SEEING IS NOT SEARCHING: GOAL-CONDITIONED EFFECTIVE COVERAGE FOR AERIAL OBJECT GOAL NAVIGATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Aerial object goal navigation (ObjectNav) requires an unmanned aerial vehicle (UAV) to actively search for a specified target using onboard visual observations. Existing methods commonly measure exploration progress through geometric coverage, treating observed regions as searched regardless of the sensing conditions under which they are acquired. In aerial navigation this assumption is problematic: altitude simultaneously affects spatial coverage and goal observability, so higher altitudes enlarge the observed area while reducing the visual resolvability of the specified target, and the same geometric coverage can provide substantially different search value across goals. To capture this goal-dependent search utility, we propose *goal-conditioned effective coverage*, which weights newly observed space by the observability of the specified goal under the current sensing geometry. Because reliable target cues are unavailable before detection, we further derive a goal-conditioned altitude prior from the target footprint and camera geometry, providing a reference sensing altitude from the start of the episode. The prior enters policy observations and observability-weighted exploration rewards, and a staged reinforcement learning scheme progressively integrates altitude control, obstacle avoidance, and spatial exploration into a unified aerial search policy. Experiments on UAV-ON achieve a success rate (SR) of 25.17%, an oracle success rate (OSR) of 41.04%, and success weighted by path length (SPL) of 19.10%, surpassing the strongest published results by 5.67, 11.74, and 8.96 percentage points, respectively.

1 INTRODUCTION

Aerial object goal navigation introduces a controllable sensing dimension that is largely absent in ground-agent object goal navigation (ObjectNav) (Anderson et al., 2018a; Batra et al., 2020): unlike ground agents, unmanned aerial vehicles (UAVs) can actively vary altitude during search—an active-perception capability (Bajcsy, 1988). Under fixed camera geometry and a locally planar ground approximation, increasing altitude enlarges the ground footprint as h^2 while reducing the projected image area of a specified goal approximately as $1/h^2$. This creates an inherent sensing trade-off: higher-altitude observations expose more space, whereas lower-altitude observations preserve stronger visual evidence of the target (Fig. 1). The trade-off is classical in aerial sensing, where flight altitude determines the ground sample distance (Cabreira et al., 2019) and target detectability has long been characterized by resolution-based criteria (Johnson, 1958; Sjaardema et al., 2015). Consequently, in aerial ObjectNav, observing more space does not necessarily provide more useful evidence for finding the specified goal: spatial exposure and goal-directed search utility can diverge.

Existing ObjectNav exploration methods largely characterize progress through spatial structure. Classical frontier exploration expands the boundary between known and unknown space (Yamauchi, 1997), while mapping-based navigation explicitly tracks explored regions and uses spatial or semantic maps to guide subsequent motion (Chaplot et al., 2020a;b). More recent methods incorporate goal information to decide where exploration should proceed: PONI predicts goal-conditioned potentials over partially observed semantic maps (Ramakrishnan et al., 2022), and zero-shot approaches bring vision–language knowledge to bear on this decision—CoW combines vision–language object localization with classical exploration (Gadre et al., 2023), while ESC and VLFM prioritize semantically

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

promising locations or frontiers (Zhou et al., 2023; Yokoyama et al., 2024). Intrinsic exploration objectives similarly encourage state novelty rather than the perceptual utility of individual observations (Pathak et al., 2017; Burda et al., 2019). These approaches substantially improve where an agent should search, but the notion of search progress remains predominantly spatial: once new space has been observed, its contribution is not explicitly conditioned on whether the specified goal was visually resolvable under the sensing geometry of that observation.

This distinction becomes consequential in aerial navigation. Aerial vision–language navigation benchmarks such as AerialVLN (Liu et al., 2023), CityNav (Lee et al., 2025), and OpenFly (Gao et al., 2026) extend embodied navigation to large-scale three-dimensional environments with altitude-varying viewpoints, while UAV-ON (Xiao et al., 2025) further formulates aerial ObjectNav as active search for a specified goal whose location is unknown. Recent methods such as APEX (Zhang et al., 2026) and OctMem-Agent (Zhou et al., 2026) improve this search through three-dimensional representations, memory, goal conditioning, and explicit exploration mechanisms. Yet these advances do not resolve how an acquired observation should be credited as search progress: geometric coverage records whether new space has been observed (Chaplot et al., 2020a; Galceran & Carreras, 2013), but not whether the specified goal was resolvable when that observation was acquired. The same newly covered region can therefore provide different search value across goals and acquisition altitudes, a discrepancy we term the *coverage–observability mismatch*. Under a spatial-search model, the utility of a new observation depends jointly on its spatial search value and goal observability, expressed as $U_t(g) \propto \Delta C_t q(h_t, g)$. Conventional geometric coverage retains only the spatial term. As altitude increases, a policy may therefore accumulate substantial geometric coverage while producing insufficient goal-specific visual evidence, creating a characteristic high-coverage, low-resolvability failure mode.

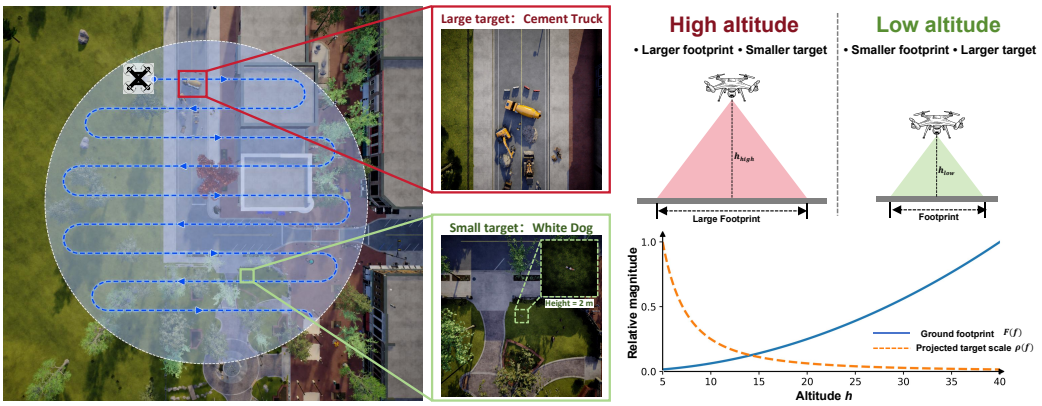
We address this mismatch with *goal-conditioned effective coverage*, which weights each coverage increment according to the observability of the specified goal under the acquisition geometry. Because reliable target-specific visual evidence is unavailable before detection, rather than learning an additional estimator we exploit quantities fixed by the task itself: the annotated goal footprint and known camera geometry yield a closed-form reference altitude $h(g)$ from a classical detectability relation (Johnson, 1958; Sjaardema et al., 2015). This reference defines an operational observability band and an altitude-dependent gain $\phi(h; g)$ that converts geometric coverage into goal-conditioned effective coverage. We inject the resulting prior into both policy observations and exploration rewards, and use staged reinforcement learning to progressively integrate altitude control, obstacle avoidance, and spatial exploration into a unified aerial search policy.

Our contributions are:

- **Goal-conditioned effective coverage.** We identify the *coverage–observability mismatch* in aerial object goal navigation: geometric coverage measures spatial exploration without accounting for whether the specified goal was resolvable under the acquisition geometry. We introduce *goal-conditioned effective coverage*, which weights geometric coverage by goal observability so that progress reflects both spatial coverage and goal-specific perceptual utility.
- **A goal-conditioned altitude prior.** We derive a closed-form reference altitude $h^*(g)$ from the annotated goal footprint and camera geometry, defining an observability band and an altitude-dependent gain $\phi(h; g)$ that makes this observability weighting available before the goal is detected.
- **Staged reinforcement learning for aerial search.** We incorporate the altitude prior into policy observations and effective-coverage rewards, and use staged reinforcement learning to progressively integrate altitude control, obstacle avoidance, and spatial exploration.

2 Related Work

Object Goal Navigation. Object Goal Navigation (ObjectNav) requires an agent to locate a specified object in an unseen environment (Anderson et al., 2018a; Batra et al., 2020), typically studied in photorealistic simulators (Kolve et al., 2017; Xia et al., 2018; Savva et al., 2019; Szot et al., 2021). Frontier-based exploration provides the classical geometric foundation (Yamauchi, 1997), and target-driven visual navigation was an early deep reinforcement learning demonstration (Zhu



(a) Same coverage, different target observability **(b) Altitude trade-off in aerial search**

Figure 1: Same coverage, different target observability. **(a)** Downward-camera observations at 20 m in the CityStreet scene: a large target remains resolvable, while a small target becomes difficult to distinguish. **(b)** Increasing altitude enlarges the ground footprint but reduces the projected target scale, revealing the coverage–observability trade-off.

et al., 2017). Mapping-based agents explore explicitly, from Active Neural SLAM (Chaplot et al., 2020a) and goal-oriented semantic exploration (Chaplot et al., 2020b) to auxiliary-task training (Ye et al., 2021), while others learn goal-conditioned priors over partial maps to bias exploration toward promising regions (Ramakrishnan et al., 2022; Zhai & Wang, 2023). Zero-shot agents replace task-specific training with web-scale models: ZSON navigates from multimodal goal embeddings (Majumdar et al., 2022), CoW drives language-based search through CLIP-scale image–text alignment (Gadre et al., 2023; Radford et al., 2021), ESC combines open-vocabulary visual grounding with language-model commonsense reasoning (Zhou et al., 2023), and VLFM ranks candidate frontiers by vision–language similarity (Yokoyama et al., 2024), relying on open-vocabulary detection. Intrinsic exploration rewards (Bellemare et al., 2016; Pathak et al., 2017; Burda et al., 2019) likewise reward the novelty of visited states rather than the perceptual utility of new observations. Vision-and-language navigation (VLN) instead conditions embodied agents on natural language: embodied question answering studies active exploration to answer questions about the environment (Das et al., 2018), Room-to-Room and its multilingual and street-level extensions follow natural-language route instructions (Anderson et al., 2018b; Ku et al., 2020; Chen et al., 2019), and ALFRED extends language-conditioned agents to household task execution (Shridhar et al., 2020); architectures range from transformer history encoders (Chen et al., 2021) to graph- and continuous-control variants (Chen et al., 2022; Hong et al., 2022) and video-based vision–language models (Zhang et al., 2024). These tasks provide language describing a route, a question, or an embodied task, whereas our setting provides only a goal specification and requires active search. On the training side we use standard deep RL with proximal policy optimization (PPO) (Schulman et al., 2017): the observability gain (§4.3) is a shaped exploration reward (Ng et al., 1999) whose weighting is goal-conditioned rather than purely spatial, and staging (§4.4) is a curriculum-style device (Bengio et al., 2009; Narvekar et al., 2020). Across these lines, progress remains tied to spatial novelty or newly observed extent—an observation's search credit never depends on whether the goal was resolvable under the acquisition geometry. We therefore define goal-conditioned effective coverage, which measures search progress by both spatial exploration and goal-specific observability.

Aerial Navigation. Aerial vision-and-language navigation extends embodied navigation to UAVs operating in large-scale three-dimensional environments, with benchmarks such as AerialVLN (Liu et al., 2023), CityNav (Lee et al., 2025), and OpenFly (Gao et al., 2026) progressively broadening scale and diversity; AerialVLN builds its city-scale environments within AirSim (Shah et al., 2018). In parallel, learning-based control has enabled increasingly agile UAV flight, from privileged and imitation learning for high-speed traversal of cluttered natural environments (Loquercio et al., 2021) to deep reinforcement learning for champion-level drone racing (Kaufmann et al., 2023), although these controllers target agile flight control rather than goal search. Unlike instruction-following VLN, aerial ObjectNav requires the agent to actively search for a specified goal whose

location is unknown; UAV-ON formalizes this setting for open-world outdoor environments (Xiao et al., 2025). Recent methods accordingly improve how such search is represented and conducted: APEX constructs dynamic 3D attraction, exploration, and obstacle maps for reinforcement-learning-based navigation (Zhang et al., 2026), while OctMem-Agent aggregates RGB-D observations into an adaptive octree memory supporting scene understanding and frontier exploration (Zhou et al., 2026); coverage-like progress has a classical counterpart in coverage path planning, surveyed for ground robots (Galceran & Carreras, 2013) and UAVs (Cabreira et al., 2019), while informative path planning instead maximizes expected information gain about the environment (Marchant & Ramos, 2014). Despite these advances, exploration progress remains largely quantified by spatial novelty, coverage, or frontier expansion, treating newly observed space as useful search progress regardless of whether the specified goal is visually resolvable under the sensing geometry—a limitation that matters especially for aerial agents, since altitude simultaneously changes the spatial footprint and the visual scale of the goal. We therefore introduce goal-conditioned effective coverage, which reweights geometric exploration progress according to goal observability, using a closed-form altitude prior derived from the goal footprint and camera geometry.

3 Problem Formulation

We follow the UAV-ON open-world aerial ObjectNav task (Xiao et al., 2025). At the beginning of each episode, the UAV is initialized at a random pose in a large outdoor environment and receives a goal specification $g = (\text{name}, \text{size}, \text{description})$. From the size annotation (a length$\times$width pair in meters) we compute the instance-level goal footprint area A_g, used to construct the observability prior and independent of online visual detection. The goal lies within a horizontal search radius of $50\,\text{m}$.

At each step, the agent receives synchronized RGB-D observations from front-, left-, right-, and downward-facing cameras, each with a 90° field of view. The learned policy has no access to GPS, a global map, or other privileged state at test time, and selects parameterized translation and rotation actions including ascent and descent. Episodes terminate upon stopping, collision, or after 150 steps; success is stopping within $\tau_d = 20\,\text{m}$ of the goal. We report success rate (SR), oracle success rate (OSR), and success weighted by path length (SPL) (Anderson et al., 2018a; Xiao et al., 2025), overall and by goal size; OSR additionally credits episodes that bring the goal within the τ_d oracle visibility range without a successful stop. During search the learned policy controls altitude, obstacle avoidance, and spatial exploration; once the target is confidently grounded, a shared post-grounding module guides the final approach and stopping for all internal variants and ablations.

4 Method

Figure 2 summarizes the proposed framework.

4.1 The coverage–observability mismatch

Geometric coverage is a common spatial measure of exploration progress. Let $\Delta C_t \geq 0$ denote the increment in geometric ground coverage obtained from the onboard observations at step t. A conventional coverage reward assigns credit in proportion to this increment, $r_t \propto \Delta C_t$, measuring how much new space has been observed but not how useful that observation is for discovering the specified goal—the same newly observed region may provide strong visual evidence for a large goal and little for a visually small one.

To formalize this distinction, we model the utility of a new observation for goal g as the product of its spatial search value and goal observability,

$$\underbrace{U_t(g)}_{\text{search utility}} \propto \underbrace{S_t}_{\text{spatial search value}} \cdot \underbrace{q(\xi_t, g)}_{\text{goal observability}}, \tag{1}$$

where S_t describes the spatial value of the newly observed region and $q(\xi_t, g)$ describes the observability of goal g under acquisition geometry ξ_t. The quantity q is an observability utility rather than a detection probability, although the two are expected to be positively related. Under a uniform

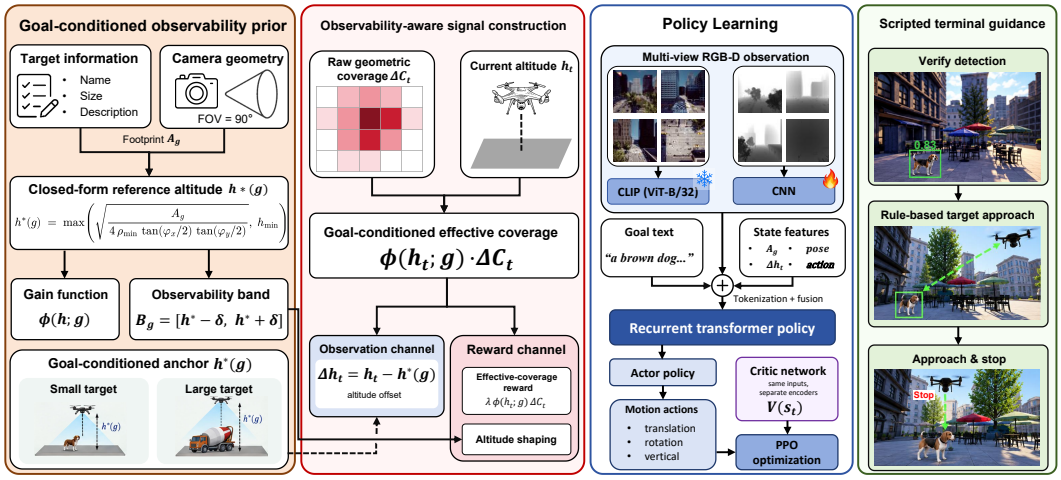


Figure 2: Method overview. From the target specification and camera geometry, we derive a goal-conditioned reference altitude $h^*(g)$, together with an observability band B_g and gain function $\phi(h; g)$. The prior constructs two learning signals: the altitude offset $h_t - h^*(g)$ for policy observation and the observability-weighted coverage $\phi(h_t; g)\Delta C_t$ for reward shaping. Multi-view RGB-D observations, goal information, and state features are fused by a recurrent transformer policy, while at evaluation time, verified detections trigger a non-learned terminal guidance procedure for target approach and stopping.

spatial-search approximation, newly observed area provides a simple proxy for the spatial term, giving $S_t \propto \Delta C_t$. For aerial search, altitude is a dominant component of the acquisition geometry, yielding

$$U_t(g) \propto \Delta C_t\, q(h_t, g). \tag{2}$$

Conventional geometric coverage effectively replaces the observability term by a constant and retains only ΔC_t. It is therefore a suitable proxy for search progress only when goal observability is approximately invariant to the sensing geometry. This assumption does not generally hold in aerial search.

To characterize this dependence, we consider a downward-looking reference camera under a planar-ground approximation. Throughout, h denotes the UAV's above-ground-level (AGL) altitude, estimated from the downward depth image. The ground footprint of the reference camera at altitude h is

$$F(h) = 4h^2 \tan(\varphi_x/2) \tan(\varphi_y/2), \tag{3}$$

where φ_x and φ_y denote the horizontal and vertical fields of view. Thus, the observed ground footprint increases quadratically with altitude. For a goal with annotated footprint area A_g, we define

$$\rho(h) = \frac{A_g}{F(h)} \tag{4}$$

as a geometric proxy for its projected visual scale under this reference model. Since $\rho(h) \propto h^{-2}$, increasing altitude simultaneously enlarges the spatial footprint of an observation and reduces the projected scale of the specified goal.

These opposing trends create the *coverage–observability mismatch*. Because the footprint grows as h^2 while the projected scale of the goal shrinks as $1/h^2$, a higher-altitude observation can receive larger geometric-coverage credit while providing lower goal-directed search utility: for two altitudes $h_1 < h_2$ and a visually demanding goal, $\Delta C(h_2)q(h_2, g) < \Delta C(h_1)q(h_1, g)$ even though $\Delta C(h_2) > \Delta C(h_1)$. The ordering induced by geometric exploration progress can therefore disagree with the ordering induced by goal-directed search utility, because a goal-independent spatial progress measure is used for a task whose perceptual utility depends jointly on the specified goal and the sensing geometry. This motivates conditioning coverage credit on goal observability.

4.2 GOAL-CONDITIONED OBSERVABILITY PRIOR

Closed-form goal-conditioned reference altitude. The search utility of Eq. 1 requires the observability term $q(h, g)$, but before the goal is reliably detected no online target-specific visual evidence is available to evaluate it. We therefore construct a *detection-independent observability prior* from quantities fixed by the task itself: the annotated goal footprint area A_g and the known camera geometry. Under the nadir-view approximation of §4.1, $\rho(h) = A_g/F(h)$ decreases monotonically with h; imposing a nominal resolvability requirement $\rho(h) \geq \rho_{\min}$ and solving at equality gives the largest altitude satisfying it,

$$\tilde{h}(g) \;=\; \sqrt{\frac{A_g}{4\,\rho_{\min}\,\tan(\varphi_x/2)\,\tan(\varphi_y/2)}}, \tag{5}$$

above which the goal's projected scale falls below $\rho_{\min}$. We further impose a reference-altitude floor $h_{\min} = 5\,\mathrm{m}$, giving the *goal-conditioned reference altitude*

$$h^*(g) \;=\; \max\Big(\tilde{h}(g),\; h_{\min}\Big)\,. \tag{6}$$

When $\tilde{h}(g) < h_{\min}$, $h^*(g)$ saturates at the floor—the reference is lower-bounded by $h_{\min}$, since such goals already have $\rho(h_{\min}) < \rho_{\min}$ (the floor regime below). Around $h^*(g)$ we define an operational *observability band* $\mathcal{B}_g = [h^* - \delta,\, h^* + \delta]$ with half-width $\delta = \max(h^*/2,\, 5\,\mathrm{m})$, used for altitude shaping rather than as a hard detectability boundary. For the 90° cameras used here and $\rho_{\min} = 0.1$, Eq. 6 simplifies to $h^*(g) = \max(1.58\sqrt{A_g},\, 5\,\mathrm{m})$: a $4\,\mathrm{m}^2$ goal yields the floor $5\,\mathrm{m}$, a $100\,\mathrm{m}^2$ one $\approx 16\,\mathrm{m}$. Because $h^*(g)$ depends only on the goal specification and camera geometry, it is available before any detection occurs; it should not be read as an optimum or a hard detectability threshold, as detector response varies continuously with altitude (§5.3).

Classical sensing interpretation. The construction follows a classical sensing principle rather than an arbitrary heuristic: the Johnson criteria (Johnson, 1958; Sjaardema et al., 2015) relate target detectability to its angular extent relative to sensor resolution, and photogrammetric mission planning likewise links flight altitude to ground sampling distance and achievable resolution (Cabreira et al., 2019). The prior contains no learned parameters; its two constants, $\rho_{\min}$ and $h_{\min}$, are fixed a priori and examined empirically in Fig. 3 and Appendix C.3. Our contribution is therefore not the observation that larger targets can be resolved from greater altitude, but the use of this classical relation as a lightweight, detection-independent surrogate for the observability term in Eq. 1, which converts geometric coverage into a goal-conditioned measure of exploration utility.

Two observability regimes. The distribution of $h^*(g)$ on UAV-ON reveals two operating regimes (Appendix C.3): small and medium goals are predominantly *floor-constrained*, so further ascent increases the sensing footprint while reducing target resolvability, whereas large goals lie mostly in an *adaptive regime* in which $h^*(g)$ varies with the specified goal and permits trading some image resolution for greater spatial coverage. The prior remains active in both regimes: in the adaptive regime $h^*(g)$ provides a goal-dependent operating reference, while in the floor regime the band and gain defined around h^* discourage excessive climbing—the primary degeneracy induced by geometric coverage.

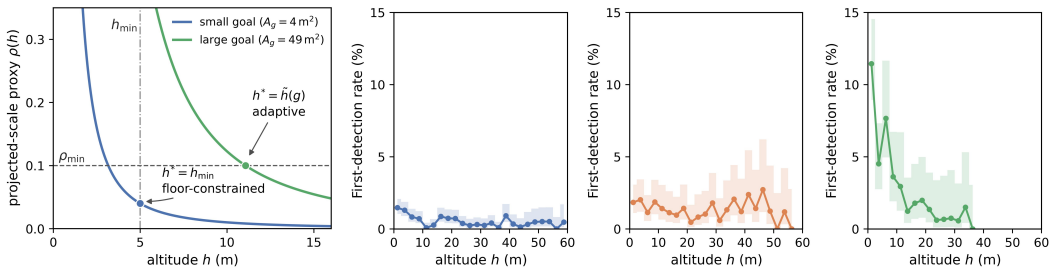
4.3 GOAL-CONDITIONED EFFECTIVE COVERAGE

Equation 1 suggests a natural goal-conditioned measure of exploration progress. Let $\Delta C_t \geq 0$ denote the newly covered ground area at step t, computed from the downward footprint. Ideal progress should weight each increment by the observability utility of the goal under the acquisition geometry,

$$\Delta C_t^{\text{useful}} \;=\; \Delta C_t \cdot q(h_t, g), \tag{7}$$

the quantity that geometric coverage ignores. Since $q(h, g)$ cannot be evaluated online before the goal is detected, we instantiate it with the goal-conditioned observability prior of §4.2: each increment is weighted by an *observability gain* $\phi(h; g)$ centered on the reference altitude $h^*(g)$,

$$\phi(h; g) \;=\; \begin{cases} \mathrm{clip}\big(2.5 - 2.5\,\frac{h-h^*}{\delta},\; 0,\; 2.5\big), & h > h^*, \\ \max\big(2.5 - 1.5\,\frac{h^*-h}{\delta},\; 1\big), & h \leq h^*, \end{cases} \tag{8}$$



(a) Reference-altitude regimes **(b) Small goals** **(c) Medium goals** **(d) Large goals**

Figure 3: Goal-conditioned observability prior and detector response. **(a)** Representative goals illustrate the floor-constrained and adaptive reference-altitude regimes induced by the resolvability constraint. **(b–d)** First-detection rate versus altitude for small, medium, and large goals, computed over pre-detection frames within 60 m of the goal (Wilson 95% CIs), showing altitude- and size-dependent detector response.

which yields the exploration reward $r_t^{\text{cov}} = \lambda\,\phi(h_t; g)\,\Delta C_t$, a tractable surrogate for Eq. 7 under the spatial-search approximation of §4.1. The gain ϕ is a utility-shaping function, not a probability or a direct estimate of q: it is unnormalized (bounded by 2.5) and preserves the qualitative preference that q expresses.

The gain is asymmetric around $h^*(g)$, and the asymmetry is deliberate. Above the reference altitude the weight decays with slope $2.5/\delta$ and reaches zero at the upper edge of the observability band, $h^* + \delta$, so observations acquired above the band receive no effective-coverage credit—suppressing the degenerate direction of collecting coverage reward by climbing to enlarge the footprint; the zero weight is a utility-shaping choice rather than a hard detectability boundary, as the measured detector response above the band is small but nonzero (Fig. 3(b–d)). Below the reference altitude the projected target scale does not deteriorate under our geometric model and the loss is confined to a smaller per-step footprint, so the weight decays more gently (slope $1.5/\delta$) and never falls below one: undershooting discounts the efficiency of exploration rather than excluding it from progress.

The underlying coverage map remains purely geometric. Goal conditioning does not modify the map and introduces no region-level observability annotation; it takes effect only when increments ΔC_t are converted into reward. By construction, this weighting removes the incentive to increase sensing footprint at the expense of goal resolvability.

4.4 Learning observability-aware exploration

Injection points. The prior enters policy learning through two main channels and is available before target detection. (i) *Observation*: the altitude offset $h_t - h^*(g)$ is provided as a policy input. (ii) *Reward*: the gain in Eq. 8 reweights geometric coverage, while an altitude-shaping term discourages operation above the observability band.

Reward composition. During policy training, the step reward is grouped as

$$r_t = \underbrace{\lambda\,\phi(h_t; g)\,\Delta C_t}_{\text{effective coverage}} + r_t^{\text{alt}} + r_t^{\text{spat}} + r_t^{\text{aux}}, \tag{9}$$

with the full decomposition, weights, and gating conditions given in Appendix B.2. The effective-coverage reward and altitude shaping r_t^{alt} form the primary observability-aware terms—the latter inactive within and below the operating band and superlinear in the excess altitude above $h^* + \delta$. The spatial term r_t^{spat} follows the staged schedule below, whereas r_t^{aux} collects standard task, safety, detection, and approach shaping.

Staged optimization. Joint optimization from scratch yielded substantially poorer navigation performance under our training configuration (§5.3): spatial exploration favors rapid expansion of observed area, collision avoidance favors clearance from nearby structures, and the observability-aware objective encourages operation near a goal-conditioned altitude regime. We therefore optimize

Table 1: Overall and size-stratified performance on the UAV-ON test set. Baseline results are obtained from previously published studies under their reported evaluation protocols; Random, CLIP-H, and AOA-F are the benchmark baselines reported with UAV-ON (Xiao et al., 2025). Results are reported in SR, OSR, and SPL (%). Best and second-best results are shown in **bold** and underlined, respectively.

Method	Small			Medium			Large			Total		
	SR↑	OSR↑	SPL↑	SR↑	OSR↑	SPL↑	SR↑	OSR↑	SPL↑	SR↑	OSR↑	SPL↑
Random	4.14	7.80	2.80	3.33	8.10	3.05	2.48	8.07	1.62	3.70	8.00	2.66
CLIP-H	2.86	8.43	1.51	10.95	16.67	7.17	13.04	19.25	10.53	6.20	11.90	4.15
AOA-F	4.45	16.38	1.61	10.48	17.62	6.36	14.29	21.74	10.66	7.30	17.50	4.06
AerialVLN	4.13	19.87	4.01	4.76	17.62	4.90	1.86	19.25	1.86	3.90	19.30	3.72
NaVId	10.02	29.55	4.41	11.43	26.44	5.95	17.39	30.73	15.03	11.50	29.10	6.44
OpenFly	12.40	26.23	6.35	10.00	20.48	4.19	13.04	31.65	7.56	12.00	25.90	6.09
APEX*	–	–	–	–	–	–	–	–	–	13.33	20.00	10.14
OctMem-Agent	**18.91**	29.57	6.03	18.09	26.66	6.27	23.60	31.67	7.80	19.50	29.30	6.37
ECO-Nav (Ours)	17.90	**36.02**	**12.73**	**32.73**	**44.85**	**26.18**	**42.02**	**54.62**	**33.22**	**25.17** (+5.67)	**41.04** (+11.74)	**19.10** (+8.96)

*APEX reports only overall results under seen/unseen splits.

these behaviors in two stages. *Stage 1* emphasizes goal-conditioned altitude behavior by combining altitude-offset observations, effective coverage, and altitude shaping with basic boundary and safety terms. *Stage 2* retains the observability-aware objective while replacing the stage-1 boundary shaping with frontier-based exploration and collision-avoidance shaping; staging changes the spatial objectives coupled with the observability-aware search objective, not the navigation task itself.

Architecture and terminal guidance. The exploration policy is trained with PPO (Schulman et al., 2017). Multi-view RGB-D observations are encoded by a frozen CLIP ViT-B/32 visual backbone and a trainable depth encoder, and a recurrent transformer policy fuses these features with goal-conditioned and navigation-state inputs including the altitude offset $h_t - h^*(g)$ (Appendix A.2), outputting seven non-terminal motion actions. Scripted terminal guidance is disabled during policy training. At evaluation time, once the goal detector (Liu et al., 2024) reports the target with sufficient confidence, control transfers to a scripted terminal-guidance module for final approach and stopping, held fixed across the reported internal comparisons (Appendix A.3).

5 Experiments

5.1 Experimental Setup

We evaluate on the UAV-ON test set (Xiao et al., 2025) following the task definition in §3. We report SR, OSR, and SPL, both overall and stratified by goal size (small/medium/large), using a 20 m success threshold and a 150-step episode budget; episodes terminate upon collision. During training on the UAV-ON training split, privileged quantities such as ground-truth goal distance and agent position are used only for reward computation, including approach/progress shaping and search-domain boundary shaping (Appendix B.2); at test time, the learned policy receives no ground-truth pose, goal location, or goal distance as input (Appendix A.1).

5.2 Main Results

Table 1 compares our method, ECO-Nav (*Effective-Coverage Observability Navigation*), with established baselines on UAV-ON. Our method achieves the best overall performance across all three navigation metrics, reaching 25.17% SR, 41.04% OSR, and 19.10% SPL. Compared with the previous state of the art, this represents improvements of 5.67, 11.74, and 8.96 percentage points in SR, OSR, and SPL, respectively, with SPL increasing from 10.14% to 19.10%. The improvement is particularly pronounced for large goals, where our method reaches 42.02% SR, exceeding the strongest baseline by 18.42 points. This result is consistent with our central motivation: once the goal-conditioned reference altitude moves beyond the reference-altitude floor, the policy can adapt its sensing geometry to the target scale and obtain substantially more effective search behavior. For

Table 2: Component-wise ablations on the UAV-ON test set. We report overall SR/OSR/SPL (%) and mean search-phase flight altitude (m) for small, medium, and large goals.

Variant	Navigation Performance			Mean Flight Altitude (m)		
	SR↑	OSR↑	SPL↑	Small	Medium	Large
ECO-Nav (full)	**25.17**	**41.04**	**19.10**	11.70	13.69	8.40
w/o observability shaping	13.27 (-11.90)	29.69 (-11.35)	9.18 (-9.92)	30.33	31.43	26.11
Fixed anchor ($h^* = 15$ m)	21.75 (-3.42)	34.06 (-6.98)	17.00 (-2.10)	16.36	17.85	13.00
Joint optimization	13.55 (-11.62)	30.03 (-11.01)	8.89 (-10.21)	23.92	22.36	19.21

small goals, where the reference altitude is largely constrained by the same floor, the improvement is correspondingly smaller.

5.3 Ablation and Behavioral Analysis

Without observability-aware weighting, geometric coverage favors altitude-driven footprint expansion at the expense of target resolvability (§4.1). Table 2 therefore reports each ablation's navigation performance together with its mean search altitude, allowing us to examine both task-level effects and the corresponding search behavior.

Observability-aware shaping. Removing observability-aware reward shaping reduces the coverage objective to geometric coverage and disables the associated altitude-conditioned shaping while retaining the goal-conditioned policy inputs (Appendix B.2). This ablation reduces SR from 25.17% to 13.27% and SPL from 19.10% to 9.18%, while substantially increasing mean search altitude from $11.70/13.69/8.40$ m to $30.33/31.43/26.11$ m for small/medium/large goals. The pronounced climbing behavior provides direct behavioral evidence for the failure mode predicted by the coverage–observability mismatch: without observability-aware weighting, geometric coverage favors footprint expansion at the expense of target resolvability. Additional altitude statistics and a qualitative trajectory comparison are provided in Appendix C.2 and Appendix D, respectively.

Goal-conditioned reference. Does the sensing reference need to depend on the goal? Detector response varies systematically with both altitude and goal size (Fig. 3(b–d)), and under the two-regime prior small and medium goals are predominantly floor-constrained whereas larger goals admit adaptive references (§4.2; Appendix C.3). Replacing the goal-conditioned anchor with a fixed 15 m reference costs 3.42 SR and 6.98 OSR, with mean search altitude rising across all three strata as the variant tracks its fixed reference. The fixed-anchor ablation thus supports adapting the sensing reference to the specified goal.

Staged optimization. Is staged optimization necessary? Optimizing all components jointly from scratch drops SR to 13.55% and SPL to 8.89%—nearly as low as removing the observability shaping—and also produces substantially higher search altitudes than the full method: directly coupling altitude control, spatial exploration, and collision avoidance is difficult under our training configuration (§4.4). Training dynamics for both schedules are shown in Appendix C.1. This result indicates that the staged schedule is important for optimizing the full objective. Together, the three ablations show complementary contributions from observability-aware shaping, goal-conditioned sensing, and staged optimization.

6 Conclusion

This work addresses a fundamental limitation of geometric exploration in aerial object goal navigation: spatial coverage alone does not reflect whether the specified goal is observable under the acquisition geometry. We introduce *goal-conditioned effective coverage*, combining a closed-form altitude prior with observability-aware staged learning. On UAV-ON, our method achieves 25.17% SR, 41.04% OSR, and 19.10% SPL, exceeding the strongest previously published results across all three metrics. Limitations include simulation-only evaluation, constrained altitude adaptivity for visually demanding goals, scripted terminal guidance, and a simplified geometric sensing model.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing. All content was reviewed and verified by the authors.

ETHICS STATEMENT

This work studies autonomous aerial object goal navigation in simulation and does not involve human subjects or private data. Potential deployment risks include safety and regulatory compliance.

REPRODUCIBILITY STATEMENT

The observability prior is closed-form and uses fixed parameters throughout all experiments. Training configurations, evaluation protocols, and hyperparameters are provided in the appendix. Source code and configurations will be released.

REFERENCES

Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, and Amir R. Zamir. On evaluation of embodied navigation agents. *arXiv preprint arXiv:1807.06757*, 2018a.

Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian Reid, Stephen Gould, and Anton van den Hengel. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3674–3683, 2018b.

Ruzena Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988.

Dhruv Batra, Aaron Gokaslan, Aniruddha Kembhavi, Oleksandr Maksymets, Roozbeh Mottaghi, Manolis Savva, Alexander Toshev, and Erik Wijmans. Objectnav revisited: On evaluation of embodied agents navigating to objects. *arXiv preprint arXiv:2006.13171*, 2020.

Marc G. Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Rémi Munos. Unifying count-based exploration and intrinsic motivation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2009.

Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. Exploration by random network distillation. In *International Conference on Learning Representations (ICLR)*, 2019.

Tau M. Cabreira, Lisandro B. Brisolara, and Paulo R. Ferreira. Survey on coverage path planning with unmanned aerial vehicles. *Drones*, 3(1):4, 2019.

Devendra Singh Chaplot, Dhiraj Gandhi, Saurabh Gupta, Abhinav Gupta, and Ruslan Salakhutdinov. Learning to explore using active neural SLAM. In *International Conference on Learning Representations (ICLR)*, 2020a.

Devendra Singh Chaplot, Dhiraj Prakashchand Gandhi, Abhinav Gupta, and Ruslan Salakhutdinov. Object goal navigation using goal-oriented semantic exploration. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020b.

Howard Chen, Alane Suhr, Dipendra Misra, Noah Snavely, and Yoav Artzi. Touchdown: Natural language navigation and spatial reasoning in visual street environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Shizhe Chen, Pierre-Louis Guhur, Cordelia Schmid, and Ivan Laptev. History aware multimodal transformer for vision-and-language navigation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Shizhe Chen, Pierre-Louis Guhur, Makarand Tapaswi, Cordelia Schmid, and Ivan Laptev. Think global, act local: Dual-scale graph transformer for vision-and-language navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Abhishek Das, Samyak Datta, Georgia Gkioxari, Stefan Lee, Devi Parikh, and Dhruv Batra. Embodied question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Samir Yitzhak Gadre, Mitchell Wortsman, Gabriel Ilharco, Ludwig Schmidt, and Shuran Song. CoWs on pasture: Baselines and benchmarks for language-driven zero-shot object navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Enric Galceran and Marc Carreras. A survey on coverage path planning for robotics. *Robotics and Autonomous Systems*, 61(12):1258–1276, 2013.

Yunpeng Gao, Chenhui Li, Zhongrui You, Junli Liu, Zhen Li, Pengan Chen, Qizhi Chen, Zhonghan Tang, Liansheng Wang, Penghui Yang, Yiwen Tang, Yuhang Tang, Shuai Liang, Songyi Zhu, Ziqin Xiong, Yifei Su, Xinyi Ye, Jianan Li, Yan Ding, Dong Wang, Xuelong Li, Zhigang Wang, and Bin Zhao. OpenFly: A comprehensive platform for aerial vision-language navigation. In *International Conference on Learning Representations (ICLR)*, 2026.

Yicong Hong, Zun Wang, Qi Wu, and Stephen Gould. Bridging the gap between learning in discrete and continuous environments for vision-and-language navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

John Johnson. Analysis of image forming systems. In *Proceedings of the Image Intensifier Symposium*, pp. 244–273, 1958.

Elia Kaufmann, Leonard Bauersfeld, Antonio Loquercio, Matthias Müller, Vladlen Koltun, and Davide Scaramuzza. Champion-level drone racing using deep reinforcement learning. *Nature*, 620:982–987, 2023.

Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Matt Deitke, Kiana Ehsani, Daniel Gordon, Yuke Zhu, Aniruddha Kembhavi, Abhinav Gupta, and Ali Farhadi. AI2-THOR: An interactive 3d environment for visual AI. *arXiv preprint arXiv:1712.05474*, 2017.

Alexander Ku, Peter Anderson, Roma Patel, Eugene Ie, and Jason Baldridge. Room-across-room: Multilingual vision-and-language navigation with dense spatiotemporal grounding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

Jungdae Lee, Taiki Miyanishi, Shuhei Kurita, Koya Sakamoto, Daichi Azuma, Yutaka Matsuo, and Nakamasa Inoue. CityNav: A large-scale dataset for real-world aerial navigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. In *European Conference on Computer Vision (ECCV)*, 2024.

Shubo Liu, Hongsheng Zhang, Yuankai Qi, Peng Wang, Yanning Zhang, and Qi Wu. AerialVLN: Vision-and-language navigation for UAVs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15338–15348, 2023.

Antonio Loquercio, Elia Kaufmann, René Ranftl, Matthias Müller, Vladlen Koltun, and Davide Scaramuzza. Learning high-speed flight in the wild. *Science Robotics*, 6(59):eabg5810, 2021.

Arjun Majumdar, Gunjan Aggarwal, Bhavika Devnani, Judy Hoffman, and Dhruv Batra. ZSON: Zero-shot object-goal navigation using multimodal goal embeddings. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Roman Marchant and Fabio Ramos. Bayesian optimisation for informative continuous path planning. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E. Taylor, and Peter Stone. Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 21(181):1–50, 2020.

Andrew Y. Ng, Daishi Harada, and Stuart Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *Proceedings of the International Conference on Machine Learning (ICML)*, 1999.

Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 2778–2787, 2017.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 8748–8763, 2021.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

Santhosh Kumar Ramakrishnan, Devendra Singh Chaplot, Ziad Al-Halah, Jitendra Malik, and Kristen Grauman. PONI: Potential functions for objectgoal navigation with interaction-free learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18890–18900, 2022.

Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A platform for embodied AI research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9339–9347, 2019.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. AirSim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics*, 2018.

Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettlemoyer, and Dieter Fox. ALFRED: A benchmark for interpreting grounded instructions for everyday tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10740–10749, 2020.

Tracy A. Sjaardema, Collin S. Smith, and Gabriel C. Birch. History and evolution of the Johnson criteria. Technical report, Sandia National Laboratories, Report SAND2015-6368; OSTI 1222446, 2015.

Andrew Szot, Alexander Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John Turner, Noah Maestre, Mustafa Mukadam, Devendra Singh Chaplot, Oleksandr Maksymets, Aaron Gokaslan, Vladimír Vondruš, Sameer Dharur, Franziska Meier, Wojciech Galuba, Angel Chang, Zsolt Kira, Vladlen Koltun, Jitendra Malik, Manolis Savva, and Dhruv Batra. Habitat 2.0: Training home assistants to rearrange their habitat. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Fei Xia, Amir R. Zamir, Zhiyang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. Gibson Env: Real-world perception for embodied agents. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9068–9079, 2018.

Jianqiang Xiao, Yuexuan Sun, Yixin Shao, Boxi Gan, Rongqiang Liu, Yanjin Wu, Weili Guan, and Xiang Deng. UAV-ON: A benchmark for open-world object goal navigation with aerial agents. In *Proceedings of the 33rd ACM International Conference on Multimedia (ACM MM)*, 2025.

Brian Yamauchi. A frontier-based approach for autonomous exploration. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, pp. 146–151, 1997.

Joel Ye, Dhruv Batra, Abhishek Das, and Erik Wijmans. Auxiliary tasks and exploration enable ObjectGoal navigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16117–16126, 2021.

Naoki Yokoyama, Sehoon Ha, Dhruv Batra, Jiuguang Wang, and Bernadette Bucher. VLFM: Vision-language frontier maps for zero-shot semantic navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

Albert J. Zhai and Shenlong Wang. PEANUT: Predicting and navigating to unseen targets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10926–10935, 2023.

Daoxuan Zhang, Ping Chen, Xiaobo Xia, Xiu Su, Ruichen Zhen, Jianqiang Xiao, and Shuo Yang. APEX: A decoupled memory-based explorer for asynchronous aerial object goal navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Jiazhao Zhang, Kunyu Wang, Rongtao Xu, Gengze Zhou, Yicong Hong, Xiaomeng Fang, Qi Wu, Zhizheng Zhang, and He Wang. NaVId: Video-based VLM plans the next step for vision-and-language navigation. In *Robotics: Science and Systems (RSS)*, 2024.

Jiacong Zhou, Jiaxu Miao, Yourun Lin, Xianyun Wang, Jun Xiao, and Jun Yu. Memory-augmented scene understanding and exploration for open-world aerial object-goal navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21616–21626, 2026.

Kaiwen Zhou, Kaizhi Zheng, Connor Pryor, Yilin Shen, Hongxia Jin, Lise Getoor, and Xin Eric Wang. ESC: Exploration with soft commonsense constraints for zero-shot object navigation. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2023.

Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, J. Jitse Lim, Abhinav Gupta, Li Fei-Fei, and Ali Farhadi. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3357–3364, 2017.

A Implementation Details

A.1 Observation and Action Spaces

The platform is equipped with four synchronized RGB-D cameras covering the front, left, right, and down views, each with a 90° field of view. For each view, RGB observations are encoded by a frozen CLIP ViT-B/32 visual backbone, while depth observations are processed by a lightweight trainable convolutional neural network (CNN). The goal description is encoded by a T5-small text encoder (Raffel et al., 2020) followed by a trainable adapter. These features are fused with a detector-derived bounding-box token and state tokens in the recurrent transformer policy described in §A.2. The state tokens include the altitude offset $h - h^*(g)$, goal footprint area A_g, collision state, elapsed step, previous action, and an estimated pose $[x, y, \text{heading}]$ obtained from dead reckoning over executed motion primitives rather than ground-truth localization. The policy receives no ground-truth agent pose, goal position, or goal-relative distance as input; privileged geometric information is used only by the environment for reward computation, success and OSR evaluation, and boundary checks. The same observation configuration is used during training and evaluation.

The environment defines eight discrete actions. The learned policy selects among seven non-terminal motion primitives: forward and lateral translations of $5\,\text{m}$, ascent and descent of $2\,\text{m}$, and yaw rotations of $\pm 15^\circ$. The remaining stop action is not part of policy training and is issued by the scripted terminal-guidance module during evaluation, as described in §A.3. Each episode is limited to 150 steps and terminates early upon collision or when the agent leaves the $70\,\text{m}$ search domain, which contains all admissible goal locations under the $50\,\text{m}$ goal-placement radius of §3.

A.2 Policy Architecture and Training

The CLIP ViT-B/32 visual backbone is frozen during training, while the T5-small text encoder, depth encoder, text adapter, bounding-box token encoder, state-token encoders, recurrent transformer policy, and actor and critic networks are optimized jointly. The actor and critic do not share parameters: each comprises its own encoders and recurrent decoder, with the actor producing logits over the seven non-terminal motion primitives and the critic a scalar state value. Training uses PPO (Schulman et al., 2017) with Adam and the hyperparameters summarized in Table 3, together with an auxiliary height-prediction head on the fused representation supervised with the altitude offset $h - h^*(g)$ under adaptive loss weighting initialized at 0.05. Following §4.4, stage 1 emphasizes goal-conditioned altitude behavior through the observability prior, effective-coverage reward, altitude shaping, and basic safety-related shaping, whereas stage 2 retains the observability-aware objective while introducing frontier-based spatial exploration and collision-avoidance shaping in place of the stage-1 boundary-focused shaping. This staged design separates altitude regulation from the subsequent joint optimization of spatial exploration and obstacle avoidance.

A.3 Detection and Terminal Guidance

The learned policy does not directly output `stop`. Scripted terminal guidance is disabled during policy training and is enabled only at evaluation time. Once the target is confidently grounded during evaluation, control is transferred from the learned exploration policy to a scripted terminal-guidance module for final approach and stopping, held fixed across the reported internal comparisons.

Detection. GroundingDINO with a Swin-T OGC checkpoint is applied to all four views at every step using the goal name as the text prompt. Candidate detections are filtered using a confidence threshold of 0.35 and an image-area-ratio criterion, and further verified using CLIP similarity between the detected crop and the goal text. During training, simulator ground-truth segmentation is used to validate detector outputs when constructing reward-related training signals. No ground-truth segmentation is used during evaluation, where target detection relies on the onboard observations and detector outputs. Detection confidence is accumulated across frames through Bayesian fusion.

Handover. During evaluation, when the fused confidence exceeds the grounding criterion, control switches from the learned exploration policy to terminal guidance. The detection condition is re-evaluated at every step, allowing control to return to the learned policy if reliable target evidence is lost.

Approach and stop. Terminal guidance estimates the relative target bearing from the detected bounding-box center and target distance from the corresponding depth region. It then aligns with the target, advances toward it, and adjusts altitude during the final approach. The module issues `stop` after at least one valid detection step when the estimated distance plus half the goal's diagonal extent is below $15\,\mathrm{m}$ and the fused detection probability exceeds 0.8. Success is evaluated using the benchmark criterion of stopping within $20\,\mathrm{m}$ of the goal. The module's internal altitude-descent tolerance uses a more permissive resolvability constant than $\rho_{\min}$; it serves only as a descent guard during approach and does not alter the reported anchor $h^*(g)$.

A.4 Evaluation Protocol and Hyperparameters

Table 3: Prior constants and core optimization hyperparameters of the full method. Reward weights and gating conditions are summarized in Table 4.

Resolvability threshold $\rho_{\min}$ / reference-altitude floor $h_{\min}$	0.1 / $5\,\mathrm{m}$
Band half-width δ	$\max(h^*/2,\ 5\,\mathrm{m})$
Effective-coverage weight λ	250
Normalized coverage-increment clip	$\Delta C_t \leq 0.05$
PPO: learning rate / γ / GAE λ / clip	$2{\times}10^{-5}$ / 0.99 / 0.95 / 0.2

Evaluation protocol. All navigation results are evaluated on the UAV-ON test set using the success threshold and 150-step budget defined in §3. A single final checkpoint is evaluated for each variant, and each variant is obtained from one training run; we therefore do not report seed-level variance. For altitude analysis, we define the search phase as frames for which the ground-truth distance to the goal exceeds 20 m. This criterion is used only for post-hoc behavioral analysis and provides a consistent proxy for the portion of the trajectory before entering the benchmark success region. We first compute the mean AGL altitude over these frames within each episode and then average the episode-level means within each goal-size stratum. Episodes containing no such frames are omitted from altitude statistics only and remain included in all navigation metrics.

B Effective-Coverage Implementation

B.1 Geometric Coverage Computation

The coverage map is a quadtree spanning the bounding square of the search region, while coverage is normalized over the valid 70 m-radius search domain, with cells outside the domain excluded (10 m leaf cells marked as *explored* according to their covered-area fraction). At every step, the visible ground regions of the active depth views are projected onto the ground plane and unioned into the map: the downward camera contributes the nadir footprint, a ground rectangle of size $2h\tan(\varphi_x/2) \times 2h\tan(\varphi_y/2)$ rotated by the current yaw, and each horizontal view (front, left, right) contributes a fan-shaped region out to its effective sensing distance. The coverage increment ΔC_t is the resulting increase in the explored fraction of the search domain, clipped to $[0, 0.05]$ per step with negative increments zeroed. Updates accumulate under union semantics, so re-observing already-explored ground yields zero increment—no ground area is counted twice. The map is purely geometric and goal-agnostic: goal conditioning enters exclusively through the reward weighting ϕ and the altitude-offset observation, never through the map. As in the anchor derivation, the ground projections assume locally planar terrain.

B.2 Effective-Coverage Reward and Ablation Definition

The exploration reward is $r_t^{\text{cov}} = \lambda\,\phi(h_t; g)\,\Delta C_t$ with $\lambda = 250$ and ϕ the asymmetric gain of Eq. 8. Two further reward-side terms are tied to the same prior: an altitude-shaping penalty (weight 0.25, tolerance δ) that penalizes flying above the band, and a height gate on the one-time detection bonus (granted only below $2h^*(g)$).

w/o observability shaping. The "w/o observability shaping" variant of Table 2 is an observability-aware *reward-side* ablation rather than a ϕ-only one: it disables the observability-conditioned reward logic—the gain weighting (reducing the term to purely geometric coverage, $\phi \equiv 1$), the altitude-shaping penalty, and the height-conditioned gate on the detection bonus—while the goal-conditioned policy inputs, including the altitude-offset observation $h_t - h^*(g)$, remain available to the policy.

Table 4 lists the reward terms of Eq. 9 used to train the learned exploration policy, together with their weights, stage activation, and gating conditions; the three groups follow the composition of §4.4. Scripted terminal guidance is enabled only during evaluation and therefore contributes no learning rewards; the terminal success and failed-stop rewards are likewise unreachable during policy training, since the policy does not emit `stop`, leaving the horizon penalty as the only step-budget termination term.

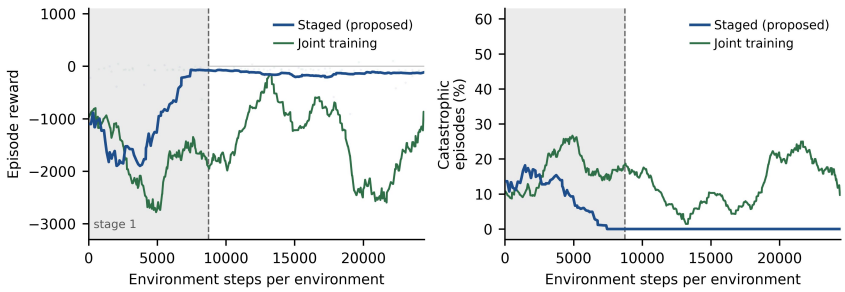
C Additional Behavioral Analysis

C.1 Training Dynamics of Staged Optimization

Figure 4 contrasts the training dynamics of the staged curriculum with joint optimization under the same reward. The staged curriculum allocates the first 10k environment steps to altitude control, followed by 20k steps of exploration and avoidance; the joint variant trains all objectives concurrently over 30k steps. Both curves are plotted over the same window of $24{,}449$ environment steps per environment.

Table 4: Full step-reward decomposition used to train the learned exploration policy; scripted terminal guidance is evaluation-only and does not contribute learning rewards.

Term	Weight	Form / gating
Prior-carrying terms (both stages)		
Effective coverage	250	$\lambda\,\phi(h;g)\,\Delta C$, normalized $\Delta C \le 0.05$ per step
Altitude shaping	0.25	0 within and below the band $[h^*-\delta,\ h^*+\delta]$; superlinear in the excess above $h^*+\delta$
Stage-scheduled spatial terms		
Search-boundary penalty	0.5 (stage 1)	linear ramp within 15 m of the 70 m search-disk edge (-0.5 past it); off in stage 2
Frontier-edge reward	1.6 (stage 2)	proximity to unexplored map frontier, clipped to $[-1,\ 0.2]$; positive only for translation primitives ($\times 1.1$ forward); off in stage 1
Collision-avoidance field	cap 0.7 (stage 2)	direction-wise repulsion from imminent obstacles, each direction capped at -0.7
Auxiliary shaping and safety (both stages, all variants)		
Detection bonus	2.0	one-time at the first verified detection, only below $2h^*(g)$
GT-distance approach	0.2	$0.2\ \min(1/d,\ 0.2)$ on steps that approach the goal by > 0.5 m
Progress	0.2	improvement of the closest-so-far distance when > 0.5 m
Step penalty	-0.0005	per step
Stagnation	$-0.005\,2^{n-1}$	per event; movement $<$ 1 m over 6 steps, capped at -2.56; extra -0.008 for ascent/descent
Collision	-75	episode terminates
Out-of-bounds	-100	episode terminates
Horizon	-0.01	terminal, on reaching the step budget



(a) Episode reward **(b) Catastrophic episodes**

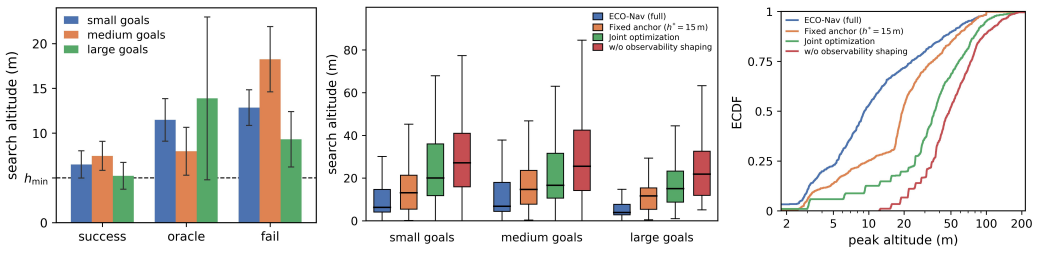
Figure 4: Training dynamics of staged curriculum learning (blue) and joint optimization (green) over the same 24,449 environment steps per environment. Lines show 3k-step rolling means. **Left:** per-episode reward. **Right:** catastrophic-episode rate (reward ≤ -2000), indicating altitude blow-through and timeout. The shaded region denotes the altitude-control stage. Staged training suppresses catastrophic failures by $\approx$ 8k steps, while joint optimization remains unstable.

Failure-mode elimination. The dominant early failure mode is ceiling hovering: the agent ascends to the $\sim$100 m ceiling and remains there until timeout, accumulating a total reward below -2000 while normal episodes center near -70. In the staged run, this failure mode disappears within the altitude-control stage, and the exploration stage then holds a stable reward plateau (rolling mean between -71 and -210) with a zero catastrophic-episode rate from $\approx$8k steps onward. Joint training never eliminates the failure mode: over the same window, its catastrophic-episode rate fluctuates between 1% and 27% and its rolling reward continues to oscillate between -165 and -2780.

Relation to the ablation. These dynamics underlie the joint-optimization ablation of Table 2: without a dedicated altitude-control stage, ceiling-hover failures persist throughout training, and the resulting policy exhibits the elevated search altitudes (23.92/22.36/19.21 m vs. 11.70/13.69/8.40 m for small/medium/large goals) and the reduced navigation performance (13.55% vs. 25.17% SR) reported there.

C.2 Why Large Goals Fly Lower on Average

Figure 5 provides additional analysis of the search-altitude behavior reported in Table 2.



(a) Search altitude by episode outcome **(b) Search altitude by variant and goal size** **(c) Distribution of per-episode peak altitude**

Figure 5: Altitude behavior of the full method and its ablations on the UAV-ON test set under the search-segment convention. **(a)** Mean search altitude by episode outcome and goal-size stratum for the full method; error bars indicate 95% confidence intervals across episodes, and the dashed line marks the reference-altitude floor $h_{\min}$. **(b)** Search-altitude distributions across the four variants and three goal-size strata; boxes show the interquartile range with the median indicated by the center line. **(c)** Empirical cumulative distribution of per-episode peak altitude. Curves shifted to the right indicate a stronger tendency to climb, with the observability-shaping ablation exhibiting the highest peak-altitude distribution.

Outcome-dependent altitude. Figure 5(a) shows that successful episodes operate at substantially lower search altitudes than oracle or failed episodes across all goal-size strata. This effect is particularly pronounced for large goals, whose successful episodes average 5.23 m during search, compared with 9.32 m for failed episodes. Large-goal episodes also enter the 20 m goal region more frequently than small and medium goals (53.8% vs. 33.1% and 43.6%, respectively). Consequently, the larger proportion of low-altitude successful episodes pulls the aggregate large-goal search altitude downward, explaining why its mean altitude is lower despite the more adaptive reference-altitude regime. The corresponding whole-episode statistics remain close to the search-segment statistics, indicating that this effect is not explained solely by the final descent.

Behavior across variants. Figure 5(b) shows a consistent upward shift in search-altitude distributions when the proposed components are removed. The full method maintains the lowest altitude distribution across goal-size strata, the fixed-anchor and joint-optimization variants operate progressively higher, and the observability-shaping ablation exhibits the strongest upward shift. Figure 5(c) further shows the same effect at the episode level: the peak-altitude distribution of the observability-shaping ablation is substantially shifted toward higher altitudes, directly exposing the excessive-climbing failure mode that observability-aware weighting is designed to suppress.

Illustrative large-goal episodes. Figure 6 shows two successful large-goal episodes. In both examples, the target provides sufficiently strong visual evidence within the first few search steps, allowing the agent to transition to detection and terminal approach before a prolonged climbing phase develops. These examples provide an intuitive illustration of the outcome-dependent behavior discussed above: early detection can shorten the search phase and thereby contribute to the lower average search altitude observed for successful large-goal episodes. They are intended as qualitative examples rather than evidence that all large-goal episodes follow the same trajectory.

C.3 Detector-Response Analysis and Prior Regimes

Detector-response analysis. The per-frame first-detection rates in Fig. 3(b–d) are computed from pre-detection frames within 60 m of the goal, using 2.5 m altitude bins with at least 100 at-risk frames per bin and Wilson 95% confidence intervals. Across the three goal-size strata, the resulting curves show that detector response varies with both altitude and target size, providing empirical support for the altitude- and size-dependent observability assumption underlying the proposed prior.



Figure 6: Qualitative examples of successful large-goal episodes. Each row shows task initialization, first detection, and successful stopping, with the UAV altitude reported above each observation. Early detection before prolonged climbing helps explain the relatively low search altitudes of successful large-goal episodes.

Prior regimes. We fix $\rho_{\min} = 0.1$ and $h_{\min} = 5\,\mathrm{m}$ throughout all experiments. Under these constants, smaller goals predominantly fall into the floor-constrained regime, whereas larger goals more often admit adaptive reference altitudes determined by their annotated footprint. The prior thus acts primarily as a lower-bounded sensing reference for small and medium goals and becomes more strongly goal-adaptive for large goals. We do not tune these constants per goal or scene.

D Qualitative Comparison

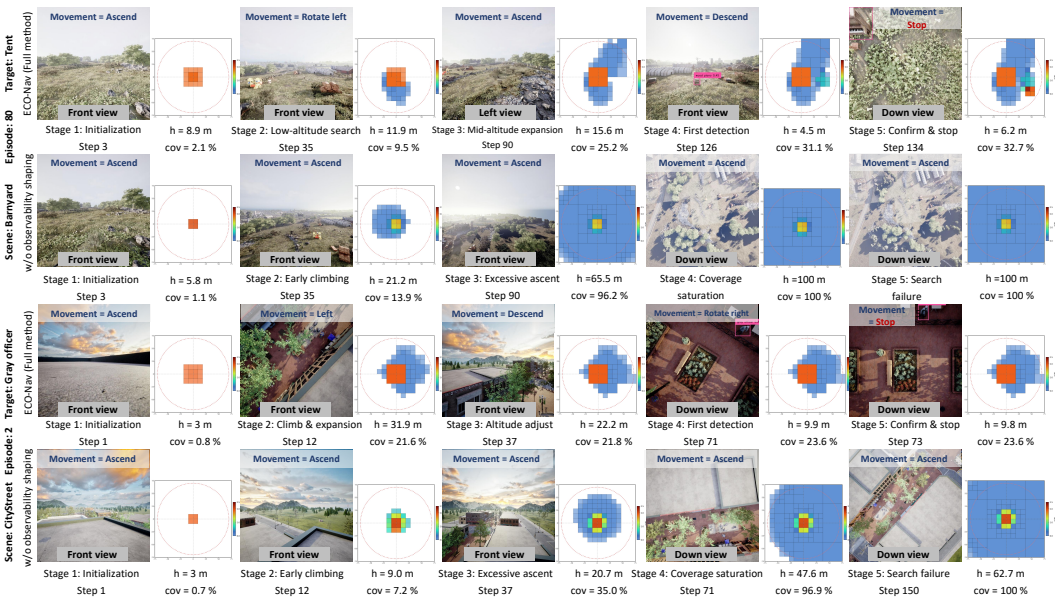


Figure 7: Qualitative comparison between the full method and the variant without observability shaping on matched UAV-ON episodes. Each row shows five representative search stages with the corresponding observation, coverage map, flight altitude, and geometric coverage. The ablation exhibits substantially stronger climbing despite high accumulated coverage.

Figure 7 further illustrates the excessive-climbing behavior discussed in Section C.2. Across both matched examples, removing observability-aware shaping causes the policy to favor sustained ascent as an efficient way to expose new ground area. This behavior rapidly increases geometric coverage, but does not necessarily improve the visual evidence available for identifying the specified goal. In the two ablation episodes, the UAV climbs to approximately $100\,\mathrm{m}$ and $63\,\mathrm{m}$, respectively, while geometric coverage approaches saturation, yet both searches terminate without successful target localization.

The full method exhibits a qualitatively different search pattern. Rather than monotonically increasing altitude, it allows moderate ascent to expand spatial coverage and subsequently adjusts altitude as the search progresses. In the first example, the UAV remains below approximately $16\,\mathrm{m}$ before detecting and approaching the target; in the second, it temporarily climbs to broaden the search area and then descends before the first detection and final stop. Thus, the benefit of observability-aware shaping is not simply to enforce low-altitude flight, but to discourage coverage-seeking ascent when the resulting observations provide limited goal-specific evidence.

These examples expose the coverage–observability mismatch underlying our formulation. A policy driven primarily by geometric coverage can appear highly exploratory—even reaching nearly complete spatial coverage—while still failing to acquire sufficiently informative observations of the target. Goal-conditioned effective coverage instead encourages the agent to trade spatial exposure against target resolvability, producing search trajectories that are more useful for goal-directed recognition rather than coverage accumulation alone.