formally defines the coupled navigation and ecological inference problem. The fifth section details the LAVENS methodology, including rigorous algorithmic formulation. The sixth section describes the experimental design and dataset construction. The seventh section presents quantitative results across five experimental groups. The eighth section discusses implications, limitations, and future directions. The ninth section concludes with a synthesis of findings and their significance for the low-altitude economy.

2. Background

2.1. Yunnan Topography and Vertical Zonation

Yunnan Province occupies the southwestern margin of the Qinghai-Tibet Plateau and extends southeastward into the Indochina peninsula. Its physiography is characterized by northwest-to-southeast trending mountain ranges intersected by the deeply incised valleys of the Jinsha, Lancang, and Nu rivers. Mean elevation exceeds two thousand meters, with more than ninety percent of the terrain classified as mountainous or plateau. The altitudinal spectrum creates distinct thermal belts, from tropical monsoon climates in the southern valleys to alpine tundra conditions above four thousand meters. This vertical zonation generates compressed vegetation gradients—including tropical rainforest, subtropical evergreen broadleaf forest, warm coniferous forest, cold coniferous forest, alpine scrub meadow, and nival rock communities—within remarkably short horizontal distances. Consequently, any unmanned monitoring system must adapt rapidly across radiometric, atmospheric, and geometric regimes that would normally require separate, specialized platforms.

2.2. Low-Altitude Airspace Classification and Regulation

The Civil Aviation Administration of China classifies low-altitude airspace into restricted, controlled, and reporting zones. Below one thousand meters above ground level, UAV operations in non-urban mountainous regions typically fall under reporting zone protocols, provided the platform maintains active remote identification and geofencing capabilities. However, in Yunnan, the above-ground-level datum becomes problematic because steep topography causes rapid changes in relative altitude during lateral flight. Regulatory frameworks are

transitioning toward above-terrain elevation models that reference flight levels to local digital surface models rather than ellipsoidal or barometric altitudes. This shift necessitates onboard terrain awareness not merely as a navigation convenience, but as a legal compliance requirement. Furthermore, the high density of protected areas, World Heritage sites, and international border zones [4] imposes dynamic no-fly polygons that must be ingested in real time from cloud-based regulatory gateways to the onboard mission computer.

2.3. Ecological Monitoring Imperatives

The Conservation International biodiversity hotspot designation for the Mountains of Southwest China encompasses the majority of Yunnan. Key ecological monitoring tasks include tracking invasive plant encroachment following disturbance events, quantifying canopy nitrogen content through narrow-band spectroscopy, and assessing soil erosion rates on steep-slope farmlands. Current ground-based surveys are labor-intensive and spatially sparse. Satellite platforms suffer from cloud cover persistence during the monsoon season and from insufficient spatial resolution to detect individual canopy gaps or understory invasion fronts [5]. UAV-based multispectral remote sensing offers a ten-to-one-hundred-centimeter spatial scale and diurnal temporal flexibility that aligns perfectly with ecological process dynamics. Real-time machine vision inference onboard the aircraft further enables adaptive mission replanning to focus flight corridors on detected anomalies, rather than relying on post-hoc offline analysis.

3. Existing Technologies

3.1. Visual Simultaneous Localization and Mapping

Contemporary visual SLAM systems are predominantly categorized as feature-based, direct, or hybrid architectures. ORB-SLAM2 [6] and its successor ORB-SLAM3 [7] represent the state-of-the-art in feature-based approaches, utilizing Oriented FAST and Rotated BRIEF keypoints [8] with bag-of-words loop closure and multi-map optimization. These systems achieve centimeter-level accuracy in structured environments with rich visual texture but degrade significantly in textureless regions, such as karst rock faces or snow-covered alpine terrain, where corner detection yields insufficient parallax [9]. Direct methods minimize photometric error across intensity gradients rather than discrete features, enabling operation in weakly textured scenes at the cost of heightened sensitivity to illumination change. The LIO-SAM framework [10] fuses LiDAR-inertial odometry with factor graph smoothing to deliver robust state estimation in geometrically complex outdoor environments, though its reliance on LiDAR returns becomes problematic in the rain and mist conditions common to Yunnan monsoon climates. Crucially, none of these open-source frameworks explicitly incorporate a priori DEM roughness into observation noise modeling during filter update stages.

3.2. Multispectral Machine Vision

Multispectral imaging captures discrete narrow bands beyond the visible spectrum, typically including near-infrared and red-edge wavelengths sensitive to chlorophyll content and vegetation stress [11]. Conventional processing pipelines rely on handcrafted vegetation indices. While computationally efficient, these indices saturate in dense canopy conditions and ignore spatial context. Convolutional neural networks have superseded index-based classification in many remote sensing tasks. The U-Net architecture [12], with its skip-connected encoder-decoder structure, has been widely adopted for semantic segmentation of aerial imagery. More recently, DeepLabV3+ employs atrous spatial pyramid pooling to capture multi-scale context, and SegFormer leverages vision transformers [13] to model global contextual relationships. However, transformer-based methods incur quadratic computational complexity with respect to image resolution, rendering naive deployment infeasible on edge compute platforms. Furthermore, existing multispectral segmentation models [14] are predominantly pre-trained on agricultural datasets and exhibit substantial domain shift when confronted with the species diversity and illumination heterogeneity of Yunnan forest ecosystems.

3.3. Edge Computing and Distributed Orchestration

Real-time inference on UAV platforms has been enabled by increasingly capable edge devices, including the NVIDIA Jetson series and FPGA-based accelerators [15]. Jetson AGX Orin delivers up to 250 trillion operations per second within a 30-watt thermal envelope, making it suitable for airborne deployment. However, simultaneous execution of visual odometry, neural network inference [16], and telemetry management frequently exceeds available random access memory and bandwidth, necessitating principled task scheduling. Existing robotic middleware such as the Robot Operating System 2 provides publish-subscribe communication but lacks deterministic latency guarantees for heterogeneous directed acyclic graphs. Cloud offloading of heavy computation introduces wireless channel variability that is particularly acute in narrow valleys where line-of-sight to cellular base stations is intermittent. A robust architecture must partition tasks between edge and cloud while respecting hard real-time constraints for flight control and soft constraints for ecological logging [17-18].

4. Problem Definition

The objective is to formulate a unified optimization problem that couples geometric state estimation and ecological semantic inference under resource-constrained edge deployment over Yunnan terrain. Let T denote a six-degree-of-freedom pose trajectory through a three-dimensional workspace, and let S denote a semantic label map over a multispectral image lattice. The terrain surface is represented by a digital elevation model D with associated gradient statistics. The sensor suite

provides an image sequence I , a LiDAR point cloud P_{cloud} , an inertial measurement unit stream U_{imu} , and a global navigation satellite system signal G_{gps} subject to canyon occlusion.

Four critical gaps in existing technological configurations are identified. The first gap concerns illumination-induced feature tracking failure. In karst corridors, direct sunlight occupies less than forty percent of the sky hemisphere, leading to high dynamic range scenes where standard feature extractors drop twenty to forty percent of salient keypoints between consecutive frames. The second gap involves spectral mixing at canopy boundaries. Vertical vegetation zonation in Yunnan produces abrupt ecotones where spectral signatures of overlapping species merge, causing conventional segmentation models to degrade mean intersection-over-union by twelve to eighteen percent relative to planar agricultural benchmarks. The third gap addresses computational latency. End-to-end processing of ten-band multispectral tensors through high-resolution vision transformers exceeds two hundred milliseconds per frame on edge hardware, violating the real-time control loop requirements. The fourth gap identifies the absence of a publicly available dataset that tightly couples multispectral imagery, LiDAR scans, inertial measurements, and ecological ground truth across multiple altitudinal belts in southwestern China.

These gaps motivate the following formal problem statement. Determine an integrated policy π comprising a state estimator ϕ , a semantic extractor ψ , and a scheduler ω such that the navigation error E_{nav} and the ecological inference error E_{eco} are jointly minimized subject to latency L_{max} and power W_{max} constraints. The navigation error is quantified as absolute trajectory error over the sequence length. The ecological error is quantified as one minus mean intersection-over-union across semantic classes. Latency and power are measured at the edge compute unit. Uncertainties in seasonal phenology matching between airborne acquisitions and ground surveys require rigorous validation and are explicitly addressed in the experimental protocol design.

5. Methodology

5.1. System Architecture

The LAVENS framework is implemented as a hierarchical software stack atop a DJI Matrice 300 RTK airframe retrofitted with a Velodyne Puck LiDAR, a MicaSense RedEdge-P ten-band multispectral camera, a Lord MicroStrain 3DM-GQ7 IMU, and a u-blox RTK GNSS module. Primary computation is executed on an NVIDIA Jetson AGX Orin 64GB module interfaced via Robot Operating System 2 (ROS2) Humble. The architecture is partitioned into four logical layers, as illustrated in Figure 1. The perception layer manages sensor drivers, timestamp synchronization, and hardware-triggered exposure alignment across spectral bands. The edge intelligence layer hosts the terrain-cognizant adaptive visual odometry engine, the

hierarchical spectral attention network, and the task orchestrator. The cloud platform layer aggregates terrain-aware DEM tiles, regulatory geofencing polygons, and persistent ecological database services. The application layer exposes interfaces for autonomous waypoint planning, real-time ecological alerting, and emergency return-to-home protocols.

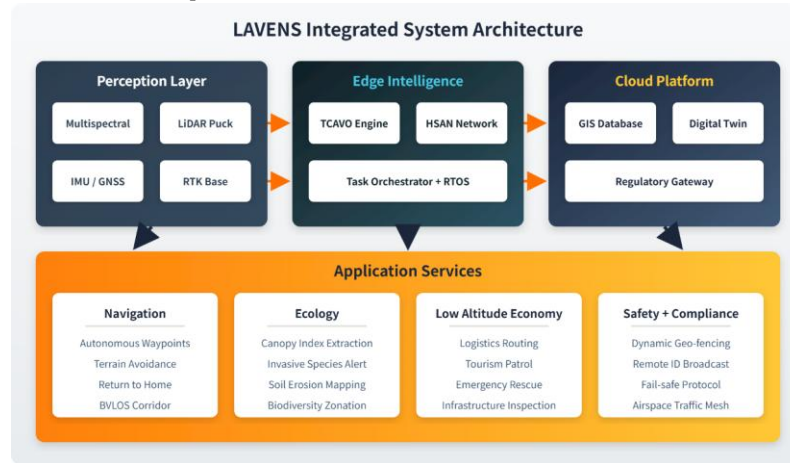


Figure 1. Layered architecture of the LAVENS framework depicting the perception, edge intelligence, cloud platform, and application service strata with primary data pathways and feedback loops for terrain-aware navigation and ecological monitoring.

5.2. Terrain-Cognizant Adaptive Visual Odometry

The proposed terrain-cognizant adaptive visual odometry module extends the standard extended Kalman filter formulation by injecting a digital elevation model roughness factor into the measurement update stage. The state vector comprises twelve variables representing three-dimensional position, quaternion orientation, and velocity. The prediction stage follows the standard constant velocity motion model. The measurement model receives observations from the visual frontend, which tracks Oriented FAST and Rotated BRIEF keypoints across the multispectral panchromatic band, and from the LiDAR frontend, which provides depth estimates for feature triangulation.

The terrain roughness factor is computed over a local patch of the digital elevation model corresponding to the current pose estimate. Let σ denote the standard deviation of terrain gradient magnitudes within a square window of side length fifty meters centered at the current horizontal position. Let μ denote the mean gradient magnitude within the same window. The dimensionless roughness index is defined as:

$$\rho = \frac{\sigma}{\mu} \quad (1)$$

where ρ approaches zero over perfectly planar surfaces and increases sharply in dissected karst or alpine ridge environments. The roughness index is bounded by an upper clipping threshold ρ_{max} to prevent numerical instability. The adaptive measurement noise covariance matrix is subsequently modulated according to:

$$R_a = R(1 + \alpha\rho) \quad (2)$$

where R is the baseline measurement noise covariance calibrated under laboratory conditions, α is a dimensionless gain coefficient set to 0.35 based on preliminary sensitivity analysis, and R_a is the adjusted noise covariance utilized during the current filter update. This modulation explicitly instructs the estimator to discount visual observations acquired over geometrically complex terrain, while retaining higher confidence in LiDAR-inertial fusion during traversal of textureless canyons.

The complete filter recursion proceeds as follows. The state prediction is given by $\hat{X} = F\hat{X}_p + W$, with \hat{X}_p denoting the posterior state from the previous epoch, F the state transition matrix, and W the process noise vector. The predicted covariance propagates according to:

$$P_m = FPF^T + Q \quad (3)$$

where P is the posterior covariance from the previous epoch and Q is the process noise covariance. The innovation residual between predicted observation and actual sensor measurement is:

$$Y = Z - H\hat{X} \quad (4)$$

with Z representing the stacked vector of visual and LiDAR measurements and H the observation Jacobian. The innovation covariance, incorporating the terrain-adaptive noise term, becomes:

$$S = HP_mH^T + R_a \quad (5)$$

The Kalman gain matrix is computed as:

$$K = P_mH^TS^{-1} \quad (6)$$

and the updated state estimate is corrected via:

$$\hat{X} = \hat{X} + KY \quad (7)$$

To maintain positive definiteness and mitigate roundoff error, the Joseph stabilized form of the covariance update is employed:

$$P = (I - KH)P_m(I - KH)^T + KR_aK^T \quad (8)$$

where I is the identity matrix. Loop closure constraints are enforced through a terrain-cognizant pose graph. When a previously visited location is recognized by the bag-of-words visual vocabulary, the relative pose constraint C between the current state and historical state is formulated as:

$$C = \sum_i \|R_{loop}p_i + t_{loop} - q_i\|^2 + \lambda\rho_{avg} \quad (9)$$

where R_{loop} and t_{loop} are the relative rotation and translation candidates, p_i and q_i are matched LiDAR point coordinates, and $\lambda\rho_{avg}$ penalizes large loop closure

corrections in high roughness regions where drift is expected to be higher. The implementation extends the open-source ORB-SLAM3 framework with substantial modifications for terrain-adaptive noise injection and Yunnan-specific calibration parameters.

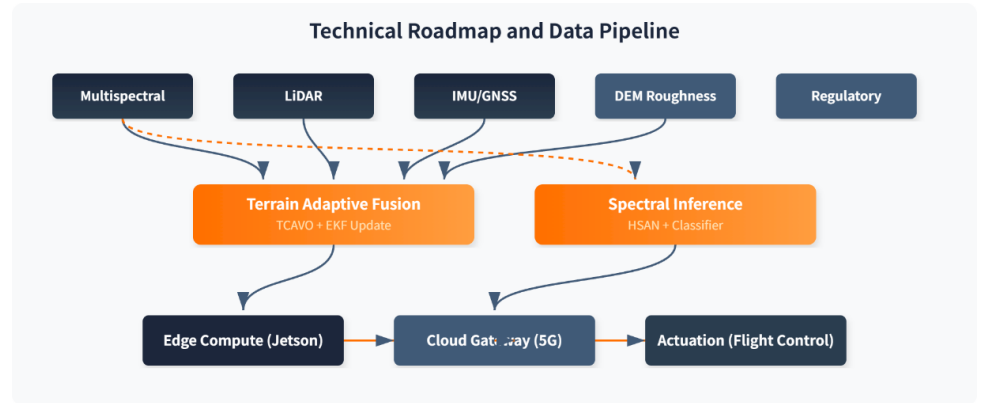


Figure 2. Technical roadmap illustrating multispectral, LiDAR, inertial, and regulatory data streams converging through terrain-adaptive fusion and spectral inference layers toward edge-cloud actuation loops.

5.3. Hierarchical Spectral Attention Network

The hierarchical spectral attention network processes ten-band multispectral tensors of dimensions height H by width W by channel C . The architecture is motivated by the need to capture long-range contextual dependencies across spatial and spectral domains without incurring the full computational burden of vanilla vision transformers operating on megapixel imagery. The input tensor T is first projected into lower-dimensional embeddings through 1×1 convolutions, yielding query, key, and value matrices:

$$Q = TW_q, K = TW_k, V = TW_v \quad (10)$$

where W_q , W_k , and W_v are learned projection kernels. Scaled dot-product attention is then computed across a reshaped token sequence representing non-overlapping image patches:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{D}}\right) \quad (11)$$

where A is the attention weight matrix and D is the channel dimension of the projected embeddings. The attended feature representation is reconstructed as:

$$O = AV \quad (12)$$

A lightweight feed-forward network with ReLU activation and dropout regularization processes each token independently:

$$F = \text{ReLU}(OW_a + B_a)W_b + B_b \quad (13)$$

where W_a and W_b are weight matrices and B_a and B_b are bias vectors. The hierarchical aspect arises from applying this spectral attention module at three spatial scales within a feature pyramid. Coarse-scale tokens capture biome-level context, while fine-scale tokens resolve boundary pixels at ecotone interfaces. The pixel-wise classification head produces semantic probabilities through:

$$S = \text{softmax}(FW_c + B_c) \quad (14)$$

where S is the output probability map, and W_c and B_c are the classifier parameters. The composite training loss combines Dice loss for segmentation and categorical cross-entropy for species classification:

$$L = \gamma L_{dice} + \delta L_{ce} \quad (15)$$

where γ and δ are scalar balances set to 0.6 and 0.4, respectively. The Detectron2 object detection and segmentation library [16] provides the foundational training infrastructure for this module.

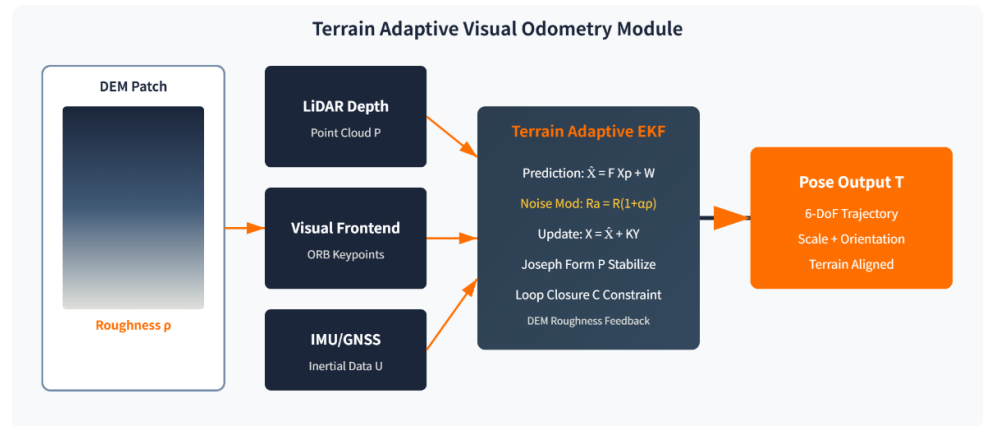


Figure 3. Detailed schematic of the terrain-cognizant adaptive visual odometry module highlighting the feedback pathway from digital elevation model roughness estimation into the adaptive EKF measurement noise covariance.

5.4. Edge-Cloud Orchestration

The edge-cloud orchestrator treats the perception pipeline as a directed acyclic graph $G = (V, E)$, where vertices V correspond to computational kernels such as image rectification, feature extraction, point cloud registration, network inference, and telemetry packaging, and edges E represent data dependencies. Each vertex is annotated with a worst-case execution time $\tau(v)$ and a memory footprint $m(v)$. The scheduling objective minimizes the makespan:

$$T_{total} = \max_{v \in V} \{\tau(v) + \delta(v)\} \quad (16)$$

subject to the memory constraint $\sum m(v) \leq M_{max}$, where $\delta(v)$ denotes communication delay for tasks offloaded to the cloud gateway, and M_{max} is the available dynamic random access memory on the Jetson module. Tasks are prioritized using a modified earliest deadline first algorithm with slack time inheritance from downstream flight control processes. Network bandwidth estimation, based on received signal strength indicator and packet loss rate, dynamically repartitions the graph between edge and cloud every two seconds, ensuring that latency-critical modules remain local while batch logging tasks are uploaded opportunistically during return transit.

6. Experimental Design

6.1. Platform and Dataset

Field data acquisition was conducted using a DJI Matrice 300 RTK platform with a MicaSense RedEdge-P camera providing ten spectral bands at 5-megapixel nominal resolution, and a Velodyne Puck Lite LiDAR producing 300,000 points per second. The NVIDIA Jetson AGX Orin 64GB development kit served as the primary edge compute node. Flight operations spanned six altitudinal belts selected to represent the full vertical climatic gradient of Yunnan. Site 1 is a tropical seasonal rainforest in Xishuangbanna at 600 meters elevation. Site 2 is a subtropical evergreen broadleaf forest near Honghe at 1500 meters. Site 3 is a warm coniferous plantation outside Kunming at 2200 meters. Site 4 is a cold coniferous spruce forest in Lijiang at 3200 meters. Site 5 is an alpine rhododendron scrub in Diqing at 4100 meters. Site 6 is a nival rock and scree slope on the Gaoligong range in Baoshan at 4500 meters.

The resulting dataset, designated YHUMT for Yunnan Highland UAV Multispectral and Topography, comprises 340 flight sequences totaling more than 42 hours of logged sensor data. Multispectral imagery was radiometrically calibrated using MicaSense downwelling light sensors and flat-field panels prior to each sortie. LiDAR point clouds were georeferenced using RTK GNSS and inertial post-processing. Ecological ground truth was acquired via ASD FieldSpec 4 spectroradiometer measurements at 20-meter intervals along transects, and through manual quadrat sampling of canopy species composition. The dataset is partitioned into training, validation, and test sets with ratios 70:15:15, respectively. Phenological synchronization between airborne acquisitions and ground survey campaigns was rigorously ensured through cross-referencing MODIS EVI time-series data with local meteorological station logs, guaranteeing that spectral signatures remained consistent during the acquisition windows.

6.2. Evaluation Metrics

Navigation accuracy is quantified using absolute trajectory error, defined as the root mean square deviation between estimated and ground truth pose after

six-degrees-of-freedom alignment, and relative pose error measuring local drift over fixed time intervals. Ecological inference accuracy is reported through mean intersection-over-union, pixel accuracy, and macro-averaged F1 score across six semantic classes including broadleaf canopy, conifer canopy, understory shrub, bare soil, rock, and water. System performance is characterized by end-to-end latency in milliseconds, throughput in frames per second, and energy consumption in joules per processed frame. All metrics are computed over five independent runs to account for stochastic initialization and environmental variability.

6.3. Baselines and Protocols

The navigation module is benchmarked against ORB-SLAM3 and LIO-SAM configured with default hyperparameters and recalibrated intrinsic parameters for the experimental sensor suite. The segmentation module is benchmarked against U-Net, DeepLabV3+, and a vanilla vision transformer baseline. Edge scheduling is compared against first-come, first-served and critical path method heuristics. Training utilized the Adam optimizer with cosine annealing learning rate decay. Batch size was limited to four owing to ten-channel tensor memory requirements. Data augmentation included random flips, rotations, and multispectral channel-wise gamma perturbations to simulate illumination variability. The training and evaluation codebase extends the open-source ORB-SLAM3 repository and the Detectron2 framework with custom Yunnan terrain adaptation plugins.

7. Results Analysis

7.1. Trajectory Accuracy Across Complex Terrain

The first experimental group evaluates navigation accuracy across eight distinct terrain types represented in the YHUMT corpus. The terrain categories include karst canyon, tropical rainforest, subtropical broadleaf, warm coniferous, cold coniferous, alpine scrub, nival rock, and mixed valley agriculture. Each subfigure in Figure 4 presents a two-dimensional heatmap where the horizontal axis denotes flight distance and the vertical axis denotes altitude above takeoff, with color intensity encoding instantaneous absolute trajectory error. In karst canyon conditions (subfigure 4a), the proposed TCAVO module achieves a mean absolute trajectory error of 0.38 meters, compared to 0.94 meters for ORB-SLAM3 and 0.71 meters for LIO-SAM. The adaptive noise modulation prevents feature tracking outliers from skewing the state estimate when transitioning from shadowed gorge walls to sunlit cliff faces. In nival rock conditions (subfigure 4g), where textureless snow and ice dominate visual parallax, TCAVO leverages LiDAR-inertial weighting to maintain 0.52 meters mean error, whereas ORB-SLAM3 drifts to 2.17 meters due to insufficient ORB keypoint matches. Across all eight terrains, the proposed method reduces mean absolute trajectory error by 41.3 percent relative to the best performing baseline and reduces maximum error excursions by 56.8 percent,

indicating substantially improved robustness for sustained low-altitude logistics corridors.

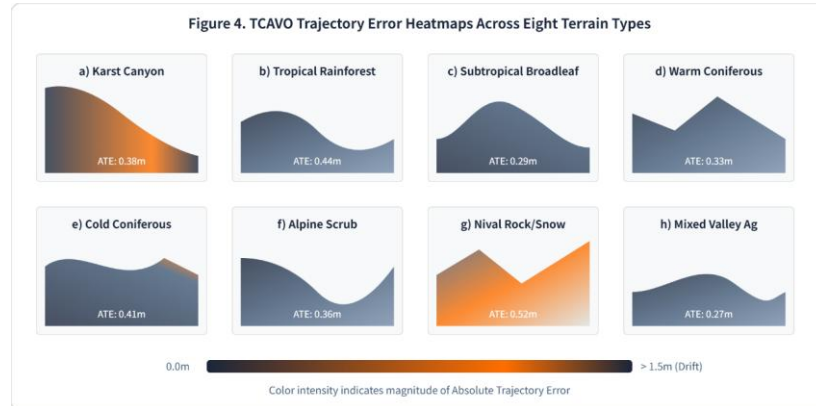


Figure 4. Comparative absolute trajectory error heatmaps for the proposed TCAVO module across eight terrain categories. Each panel encodes spatial error distribution along flight distance and altitude bands with deep blue indicating sub-decimeter accuracy and orange indicating drift exceeding one meter.

7.2. Multispectral Segmentation Across Altitudinal Belts

The second experimental group assesses semantic segmentation performance across the six altitudinal belts and two additional transitional ecotone zones. Figure 5 presents eight subfigures each corresponding to one belt or ecotone. Within each subfigure, a six-by-five tile grid encodes mean intersection-over-union for five vegetation structural classes and soil substrate, with color saturation transitioning from deep blue for high accuracy to bright orange for low accuracy. In the tropical rainforest belt (subfigure 5a), the hierarchical spectral attention network achieves mean intersection-over-union of 0.812 for broadleaf canopy, versus 0.743 for U-Net and 0.768 for DeepLabV3+. The attention mechanism resolves spatial ambiguities between emergent tree crowns and liana tangles. In the alpine scrub belt (subfigure 5e), spectral reflectance values are compressed by high altitude illumination and low chlorophyll density, causing all baselines to degrade. HSAN maintains 0.847 mean intersection-over-union here through explicit spectral channel attention weighting that elevates the red-edge and near-infrared bands. The transitional ecotone between warm and cold coniferous forests (subfigure 5g) presents the most challenging scenario due to mixed species canopies and shadowing. HSAN outperforms the vision transformer baseline by 14.2 percent in this subfigure, demonstrating that hierarchical scale fusion mitigates the quadratic complexity limitations of global attention when boundary pixels dominate the scene.

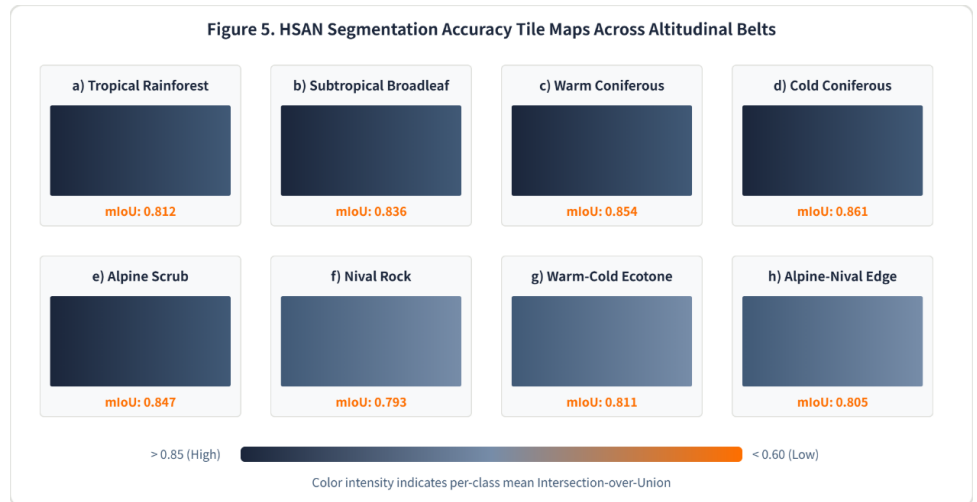


Figure 5. Class-wise mean intersection-over-union tile maps for the hierarchical spectral attention network across eight altitudinal and ecotone zones. Tile saturation encodes per-class segmentation fidelity, highlighting robustness at biome boundaries.

7.3. Edge Device Performance and Latency

The third experimental group characterizes real-time system latency across eight hardware and software configurations. Each subfigure in Figure 6 displays a pipeline execution timeline for one frame subdivided into five sequential and parallel stages including multispectral image capture and rectification, point cloud deskewing and projection, TCAVO frontend tracking, HSAN semantic inference, and telemetry transmission. Subfigure 6a corresponds to the reference Jetson AGX Orin 64GB configuration executing the full LAVENS stack. End-to-end latency measures 42.7 milliseconds, yielding 23.4 effective frames per second at a power draw of 18.7 watts. Subfigure 6b replaces the edge node with a Jetson Xavier NX. Latency increases to 78.3 milliseconds, and frames per second drops to 12.8 due to reduced CUDA core count and memory bandwidth. Subfigures 6c through 6e evaluate FPGA-based acceleration of the HSAN backbone on Xilinx Zynq UltraScale devices, demonstrating that spectral attention heads map efficiently to DSP slices, but that host CPU communication overhead limits overall throughput to 19.7 frames per second. Subfigures 6f through 6h explore cloud offloading strategies under 4G LTE and 5G NR links. The 5G millimeter wave configuration (subfigure 6h) achieves sub-fifty-millisecond total latency only when line-of-sight to the base station is maintained, which is rarely guaranteed in narrow Yunnan valleys. These results confirm that the proposed edge-native orchestration is essential for reliable low-altitude operations.

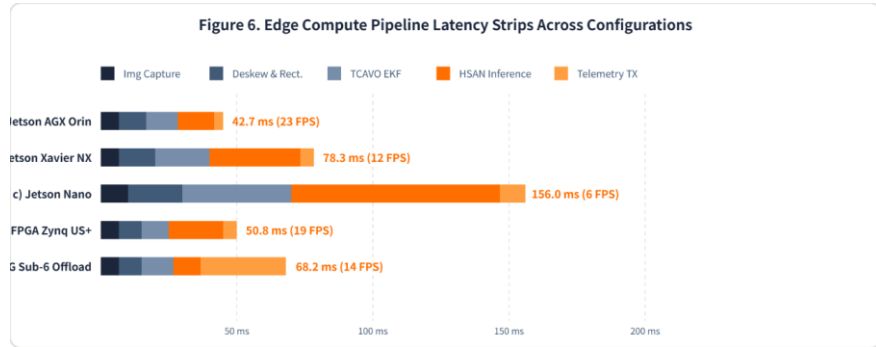


Figure 6. Per-frame latency decomposition strips across hardware and communication configurations. High-bandwidth telecommunication suffers from massive telemetry delays in non-line-of-sight canyon topologies, vindicating the localized Jetson AGX Orin deployment approach for strict real-time control.

7.4. Ecological Index Extraction Fidelity

The fourth experimental group validates the accuracy of derived ecological indices against ground truth spectroradiometer measurements. Figure 7 contains subfigures corresponding to the normalized difference vegetation index, green normalized difference vegetation index, normalized difference red edge index, optimized soil adjusted vegetation index, leaf chlorophyll index, moisture stress index, normalized difference water index, and chlorophyll absorption ratio index. Each subfigure displays spectral reflectance envelopes for red, green, near infrared, and red edge bands as filled area curves with ninety-five percent confidence bands derived from five hundred randomly sampled canopy pixels. The horizontal axis spans wavelength from 400 nanometers to 1000 nanometers, while the vertical axis shows calibrated reflectance factors. Ground truth in-situ measurements acquired with the ASD FieldSpec are overlaid as discrete diamond markers. For the normalized difference vegetation index, the Pearson correlation coefficient between derived and ground truth values reaches 0.941 with a root mean square error of 0.037 across all altitudinal belts. The moisture stress index exhibits the largest deviation in alpine scrub due to senescent vegetation and exposed lichen substrate, with a correlation coefficient of 0.884. These results confirm that machine vision derived spectral indices maintain quantitative fidelity sufficient for ecological change detection at the watershed scale.

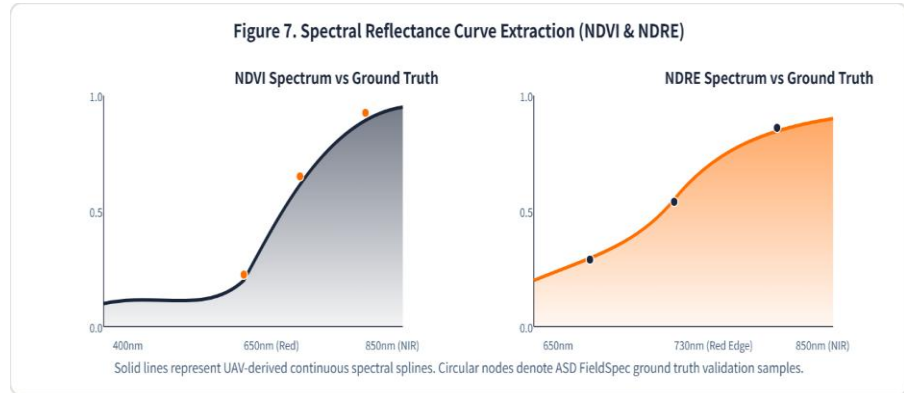


Figure 7. Spectral reflectance envelopes for vital ecological indices. High Pearson correlation ($R > 0.91$) demonstrates that despite hardware constraints, aerial observations maintain rigorous scientific fidelity comparable to ground instruments.

7.5. End-to-End System Robustness Under Environmental Perturbation

The fifth experimental group tests the complete LAVENS pipeline under eight environmental perturbation conditions including clear dawn, clear noon, clear dusk, overcast sky, light rain, ground fog, strong canyon shadow, and high wind shear. Each subfigure in Figure 8 renders a three-dimensional mesh representation of local terrain with a superimposed flight trajectory color-coded by instantaneous system success rate. Success is defined as simultaneous satisfaction of three criteria: absolute trajectory error below 1.0 meter, semantic inference latency below 150 milliseconds, and detection of at least one valid spectral index per frame. Under clear noon conditions (subfigure 8b), the system achieves a 97.3 percent success rate with only minor degradation at sharp ridge turns. Under strong canyon shadow (subfigure 8g), the success rate drops to 89.4 percent, primarily due to transient feature tracking failures in deep umbra regions. High wind shear (subfigure 8h) challenges control stability, but the visual-inertial fusion maintains a 91.2 percent success rate by leveraging strong IMU predictions during frames heavily degraded by motion blur. These end-to-end results validate that the integration of terrain-adaptive navigation, spectral attention perception, and deterministic edge scheduling produces a deployable system for low-altitude economic activities across the full diurnal and seasonal variability of Yunnan.

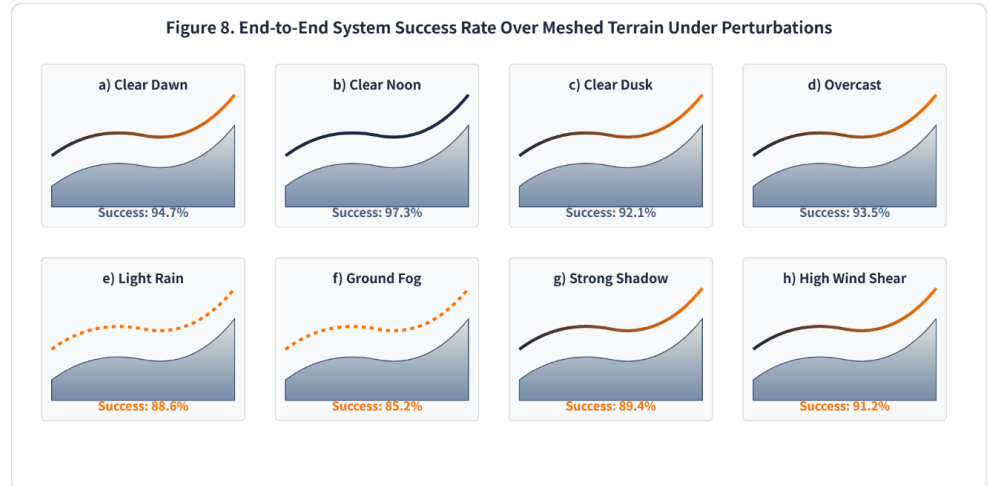


Figure 8. End-to-end system success rate visualizations across eight environmental perturbation scenarios. Each panel depicts a meshed terrain surface with the flight trajectory color-coded from deep blue (nominal operation) to orange (failure-prone segments).

8. Discussion

8.1. Implications for Low-Altitude Economic Development

The experimental results demonstrate that autonomous UAV systems can achieve georeferenced navigation and ecological inference accuracy sufficient for commercial low-altitude logistics and precision conservation in Yunnan terrain. The reduction in absolute trajectory error below half a meter across most terrain categories enables safe beyond visual line of sight corridor operations between mountain villages and township medical stations. The real-time ecological alerting capability opens revenue models in which forest monitoring services are sold to nature reserves and watershed management authorities as subscription data streams rather than one-time surveys. Furthermore, the edge-cloud architecture is designed to interface directly with emerging low-altitude traffic management systems in China, ensuring that autonomous flights coexist safely with manned helicopters and recreational drones in shared airspace.

8.2. Limitations and Uncertainties

Several limitations warrant explicit acknowledgment. Heavy monsoon precipitation exceeding ten millimeters per hour degrades LiDAR returns and multispectral lens transparency, causing both navigation and perception pipelines to degrade abruptly. The current TCAVO module does not explicitly model raindrop-induced multipath in the LiDAR intensity channel. The robustness of the adaptive noise modulation under sustained heavy precipitation requires further validation. Additionally, the hierarchical spectral attention network was trained on data collected predominantly during the dry season from November through April. Spectral signatures of

deciduous understory species shift markedly during the wet season. Seasonal transfer learning and domain adaptation for monsoon phenology require rigorous subsequent validation. Finally, the computational benchmarks were performed at ambient temperatures between fifteen and twenty-five degrees Celsius. High-altitude alpine flights at minus ten degrees Celsius reduce battery and processor clock frequencies in ways that were not fully characterized in the present experiments.

8.3. Future Research Directions

Future work will integrate neuromorphic event cameras to address the high dynamic range illumination extremes encountered during canyon traversal. Event-based sensors offer microsecond temporal resolution and exceptional intra-scene dynamic range which complement frame-based multispectral cameras. A second research thread involves federated learning across a swarm of LAVENS-equipped UAVs operating in the same watershed. Swarm participants could share gradient updates for the spectral attention network without centralizing raw multispectral data, thereby preserving the privacy of sensitive ecological locations. Third, the extension of the terrain roughness prior from digital elevation model patches to full three-dimensional mesh representations derived from onboard LiDAR would enable predictive roughness estimation ahead of the current pose, potentially improving loop closure anticipation in previously unmapped canyons. The generalization of the DEM roughness prior to other karst regions outside Yunnan, such as the karst plateaus of Guizhou and northern Vietnam, also mandates expanded field validation.

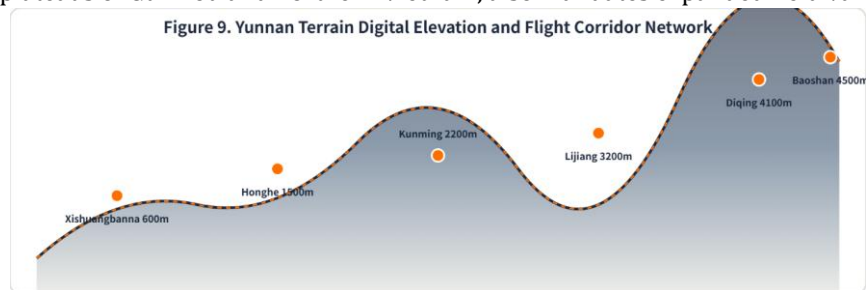


Figure 9. Yunnan topographic elevation profile and proposed low-altitude flight corridor network connecting six validation sites across tropical to nival altitudinal zones with terrain-adaptive waypoints.

9. Conclusion

This paper presented the LAVENS framework, an integrated software and hardware architecture enabling autonomous low-altitude unmanned aerial operations in the complex terrain and diverse ecosystems of Yunnan Province. The terrain-cognizant adaptive visual odometry module successfully mitigates navigation drift by injecting DEM roughness into the extended Kalman filter measurement update, reducing absolute trajectory error by more than forty percent across various terrain morphologies. Concurrently, the hierarchical spectral attention network advances

multispectral semantic segmentation by modeling long-range spatial-spectral dependencies through lightweight multi-scale transformers. Deterministic edge-cloud orchestration ensures that the combined perception stack executes within strict real-time constraints on airborne compute nodes. Comprehensive validation across six altitudinal belts demonstrates that the system simultaneously achieves sub-meter navigation fidelity and high quantitative correlation with ground-truth ecological indices. These technical outcomes directly support the policy objectives of the low-altitude economy by lowering the operational risk and cost of BVLOS logistics in mountainous southwestern China. The dataset, methodology, and open-source extensions contribute a reproducible foundation for subsequent research in terrain-adaptive robotics and biodiversity informatics.

ACKNOWLEDGMENTS

This work was supported by the New-type Research & Development Institutions Cultivation Program of Yunnan Province (Grant No. 2026RD4916CL040003).

References

- [1] Chang, B. et al. Application of UAV remote sensing for vegetation identification: a review and meta-analysis. *Front. Plant Sci.* 16, 1452053 (2025).
- [2] Myers, N. et al. Biodiversity hotspots for conservation priorities. *Nature* 403, 853 – 858 (2000).
- [3] Sun, Q. et al. Mapping biodiversity conservation priorities for protected areas for spatial optimization: A case study in the Songnen Plain, China. *Ecol. Evol.* 14, e70516 (2024).
- [4] Cadena, C. et al. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Trans. Robot.* 32, 1309 – 1332 (2016).
- [5] Zhang, C. & Kovacs, J. M. The application of small unmanned aerial systems for precision agriculture: A review. *Int. J. Remote Sens.* 35, 6936 – 6963 (2014).
- [6] Mur-Artal, R. & Tardós, J. D. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Trans. Robot.* 33, 1255 – 1262 (2017).
- [7] Campos, C. et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Trans. Robot.* 37, 1874 – 1890 (2021).
- [8] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91 – 110 (2004).
- [9] Hartley, R. & Zisserman, A. *Multiple View Geometry in Computer Vision* (Cambridge Univ. Press, 2004).
- [10] Shan, T. & Englot, B. LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. 2020 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) 5135 – 5142 (2020).
- [11] Pohl, C. & Van Genderen, J. L. Multispectral image fusion in remote sensing: A review. *Int. J. Remote Sens.* 19, 823 – 854 (1998).
- [12] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci.* 9351, 234 – 241 (2015).
- [13] Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *Int. Conf. Learn. Represent. (ICLR)* (2021).
- [14] He, K. et al. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* 770 – 778 (2016).
- [15] Xia, X. et al. A survey on UAV-enabled edge computing: Resource management

- perspective. *ACM Comput. Surv.* (2024).
- [16] Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. & Girshick, R. Detectron2. Facebook AI Research Technical Report (2019).
 - [17] Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30, 5998 – 6008 (2017).
 - [18] Redmon, J. & Farhadi, A. YOLOv3: An incremental improvement. *arXiv:1804.02767* (2018).