

SASGeo: Stability-Aware Semantic Map Localization for GNSS-Denied UAVs – A Framework and Synthetic Proof of Concept

Natalia Trukhina and Vadim Vashkelis
Embedded Intelligence Lab (emilab.org)
{ntrukhina, vvashkelis}@emilab.org

Abstract—GNSS-denied unmanned aerial vehicles require occasional absolute position fixes to bound the drift of visual-inertial odometry. Cross-view image retrieval can provide such fixes, but raw appearance is sensitive to season, illumination, viewpoint, map age, and sensor modality. We propose SASGeo, a semantic map-localization framework that represents the environment through persistent structures such as roads, buildings, waterways, railways, intersections, and field boundaries. The method combines semantic raster alignment, relational graph evidence, feature stability and geographic distinctiveness, explicit positive/contradictory/unknown observations, and integrity-aware rejection of ambiguous fixes. Unlike a broad architecture-only proposal, this paper specifies concrete weighting and decision models and reports a reproducible synthetic proof of concept. In 220 randomized retrieval trials with rotation, scale changes, partial crops, occlusion, simulated map changes, and hard semantic decoys, a global semantic descriptor achieved 58.6% Recall@1, while spatial semantic matching variants achieved 94.5–95.5%. Wilson 95% intervals separate the global descriptor from the spatial variants but overlap among the spatial variants, so the experiment supports semantic geometry rather than a definitive benefit from each proposed module. The preliminary experiment does not validate real-flight navigation; rather, it demonstrates that structured semantic geometry can discriminate locations under controlled cross-view perturbations and identifies the harder aliasing, map-aging, and rejection tests required next.

Index Terms—UAV localization, GNSS-denied navigation, semantic maps, cross-view geo-localization, map matching, integrity monitoring, visual-inertial odometry.

I. INTRODUCTION

GNSS is lightweight and globally referenced, but can become unreliable under multipath, obstruction, jamming, or spoofing. Visual-inertial odometry (VIO) supplies high-rate relative motion, yet its error grows without absolute observations [1]. A UAV can obtain global corrections by matching onboard imagery to georeferenced satellite or aerial imagery. Learned cross-view retrieval has progressed through datasets such as University-1652 and SUES-200 and practical UAV-satellite matching methods [2], [3], [4], but image appearance can differ substantially with altitude, season, illumination, viewpoint, sensor modality, and map age.

Many geographic structures are more persistent than their pixels. Road topology, rivers, railways, bridges, building footprints, and coastlines may remain identifiable even when color and texture change. Prior studies have consequently explored

OpenStreetMap (OSM), vector maps, semantic embeddings, object graphs, road-geometry bird’s-eye-view (BEV) calibration, and sequential filters [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. However, the literature generally covers subsets of five requirements that matter for safety-critical map fixes: dense semantic evidence, relational verification, temporal consistency, explicit persistence modeling, and a calibrated option to reject ambiguous observations.

This paper proposes SASGeo, whose technical thesis is that an absolute map fix should be based on the joint evidence of (i) dense semantic alignment, (ii) semantic-object relationships, (iii) temporal and map-age persistence, and (iv) an integrity decision that can withhold the fix. The contributions are:

- an operational stability-and-distinctiveness model rather than an unspecified semantic weight;
- a joint raster-graph-temporal localization objective with separate positive, contradictory, and unknown evidence;
- a decision-theoretic acceptance rule and risk-coverage evaluation for false absolute fixes;
- a reproducible synthetic proof of concept with controlled perturbations, hard decoys, confidence intervals, and ablated risk-coverage curves;
- an embedded implementation and real-world evaluation plan that clearly distinguishes design targets from measured results.

The experiment is intentionally modest. It tests the central semantic-discrimination mechanism, not end-to-end flight performance, learned perception, or calibrated field integrity.

II. RELATED WORK AND POSITIONING

A. Cross-view and semantic map localization

Cross-view methods commonly retrieve a satellite tile from a UAV image and optionally refine the pose. University-1652, SUES-200, and location-classification matching established strong benchmark-driven formulations for UAV-satellite retrieval [2], [3], [4]. Hierarchical matching adds semantic or structural constraints before fine correspondences [13]. Semantic-map methods instead align camera observations with vector or rasterized geographic information: Vojir et al. used building instances [5]; Zhou et al. used image-to-OSM likelihoods with sequential Monte Carlo [6]; Schmidt et al. and Zilke matched semantic UAV observations to OSM/map

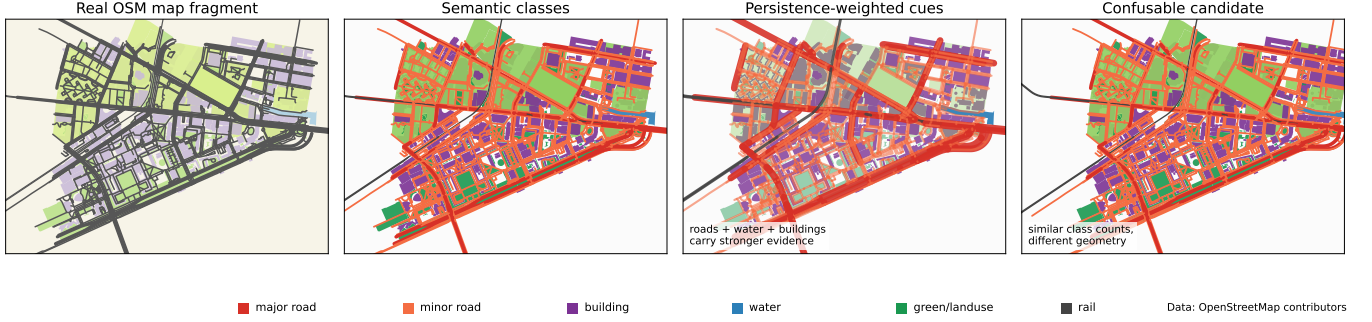


Fig. 1. Reader-level intuition for SASGEO using a real OpenStreetMap vector extract near Cambridge, Massachusetts. A map fragment is converted into semantic classes such as roads, buildings, water, green/land-use areas, and railways. Persistent and distinctive structures receive stronger evidence weights, while a confusable candidate can share similar class counts but differ in semantic geometry. Data: © OpenStreetMap contributors, ODbL; rendered locally from vector features rather than from a map tile.

data [7], [8]; Ouyang et al. combined vector-shape matching and particle filtering [9]; and VecMapLocNet learned UAV-to-vector-map features [11]. Object-graph matching and semantic-weighted particle filtering are recent complementary directions [14], [15].

Table I clarifies that the claimed novelty is not any individual component. It is the integrated probabilistic role assigned to each component: raster evidence proposes and refines poses; object relations verify structure; persistence weights determine evidential reliability; and an integrity model decides whether the result may enter the navigation estimator.

III. PROBLEM FORMULATION

Let the map contain semantic raster layers $M^c(\mathbf{r})$, vector entities, and a graph $G_M = (V_M, E_M)$ for classes $c \in \mathcal{C}$. A UAV observation window produces a local bird's-eye semantic map $L_t^c(\mathbf{r})$ and graph G_t . The candidate map-relative pose is

$$\xi_t = [x_t, y_t, \psi_t, s_t]^T, \quad (1)$$

where s_t optionally absorbs residual scale uncertainty. The objective is not merely to rank tiles but to estimate a posterior

$$p(\xi_t \mid L_{1:t}, G_{1:t}, U_{1:t}, \mathcal{M}) \quad (2)$$

and determine whether its best mode is safe to use as an absolute correction.

For class c and local cell $\mathbf{r}$, the observation state is

$$z^c(\mathbf{r}) \in \{\text{present, absent, unknown}\}. \quad (3)$$

Unknown means that occlusion, truncation, or perception uncertainty prevents a conclusion; it must not be counted as negative evidence.

IV. SASGEO METHOD

A. Semantic BEV and temporal accumulation

Camera semantic predictions are projected to a local ground plane using calibration, attitude, altitude, and optionally a DEM; road-geometry calibration can estimate such BEV homographies from lanes, borders, and crosswalks in oblique

UAV video [12]. Multiple frames are warped by VIO and accumulated:

$$L_t^c(\mathbf{r}) = \frac{\sum_{\tau=t-K}^t \omega_\tau(\mathbf{r}) \mathcal{W}(S_\tau^c, T_{t \leftarrow \tau})}{\sum_{\tau=t-K}^t \omega_\tau(\mathbf{r}) + \epsilon}, \quad (4)$$

where ω_τ reflects perception and projection uncertainty.

B. Stability and distinctiveness

Persistence and geographic informativeness are separated. For feature i ,

$$\sigma_i = p_{c_i} \exp(-\lambda_{c_i} a_i) d_{\text{season}}(c_i, s_i) v_i q_i, \quad (5)$$

where p_c is a class prior, a_i is map age, λ_c is a class-specific decay rate, d_{season} is seasonal compatibility, v_i is visibility, and q_i is perception confidence. Distinctiveness is

$$\nu_i = -\log(P(\text{similar configuration to } i) + \epsilon), \quad (6)$$

and the final evidence weight is $w_i = \sigma_i \nu_i$. A straight road can therefore be highly persistent but weakly distinctive; a bridge near a rare junction can be both.

C. Three-state semantic raster likelihood

A compatibility function distinguishes agreement, contradiction, and missing evidence:

$$\ell(z, m) = \begin{cases} +\lambda_c^+, & z = \text{present}, m = \text{present}, \\ -\lambda_c^-, & z = \text{present}, m = \text{absent}, \\ -\bar{\lambda}_c^-, & z = \text{absent}, m = \text{present}, \\ 0, & z = \text{unknown}. \end{cases} \quad (7)$$

The raster objective is

$$J_R(\xi) = \sum_{c, \mathbf{r}} w^c(\mathbf{r}) \ell(z^c(\mathbf{r}), M^c(T_\xi \mathbf{r})). \quad (8)$$

Distance-transform compatibility is used for thin lines so small map-width errors do not dominate.

TABLE I

POSITIONING RELATIVE TO REPRESENTATIVE SEMANTIC, VECTOR-MAP, AND CROSS-VIEW UAV LOCALIZATION WORK. “PERSIST.” DENOTES EXPLICIT PERSISTENCE MODELING UNDER MAP AGE, SEASON, VISIBILITY, OR TRANSIENT CONTENT; “REJECT” DENOTES AN EXPLICIT DECISION TO WITHHOLD AN UNCERTAIN ABSOLUTE FIX. SASGEO ENTRIES ARE PROPOSED CAPABILITIES, NOT REAL-FLIGHT VALIDATION CLAIMS.

Method	Main reference cue	Dense	Graph	Temporal	Persist.	Reject	Unknown
Vojir et al. [5]	OSM building instances	partial	partial	—	—	—	—
Zhou et al. [6]	OSM semantic likelihood	✓	—	SMC	—	partial	—
Ouyang et al. [9]	Vector-map shapes	partial	partial	PF	—	partial	—
Hu et al. [10]	OSM with satellite imagery	✓	—	—	—	—	—
VecMapLoc. [11]	UAV-to-vector-map features	✓	partial	—	—	—	—
Zhang et al. [13]	Satellite coarse-to-fine matching	✓	partial	—	—	margin	—
Liu et al. [14]	Object graph matching	—	✓	—	—	—	—
SWA-PF [15]	Semantic-weighted particle filter	✓	—	PF	class weights	partial	—
SASGEO	Raster, vector, and graph semantics	✓	✓	proposed	explicit	calibrated	explicit

D. Relational graph verification

Nodes encode intersections, building clusters, river bends, crossings, and polygons. Edges encode relative distance, bearing, adjacency, connectivity, and crossing. Given correspondence hypothesis π ,

$$J_G(\xi, \pi) = \sum_{i \leftrightarrow j} w_i \phi_V(i, j) + \sum_{(i,k) \leftrightarrow (j,l)} w_{ik} \phi_E((i,k), (j,l)). \quad (9)$$

Raster matching efficiently generates pose candidates; graph matching suppresses candidates with similar occupancy but incompatible topology.

E. Temporal consistency and joint score

A VIO increment $\Delta \hat{T}_t$ induces

$$J_T(\xi_t) = -\|\text{Log}(\xi_{t-1}^{-1} \xi_t \Delta \hat{T}_t^{-1})\|_{\Sigma_T^{-1}}^2. \quad (10)$$

The joint objective is

$$J = \alpha_t J_R + \beta_t J_G + \gamma_t J_T - \eta J_{\text{contradiction}}, \quad (11)$$

with evidence-adaptive weights rather than fixed constants.

V. INTEGRITY-AWARE POSE ACCEPTANCE

A false fix is a candidate whose error exceeds mission tolerance $e_{\max}$. Let F_t denote this event. Accepting and rejecting incur

$$C_{\text{accept}} = C_{\text{FF}} P(F_t = 1 \mid E_t), \quad (12)$$

$$C_{\text{reject}} = C_{\text{MR}} P(F_t = 0 \mid E_t). \quad (13)$$

The fix is accepted only when $C_{\text{accept}} < C_{\text{reject}}$, equivalently

$$P(F_t = 1 \mid E_t) < \frac{C_{\text{MR}}}{C_{\text{FF}} + C_{\text{MR}}}. \quad (14)$$

The probability can be calibrated from candidate margin, posterior entropy, cross-class agreement, graph consistency, unknown-evidence fraction, and VIO residual. The operating point is reported using a risk–coverage curve, where coverage is the fraction of accepted fixes and risk is the false-fix rate among accepted fixes.

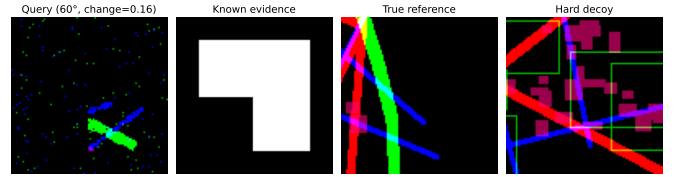


Fig. 2. Example synthetic query, observability mask, true reference, and hard semantic decoy.

VI. SYNTHETIC PROOF OF CONCEPT

A. Purpose and protocol

The experiment asks whether structured semantic geometry can retrieve a transformed local observation from hard semantic decoys. It does not model real aerial image segmentation or closed-loop flight.

We procedurally generated 30 semantic regions of 256×256 cells with five layers: major roads, buildings, water, minor roads, and field boundaries. Candidate crops measured 96×96 cells. Each query was transformed by a randomly selected rotation in $\{0, 30, 60, 90\}^\circ$, scale in $\{0.85, 1.0, 1.15\}$, retained crop fraction in $\{0.70, 0.85, 1.0\}$, rectangular occlusion up to 24%, and class-dependent map-change probability up to 0.32. Decoys were selected from the nearest semantic descriptors rather than uniformly, making them intentionally confusable. We ran 220 trials with ten candidates each using seed 42; the released script accepts additional seeds for longer aggregate runs.

Compared variants were: (1) global semantic descriptor; (2) uniform raster distance-transform matching; (3) raster plus graph descriptor; (4) stability-weighted raster plus graph; and (5) the full model that additionally ignores unknown regions. The graph descriptor used connected-component counts and rotation-invariant pairwise-distance histograms. These simplified components test the formulation, not a production implementation.

B. Results

Table III shows that a global semantic description is insufficient for hard decoys, while spatial raster alignment raises Recall@1 from 58.6% to 94.5%. The Wilson intervals for

TABLE II

SCOPE OF THE CURRENT SYNTHETIC IMPLEMENTATION RELATIVE TO THE PROPOSED FORMULATION. THIS TABLE IS INCLUDED TO MAKE CLEAR WHICH ELEMENTS ARE EXPERIMENTALLY EXERCISED AND WHICH REMAIN ENGINEERING TARGETS FOR THE REAL-WORLD STUDY.

Component	Proposed formulation	Synthetic proof of concept
Semantic BEV, Eq. (4)	Multi-frame projection from camera semantics using calibration, attitude, altitude, and VIO.	Starts from synthetic semantic layers; no raw imagery, segmentation model, projection error, or VIO drift.
Stability, Eq. (5)	Class prior, map age, seasonal compatibility, visibility, and perception confidence.	Uses hand-set class weights; no longitudinal calibration of age or season.
Distinctiveness, Eq. (6)	Geographic rarity of local semantic configurations.	Hard decoys are chosen by nearest global semantic descriptors; no learned rarity model.
Raster likelihood, Eqs. (7)–(8)	Three-state positive, contradictory, and unknown evidence with semantic weights.	Distance-transform raster score with contradiction penalty; unknown mask removes unobserved cells from contradiction.
Graph term, Eq. (9)	Correspondence-hypothesis graph matching over objects and relations.	Connected-component counts and rotation-invariant pairwise-distance histograms.
Temporal term, Eq. (10)	Multi-frame VIO consistency and filter/factor-graph fusion.	Not evaluated; each retrieval trial is independent.
Integrity, Eqs. (12)–(14)	Calibrated false-fix probability and cost-sensitive acceptance.	Candidate-score margin is used as an uncalibrated rejection statistic for risk-coverage curves.

TABLE III

SYNTHETIC RETRIEVAL RESULTS OVER 220 RANDOMIZED TRIALS. RECALL@1 INTERVALS ARE WILSON 95% CONFIDENCE INTERVALS.

Method	Recall@1 (95% CI)	Recall@5	MRR
Global descriptor	0.586 [0.520, 0.649]	0.886	0.701
Raster, uniform	0.945 [0.907, 0.969]	1.000	0.966
Raster + graph	0.950 [0.913, 0.972]	0.995	0.968
+ stability	0.950 [0.913, 0.972]	0.995	0.967
Full + unknown handling	0.955 [0.918, 0.975]	1.000	0.971

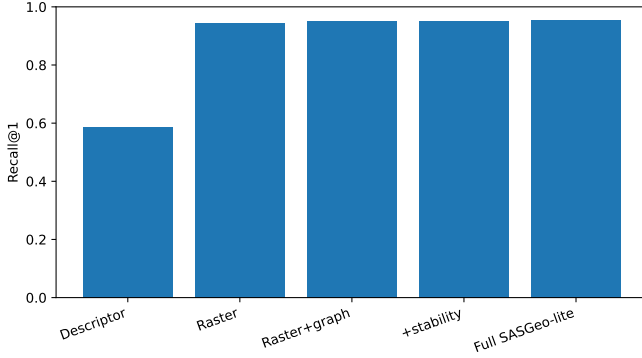


Fig. 3. Recall@1 ablation. Dense semantic geometry provides the dominant gain in this controlled setup; stability and unknown handling produce smaller changes.

the descriptor and raster variants are well separated, but the intervals for the four spatial variants overlap. Therefore the supported conclusion is that dense semantic geometry is highly informative in this controlled setting; the present benchmark does not statistically distinguish graph verification, stability weighting, or unknown-evidence handling. This is a useful negative result because it exposes a ceiling effect: raster alignment is already strong on these synthetic maps, leaving too little room for the distinctive SASGeo modules to show aggregate Recall@1 gains. Targeted aliasing tests with repetitive topology, realistic map aging, and absent-match cases are required before claiming that the stability or graph

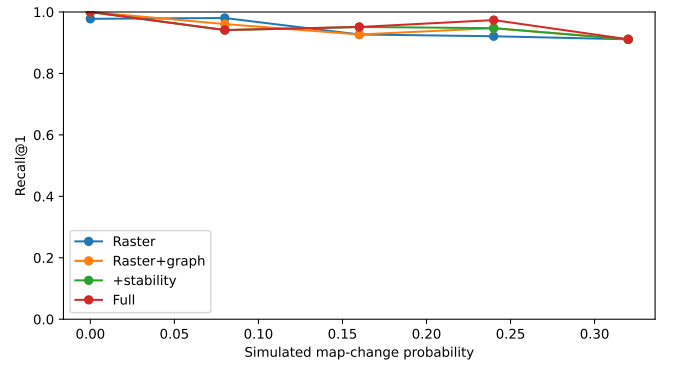


Fig. 4. Retrieval robustness versus simulated map-change probability.

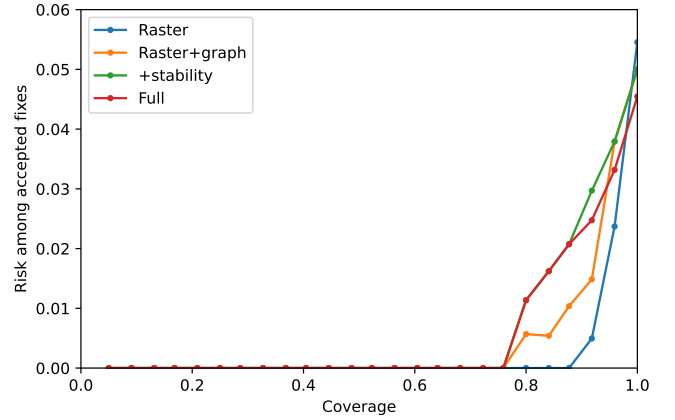


Fig. 5. Risk-coverage curves using candidate-score margin as an uncalibrated rejection statistic. Curves are shown for all spatial ablations to test whether the added terms reduce confident false fixes.

terms improve localization reliability.

The experiment therefore validates the feasibility of the scoring pipeline and motivates a second-stage study using OSM vectors and public aerial imagery, where map age, seasonal mismatch, segmentation confidence, and true unknown

regions can be measured rather than simulated.

VII. EMBEDDED IMPLEMENTATION TARGETS

The intended architecture separates a certified flight controller from a companion computer. A Jetson Orin NX-class module is a plausible target, but the following are design budgets, not measurements: VIO at camera rate; semantic segmentation and BEV accumulation at 5–10 Hz; coarse retrieval at 1–2 Hz; raster alignment over a small top- N set; graph verification on the best candidates; and absolute correction only after multi-frame confirmation. The map pyramid stores low-resolution semantic rasters for search and compact vectors/graphs for refinement. Hybrid telemetry can preserve low-rate context while transmitting selected high-detail semantic ROIs under bandwidth limits [16]; SASGEO treats this as an implementation layer. Real profiling must report latency, memory, energy, map size, communication load, and correction frequency.

VIII. REAL-WORLD EVALUATION PLAN

A reproducible next step uses OSM vectors as reference semantics and publicly distributable aerial imagery as simulated UAV observations across urban, suburban, industrial, riverside, and agricultural scenes. Evaluation must use geographic holdout regions and transformations in scale, rotation, crop, occlusion, season, and map age. Besides broad retrieval, the benchmark should include targeted aliasing: repetitive road grids, similar parcels, structurally similar intersections, absent-match queries, and map-age conflicts where raster occupancy alone is ambiguous. Required metrics include Recall@1/5, metric translation and yaw error, false-fix probability, risk-coverage area, latency, memory, and map storage. Baselines should include appearance retrieval, global semantic descriptors, raster-only matching, graph-only matching, and standard sequential filtering. Real UAV sequences are ultimately required for oblique views, motion blur, VIO coupling, and navigation recovery.

IX. LIMITATIONS

The current experiment begins from semantic layers rather than raw images and therefore does not measure perception errors. It uses discrete rotations, planar maps, simplified graphs, single-frame retrieval trials, and synthetic changes. The stability parameters are hand-set rather than calibrated from longitudinal geographic data. The confidence intervals quantify binomial uncertainty for the released 220-trial seed, not multi-region, multi-seed performance. The gains do not establish real-time feasibility or safe closed-loop operation. Feature-poor environments such as forests, deserts, snow fields, and open water may remain weakly observable.

X. CONCLUSION

SASGEO reframes UAV geo-localization as integrity-aware matching of persistent geographic structure rather than appearance retrieval alone. The proposed method integrates semantic rasters, relational graphs, temporal motion, persistence, unknown evidence, and explicit rejection. A controlled proof

of concept shows that spatial semantic alignment strongly outperforms global semantic description under hard decoys. It also reveals the current ceiling effect: the benchmark validates semantic geometry but does not yet validate the distinctive persistence and graph claims. The resulting preprint is a concrete, falsifiable starting point for OSM-to-aerial and real-flight evaluation rather than a claim of a completed navigation system.

REPRODUCIBILITY STATEMENT

The source package includes the generator, transformations, scoring functions, fixed seed, raw trials, summary metrics, figure code, and OSM vector extract for Fig. 1. The complete source code is publicly available at <https://gitlab.com/emilab-group/sasgeo>. It avoids nonstandard compiled image-processing dependencies and regenerates artifacts under figures/ and results/. Run `python run_synthetic_experiment.py`; longer checks can use `--seeds 42,43,44,45,46`.

REFERENCES

- [1] T. Qin, P. Li, and S. Shen, “VINS-Mono: A robust and versatile monocular visual-inertial state estimator,” *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [2] Z. Zheng, Y. Wei, and Y. Yang, “University-1652: A multi-view multi-source benchmark for drone-based geo-localization,” in *ACM Multimedia*, 2020, pp. 1395–1403.
- [3] R. Zhu et al., “SUES-200: A multi-height multi-scene cross-view image benchmark across drone and satellite,” *IEEE TCSVT*, vol. 33, no. 9, pp. 4825–4839, 2023.
- [4] L. Ding, J. Zhou, L. Meng, and Z. Long, “A practical cross-view image matching method between UAV and satellite for UAV-based geo-localization,” *Remote Sens.*, vol. 13, no. 1, art. 47, 2021.
- [5] T. Vojir, I. Budvytis, and R. Cipolla, “Efficient large-scale semantic visual localization in 2D maps,” in *ACCV*, 2020.
- [6] M. Zhou, X. Chen, N. Samano, C. Stachniss, and A. Calway, “Efficient localisation using images and OpenStreetMaps,” in *IROS*, 2021.
- [7] R. Schmidt, J. Rüter, S. Krause, and S. Schubert, “Vision-based self-localization for UAVs using semantic features and OpenStreetMap,” in *Proc. IEEE Aerospace Conf.*, 2025.
- [8] M. Zilke, “Visual localization of a UAV using semantic environment information and map data,” DLR Internal Report DLR-IB-FT-BS-2025-156, University of Göttingen, 2025.
- [9] C. Ouyang et al., “A semantic vector map-based approach for aircraft geo-localization,” *Results in Engineering*, vol. 21, 2024.
- [10] Y. Hu et al., “Combining OpenStreetMap with satellite imagery to enhance cross-view geo-localization,” *Sensors*, vol. 25, no. 1, art. 44, 2025.
- [11] Z. Wang et al., “VecMapLocNet: Vision-based UAV localization using vector maps in GNSS-denied environments,” *ISPRS J. Photogramm. Remote Sens.*, vol. 225, pp. 362–381, 2025.
- [12] A. Popov, N. Trukhina, and V. Vashkelis, “Mobile traffic camera calibration from road geometry for UAV-based traffic surveillance,” arXiv:2605.11900, 2026.
- [13] X. Zhang et al., “Hierarchical image matching for UAV absolute visual localization via semantic and structural constraints,” arXiv:2506.09748, 2025.
- [14] T. Liu, K. Ren, and Q. Chen, “Object detection as an optional basis: A graph matching network for cross-view UAV localization,” arXiv:2511.02489, 2025.
- [15] Y. Chen et al., “SWA-PF: Semantic-weighted adaptive particle filter for memory-efficient 4-DoF UAV localization in GNSS-denied environments,” arXiv:2509.13795, 2025.
- [16] N. Trukhina and V. Vashkelis, “Hybrid visual telemetry for bandwidth-constrained robotic vision: HEVC base video and JPEG ROI stills,” arXiv:2605.01826, 2026.