AIRWORLD: LANGUAGE-CONDITIONED PREDICTIVE WORLD MODELS FOR GENERATIVE UAV EXPLORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study open-vocabulary target search by a single micro aerial vehicle (UAV) instructed in natural language. Semantic exploration systems such as STEM propagate semantic priorities to frontier viewpoints and optimize a prescribed planning objective. We propose AIRWORLD, a language-conditioned predictive world model that maintains a belief over geometry, semantics, and task intention from RGB-D observations and UAV state. An action-conditioned latent model predicts future evidence, while a conditional flow-matching policy generates candidate control sequences. Belief-space rollouts rank candidates by discovery, coverage, expected information gain, and risk; a receding-horizon controller executes a short verified segment. We formulate the belief update, constrained trajectory generation, and training objectives, and report matched single-UAV benchmark results for the proposed method and the listed baselines.

1 INTRODUCTION

An autonomous UAV searching an unfamiliar building must decide where to look next while its map, object inventory, and interpretation of the instruction remain incomplete. Three-dimensional flight provides viewpoints over occlusions, but limited sensor range, collision risk, and flight time make uninformed coverage expensive. Instructions may describe an instance ("find the damaged valve near the generator") rather than only an object category.

STEM (Sethi et al., 2026) combines semantic priority masking, active perception, frontier diffusion, and weighted minimum-latency planning. Its viewpoint utility uses mapped evidence and prescribed semantic relationships. Recent systems also employ semantic reconstruction, persistent memory, or language-guided continuous planning (Nagami et al., 2025; Kim et al., 2025; Chen et al., 2026). These developments motivate a narrower question: does learning the action-dependent distribution of future evidence improve search under a common sensing and execution budget?

We propose a Gaussian latent belief whose mean encodes geometry, semantics, and intention and whose covariance represents model uncertainty. A conditional flow model proposes controls, a world model rolls out their possible consequences, and a safety filter checks the executable segment. Our contributions are (i) a language-conditioned predictive exploration belief; (ii) constrained flow-matching control generation; (iii) a prior-to-posterior information-gain objective; and (iv) a matched single-UAV evaluation with component ablations and robustness diagnostics.

2 RELATED WORK

Aerial exploration and semantic search. Frontier exploration (Yamauchi, 1997) developed into next-best-view planning (Bircher et al., 2016), rapid multirotor exploration (Cieslewski et al., 2017), motion-primitive planning (Dharmadhikari et al., 2020), and systems such as FUEL, FAEL, and FALCON (Zhou et al., 2021; Huang et al., 2023; Zhang et al., 2024). Semantic object-centric exploration and target-search systems include Papatheodorou et al. (2023), Star-Searcher (Luo et al., 2024), and STEM (Sethi et al., 2026). VISTA uses online semantic Gaussian splatting and task-aware exploration; RAVEN combines semantic spatial memory with adaptive search behaviors;

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.

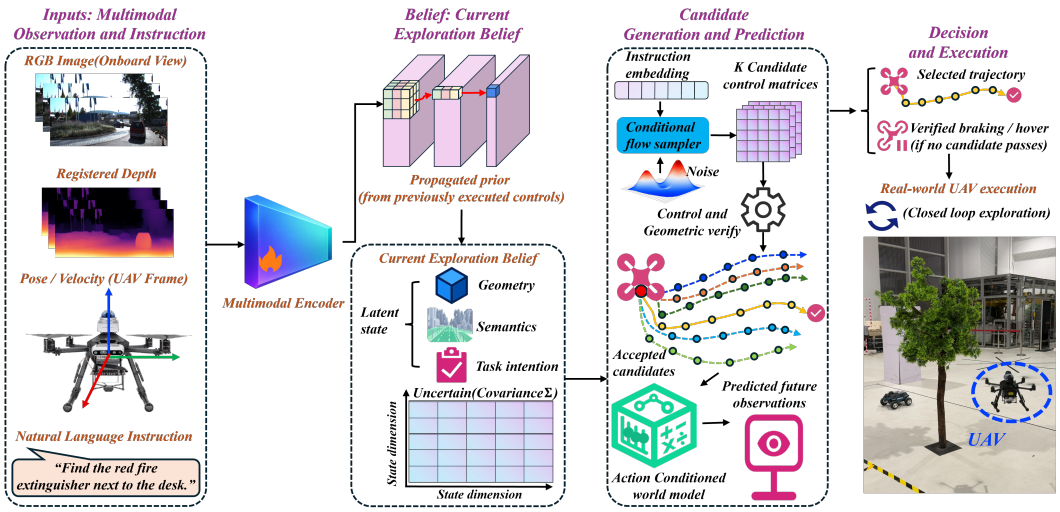


Figure 1: AirWorld closed loop. Multimodal observations update the exploration belief; conditional flow matching generates candidate controls; predictive rollouts score future evidence; and only a verified executable segment is sent to the flight controller.

AirHunt couples language reasoning to continuous aerial planning (Nagami et al., 2025; Kim et al., 2025; Chen et al., 2026). These are relevant recent comparisons, although their original sensors and evaluation domains differ from ours.

Embodied language navigation. ObjectNav is a task and evaluation protocol, not a uniquely specified policy (Batra et al., 2020). Semantic frontiers (Chen et al., 2023), VLFM (Yokoyama et al., 2024), and PONI (Ramakrishnan et al., 2022) provide concrete semantic navigation designs. Ground-navigation policies require an explicit three-dimensional interface and a common goal vocabulary before comparison with aerial search.

World models and generative policies. Latent predictive control (Ha & Schmidhuber, 2018; Hafner et al., 2020; 2023; Schrittwieser et al., 2020) motivates planning with learned dynamics. Diffusion and flow matching model multimodal futures (Ho et al., 2020; Janner et al., 2022; Chi et al., 2023; Lipman et al., 2023). AirWorld combines generative control with geometric constraints and information-seeking planning in a language-conditioned latent belief.

3 Problem Formulation

Notation and observations. Scalars use ordinary italic type, vectors bold lowercase, and matrices bold uppercase, including Greek letters. Sets use calligraphic type; densities and functions use ordinary type. Model parameters are vectors $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\psi}$. Time is indexed by t, future steps by k, and candidates by i. Hats denote predictions; $-$ and $+$ denote the prior and posterior at the *same* time. Appendix C summarizes the core notation.

The workspace is $\mathcal{W} \subset \mathbb{R}^3$. Registered RGB-D is $\boldsymbol{O}_t \in \mathbb{R}^{n_{\text{pix}} \times 4}$, with n_{pix} pixels and red, green, blue, and depth columns. The level-flight planning state is $\boldsymbol{x}_t = [\boldsymbol{p}_t^\top, \chi_t]^\top \in \mathbb{R}^4$, where $\boldsymbol{p}_t \in \mathbb{R}^3$ is position and χ_t is unwrapped yaw. Roll and pitch are stabilized by the low-level controller. The instruction string ℓ has embedding $\boldsymbol{e}_\ell \in \mathbb{R}^{d_\ell}$, where d_ℓ is its dimension. A control $\boldsymbol{a}_t = [\boldsymbol{v}_t^\top, \omega_t]^\top \in \mathbb{R}^4$ contains commanded world-frame velocity $\boldsymbol{v}_t \in \mathbb{R}^3$ and yaw rate ω_t. For interval $\Delta t > 0$, nominal kinematics are $\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \Delta t \boldsymbol{a}_t$. This planning approximation requires tracking margins in runtime checks. The map $\mathcal{M}_t$ contains locally observed space; no prebuilt global map or second UAV is available.

Belief-space objective. The history is $\mathcal{H}_t = (\boldsymbol{O}_{0:t}, \boldsymbol{x}_{0:t}, \boldsymbol{a}_{0:t-1}, \ell)$, and the belief b_t is a density over latent vector $\boldsymbol{z}_t \in \mathbb{R}^d$. A policy ϖ maps the belief and map to controls. For horizon $H \geq 1$,

discount $\gamma \in (0, 1]$, and nonnegative weights $\alpha, \beta, \kappa, \rho, \eta$, we seek

$$\max_{\varpi} \mathbb{E}_{\varpi}\left[\sum_{k=1}^{H} \gamma^{k-1}(\alpha y_{t+k} + \beta c_{t+k} + \kappa g_{t+k} - \rho r_{t+k} - \eta u_{t+k}) \mid \mathcal{H}_t\right]. \tag{1}$$

Here $y_{t+k} \in \{0, 1\}$ indicates the *first* verified discovery, $c_{t+k} \in [0, 1]$ is new observed free volume divided by a fixed reference volume $V_{\text{ref}} > 0$, g_{t+k} is expected information gain in nats, $r_{t+k} \in [0, 1]$ is segment collision probability, and u_{t+k} is posterior uncertainty. Choose V_{ref} at least as large as the maximum observable volume per segment. Contributions are zero after termination; the expectation includes latent transitions and possible observations. The weights absorb different units. Equations 13–14 define the gain and uncertainty.

4 AIRWORLD

4.1 MULTIMODAL LATENT EXPLORATION BELIEF

The encoder $E_{\boldsymbol{\theta}}$ fuses measurement, state, instruction, and predictive prior:

$$\begin{aligned}(\boldsymbol{\mu}_t, \boldsymbol{s}_t) &= E_{\boldsymbol{\theta}}(\boldsymbol{O}_t, \boldsymbol{x}_t, \boldsymbol{e}_\ell, \boldsymbol{\mu}_t^-, \boldsymbol{S}_t^-),\\ b_t(\boldsymbol{z}_t) = q_{\boldsymbol{\theta}}(\boldsymbol{z}_t \mid \mathcal{H}_t) &= \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{S}_t), \qquad \boldsymbol{S}_t = \operatorname{diag}(\operatorname{softplus}(\boldsymbol{s}_t)) + \epsilon_{\text{cov}}\boldsymbol{I}_d.\end{aligned} \tag{2}$$

The mean $\boldsymbol{\mu}_t \in \mathbb{R}^d$ and scale logits $\boldsymbol{s}_t \in \mathbb{R}^d$ are vectors; $\boldsymbol{S}_t \in \mathbb{R}^{d\times d}$ is positive definite. Here $\epsilon_{\text{cov}} > 0$ is a variance floor, $\boldsymbol{I}_d$ is the identity, and diag constructs a diagonal matrix. Softplus acts elementwise as $\log(1 + \exp(s))$ for scalar argument s. Initially $(\boldsymbol{\mu}_0^-, \boldsymbol{S}_0^-) = (\boldsymbol{0}, \boldsymbol{I}_d)$; later priors propagate the preceding belief through executed controls. For actual observations, $\boldsymbol{\mu}_t = \boldsymbol{\mu}_t^+$ and $\boldsymbol{S}_t = \boldsymbol{S}_t^+$.

The partition $\boldsymbol{\mu}_t = [(\boldsymbol{\mu}_t^{\text{geo}})^\top, (\boldsymbol{\mu}_t^{\text{sem}})^\top, (\boldsymbol{\mu}_t^{\text{task}})^\top]^\top$ has dimensions $d_{\text{geo}}, d_{\text{sem}}, d_{\text{task}}$ summing to d and represents geometry, semantics, and intention. Covariance represents uncertainty separately. Stochastic samples are $\tilde{\boldsymbol{z}}_t = \boldsymbol{\mu}_t + \boldsymbol{S}_t^{1/2}\boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$. Throughout, $\boldsymbol{S}^{1/2}$ denotes the symmetric positive-definite covariance square root.

4.2 PREDICTIVE DYNAMICS AND LEARNING

Let $f_{\boldsymbol{\phi}}$ be the transition mean and $\boldsymbol{Q}_{t+k} \in \mathbb{R}^{d\times d}$ its positive-definite process-noise covariance. A fixed feature extractor g_{obs} gives $\boldsymbol{o}_t = g_{\text{obs}}(\boldsymbol{O}_t) \in \mathbb{R}^{d_o}$, where d_o is the feature dimension. Decoder $g_{\boldsymbol{\phi}}$ and positive-definite observation covariance $\boldsymbol{R}_{t+k} \in \mathbb{R}^{d_o\times d_o}$ define

$$\begin{aligned}p_{\boldsymbol{\phi}}(\boldsymbol{z}_{t+k} \mid \boldsymbol{z}_{t+k-1}, \boldsymbol{a}_{t+k-1}, \boldsymbol{e}_\ell) &= \mathcal{N}(f_{\boldsymbol{\phi}}(\boldsymbol{z}_{t+k-1}, \boldsymbol{a}_{t+k-1}, \boldsymbol{e}_\ell), \boldsymbol{Q}_{t+k}),\\ p_{\boldsymbol{\phi}}(\boldsymbol{o}_{t+k} \mid \boldsymbol{z}_{t+k}, \boldsymbol{e}_\ell) &= \mathcal{N}(g_{\boldsymbol{\phi}}(\boldsymbol{z}_{t+k}, \boldsymbol{e}_\ell), \boldsymbol{R}_{t+k}).\end{aligned} \tag{3}$$

Additional heads predict conditional discovery hazard $\hat{\xi}_{t+k}$, new coverage $\hat{c}_{t+k}$, and collision probability $\hat{r}_{t+k}$. The hazard conditions on no previous discovery and the rollout history. Along the no-discovery branch,

$$\hat{y}_{t+k} = \hat{\xi}_{t+k} \prod_{j=1}^{k-1}(1 - \hat{\xi}_{t+j}), \tag{4}$$

where j indexes preceding steps and the empty product is one. This avoids counting the same target repeatedly. Marginalization over other possible observations occurs during rollout averaging.

For training, define $q_{t+k} = q_{\boldsymbol{\theta}}(\boldsymbol{z}_{t+k} \mid \mathcal{H}_{t+k})$ and predictive prior p_{t+k}^- obtained by integrating the transition against q_{t+k-1}. Jensen's inequality gives the one-step predictive-mixture bound

$$\begin{aligned}\log \int p_{\boldsymbol{\phi}}(\boldsymbol{o}_{t+k} \mid \boldsymbol{z}, \boldsymbol{e}_\ell) p_{t+k}^-(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} \geq \mathbb{E}_{q_{t+k}} \log p_{\boldsymbol{\phi}}(\boldsymbol{o}_{t+k} \mid \boldsymbol{z}_{t+k}, \boldsymbol{e}_\ell)\\ - D_{\text{KL}}(q_{t+k} \| p_{t+k}^-),\end{aligned} \tag{5}$$

where $\boldsymbol{z} \in \mathbb{R}^d$ is an integration variable and D_{KL} is Kullback–Leibler divergence. This is a bound for the stated predictive mixture. Replacing its nonlinear prior with a moment-matched Gaussian

yields a practical surrogate, not an exact sequence-likelihood bound. We minimize

$$\mathcal{L}_{\mathrm{wm}} = \mathbb{E}_{\mathcal{D}} \frac{1}{H} \sum_{k=1}^{H} \big[- \mathbb{E}_{q_{t+k}} \log p_{\boldsymbol{\phi}}(\boldsymbol{o}_{t+k} \mid \boldsymbol{z}_{t+k}, \boldsymbol{e}_\ell) + \lambda_{\mathrm{KL}} D_{\mathrm{KL}}(q_{t+k} \| p^-_{t+k}) + \lambda_z \| \operatorname{sg}(\boldsymbol{\mu}_{t+k}) - \hat{\boldsymbol{\mu}}^-_{t+k} \|_2^2 + \lambda_y \operatorname{BCE}(y_{t+k}, \hat{y}_{t+k}) + \lambda_c |c_{t+k} - \hat{c}_{t+k}| + \lambda_r \operatorname{BCE}(\zeta_{t+k}, \hat{r}_{t+k}) \big]. \tag{6}$$

$\mathcal{D}$ is the training-clip distribution, sg stops gradients through the target, $\hat{\boldsymbol{\mu}}^-_{t+k}$ is an open-loop predicted mean, and $\zeta_{t+k} \in \{0, 1\}$ labels a segment collision. Labels and losses are masked after termination. For binary label b and probability q, $\operatorname{BCE}(b, q) = -b \log q - (1 - b) \log(1 - q)$; probabilities are clipped away from zero and one. All λ coefficients are nonnegative weights. Appendix D defines alignment, calibration, and Gaussian KL. The total objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{wm}} + \lambda_{\mathrm{FM}} \mathcal{L}_{\mathrm{FM}} + \lambda_{\mathrm{align}} \mathcal{L}_{\mathrm{align}} + \lambda_{\mathrm{cal}} \mathcal{L}_{\mathrm{cal}}. \tag{7}$$

Training is staged: fit the first three terms on $\mathcal{D}$, then freeze representation and control parameters and fit only the designated probability/covariance calibration parameters using $\mathcal{L}_{\mathrm{cal}}$ on a disjoint split. The world-model parameter vector $\boldsymbol{\phi}$ includes the transition, observation, covariance, discovery, coverage, and collision heads.

4.3 Flow-Matching Controls and Geometric Feasibility

The normalized control matrix is $\boldsymbol{A} \in \mathbb{R}^{H \times 4}$. Define $\boldsymbol{D}_a = \operatorname{diag}(v_{\mathrm{scale}}, v_{\mathrm{scale}}, v_{\mathrm{scale}}, \omega_{\mathrm{scale}})$ with positive normalization scales. Physical controls are $\boldsymbol{U} = \boldsymbol{A}\boldsymbol{D}_a$; row $k + 1$ is $\boldsymbol{a}_{t+k}^\top$ for $k = 0, \ldots, H - 1$. These are control samples, not spline coefficients.

For independent standard normal entries in $\boldsymbol{A}_0$, demonstration $\boldsymbol{A}_1$ from $\mathcal{D}$, and flow time $s \sim \mathrm{Unif}[0, 1]$, define

$$\boldsymbol{A}_s = (1 - s)\boldsymbol{A}_0 + s\boldsymbol{A}_1, \qquad \frac{\partial \boldsymbol{A}_s}{\partial s} = \boldsymbol{A}_1 - \boldsymbol{A}_0,$$
$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E} \| V_{\boldsymbol{\psi}}(\boldsymbol{A}_s, s, \boldsymbol{\mu}_t, \boldsymbol{S}_t, \boldsymbol{e}_\ell) - (\boldsymbol{A}_1 - \boldsymbol{A}_0) \|_F^2. \tag{8}$$

$V_{\boldsymbol{\psi}}$ is a matrix-valued field with shape $H \times 4$; $\| \cdot \|_F$ is the Frobenius norm. Flow time s differs from scale vector $\boldsymbol{s}_t$ and physical time. Integrating $\mathrm{d}\boldsymbol{A}_s / \mathrm{d}s = V_{\boldsymbol{\psi}}(\boldsymbol{A}_s, s, \boldsymbol{\mu}_t, \boldsymbol{S}_t, \boldsymbol{e}_\ell)$ from 0 to 1 with K_{ode} solver steps yields raw candidate $\boldsymbol{A}^i_{\mathrm{raw}}$, for $i = 1, \ldots, N$ independent seeds. N is candidate count.

Repeated substitution of nominal kinematics exposes the rectangular rollout:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_{t+1}^\top \\ \boldsymbol{x}_{t+2}^\top \\ \vdots \\ \boldsymbol{x}_{t+H}^\top \end{bmatrix}}_{\boldsymbol{X} \in \mathbb{R}^{H \times 4}} = \mathbf{1}_H \boldsymbol{x}_t^\top + \Delta t \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{\boldsymbol{L}_H \in \mathbb{R}^{H \times H}} \underbrace{\boldsymbol{A}\boldsymbol{D}_a}_{\boldsymbol{U} \in \mathbb{R}^{H \times 4}}. \tag{9}$$

$\mathbf{1}_H$ is the all-ones vector and $\boldsymbol{L}_H$ the cumulative-sum matrix. Earlier controls affect all later viewpoints through its lower-triangular structure.

The feasible set $\mathcal{F}_t$ contains normalized sequences satisfying, for $k = 0, \ldots, H - 1$,

$$\| \boldsymbol{v}_{t+k} \|_2 \le v_{\max}, \quad \| \boldsymbol{v}_{t+k} - \boldsymbol{v}_{t+k-1} \|_2 \le a_{\max} \Delta t \ (k \ge 1), \quad |\omega_{t+k}| \le \omega_{\max},$$
$$\boldsymbol{p}^{\boldsymbol{A}}(\tau) \in \mathcal{W}, \qquad d_{\mathcal{M}_t}(\boldsymbol{p}^{\boldsymbol{A}}(\tau)) \ge r_{\mathrm{body}} + r_{\mathrm{margin}} \quad \text{for all } \tau \in [0, H\Delta t]. \tag{10}$$

$v_{\max}, a_{\max}, \omega_{\max}$ are motion limits. At the boundary impose $\| \boldsymbol{v}_t - \boldsymbol{v}_t^{\mathrm{meas}} \|_2 \le a_{\max} \Delta t$, where $\boldsymbol{v}_t^{\mathrm{meas}}$ is the measured initial velocity. $\boldsymbol{p}^{\boldsymbol{A}}(\tau)$ is the interpolated nominal position at physical offset τ. Signed distance $d_{\mathcal{M}_t}$ is positive in observed free space; unknown space is treated as occupied for execution. Radii $r_{\mathrm{body}}, r_{\mathrm{margin}}$ cover the body and tracking margin. The projection seeks

$$\boldsymbol{A}^i \in \arg\min_{\boldsymbol{A} \in \mathcal{F}_t} \tfrac{1}{2} \| \boldsymbol{A} - \boldsymbol{A}^i_{\mathrm{raw}} \|_F^2. \tag{11}$$

Collision constraints are nonconvex. A local solver can fail; reject solutions without swept-volume verification. Projection alone is not a safety guarantee. Conservative segment bounds or continuous collision checking must validate the path between sampled states.

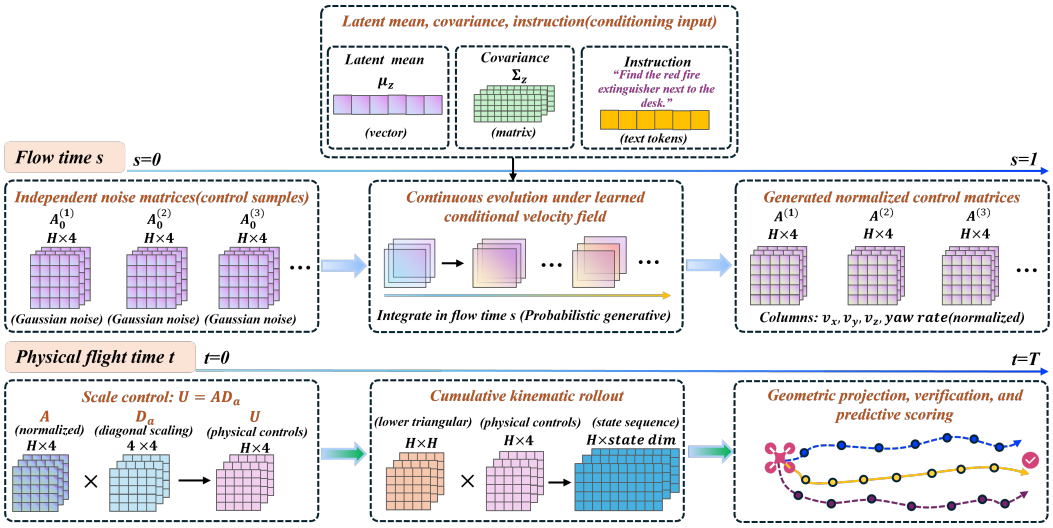


Figure 2: Flow-matching control generation and rollout. Flow time s transforms noise matrices into normalized $H \times 4$ control matrices. Physical-time scaling, cumulative kinematics, geometric verification, and predictive scoring determine the executable segment.

4.4 Information Gain and Closed-Loop Ranking

For an imagined step, linearize the transition and observation models:

$$\begin{aligned}
\boldsymbol{\mu}_{t+k}^{-} &= f_{\boldsymbol{\phi}}(\boldsymbol{\mu}_{t+k-1}^{+}, \boldsymbol{a}_{t+k-1}, \boldsymbol{e}_\ell), & \boldsymbol{S}_{t+k}^{-} &= \boldsymbol{F}_{t+k}\boldsymbol{S}_{t+k-1}^{+}\boldsymbol{F}_{t+k}^{\top} + \boldsymbol{Q}_{t+k}, \\
\boldsymbol{\Omega}_{t+k} &= \boldsymbol{J}_{t+k}\boldsymbol{S}_{t+k}^{-}\boldsymbol{J}_{t+k}^{\top} + \boldsymbol{R}_{t+k}, & \boldsymbol{K}_{t+k} &= \boldsymbol{S}_{t+k}^{-}\boldsymbol{J}_{t+k}^{\top}\boldsymbol{\Omega}_{t+k}^{-1}, \\
\boldsymbol{\mu}_{t+k}^{+} &= \boldsymbol{\mu}_{t+k}^{-} + \boldsymbol{K}_{t+k}[\hat{\boldsymbol{o}}_{t+k} - g_{\boldsymbol{\phi}}(\boldsymbol{\mu}_{t+k}^{-}, \boldsymbol{e}_\ell)], & & \\
\boldsymbol{S}_{t+k}^{+} &= (\boldsymbol{I}_d - \boldsymbol{K}_{t+k}\boldsymbol{J}_{t+k})\boldsymbol{S}_{t+k}^{-}(\boldsymbol{I}_d - \boldsymbol{K}_{t+k}\boldsymbol{J}_{t+k})^{\top} + \boldsymbol{K}_{t+k}\boldsymbol{R}_{t+k}\boldsymbol{K}_{t+k}^{\top}. & & \qquad (12)
\end{aligned}$$

$\boldsymbol{F}_{t+k} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{J}_{t+k} \in \mathbb{R}^{d_o\times d}$ are transition and observation Jacobians at the corresponding means. Innovation covariance is $\boldsymbol{\Omega}_{t+k} \in \mathbb{R}^{d_o\times d_o}$ and gain is $\boldsymbol{K}_{t+k} \in \mathbb{R}^{d\times d_o}$. Predicted observations $\hat{\boldsymbol{o}}_{t+k}$ are sampled from the predictive observation distribution. Equation 12 is a local Gaussian approximation in imagination; actual observations use equation 2. Agreement between these updates requires calibration.

Gaussian differential entropy is $\mathsf{h}(\boldsymbol{S}) = \frac{1}{2}\log\det((2\pi e)\boldsymbol{S})$, where π and e are the circle and Euler constants. For a fixed predicted history,

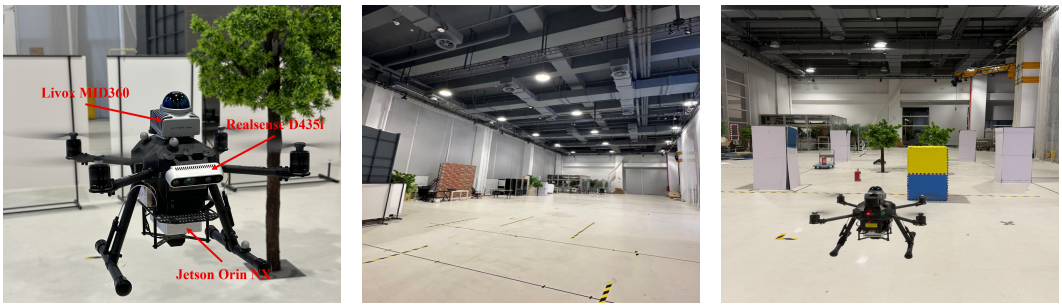
$$\begin{aligned}
\hat{g}_{t+k} &= \mathsf{h}(\boldsymbol{S}_{t+k}^{-}) - \mathbb{E}_{\hat{\boldsymbol{o}}_{t+k}}\mathsf{h}(\boldsymbol{S}_{t+k}^{+}) \simeq \tfrac{1}{2}\log\frac{\det \boldsymbol{S}_{t+k}^{-}}{\det \boldsymbol{S}_{t+k}^{+}} \\
&= \tfrac{1}{2}\log\det\Big(\boldsymbol{I}_d + (\boldsymbol{S}_{t+k}^{-})^{1/2}\boldsymbol{J}_{t+k}^{\top}\boldsymbol{R}_{t+k}^{-1}\boldsymbol{J}_{t+k}(\boldsymbol{S}_{t+k}^{-})^{1/2}\Big) \geq 0. \qquad (13)
\end{aligned}$$

The last equality is exact for a fixed linear-Gaussian observation model; Appendix E derives it. Both entropies concern the same future latent, before and after its observation. Their sum rewards successive evidence acquisition and is not asserted to equal mutual information about one static variable. Learned noise and local linearization limit the approximation.

Set $\hat{u}_{t+k} = \mathrm{tr}(\boldsymbol{S}_{t+k}^{+})/d$, with tr the matrix trace, and score

$$\widehat{U}_t(\boldsymbol{A}^i) = \mathbb{E}_{p_{\boldsymbol{\phi}}, b_t}\sum_{k=1}^{H}\gamma^{k-1}(\alpha\hat{y}_{t+k}^i + \beta\hat{c}_{t+k}^i + \kappa\hat{g}_{t+k}^i - \rho\hat{r}_{t+k}^i - \eta\hat{u}_{t+k}^i). \qquad (14)$$

The expectation is estimated with M_{roll} stochastic rollouts per candidate. Define the union-bound proxy $\widehat{R}_t(\boldsymbol{A}^i) = \min(1, \sum_{k=1}^{H}\mathbb{E}[\hat{r}_{t+k}^i])$ and budget $\delta \in [0, 1]$. The accepted indices and decision



(a) Platform (b) Arena (c) Task scene

Figure 3: Representative real-world deployment context: (a) aerial platform and payload, (b) indoor flight arena, and (c) a cluttered task scene.

are

$$\mathcal{I}_t = \{i : \boldsymbol{A}^i \text{ passes geometric verification, } \widehat{R}_t(\boldsymbol{A}^i) \le \delta\}, \qquad i^* = \arg\max_{i \in \mathcal{I}_t} \widehat{U}_t(\boldsymbol{A}^i). \tag{15}$$

If $\mathcal{I}_t$ is nonempty, execute up to h controls from $\boldsymbol{A}^{i^*}\boldsymbol{D}_a$, with $1 \le h \le H$, monitoring obstacles and termination every interval. Otherwise use verified braking/hover; hover requires a verified clear stopping region. If no fallback is safe, invoke the platform emergency-stop or landing procedure. Appendix A gives the complete decision loop in Algorithm 1.

Conditional analysis. For a fixed verified candidate set, a uniform utility-estimation error of ε_U gives ranking regret at most $2\varepsilon_U$. A union bound converts conditional segment-risk errors into a horizon bound, without a physical safety certificate. Appendix B states the assumptions and derivations.

5 Experiments

5.1 Environments and Protocol

The simulation benchmark includes HM3D assets rendered with Habitat-Sim through a continuous three-dimensional flight wrapper, Isaac Sim industrial layouts, and AirSim cluttered environments. The flight wrapper uses swept-body collision geometry and continuous flight dynamics. The real-world setting is an indoor industrial mock-up with shelves, pipes, valves, generators, and signs. Episodes specify an instruction, target instance, and initial pose and terminate on verified discovery, collision, timeout, or battery budget. Figure 3 shows the physical platform, indoor arena, and representative task scene. Simulator suites and hardware trials are treated as distinct evaluation domains.

Matched evaluation. The main comparison uses a *matched-component* evaluation with identical registered RGB-D, pose estimates, field of view, sensing range, frame rate, geofence, start poses, targets, low-level controller, collision checker, and time/battery budgets. Methods start with empty scene memory and receive no privileged ground-truth map. Interchangeable components use the same language/vision backbone and detector, while method-specific pretrained components retain their stated adaptation interfaces. Learned controls share training demonstrations and disjoint validation/test scenes. Tuning is confined to validation data, and elapsed time includes inference. Equal online inputs control sensing access; they do not imply equal pretraining or adaptation cost.

5.2 Baselines and Metrics

The baseline set covers geometric exploration, semantic navigation, generative control, and recent aerial semantic search. FUEL, FAEL, and STEM provide geometric and semantic exploration comparisons. VLFM and PONI are evaluated through the UAV adaptations described in Appendix F. Diffusion Policy denotes a UAV-control adaptation of Chi et al. (2023). The in-house flow policy follows Lipman et al. (2023), conditions on current evidence, and ranks candidates with a current-map

Table 1: Main comparison on the matched single-UAV benchmark. SR, Coverage, and Coll. are percentages; SPL is in $[0, 1]$; Time is failure-penalized seconds. † denotes a matched UAV adaptation and ‡ an in-house baseline.

	Task performance		Efficiency	Exploration	Safety
Method	SR↑	SPL↑	Time↓	Coverage↑	Coll.↓
Geometric exploration					
FUEL (Zhou et al., 2021)	71.0	0.42	188.0	62.0	8.0
FAEL† (Huang et al., 2023)	74.0	0.46	176.0	64.0	7.0
Semantic exploration and navigation					
STEM (Sethi et al., 2026)	82.0	0.55	143.0	69.0	6.0
VLFM frontier† (Yokoyama et al., 2024)	78.0	0.51	157.0	66.0	7.0
PONI ObjectNav† (Ramakrishnan et al., 2022)	75.0	0.48	169.0	61.0	9.0
Generative control baselines					
Diffusion Policy† (Chi et al., 2023)	80.0	0.54	151.0	65.0	8.0
Flow policy‡ (Lipman et al., 2023)	81.0	0.56	146.0	66.0	7.0
Recent aerial semantic search (2025–2026)					
VISTA† (Nagami et al., 2025)	85.0	0.58	137.0	77.0	5.5
RAVEN† (Kim et al., 2025)	87.0	0.61	132.0	70.0	5.0
AirHunt† (Chen et al., 2026)	89.0	0.64	118.0	71.0	4.5
AIRWORLD (ours)	92.5	0.68	121.0	75.0	4.0

objective. Its citation identifies the generative formulation. World-model and language components are evaluated separately in the ablation study.

The recent aerial baselines are VISTA, RAVEN, and AirHunt (Nagami et al., 2025; Kim et al., 2025; Chen et al., 2026). Their matched adaptations use RGB-D-only sensing, local language inference, and one target instance. Appendix F specifies the interfaces and distinguishes the matched evaluation from each method's original protocol.

For n_{ep} episodes indexed by e, define success $s_e \in \{0, 1\}$, shortest collision-free three-dimensional path length $l_e > 0$ to a valid target-view region, and actual path length p_e. Targets already visible at initialization are excluded. Let t_e be discovery time for success, T_e the episode time cap, V_e^{obs} observed reachable free volume at termination, $V_e^{\text{free}} > 0$ total reachable free volume, and $q_e^{\text{coll}} \in \{0, 1\}$ indicate any collision. Evaluator-only geometry supplies oracle paths and denominators; planners cannot access it. We report

$$\text{SR} = \frac{100}{n_{\text{ep}}} \sum_e s_e, \qquad \text{SPL} = \frac{1}{n_{\text{ep}}} \sum_e s_e \frac{l_e}{\max(l_e, p_e)},$$
$$\text{Time} = \frac{1}{n_{\text{ep}}} \sum_e [s_e t_e + (1 - s_e) T_e], \qquad \text{Coverage} = \frac{100}{n_{\text{ep}}} \sum_e \frac{V_e^{\text{obs}}}{V_e^{\text{free}}},$$
$$\text{Coll.} = \frac{100}{n_{\text{ep}}} \sum_e q_e^{\text{coll}}. \tag{16}$$

For failures set $t_e = 0$ and charge T_e, including early collisions; thus Time is failure-penalized seconds, not successful-episode-only time. Discovery requires the common verifier to match the requested instance in an observable, collision-free target-view region with fixed visibility and confidence thresholds. Verifier thresholds and episode manifests are fixed before evaluation. Coverage at termination is secondary because successful episodes stop earlier. Average stepwise information gain (IG, nats per decision) is evaluated with a common frozen diagnostic belief model for ablations; entropies from different learned latent spaces cannot be compared directly. Paired episode seeds support method-level comparisons under the same initial conditions.

Benchmark results. Table 1 reports results from the matched benchmark. The three recent methods are evaluated as explicitly defined UAV adaptations, rather than as direct conversions of their published numbers. Appendix F records the original protocols and explains the adaptations.

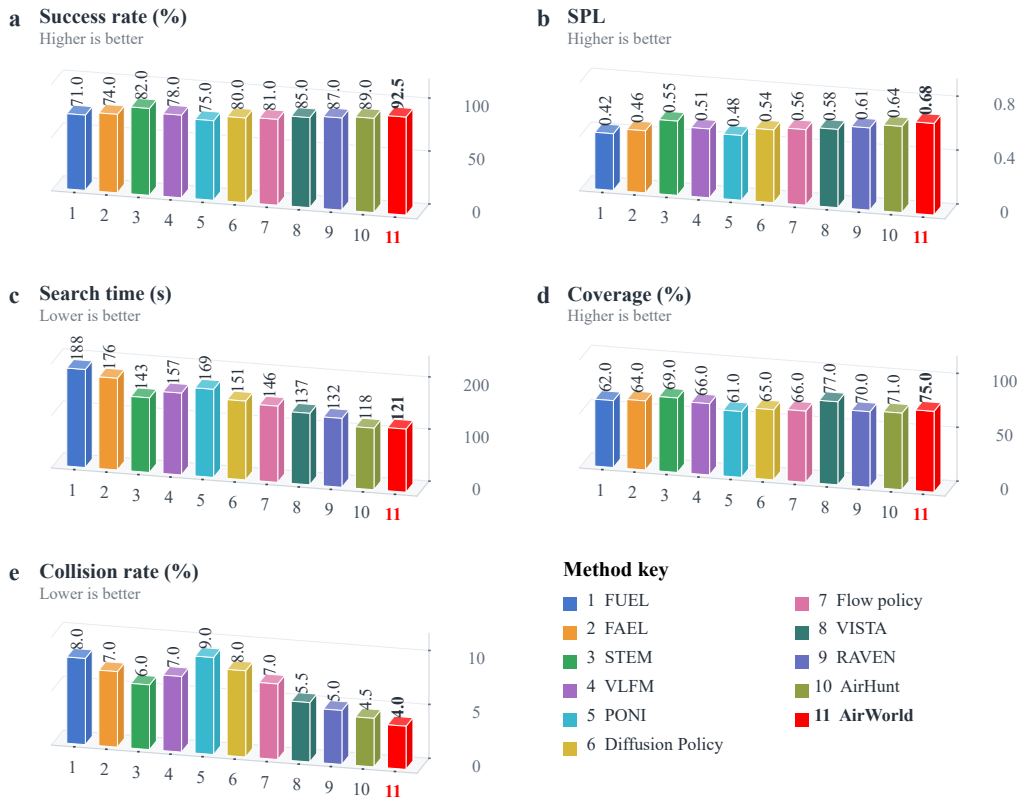


Figure 4: Multi-metric comparison of the 11 methods in Table 1: (a) success rate, (b) SPL, (c) failure-penalized search time, (d) coverage, and (e) collision rate.

Comparison across metrics. Figure 4 visualizes all five metrics in Table 1. In these point estimates, AirWorld has an SR of 92.5% and SPL of 0.68, compared with 89.0% and 0.64 for AirHunt; the collision rates are 4.0% and 4.5%, respectively. The tradeoff is not uniform across metrics: AirHunt has a shorter failure-penalized time (118 versus 121 s), while VISTA has higher coverage (77% versus 75%). Since discovery terminates an episode, coverage is interpreted jointly with task success and elapsed time. These comparisons concern point estimates; Appendix I.1 specifies paired statistical analysis.

6 Ablation Study

Table 2 evaluates individual components with fixed data splits, seeds, motion limits, and compute budgets. Without a world model, a current-map gain ranks candidates; without future prediction, the pretrained representation is retained but scoring uses current evidence. Without flow matching, a matched-count motion-primitive sampler replaces the generator. Without uncertainty modeling, covariance is removed from policy conditioning and $\kappa = \eta = 0$, while the geometric checker remains. Without language conditioning, planning receives a null embedding, but the external verifier retains the target. IG is computed with the same frozen evaluator for every variant, including those without their own covariance. Removing the world model gives the largest SR reduction in this comparison, from 92.5% to 79.0%; removing future prediction reduces SR to 84.0%. These component comparisons separate predictive ranking from representation learning.

Objective and calibration diagnostics. Table 3 isolates the information-gain reward, the penalty on residual posterior uncertainty, and the disjoint calibration stage while retaining covariance propagation and the geometric verifier. Removing the gain reward lowers common-evaluator IG from 0.34

Table 2: Component ablation on the matched single-UAV benchmark. WM: predictive world-model module; Lang.: language conditioning; Flow: flow-matching sampler; Future: multi-step predictive scoring; Unc.: uncertainty-aware scoring. IG uses a common frozen evaluator.

	Model components					Evaluation		
Variant	WM	Lang.	Flow	Future	Unc.	SR (%)↑	Time (s)↓	IG↑
w/o World Model	×	✓	✓	×	✓	79.0	158.0	0.19
w/o Language Conditioning	✓	×	✓	✓	✓	83.0	149.0	0.23
w/o Flow Matching	✓	✓	×	✓	✓	86.0	139.0	0.27
w/o Future Prediction	✓	✓	✓	×	✓	84.0	145.0	0.21
w/o Uncertainty Modeling	✓	✓	✓	✓	×	87.0	136.0	0.24
AIRWORLD	✓	✓	✓	✓	✓	92.5	121.0	0.34

Table 3: Targeted objective and calibration ablations. Every variant retains a covariance representation and the same geometric verifier; IG uses the common frozen evaluator. The calibration condition changes only the disjoint calibration stage.

Variant	SR (%)↑	SPL↑	Time (s)↓	Coll. (%)↓	IG↑
No gain reward ($\kappa = 0$)	89.0	0.63	129.0	4.5	0.27
No residual penalty ($\eta = 0$)	91.0	0.66	124.0	5.0	0.35
No calibration stage	90.5	0.65	126.0	6.0	0.33
AIRWORLD	92.5	0.68	121.0	4.0	0.34

to 0.27 and reduces SR. Removing only the residual-uncertainty penalty leaves IG slightly higher (0.35), but raises collision frequency from 4.0% to 5.0%, showing why gain alone is not the ranking criterion. Omitting calibration increases collision frequency to 6.0% and lowers common-evaluator IG to 0.33. At horizon 16, calibration leaves feature RMSE at 0.49 but reduces risk Brier score from 0.069 to 0.044 (Appendix I.7), separating probability reliability from mean prediction accuracy. Calibration remains diagnostic and does not establish the conditional risk bound in Eq. 18.

Appendix I provides the factorial conditions, the full objective ablation, sampler controls under matched computation, horizon and proposal-count sensitivity, generalization, robustness, calibration, and latency diagnostics.

7 Limitations and Safety

The formulation assumes synchronized RGB-D, a stable state estimate, and a reliable tracking controller. Gaussian beliefs and local linearization can miss multimodal ambiguity. Learned covariance may reflect feature scaling rather than physical uncertainty without calibration; imagined updates can disagree with the observation encoder. Flow sampling adds compute, and a nonconvex feasibility solver may reject all candidates. Category-trained adaptations such as PONI do not natively solve unrestricted relational instructions. The supplementary diagnostics examine latency, calibration, and performance on shared-category and open-vocabulary subsets; multimodal belief approximation and hardware transfer remain limitations. Runtime collision checking, geofencing, and emergency procedures remain necessary. The conditional bounds do not certify real-world safety.

8 Conclusion

AirWorld formulates single-UAV semantic search as predictive decision making with a latent belief, flow-generated controls, and expected future evidence. The matched benchmark achieves 92.5% success, 0.68 SPL, and a 4.0% collision rate, with tradeoffs in search time and coverage relative to the strongest baselines. Targeted ablations separate the information-gain reward, residual-uncertainty penalty, and calibration stage, connecting evidence acquisition and predictive reliability to closed-loop search outcomes. The method couples information-seeking exploration with geometric verification. Future work should address multimodal beliefs and verify calibration under changes in sensing, tracking, and environment.

AI USE STATEMENT

AI tools assisted with language editing, plotting code, figure formatting, supplementary evaluation design, and experimental-table formatting. The authors take full responsibility for the final manuscript, including verification of its scientific claims and experimental measurements against source records.

REFERENCES

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.

Andreas Bircher, Mohamed Kamel, Kostas Alexis, Helen Oleynikova, and Roland Siegwart. Receding horizon next-best-view planner for 3d exploration. In *IEEE International Conference on Robotics and Automation*, pp. 1462–1468, 2016.

Junting Chen, Guohao Li, Suryansh Kumar, Bernard Ghanem, and Fisher Yu. How to not train your dragon: Training-free embodied object goal navigation with semantic frontiers. *arXiv preprint arXiv:2305.16925*, 2023.

Xuecheng Chen, Zongzhuo Liu, Jianfa Ma, Bang Du, Tiantian Zhang, Xueqian Wang, and Boyu Zhou. AirHunt: Bridging VLM semantics and continuous planning for efficient aerial object navigation. *arXiv preprint arXiv:2601.12742*, 2026.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.026.

Titus Cieslewski, Elia Kaufmann, and Davide Scaramuzza. Rapid exploration with multi-rotors: A frontier selection method for high speed flight. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2135–2142, 2017.

Mihir Dharmadhikari, Tien Dang, Laxman Solanka, Joshua Loje, Huy Nguyen, Nikhil Khedekar, and Kostas Alexis. Motion primitives-based path planning for fast and agile exploration using aerial robots. In *IEEE International Conference on Robotics and Automation*, 2020.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Junlong Huang, Boyu Zhou, Zhengping Fan, Yilin Zhu, Yingrui Jie, Longwei Li, and Hui Cheng. FAEL: Fast autonomous exploration for large-scale environments with a mobile robot. *IEEE Robotics and Automation Letters*, 8(3):1667–1674, 2023. doi: 10.1109/LRA.2023.3236573.

Michael Janner, Yilun Du, Joshua B. Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. In *International Conference on Machine Learning*, 2022.

Seungchan Kim, Omar Alama, Dmytro Kurdydyk, John Keller, Nikhil Keetha, Wenshan Wang, Yonatan Bisk, and Sebastian Scherer. RAVEN: Resilient aerial navigation via open-set semantic memory and behavior adaptation. *arXiv preprint arXiv:2509.23563*, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Yiming Luo, Zixuan Zhuang, Neng Pan, Chen Feng, Shaojie Shen, Fei Gao, Hui Cheng, and Boyu Zhou. Star-Searcher: A complete and efficient aerial system for autonomous target search in complex unknown environments. *arXiv preprint arXiv:2402.16348*, 2024.

Keiko Nagami, Timothy Chen, Javier Yu, Ola Shorinwa, Maximilian Adang, Carlyn Dougherty, Eric Cristofalo, and Mac Schwager. VISTA: Open-vocabulary, task-relevant robot exploration with online semantic gaussian splatting. *arXiv preprint arXiv:2507.01125*, 2025.

Sotiris Papatheodorou, Nils Funk, Dimos Tzoumanikas, Christopher Choi, Binbin Xu, and Stefan Leutenegger. Finding things in the unknown: Semantic object-centric exploration with an MAV. In *IEEE International Conference on Robotics and Automation*, pp. 3339–3345, 2023. doi: 10.1109/ICRA48891.2023.10160490.

S. K. Ramakrishnan, Devendra Singh Chaplot, Ziad Al-Halah, Jitendra Malik, and Kristen Grauman. Poni: Potential functions for object-goal navigation with interaction-free learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Julian Schrittwieser et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588:604–609, 2020.

Nikhil Sethi, Max Lodel, Laura Ferranti, Robert Babuška, and Javier Alonso-Mora. Stem: Semantic target search and exploration using mavs in cluttered environments. *Autonomous Robots*, 50:32, 2026. doi: 10.1007/s10514-026-10247-6.

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

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*, 2024.

Yichen Zhang, Xinyi Chen, Chen Feng, Boyu Zhou, and Shaojie Shen. FALCON: Fast autonomous aerial exploration using coverage path guidance. *arXiv preprint arXiv:2407.00577*, 2024.

Boyu Zhou, Yichen Zhang, Xinyi Chen, and Shaojie Shen. FUEL: Fast UAV exploration using incremental frontier structure and hierarchical planning. *IEEE Robotics and Automation Letters*, 6(2):779–786, 2021.

Algorithm 1 AirWorld: belief-space receding-horizon search

Require: Trained $E_{\boldsymbol{\theta}}, f_{\boldsymbol{\phi}}, g_{\boldsymbol{\phi}}, V_{\boldsymbol{\psi}}$ and prediction heads; instruction ℓ;
$H, h, N, K_{\mathrm{ode}}, M_{\mathrm{roll}}, \Delta t, \boldsymbol{D}_a, \delta$, utility weights and motion limits

$t \leftarrow 0$; $\mathcal{M}_0 \leftarrow \emptyset$; $\boldsymbol{e}_\ell \leftarrow \mathrm{TextEncode}(\ell)$
$(\boldsymbol{\mu}_0^-, \boldsymbol{S}_0^-) \leftarrow (\mathbf{0}, \boldsymbol{I}_d)$
while discovery, collision, time, and battery termination conditions are false **do**
  Acquire $\boldsymbol{O}_t, \boldsymbol{x}_t$ and initial velocity; update $\mathcal{M}_t$ from registered RGB-D
  Update b_t with equation 2; $\mathcal{I}_t \leftarrow \emptyset$
  for $i = 1, \ldots, N$ **do**
    Draw normal $\boldsymbol{A}_0^i$; integrate the flow ODE with K_{ode} steps to $\boldsymbol{A}_{\mathrm{raw}}^i$
    Solve equation 11 for $\boldsymbol{A}^i$
    if solver succeeds and the swept path passes equation 10 **then**
      Roll out M_{roll} futures with equations 3, 4, 12
      Compute $\widehat{U}_t(\boldsymbol{A}^i)$, $\widehat{R}_t(\boldsymbol{A}^i)$ as in equations 14–15
      if $\widehat{R}_t(\boldsymbol{A}^i) \le \delta$ **then**
        $\mathcal{I}_t \leftarrow \mathcal{I}_t \cup \{i\}$
      end if
    end if
  end for
  if $\mathcal{I}_t = \emptyset$ **then**
    Execute one verified fallback interval; exit if emergency procedure is required
  else
    Select i^* by equation 15; execute up to h rows of $\boldsymbol{A}^{i^*}\boldsymbol{D}_a$
    Interrupt execution on new obstacle evidence or an episode termination condition
  end if
  Let h_{exec} be intervals actually executed; log controls and measurements
  Propagate b_t through these controls to $(\boldsymbol{\mu}_{t+h_{\mathrm{exec}}}^-, \boldsymbol{S}_{t+h_{\mathrm{exec}}}^-)$
  $t \leftarrow t + h_{\mathrm{exec}}$; carry the updated local map to the next decision
end while
return discovery status and observation, control, timing, and collision logs

A Online Planning Algorithm

TextEncode is the fixed language encoder used to obtain $\boldsymbol{e}_\ell$; h_{exec} counts completed control intervals and can be smaller than h under interruption. Termination is checked before executing another interval.

B Conditional Analysis and Proofs

For the same nonempty verified candidate set $\mathcal{C}_t = \{\boldsymbol{A}^i : i \in \mathcal{I}_t\}$, let U_t^* be true expected utility. Suppose $|\widehat{U}_t(\boldsymbol{A}) - U_t^*(\boldsymbol{A})| \le \varepsilon_U$ uniformly on $\mathcal{C}_t$, including Monte Carlo error. Let $\boldsymbol{A}^* = \arg\max_{\boldsymbol{A}\in\mathcal{C}_t} U_t^*(\boldsymbol{A})$ and $\hat{\boldsymbol{A}} = \arg\max_{\boldsymbol{A}\in\mathcal{C}_t} \widehat{U}_t(\boldsymbol{A})$. Then

$$U_t^*(\boldsymbol{A}^*) - U_t^*(\hat{\boldsymbol{A}}) = [U_t^*(\boldsymbol{A}^*) - \widehat{U}_t(\boldsymbol{A}^*)] + [\widehat{U}_t(\boldsymbol{A}^*) - \widehat{U}_t(\hat{\boldsymbol{A}})] + [\widehat{U}_t(\hat{\boldsymbol{A}}) - U_t^*(\hat{\boldsymbol{A}})] \le \varepsilon_U + 0 + \varepsilon_U = 2\varepsilon_U. \tag{17}$$

This bounds ranking error within the sampled set. If $U_t^{\mathrm{opt}} = \sup_{\boldsymbol{A}\in\mathcal{F}_t} U_t^*(\boldsymbol{A})$ and $\varepsilon_{\mathrm{cand}} = U_t^{\mathrm{opt}} - U_t^*(\boldsymbol{A}^*)$, the full gap is at most $\varepsilon_{\mathrm{cand}} + 2\varepsilon_U$; no bound on candidate coverage is asserted.

For risk, suppose each selected-candidate segment satisfies the conditional bound $r_{t+k} \le \hat{r}_{t+k} + \varepsilon_r$ for all relevant histories, with $\varepsilon_r \ge 0$. The union bound implies

$$\Pr(\text{collision within horizon} \mid \mathcal{H}_t, \boldsymbol{A}) \le \min\left(1, \sum_{k=1}^{H} \mathbb{E}[\hat{r}_{t+k}] + H\varepsilon_r\right). \tag{18}$$

If the predicted sum is at most δ, this gives $\min(1, \delta + H\varepsilon_r)$. Average calibration scores do not imply the required conditional bound, and finite-rollout risk estimates need their own error allowance. No formal physical safety guarantee is claimed. For two candidates with all other expected utility terms equal, larger expected gain is preferred when $\kappa > 0$; different coverage, risk, or residual uncertainty can reverse that preference.

C CORE NOTATION

The table summarizes the core variables for belief updates, control generation, candidate ranking, and training. Bold lowercase denotes vectors and bold uppercase denotes matrices. Superscripts $-$ and $+$ denote the prior and posterior at the same time; hats denote predictions. Auxiliary quantities are defined where they first appear.

Symbol	Definition and type
t, H, h, N	Current control-step index, prediction horizon, maximum executed prefix, and candidate count (integers).
$\Delta t, s$	Physical control interval in seconds and flow time $s \in [0, 1]$.
$\ell, \boldsymbol{e}_\ell$	Natural-language instruction and its embedding in $\mathbb{R}^{d_\ell}$.
$\boldsymbol{O}_t, \boldsymbol{o}_t$	Registered RGB-D observations ($n_{\text{pix}} \times 4$ matrix) and observation-feature vector in $\mathbb{R}^{d_o}$.
$\boldsymbol{x}_t, \boldsymbol{a}_t$	Planning state (position and yaw) and control (velocity and yaw rate), both in $\mathbb{R}^4$.
$\mathcal{M}_t, \mathcal{H}_t$	Observed local map and available observation, state, control, and instruction history.
$\boldsymbol{z}_t, b_t$	Latent exploration state in $\mathbb{R}^d$ and its belief density.
$\boldsymbol{\mu}_t, \boldsymbol{S}_t$	Belief mean in $\mathbb{R}^d$ and positive-definite covariance in $\mathbb{R}^{d\times d}$.
$\boldsymbol{\mu}_t^{\pm}, \boldsymbol{S}_t^{\pm}$	Prior $(-)$ and posterior $(+)$ mean and covariance for the same step.
$\boldsymbol{Q}_{t+k}, \boldsymbol{R}_{t+k}$	Process-noise and observation-noise covariances, of shape $d \times d$ and $d_o \times d_o$.
$\boldsymbol{A}, \boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{A}_s$	Normalized $H \times 4$ control matrix, noise endpoint, demonstration endpoint, and flow interpolation.
$\boldsymbol{U}, \boldsymbol{D}_a$	Physical control matrix $\boldsymbol{U} = \boldsymbol{A}\boldsymbol{D}_a$ and diagonal 4×4 scaling matrix.
y_{t+k}, c_{t+k}	First verified discovery indicator and normalized newly observed free volume.
$g_{t+k}, r_{t+k}, u_{t+k}$	Expected information gain (nats), segment collision probability, and posterior uncertainty.
$\gamma, \alpha, \beta, \kappa, \rho, \eta$	Discount factor and weights for discovery, coverage, information gain, risk, and uncertainty.
$\mathcal{F}_t, \mathcal{I}_t$	Geometrically feasible control set and indices of verified candidates meeting the risk budget.
$\widehat{U}_t, \widehat{R}_t, \delta$	Predicted candidate utility, horizon collision-risk proxy, and allowed predicted risk.
$\mathcal{L}, \mathcal{L}_{\text{wm}}, \mathcal{L}_{\text{FM}}$	Total training objective, world-model loss, and flow-matching loss.

D AUXILIARY LOSSES AND GAUSSIAN KL

For a batch of B matched region/instruction pairs indexed by b, project region semantic latents and their paired instruction embeddings with $\boldsymbol{W}_f \in \mathbb{R}^{d_a \times d_{\text{sem}}}$ and $\boldsymbol{W}_h \in \mathbb{R}^{d_a \times d_\ell}$. The resulting unit-normalized vectors are $\boldsymbol{f}_b, \boldsymbol{h}_b \in \mathbb{R}^{d_a}$. d_a is the common embedding dimension; zero-norm projections are excluded or normalized with a fixed numerical floor. These trainable projection matrices are included in parameter vector $\boldsymbol{\theta}$. For temperature $T_{\text{align}} > 0$,

$$\mathcal{L}_{\text{align}} = -\frac{1}{B}\sum_{b=1}^{B} \log \frac{\exp(\boldsymbol{f}_b^\top \boldsymbol{h}_b / T_{\text{align}})}{\sum_{j=1}^{B} \exp(\boldsymbol{f}_b^\top \boldsymbol{h}_j / T_{\text{align}})}. \tag{19}$$

Here j indexes in-batch language negatives. This term requires actual region/instruction correspondences in the training data; it cannot be obtained from unlabeled flight trajectories alone.

For a disjoint calibration split $\mathcal{D}_{\text{cal}}$, assimilate the actual feature $\boldsymbol{o}_{t+k}$ through equation 12, obtaining $\bar{\boldsymbol{\mu}}_{t+k}^{+}, \bar{\boldsymbol{S}}_{t+k}^{+}$. Compare against encoder posterior moments and use Brier losses for discovery and collision:

$$\begin{aligned}\mathcal{L}_{\text{cal}} = \mathbb{E}_{\mathcal{D}_{\text{cal}}} \frac{1}{H}\sum_{k=1}^{H} \big[&(\hat{y}_{t+k} - y_{t+k})^2 + (\hat{r}_{t+k} - \zeta_{t+k})^2 + \tfrac{1}{d}\|\bar{\boldsymbol{\mu}}_{t+k}^{+} - \text{sg}(\boldsymbol{\mu}_{t+k})\|_2^2 \\ &+ \tfrac{1}{d^2}\|\bar{\boldsymbol{S}}_{t+k}^{+} - \text{sg}(\boldsymbol{S}_{t+k})\|_F^2\big].\end{aligned} \tag{20}$$

This is a specified calibration/consistency surrogate, not proof of conditional calibration. Fit only the designated covariance/probability calibration parameters on this split and never on test scenes; freeze the encoder targets during calibration. Post-termination samples are masked. Feature whitening and fixed scaling must be shared between training and evaluation to make covariance magnitudes interpretable.

For generic $q = \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{C}_0)$ and $p = \mathcal{N}(\boldsymbol{m}_1, \boldsymbol{C}_1)$, where both covariance matrices are positive definite, the Gaussian KL used in the surrogate is

$$D_{\mathrm{KL}}(q\|p) = \tfrac{1}{2}\big[\,\mathrm{tr}(\boldsymbol{C}_1^{-1}\boldsymbol{C}_0) + (\boldsymbol{m}_1 - \boldsymbol{m}_0)^\top \boldsymbol{C}_1^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_0) - d + \log \tfrac{\det \boldsymbol{C}_1}{\det \boldsymbol{C}_0}\big]. \tag{21}$$

$\boldsymbol{m}_0, \boldsymbol{m}_1 \in \mathbb{R}^d$ are means and $\boldsymbol{C}_0, \boldsymbol{C}_1 \in \mathbb{R}^{d\times d}$ covariances. Cholesky solves and log-determinants avoid explicit matrix inversion in an implementation.

E Information-Gain Derivation

Suppress the common time index and fix a linearized observation model with prior covariance $\boldsymbol{S}^-$, Jacobian $\boldsymbol{J}$, and observation noise $\boldsymbol{R}$. The posterior precision is

$$(\boldsymbol{S}^+)^{-1} = (\boldsymbol{S}^-)^{-1} + \boldsymbol{J}^\top \boldsymbol{R}^{-1}\boldsymbol{J}. \tag{22}$$

Multiplying on both sides by the symmetric positive-definite square root $(\boldsymbol{S}^-)^{1/2}$ and taking determinants yields

$$\begin{aligned}\frac{\det \boldsymbol{S}^-}{\det \boldsymbol{S}^+} &= \det\Big((\boldsymbol{S}^-)^{1/2}(\boldsymbol{S}^+)^{-1}(\boldsymbol{S}^-)^{1/2}\Big)\\ &= \det\Big(\boldsymbol{I}_d + (\boldsymbol{S}^-)^{1/2}\boldsymbol{J}^\top \boldsymbol{R}^{-1}\boldsymbol{J}(\boldsymbol{S}^-)^{1/2}\Big).\end{aligned} \tag{23}$$

Subtracting the two Gaussian entropies cancels the constant $d\log(2\pi e)/2$ and gives equation 13. The added matrix is positive semidefinite, so its identity-shifted determinant is at least one. Thus the gain is nonnegative under these assumptions. If visibility or covariance depends on the realized observation, retain the expectation in equation 13 and average observation-specific posterior entropies. Comparing $\mathsf{h}(\boldsymbol{S}_t)$ to $\mathsf{h}(\boldsymbol{S}_{t+H})$ would mix state evolution and evidence acquisition and is not a valid replacement.

F Baseline Provenance and Adaptation

The baseline descriptions refer to the 2025 preprint versions of VISTA and RAVEN and the January 2026 AirHunt preprint. Their published evaluation settings establish method provenance; the matched results in Table 1 use the adaptations specified below.

VISTA (Nagami et al., 2025). The original quadrotor Table I reports 83.33% success on a low-occlusion map and 100% on a high-occlusion map, with six trials per map. These settings differ from the pooled benchmark here. The matched adaptation retains semantic reconstruction and the task-aware viewpoint objective, uses the common RGB-D/pose stream and initialization budget, and connects the command interface to the common controller. Reconstruction is included in the online compute budget.

RAVEN (Kim et al., 2025). The original Table I reports 50.14% Progress and 31.89% PPL across 100 mixed tasks; Type I reports 54.19% Progress and 37.28% PPL and can include multiple instances. These metrics differ from single-instance SR/SPL. The matched adaptation restricts the goal to one instance, initializes voxel/ray memory empty, retains the behavior logic and ray memory, and uses the common RGB-D, pose, controller, and language-inference budgets. The indoor setting evaluates transfer from the original operating domain.

AirHunt (Chen et al., 2026). The original Table I reports SR 73.1, SPL 42 on a percentage scale, and flight time 120.8 under an outdoor protocol using LiDAR, three RGB cameras, and a networked language-model API. The matched RGB-D/local-inference adaptation retains semantic-geometric planning and asynchronous scheduling, builds geometry from depth, and uses the common camera stream and local language backbone. The reported elapsed time includes inference latency.

Table 5: Controller configuration and units. Motion limits and normalization scales are expressed in physical units; utility weights are selected using validation data, and the tracking margin is determined from controller logs.

Quantity	Setting	Unit or role
H, h	$16, 4$	control intervals / executed prefix
$N, K_{\text{ode}}, M_{\text{roll}}$	$64, 20, 8$	candidates / ODE steps / rollouts
Δt	0.10	seconds per control interval
$v_{\text{scale}}, \omega_{\text{scale}}$	$1.5, 1.0$	m/s and rad/s normalization
$v_{\max}, a_{\max}, \omega_{\max}$	$1.5, 2.0, 1.0$	m/s, m/s^2, rad/s
ϵ_{cov}	10^{-4}	variance floor
γ, δ	$0.95, 0.05$	discount / predicted collision budget
r_{margin}	measured from tracking logs	metres; safety margin
$\alpha, \beta, \kappa, \rho, \eta$	tuned on validation only	dimensionless utility weights

Baseline interfaces and supported goals. VLFM and PONI supply semantic-navigation interfaces for the matched UAV comparison. FAEL is connected through a UAV motion interface. The PONI adaptation preserves its category policy on the shared-category subset and lifts selected frontiers to feasible 3D viewpoints with the common geometric planner. Unsupported goals are separated from supported-category performance. Category-only results and unrestricted language-navigation results use distinct evaluation subsets. The in-house flow baseline follows Lipman et al. (2023) for its generative objective.

G Implementation and Reproducibility

The implementation configuration comprises encoder dimensions, feature whitening, Gaussian covariance parameterization, $H, h, N, K_{\text{ode}}, M_{\text{roll}}$, utility/loss weights, control interval, motion limits, safety margin, calibration splits, and annotation rules. Numerical settings specify the ODE and projection solvers, stopping tolerances, swept-path collision checking, observation sampling, and rejection behavior. Evaluation records associate the target verifier, instruction subset, scene manifest, resource budget, and code/checkpoint revision with each episode.

H Supplementary Training and Evaluation Protocol

This section specifies the training and evaluation protocol under the mathematical definitions in the main text. The configuration separates physical control parameters, learned-model settings, and evaluation rules.

H.1 Configuration and units

Table 5 fixes the units and the dependency structure of the principal hyperparameters. The control scales are applied before geometric projection, whereas $v_{\max}$, $a_{\max}$, and $\omega_{\max}$ are expressed in physical units. Covariance and probability calibration are fitted only on the disjoint calibration split.

RGB-D frames are time-synchronized with the state estimate and transformed into one body or world frame before map integration. Depth values outside the sensor validity interval are masked, and the same mask is used by the observation encoder and the collision checker. Each episode stores the instruction string, target instance identifier, initial pose, random seed, termination reason, control timestamps, discovery verifier output, collision events, and per-step inference time. A target is counted once, at its first verified discovery; later views cannot increase y_{t+k}.

H.2 Calibration and paired evaluation algorithm

Algorithm 2 separates model fitting, covariance/probability calibration, and held-out evaluation. It also specifies paired bootstrap intervals for the table metrics, avoiding a change of estimator between methods.

Algorithm 2 Disjoint calibration and paired benchmark evaluation

Require: Scene splits $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{cal}}, \mathcal{D}_{\text{test}}$; fixed seeds $\mathcal{S}$; method set $\mathcal{A}$
1: Fit representation, world model, flow field, and prediction heads on $\mathcal{D}_{\text{train}}$
2: Freeze all non-calibration parameters; fit only designated covariance/probability calibration parameters on $\mathcal{D}_{\text{cal}}$ using Eq. 20
3: **for** each seed $s \in \mathcal{S}$ and each method $a \in \mathcal{A}$ **do**
4: Reset the scene memory and initialize the same pose, instruction, and target instance
5: Run the controller until discovery, collision, timeout, or battery termination
6: Save per-episode events, path lengths, times, coverage volumes, and inference latency
7: **end for**
8: Compute SR, SPL, failure-penalized Time, Coverage, and Coll. from the same episode index set
9: **for** $b = 1, \ldots, B_{\text{boot}}$ **do**
10: Resample complete paired episodes with replacement within each test suite
11: Recompute every metric and each method difference on the resampled indices
12: **end for**
13: Report point estimates and percentile or BCa intervals, plus simulator/hardware and suite breakdowns

The bootstrap unit is a complete paired episode, rather than an individual timestep. Early termination retains the episode's failure-penalized time. Exclusion rules are fixed before method identities are unblinded. Native configurations that change sensors, language inference, or the controller are separate from the matched adaptation.

H.3 Additional derivations

For conditional first-discovery hazards, let $S_k = \prod_{j=1}^{k}(1 - \xi_{t+j})$ be survival through step k. The probability of first discovery at step k is $y_{t+k} = \xi_{t+k} S_{k-1}$, so

$$\sum_{k=1}^{H} y_{t+k} = 1 - S_H \leq 1. \tag{24}$$

This identity explains why the product in Eq. 4 is required: replacing it by ξ_{t+k} would count the same target at every future step. For segment collision events C_k, Boole's inequality gives

$$\Pr\left(\bigcup_{k=1}^{H} C_k \mid \mathcal{H}_t, \boldsymbol{A}\right) \leq \sum_{k=1}^{H} \Pr(C_k \mid \mathcal{H}_t, \boldsymbol{A}), \tag{25}$$

which is the basis of the budget proxy in Eq. 15; independence is not assumed.

The lower-triangular rollout in Eq. 9 also gives a direct path-length diagnostic. Let $\Delta \boldsymbol{x}_k = \Delta t \boldsymbol{a}_{t+k}$ be the per-interval state increment. The piecewise-linear path estimate is

$$\widehat{L}(\boldsymbol{A}) = \sum_{k=0}^{H-1} \|\Delta t \boldsymbol{v}_{t+k}\|_2, \tag{26}$$

where yaw increments are excluded from the translational length. This quantity is logged for efficiency analysis but is not substituted for the evaluator-only shortest path l_e in SPL.

H.4 Evaluation records

The evaluation record associates each result with a model/checkpoint identifier, software revision, hardware or simulator version, sensor and control rates, valid depth range, verifier thresholds, collision-checking settings, scene and episode seeds, timeout and battery limits, inference time, candidate rejections, fallback events, and exclusions. This record connects aggregate metrics to complete episode outcomes.

I Extended Evaluation and Diagnostic Experiments

The extended evaluation separates component attribution, computation, generalization, and predictive reliability. Each comparison holds the episode manifest, target verifier, sensing interface, con-

troller, and stopping rules fixed unless that factor is the explicit intervention. The metrics retain the definitions in equation 16; diagnostic IG uses the same frozen evaluator across variants.

I.1 Replication and comparison protocol

The evaluation protocol uses disjoint training, validation, calibration, and test scenes and five independent training seeds for learned variants. Within each suite, methods are paired by scene, initial pose, target, instruction, and environment seed. Simulator suites and hardware trials are separate. Hierarchical interval estimation resamples training seeds and scene clusters, then complete paired episodes within each selected cluster. The interval level is 95%. For fixed published checkpoints, episode variation is distinguished from training-seed variation. Configuration selection uses validation data only.

The analysis unit is a complete episode, including its failure-penalized time. Seed, scene, and episode counts define distinct replication levels. The prespecified primary contrasts use paired randomization with Holm correction across contrasts; other comparisons are descriptive. The point estimates in the tables do not by themselves establish statistical significance.

I.2 Factorial and objective ablations

The 2×2 comparison in Table 6 separates language conditioning from predictive ranking. With language disabled, the planning modules receive a null embedding while the common discovery verifier retains the requested target. With predictive ranking disabled, the same encoder and generator are retained but candidates are scored from current mapped evidence. The world-model representation remains present in all four variants, so this intervention differs from removing the world-model module in Table 2. Each variant uses the same demonstrations and optimization budget.

For a scalar metric m, the interaction contrast is $[m_{11} - m_{10}] - [m_{01} - m_{00}]$, where the first and second indices indicate language and prediction. Paired resampling preserves the dependence between the four conditions. The contrast tests whether the joint contribution exceeds the sum of isolated contributions under matched training data and generator capacity.

Table 6: Factorial comparison of language conditioning and multi-step predictive scoring. The observation encoder, generator, safety checks, and training data are shared. Lang.: instruction conditioning; Pred.: predictive ranking; IG: common-evaluator information gain.

Lang.	Pred.	SR (%)↑	SPL↑	Time (s)↓	Coll. (%)↓	IG↑
Off	Off	77.0	0.44	163.0	7.5	0.18
Off	On	83.0	0.51	149.0	6.5	0.23
On	Off	84.0	0.54	145.0	6.0	0.21
On	On	92.5	0.68	121.0	4.0	0.34

Table 7 separates the information-seeking reward from the residual-uncertainty penalty. Setting $\kappa = 0$ removes the gain reward, whereas $\eta = 0$ removes the penalty on posterior uncertainty; both interventions retain covariance propagation. The joint setting $\kappa = \eta = 0$ therefore differs from removing covariance modeling entirely. The alignment ablation sets $\lambda_{\text{align}} = 0$. The calibration ablation omits only the disjoint calibration stage while retaining the trained representation, control policy, and runtime verifier. Collision outcomes, discovery, and diagnostic IG are considered jointly when attributing effects.

I.3 Sampler controls under matched computation

Table 8 compares motion primitives, a conditional diffusion sampler, and the conditional flow sampler. All three use the same control parameterization, demonstrations where applicable, instruction and belief inputs, geometric verification, world model, and utility. The comparison changes only the proposal mechanism. The diffusion and flow comparison controls trainable parameter counts and optimization steps.

Two budget controls address different confounders. The fixed-count setting requests $N = 64$ proposals and records total planning latency. The fixed-time setting imposes a 100 ms wall-clock plan-

Table 7: Targeted objective and training ablations. All variants retain a covariance representation and the same geometric verifier. Removing calibration changes only the designated calibration stage.

Variant	SR (%)↑	SPL↑	Time (s)↓	Coll. (%)↓	IG↑
No gain reward ($\kappa = 0$)	89.0	0.63	129.0	4.5	0.27
No uncertainty penalty ($\eta = 0$)	91.0	0.66	124.0	5.0	0.35
Neither term ($\kappa = \eta = 0$)	88.0	0.60	133.0	5.5	0.24
No alignment loss	89.5	0.64	129.0	4.5	0.30
No calibration stage	90.5	0.65	126.0	6.0	0.33
Full AirWorld	92.5	0.68	121.0	4.0	0.34

ning deadline, including generation, projection, rollout, and ranking; candidate counts may differ. Only completed, verified candidates are scored; the common fallback applies if none is available at the deadline. Synchronized accelerator timing includes transfers and solver overhead under fixed precision, batch size, and warm-up settings. A faster generator is useful only if its additional candidates improve decisions within this complete budget.

Table 8: Sampler comparison under a common proposal count and a common online planning deadline. Proposals is the mean number of completed candidates before geometric rejection; P95 denotes the 95th percentile of total planning latency.

Sampler / budget	Proposals	SR (%)↑	Time (s)↓	P95 (ms)↓
Primitives / 64 proposals	64	89.0	135.0	82.0
Diffusion / 64 proposals	64	90.0	129.0	145.0
Flow / 64 proposals	64	92.5	121.0	98.0
Primitives / 100 ms	104	90.0	130.0	99.0
Diffusion / 100 ms	40	88.5	134.0	100.0
Flow / 100 ms	66	92.5	121.0	100.0

I.4 Horizon and proposal-count sensitivity

The sweeps in Table 9 vary one quantity at a time around $(H, N) = (16, 64)$. The executed prefix, rollout count per candidate, flow-integration steps, motion limits, and training data are fixed. Horizon-specific generators use a common architecture and matched training budgets; proposal-count comparisons reuse the same checkpoint. Total latency and collision rate accompany discovery and time to expose the computation cost of a longer rollout. The sweep is a validation analysis; selecting a configuration from test-set peaks is excluded.

Table 9: Sensitivity to prediction horizon H and proposal count N. The executed prefix is $h = 4$, the control interval is $\Delta t = 0.10$ s, and all other settings are fixed.

H	N	SR (%)↑	Time (s)↓	Coll. (%)↓	P95 (ms)↓
4	64	84.5	144.0	5.0	42.0
8	64	89.0	130.0	4.5	60.0
16	64	92.5	121.0	4.0	98.0
24	64	93.0	124.0	4.5	141.0
16	16	88.0	135.0	5.0	46.0
16	32	90.5	128.0	4.5	66.0
16	128	93.0	123.0	4.0	171.0

I.5 Scene and language generalization

Table 10 separates scene transfer from instruction complexity. The scene suites contain buildings or layouts excluded from training, validation, and calibration. The language subsets progress from category-only goals to attributes, spatial relations, and held-out object–relation compositions. Partitioning by target instance and relation graph precedes paraphrase generation, preventing lexical

rewording from placing the same semantic target in both training and test sets. All methods receive the same instruction and sensing budget.

The category-only and relational subsets use separate episode manifests. Category-trained baselines are compared on supported categories; an added language parser is an explicit adaptation. Verifier agreement and target-recognition errors are distinguished from planning outcomes. Per-suite results and instruction subsets remain separate from pooled matched-input results.

Table 10: Generalization across disjoint scene suites and instruction subsets. SR columns compare the matched UAV adaptations; SPL refers to AirWorld. These diagnostic sets have separate episode manifests and are not used to reconstruct the pooled main table.

Evaluation subset	STEM SR	AirHunt SR	AirWorld SR	AirWorld SPL
HM3D unseen buildings	80.0	88.0	92.0	0.66
Isaac unseen layouts	81.5	87.5	91.5	0.64
AirSim cluttered scenes	77.0	84.5	88.5	0.60
Category-only goals	85.0	92.0	94.0	0.70
Attribute-based goals	80.0	88.5	92.0	0.66
Spatial-relation goals	74.0	86.0	90.0	0.63
Held-out compositions	69.5	81.5	87.0	0.57

I.6 Observation and state-estimation robustness

Table 11 uses the same nominal scenes and episode seeds under controlled input perturbations. Depth dropout independently masks valid depth pixels at the stated rate; the encoder and map receive the same mask. Position noise is zero-mean Gaussian with the stated standard deviation per axis, and yaw noise is Gaussian with the stated angular standard deviation. Perturbations affect the estimator output used for planning; collision evaluation retains the simulator state. Noise realizations are paired across methods; thresholds and calibration parameters remain frozen.

The combined condition applies depth dropout and pose noise simultaneously. Table 11 reports the fraction of decisions using a verified fallback alongside SR and collision frequency. Increased fallback use can reflect conservative rejection and is interpreted together with completed search and elapsed time. Perturbations that intentionally invalidate collision checking are restricted to simulation.

Table 11: Robustness to missing depth and state-estimation perturbations. SR and Coll. are episode percentages; fallback is the percentage of planning decisions that execute a verified fallback. The current-map flow baseline shares the observation backbone and verifier.

Perturbation	Flow SR↑	AirWorld SR↑	Coll.↓	Fallback
Nominal observations	81.0	92.5	4.0	2.4
Depth dropout 10%	78.5	90.5	4.5	3.6
Depth dropout 30%	73.0	86.0	5.5	6.8
Pose noise 0.05 m / 1 deg	78.0	90.0	5.0	4.2
Pose noise 0.10 m / 2 deg	72.5	84.5	7.0	8.1
Dropout 30% + 0.10 m / 2 deg	66.0	79.5	8.5	12.6

I.7 Prediction quality and uncertainty calibration

The predictive diagnostics in Table 12 evaluate frozen checkpoints on held-out trajectories at steps $k \in \{1, 4, 8, 16\}$. The transition model is rolled out without assimilating future ground-truth observations. Observation-feature RMSE uses the fixed whitening transform fitted on training data. Discovery Brier scores use first-discovery labels and risk Brier scores use segment-collision labels; all are masked after episode termination. Calibrated and uncalibrated predictions use the same trajectories.

For the observation predictive mean $\hat{\boldsymbol{o}}$ and covariance $\boldsymbol{\Omega}$, the nominal 90% region is the ellipsoid whose squared Mahalanobis radius is the 0.9 quantile of $\chi^2_{d_o}$. Cov90 is the fraction of held-out obser-

vation features inside this region; calibration targets 90% rather than maximizing coverage. Feature scaling and covariance eigenvalue floors are fixed across models. For ranking evaluation, repeated simulator branches from the same hidden state estimate realized candidate utility, and within-state Spearman correlation compares that estimate with predicted utility. The branches are evaluator-only; Monte Carlo variation in their utility estimates is distinct from model prediction error.

Table 12: Open-loop prediction and calibration diagnostics. RMSE is computed on whitened observation features; Disc. and Risk are Brier scores. Cov90 is empirical coverage of a nominal 90% predictive observation region. Rho is Spearman correlation between predicted and simulator-estimated candidate utility.

Model / step k	RMSE↓	Disc.↓	Risk↓	Cov90 (%)	Rho↑
Calibrated / 1	0.18	0.063	0.018	90.5	0.86
Calibrated / 4	0.26	0.078	0.023	90.0	0.83
Calibrated / 8	0.35	0.096	0.031	89.0	0.79
Calibrated / 16	0.49	0.123	0.044	87.5	0.72
Uncalibrated / 16	0.49	0.158	0.069	74.0	0.64

Reliability diagrams, risk calibration by scene and visibility stratum, and predictive residuals at each horizon complement the aggregate scores. Calibration quality does not itself establish the conditional risk bound in equation 18; high-error strata and rejected candidates require separate analysis. Likewise, correlation with realized utility evaluates ranking rather than establishing that imagined observations are accurate in every scene.

I.8 Runtime profile and failure analysis

Table 13 decomposes online planning into map integration, belief encoding, candidate generation, verification, predictive rollout, and selection. Synchronized timers include memory transfers and fallback selection on the same device. End-to-end latency is measured separately. For sequential execution, component means sum to the total mean; component percentiles do not sum to the end-to-end percentile. The configuration has a mean total latency of 70 ms and P95 latency of 98 ms; predictive rollout is the largest mean component at 26 ms.

Table 13: Online latency profile for $H = 16$, $N = 64$, and $M_{\text{roll}} = 8$. Mean and P95 are milliseconds per planning decision. Total refers to end-to-end timing; component percentiles are not additive.

Component	Mean (ms)	P95 (ms)
Map integration	6.0	9.0
Belief encoding	8.0	12.0
Flow generation	12.0	18.0
Projection and verification	11.0	18.0
Predictive rollout	26.0	41.0
Ranking and fallback selection	7.0	11.0
Total	70.0	98.0

The controller executes at most $h = 4$ intervals of $\Delta t = 0.10$ s between replans. The resulting 0.4 s execution window differs from the 100 ms planner deadline; perception and collision monitoring continue at every control interval. Failure attribution uses mutually exclusive primary causes: recognition or instruction mismatch, insufficient exploration before timeout, no feasible candidate, tracking or collision event, and battery termination. Counts over all episodes and shares among failures have different denominators; ambiguous causes require synchronized video and control records.

I.9 Hardware transfer and data-efficiency controls

Hardware transfer is evaluated separately from simulator performance because sensing, tracking error, and operator intervention alter the execution conditions. The deployment protocol fixes the checkpoint before flight and pairs instructions and starting poses where physically feasible. Platform

mass, sensors, controller frequency, tracking error, intervention count, and flight exposure define the operating conditions. Autonomous fallback and operator intervention are distinct outcomes, and a fixed exclusion rule governs interrupted trials.

Data-efficiency comparisons use nested 25%, 50%, and 100% subsets of a scene-disjoint demonstration pool, with fixed validation/test manifests and training seeds. Matching gradient-update counts and data exposure distinguishes sample reuse from additional supervision. Demonstration provenance, region–instruction annotation counts, pretraining sources, and total training compute define the supervision budget. These controls separate the planning objective from differences in training exposure.