WORLD-VALUE-ACTION MODEL WITH FUTURE TARGET GROUNDING FOR LANGUAGE-GOAL AERIAL OBJECT NAVIGATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Language-goal aerial object navigation is a formidable task that requires an unmanned aerial vehicle (UAV) to find a language-specified object without knowing its location in advance. World-action models (WAMs), capable of coupling the prediction of future observations (video frames) with action generation, serve as one possible approach to aerial object navigation. However, directly deploying WAMs to this task is non-trivial due to two key challenges: (i) it is challenging to ensure that the predicted future observations contain information about the target object; and (ii) it is challenging to evaluate how the predicted future observations contribute to target search given the limited supervision provided by sparse navigation rewards. To address these challenges, we present FG-WVAM, the first World-Value-Action Model for language-goal aerial object navigation, to the best of our knowledge. Specifically, FG-WVAM employs Future Target Grounding to jointly predict future observations and heatmaps indicating the target's expected location. During training, a frozen Grounding DINO teacher provides target annotations from ground-truth future frames, teaching the model to predict both scene changes and the target's location in future observations. Furthermore, a Future Value Expert evaluates the predicted visual and target-location representations for their contribution to the search goal, using dense supervision derived from successful expert trajectories and unsuccessful search rollouts. During inference, FG-WVAM generates multiple predictions in latent space and ranks them by their value scores. The Action Expert then uses the representations associated with the selected prediction through internal cross-attention to predict the corresponding actions. Experiments on UAV-ON and AerialDojo demonstrate improved navigation success, including in unseen scenes and for unseen target categories.

1 INTRODUCTION

Language-goal aerial object navigation requires an unmanned aerial vehicle (UAV) to find a language-specified object and stop at a valid position without prior knowledge of the target's location beforehand (Xiao et al., 2025; Zhang et al., 2026a). A central difficulty is to decide the direcitons and paths of searching given that the target objects remain unseen — limited observations leave the outcomes of different actions uncertain, and three-dimensional flight continually changes the visible scene.

On the other hand, world-action models (WAMs) is designed to jointly learn visual prediction and action generation, incorporating future-observation prediction into navigation policies (Ye et al., 2026). WAMs have shown promise in learning control policies through video prediction (Bi et al., 2026; Liao et al., 2026). For example, ANWM predicts future views along candidate flight trajectories to inform trajectory selection (Zhang et al., 2026b), while WorldFly jointly models future video and navigation actions (Zheng et al., 2026). AeroAct uses future visual prediction as training supervision for language-conditioned aerial navigation and demonstrates real-world quadrotor flight (Zhang et al., 2026c). These studies show that visual prediction can support flight-trajectory selection and provide supervision for learning navigation policies. As such, WAMs serve as one promising solution for language-goal aerial object navigation.

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.

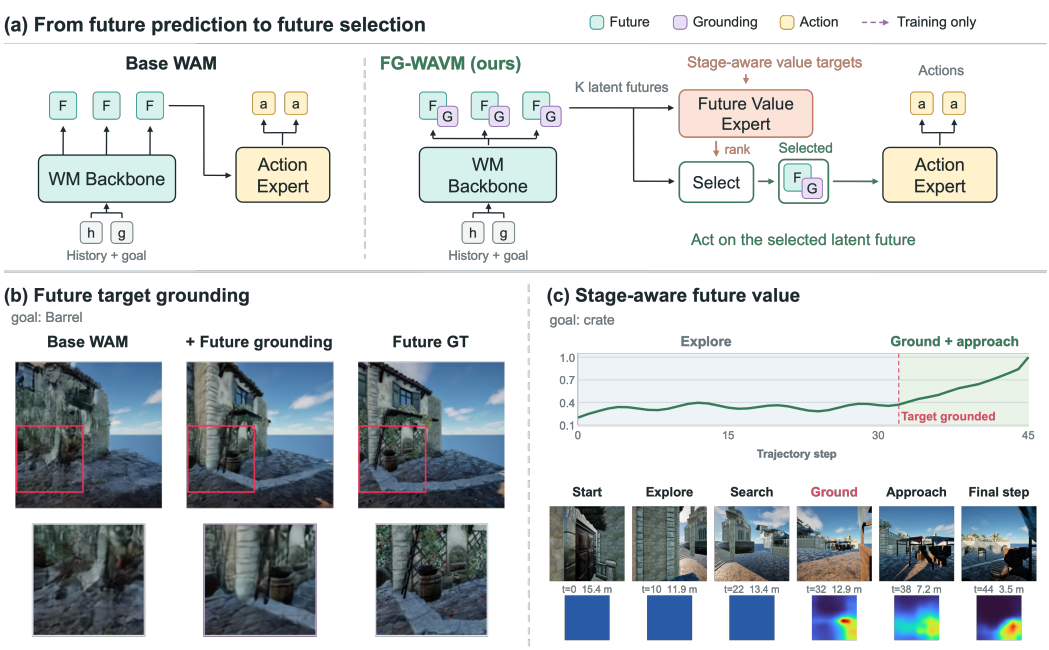


Figure 1: Target-consistent future prediction and value-guided selection. (a) Base WAM and FG-WVAM, with training supervision shown by dashed arrows. (b) Future predictions with and without target-grounding supervision. (c) Stage-aware value across exploration and approach.

However, directly deploying WAMs to language-goal aerial object navigation faces two key challenges. First, future predictions may overlook important information and training objectives centered on overall scene quality and temporal consistency may give insufficient emphasis to changes in the target's visibility and position in future observations. Errors in predicting these changes can affect the UAV's decisions about where to search next. Figure 1(b) illustrates differences in target-region predictions. Therefore it is challenging to ensure that the predicted future observations contain information about the target object. Second, sparse task feedback makes it difficult to evaluate how predicted observations contribute to finding the target object. Long searches require a sequence of movement decisions, but task feedback provides limited information about whether each intermediate step helps find the target. Consequently, it is challenging to evaluate how the predicted future observations contribute to target search given the limited supervision provided by sparse navigation rewards.

To address these challenges, we propose FG-WVAM, the first World-Value-Action Model for language-goal aerial object navigation. Figure 1(a) summarizes how the model connects future prediction, value evaluation, and action generation. To preserve target information, we introduce **Future Target Grounding**, which jointly predicts future observations and heatmaps of the target's location. A frozen Grounding DINO teacher (Liu et al., 2023) provides annotations from ground-truth future frames to supervise heatmap prediction and give target regions greater weight in the visual-generation loss (Figure 1(b)). To learn task value despite sparse feedback, we train a **Future Value Expert** using dense labels from successful expert trajectories and unsuccessful search rollouts collected in training scenes. These labels describe exploration before target discovery and distance-based progress afterward (Figure 1(c)). The expert evaluates the predicted visual and target-location representations for their contribution to finding the target. At inference stage, the model generates and scores multiple predictions in a batch; its Action Expert uses the selected visual, grounding, and value hidden states through internal cross-attention to generate actions. Deployment requires neither Grounding DINO nor decoding future RGB frames or heatmaps. Our contributions are as follows:

1. We propose FG-WVAM, the first World-Value-Action Model for language-goal aerial object navigation, which integrates future prediction, task-value evaluation, and action generation in a unified architecture.

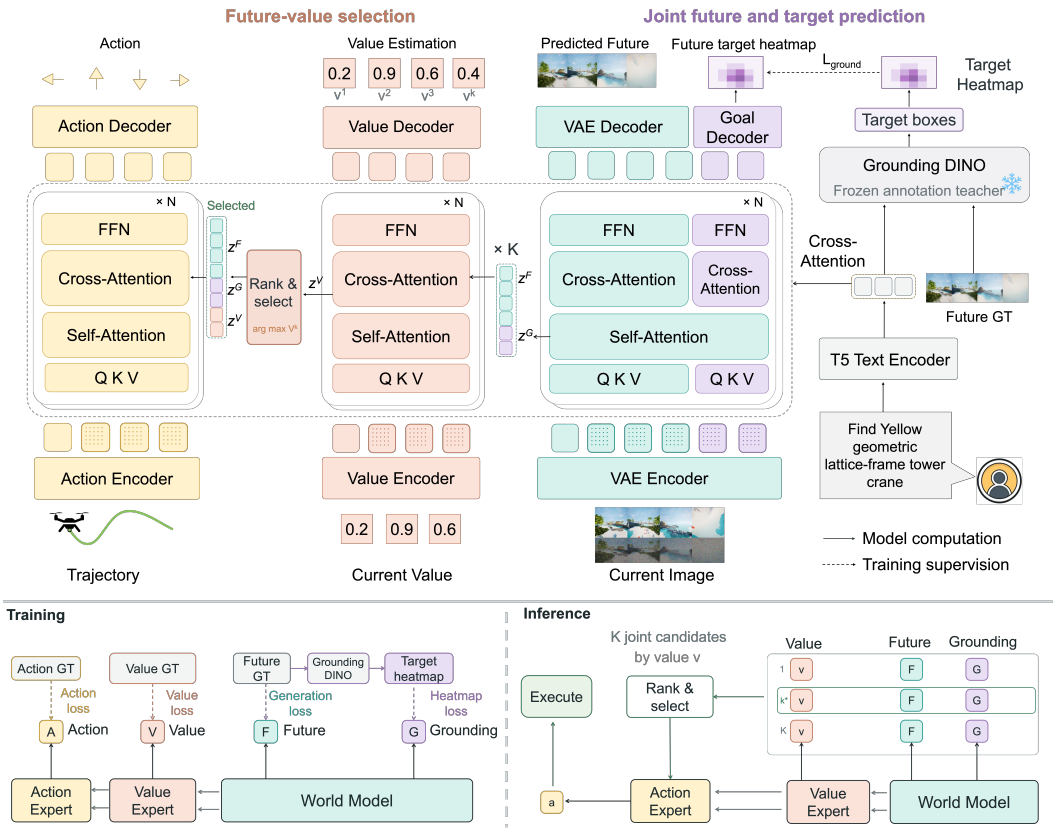


Figure 2: FG-WVAM architecture and training/inference workflows. **Top:** Three coupled experts connect future modeling, value estimation, and action generation. The world expert produces visual-future and grounding representations; the Future Value Expert evaluates candidate representations, and the Action Expert uses the selected visual, grounding, and value states through cross-attention. The VAE Decoder and Goal Decoder map the predicted representations to future RGB frames and target heatmaps. **Bottom left (training):** Frozen Grounding DINO supplies target annotations from ground-truth future frames and the language goal. These annotations supervise future grounding and define target-region weights for visual generation; trajectory annotations supervise value and action learning. **Bottom right (inference):** Candidate evaluation and selection take place inside the joint model, whose action output is used for execution. Future RGB and heatmap decoding are optional, and Grounding DINO is not used.

2. We introduce Future Target Grounding to jointly predict future observations and target heatmaps, with target-weighted generation strengthening target information. Future Value Expert is presented to learn from dense stage-aware supervision to score the predicted visual and grounding representations and guide action generation.
3. On UAV-ON and AerialDoJo, FG-WVAM achieves the best SR, OSR, SPL, and DTS among the compared methods, improving SR by **12.3 and 10.2 percentage points**, respectively, over the strongest evaluated aerial WAM baseline. Ablations and unseen-scene evaluations further support its effectiveness and generalization.

2 Related Work

Aerial object-goal navigation. UAV-ON formalizes open-world aerial ObjectNav with semantic target descriptions and long-range search (Xiao et al., 2025). AION separates exploration from goal reaching in indoor aerial ObjectNav (Yan et al., 2026); APEX combines spatio-semantic memory,

target grounding, and asynchronous control (Zhang et al., 2026a), while OctMem-Agent builds hierarchical scene memory for exploration on UAV-ON (Zhou et al., 2026). AirHunt connects vision-language semantics to continuous aerial planning (Chen et al., 2026), and USS-Nav organizes spatial and semantic evidence in an incremental scene graph (Gai et al., 2026). These approaches improve goal-conditioned perception, memory, and exploration; our work studies how a predictive aerial policy can represent target-related changes in its imagined futures and evaluate those same futures before acting.

Aerial world models and world-action models. OpenFly-Agent provides a keyframe-aware aerial vision-language navigation policy (Gao et al., 2026). ANWM predicts action-conditioned future views to rank candidate UAV trajectories (Zhang et al., 2026b); WorldFly couples future-video and action generation (Zheng et al., 2026), and WorldVLN autoregressively predicts latent transitions for waypoint control (Zhao et al., 2026). These works establish future prediction as a useful representation for aerial navigation, although their original evaluations target image-goal navigation, instruction following, or flight control rather than open-world aerial ObjectNav.

3 Problem Formulation

Given a language goal g, a UAV searches for the target, approaches it, and stops at a valid position under partial observability. With V synchronized cameras and a history length L, its input and policy are

$$\mathcal{H}_t = \{\mathbf{o}^{1:V}_{t-L+1:t}, \mathbf{p}_{t-L+1:t}\}, \qquad \mathbf{a}_{t:t+T_a-1} \sim \pi(\cdot \mid \mathcal{H}_t, g), \tag{1}$$

where $\mathbf{o}$ denotes RGB observations, $\mathbf{p}$ the UAV pose, and T_a the action-chunk length. The discrete action set includes forward, ascend, descend, left/right rotation, and stop. The UAV executes a short action prefix and replans from new observations. We evaluate task completion using each benchmark's success protocol; the model predicts T_f future steps to support its decisions. The UAV uses eight discrete actions: forward, lateral left/right, up/down, yaw left/right, and stop, with a translation step of $1\,\mathrm{m}$ and a yaw increment of 15°.

4 Method

FG-WVAM couples a world expert, a Future Value Expert, and an Action Expert, conditioned on $\mathcal{H}_t$ and the language embedding $\mathbf{e}_g$. The model jointly represents visual futures, future target grounding, task values, and actions. The Future Value Expert scores generated visual and grounding latents, and value-guided selection determines the internal representations used by the Action Expert through cross-attention. Joint modeling retains this internal dependency between future evaluation and action generation. The visual and grounding representations support auxiliary decoding into RGB frames and heatmaps, but these decoded outputs are not required for action-only deployment.

4.1 Future Target Grounding

A video VAE encodes multi-view future frames into $\mathbf{Z}^F_0$. We learn visual dynamics with conditional flow matching, whose target velocity is $\mathbf{u}^*_F = \boldsymbol{\epsilon}^F - \mathbf{Z}^F_0$ for Gaussian noise $\boldsymbol{\epsilon}^F$. The world expert combines within-view and cross-view attention and jointly predicts visual latents and grounding tokens. The Goal Decoder maps the grounding tokens to future heatmaps $\widehat{H}^v_s$ for step s and view v.

During training, a frozen Grounding DINO teacher obtains target boxes from future ground-truth frames and g. These boxes define soft heatmaps $H^{v,*}_s$. Normalizing over views and locations gives target and predicted spatial distributions P^*_s and $\widehat{P}_s$. We supervise future grounding by

$$\mathcal{L}_{\text{ground}} = \frac{\sum_s m_s D_{\mathrm{KL}}(P^*_s \,\|\, \widehat{P}_s)}{\sum_s m_s + \varepsilon}, \tag{2}$$

where m_s indicates a reliable target annotation. This objective teaches the model where the language target will appear across future views.

To align visual generation with the same target, we downsample H^* to latent resolution and weight the flow-matching error:

$$\mathcal{L}_{\text{video}}^{\text{tw}} = \mathbb{E}\left[\frac{\sum_i w_i \|\widehat{\mathbf{u}}_{F,i} - \mathbf{u}_{F,i}^*\|_2^2}{\sum_i w_i}\right], \qquad w_i = 1 + \lambda_h \widetilde{H}_i^*. \tag{3}$$

Here i indexes future time, view, and spatial location. Grounding thus shapes both explicit target prediction and visual-future generation. The teacher and ground-truth future frames are used only for training; heatmap construction and flow-matching details appear in Appendix A.

4.2 Future Value Expert

The Future Value Expert cross-attends to predicted visual and grounding latents to estimate a task-value sequence. We construct dense supervision from successful expert trajectories and unsuccessful rollouts collected in training scenes. Grounding separates target search from approach: before discovery, labels reflect inter-step motion and target-evidence changes; afterward, they measure progress toward the located target.

Let t_g be the first reliably grounded step, $e_t \in [0,1]$ the normalized exploration score, and d_t the distance to the target. With a stage boundary $\beta = 0.25$, we define

$$y_t^* = \begin{cases} \beta e_t, & t < t_g, \\ \beta + (1-\beta)\left[1 - \dfrac{d_t}{\max(d_{t_g}, \varepsilon_d)}\right]_0^1, & t \geq t_g, \end{cases} \tag{4}$$

where $[x]_0^1 = \min(1, \max(0, x))$ and $\varepsilon_d > 0$ prevents division by zero. Exploration labels lie in $[0, 0.25]$, while approach labels lie in $[0.25, 1]$, preserving stage ordering without requiring monotonic exploration scores. The expert learns these targets with conditional flow matching and evaluates generated latent futures at both training and inference. Annotation details and expert-training equations are given in Appendices A.5 and A.6.

4.3 Joint Inference and Training

Value-guided candidate selection. At each decision step, the coupled model generates K candidate latent futures in a single batch. For candidate k, the Future Value Expert produces a value representation $\widehat{\mathbf{Z}}^{V,k}$ from the corresponding visual-future and grounding representations. A value decoder maps this representation to a scalar score,

$$S_k = D_V\left(\widehat{\mathbf{Z}}^{V,k}\right), \qquad k^* = \arg\max_{1 \leq k \leq K} S_k. \tag{5}$$

The scalar score is used only for candidate ranking. The Action Expert instead cross-attends to the selected vector representations $\widehat{\mathbf{Z}}^{F,k^*}$, $\widehat{\mathbf{Z}}^{G,k^*}$, and $\widehat{\mathbf{Z}}^{V,k^*}$ to predict an action chunk.

All K candidates are generated and scored once per decision step; no iterative refinement, elite resampling, or repeated update of the future-noise distribution is performed.

Training objective. We jointly optimize target-weighted generation, future grounding, value learning, and action behavior cloning:

$$\mathcal{L} = \mathcal{L}_{\text{video}}^{\text{tw}} + \lambda_g \mathcal{L}_{\text{ground}} + \lambda_v \mathcal{L}_{\text{value}} + \lambda_a \mathcal{L}_{\text{action}}. \tag{6}$$

Training first learns visual dynamics, then introduces future grounding and target-weighted generation, and finally trains the value and action experts. Appendix A.6 provides the expert-training objectives and conditioning details.

5 Experiments

5.1 Experimental Setup

Benchmarks and evaluation settings. We evaluate FG-WVAM on UAV-ON and AerialDoJo (Feng et al., 2026) for language-conditioned aerial object-goal search. In the main evaluation,

Table 1: Same-scene comparison. SR, OSR, and SPL are in percentages (%); DTS is in meters (m). ± denotes std over three runs.

	UAV-ON				AerialDoJo			
Method	SR ↑	OSR ↑	SPL ↑	DTS ↓	SR ↑	OSR ↑	SPL ↑	DTS ↓
APEX	12.36	19.10	9.1	55.6				
GPT5.6sol	10.2±2.2	16.5±2.7	5.8	47.2	6.9±1.9	12.2±2.0	4.4	16.1
OpenFly	11.7±2.4	20.1±3.5	7.2	61.2	8.4±2.1	15.0±2.5	7.7	14.8
ANWM	14.6±2.3	18.4±2.6	6.9	43.4	9.2±3.1	14.2±3.9	9.2	18.1
AeroAct	25.0±2.7	33.2±2.9	15.6	36.5	18.6±2.5	29.1±2.8	13.6	13.4
FG-WVAM (ours)	**37.3±2.5**	**42.9±3.0**	**23.7**	**27.3**	**28.8±1.8**	**34.4±1.3**	**17.8**	**11.8**

training and testing use the same scenes, while test targets include different object instances and categories unseen during training. We additionally evaluate generalization to unseen scenes on UAV-ON. We follow the respective benchmark protocols for Success Rate (SR) and Oracle Success Rate (OSR), with SR/OSR distance thresholds of 20 m on UAV-ON and 3 m on AerialDoJo. Main and unseen-scene results are reported as mean ± standard deviation over three runs. All navigation results are expressed as percentages; differences between success rates are reported in percentage points.

Compared methods. We compare with GPT-5.6, OpenFly (Gao et al., 2026), ANWM (Zhang et al., 2026b), and AeroAct (Zhang et al., 2026c). For a consistent comparison, we retrain these baselines on the corresponding training split of each benchmark using four synchronized camera views and the same eight-action space as FG-WVAM, and evaluate all models under the same protocol. AeroAct is the strongest evaluated baseline and serves as our primary reference. We additionally report published APEX results (Zhang et al., 2026a), marked with †.

Implementation settings. We use $V = 4$ synchronized 256×256 camera views, a history length of $L = 2$, and a prediction horizon of $T_f = 10$. The action chunk has length $T_a = 2$; only the first action is executed before replanning. The eight-action interface uses 1 m translations and 15° yaw increments. We set $(\lambda_h, \lambda_g, \lambda_v, \lambda_a) = (3.0, 1.0, 0.3, 1.0)$, with unit weight on $\mathcal{L}^{\text{tw}}_{\text{video}}$, and use $\beta = 0.25$, $\varepsilon = 10^{-8}$, and $\varepsilon_d = 10^{-3}$. For $\widetilde{H}^*_i \in [0, 1]$, the spatial weights satisfy $w_i \in [1, 4]$. Inference uses 10 FlowMatch Euler steps for the world and value branches and a single batched selection over $K = 3$ candidates; we test $K \in \{1, 2, 3, 4\}$. Results from our repeated evaluations are reported as mean ± standard deviation over three runs. Architecture, training, annotation, and further evaluation settings are provided in Appendix A.1.

5.2 Comparison on Two Aerial Search Benchmarks

Table 1 shows that FG-WVAM achieves the best results across all four metrics among the compared methods on both benchmarks. Compared with AeroAct, it improves SR and OSR by 12.3 and 9.7 percentage points on UAV-ON, and by 10.2 and 5.3 points on AerialDoJo, respectively. SPL increases by 8.1 and 4.2 percentage points, while DTS decreases from 36.5 to 27.3 and from 13.4 to 11.8 on the two benchmarks, respectively. These gains reflect higher task success, improved success-weighted path efficiency, and smaller final distances to the target.

5.3 Generalization to Unseen UAV-ON Scenes

Unseen UAV-ON scenes. Table 2 reports the unseen-scene evaluation on UAV-ON. FG-WVAM obtains 34.0% SR and 38.7% OSR, compared with 19.8% and 27.8% for AeroAct. The corresponding gains are **14.2** and **10.9** percentage points. Relative to the same-scene evaluation, the SR of FG-WVAM decreases by 3.3 points, while AeroAct decreases by 5.2 points. FG-WVAM thus retains its advantage when evaluation moves beyond the training scenes. This comparison supports generalization across the evaluated scene splits; it does not by itself establish robustness to every type of environmental change.

Table 2: Unseen-scene results on UAV-ON (%). Results from our evaluations are mean ± standard deviation over three runs. †: APEX's Unseen results quoted from the original paper, which does not report standard deviations. Bold and underline indicate the highest and second-highest values in each column.

Method	SR ↑	OSR ↑
GPT5.6sol	7.9 ± 2.7	13.6 ± 3.0
APEX† (Zhang et al., 2026a)	16.13	22.58
OpenFly	8.3 ± 2.9	15.1 ± 3.2
ANWM	11.8 ± 2.8	15.8 ± 3.0
AeroAct	$\underline{19.8 \pm 3.4}$	$\underline{27.8 \pm 3.8}$
FG-WVAM (ours)	$\mathbf{34.0 \pm 2.6}$	$\mathbf{38.7 \pm 2.1}$

Table 3: Unseen-scene results on AerialDoJo (%). Results from our evaluations are mean ± standard deviation over three runs. †: quoted from the original paper, which does not report a standard deviation. Bold and underlined values denote the best and second-best results, respectively.

Method	SR ↑	OSR ↑
GPT5.6sol†	6.1	12.2
OpenFly	21.5 ± 2.5	28.7 ± 2.8
ANWM	27.4 ± 3.0	30.2 ± 2.1
AeroAct	$\underline{32.4 \pm 3.6}$	$\underline{35.1 \pm 2.5}$
FG-WVAM (ours)	$\mathbf{43.2 \pm 2.3}$	$\mathbf{47.7 \pm 2.0}$

Unseen AerialDoJo scenes. Table 3 extends the unseen-scene evaluation to AerialDoJo. FG-WVAM achieves 43.2% SR and 47.7% OSR, outperforming AeroAct, the strongest evaluated baseline, by **10.8** and **12.6** percentage points, respectively. The improvements on both metrics show that the advantage of FG-WVAM extends to scenes not used for training. Together with the UAV-ON results, this comparison supports the generalization of the combined future target grounding and value-guided selection framework across the unseen-scene settings of both benchmarks.

5.4 Component Ablations

Table 4 evaluates the contributions of our main components. Base WAM retains the world and action experts of FG-WVAM but excludes future target grounding, target-weighted generation, and the Future Value Expert with value-guided selection. All variants use the same backbone, observation and action interfaces, and training data. Adding future-heatmap prediction to base WAM raises SR from 25.6% to 32.8% on UAV-ON and from 19.1% to 25.2% on AerialDoJo, yielding gains of **7.2** and **6.1** points. This variant does not use target-weighted generation, so the improvement supports the utility of future-grounding supervision even without that additional objective.

The future-value-only configuration adds the Future Value Expert together with value-guided candidate selection, improving SR by **5.0** points on UAV-ON and **3.8** points on AerialDoJo. These gains demonstrate the benefit of value-guided selection for task completion. The complete FG-WVAM achieves the highest SR, exceeding the grounding-only variant by 4.5 and 3.6 points and the value-only variant by 6.7 and 5.9 points.

5.5 Future-Generation Quality

Table 5 compares visual-future generation using PSNR, SSIM, LPIPS, and FID. FG-WVAM improves all four metrics relative to Base WAM. PSNR increases from 18.71 to 24.46 dB and SSIM from 0.541 to 0.775, indicating improved reconstruction fidelity and structural similarity. LPIPS decreases from 0.288 to 0.094, an absolute reduction of 0.194, while FID decreases from 167.96 to 158.12. Together, these results show improvements in the paired-image metrics and in the distribution-level FID metric under the reported evaluation.

These observations are consistent with the objective of preserving useful visual information in predicted futures. However, the reported whole-image metrics do not isolate target-region fidelity or

Table 4: Component ablations. Heatmap denotes future target-heatmap prediction; Value denotes the Future Value Expert.

Variant	Heatmap	Value	UAV-ON SR (%)	AerialDoJo SR (%)
Base WAM	No	No	25.6	19.1
+ Future grounding	Yes	No	32.8	25.2
+ Future value only	No	Yes	30.6	22.9
FG-WVAM (full)	Yes	Yes	**37.3**	**28.8**

Table 5: Visual-future generation quality. Higher PSNR and SSIM and lower LPIPS and FID are better.

Model	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	FID ↓
Base WAM	18.71	0.541	0.288	167.96
FG-WVAM (ours)	**24.46**	**0.775**	**0.094**	**158.12**

temporal consistency. We therefore treat this comparison as evidence of improved overall generation quality, and use the candidate visualizations below to examine how future grounding and value relate to selection.

5.6 Qualitative Analysis of Future-Value Selection

Figure 3 examines whether value ranking favors futures that provide evidence about the requested target. Each row specifies the target object and compares a selected candidate with a rejected one through their predicted visual futures, grounding responses, and value scores.

Crate search. For the *Crate* goal, the right-rotation candidate (`rotr`) receives a value of 0.328, compared with 0.013 for forward motion. The selected future depicts the target region near the foreground object cluster, and its grounding response concentrates on that region. In the rejected candidate, the strongest grounding response instead appears on the upper building facade, away from the crate. The higher score therefore accompanies a future whose visual and grounding predictions provide spatially aligned evidence about the requested object.

Statue search. For the *Statue* goal, the left-rotation candidate (`rotl`) receives 0.254, while forward motion receives 0.033. The selected future reveals the statue at the left edge of the image, with a localized grounding response at its predicted position. The rejected future is dominated by the nearby wall, and its grounding response spreads across regions that do not identify the statue. Value ranking selects the viewpoint change that exposes target evidence relevant to continuing the search.

Together, these examples show how the Future Value Expert prioritizes task-relevant candidates: futures that depict and ground the requested object receive higher scores. Combined with the SR improvements of the value-only configuration in Table 4, the visualizations support both the contribution of value-guided selection to task completion and its role in selecting target-relevant predicted futures.

5.7 Candidate Count and Inference Efficiency

We vary the candidate count K from 1 to 4 and measure median latency per action request on an NVIDIA A100 GPU. Candidates are packed along the batch dimension and processed together, and value scoring is also batched. Increasing K therefore enlarges the inference batch rather than requiring K separate sequential generation runs.

As shown in Figure 4, increasing K from 1 to 3 improves SR from 32.8% to 37.3% on UAV-ON and from 25.2% to 28.8% on AerialDoJo, with median latency increasing from 0.47 to 0.54 s per action request ($+70$ ms; $+14.9\%$). Increasing K further to 4 yields smaller gains of 0.7 and 0.5 percentage points, respectively, while adding another 70 ms. Thus, $K = 3$ offers a favorable empirical trade-off, whereas $K = 4$ achieves the highest SR among the tested settings. Across this



Figure 3: Future-value ranking for *Crate* (top) and *Statue* (bottom) with a 10-step prediction horizon. Each row shows the target, current observation, and selected (green) versus rejected (red) futures with grounding responses. Higher-scoring futures depict and localize the requested object, illustrating task-relevant selection. Scores are predicted by the value model.

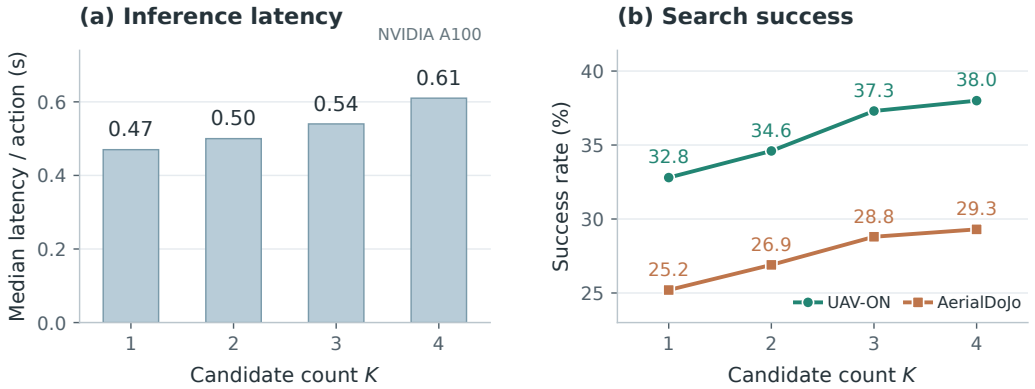


Figure 4: Candidate-count trade-off on NVIDIA A100. (a) Median latency per action request. (b) Success rate on UAV-ON and AerialDoJo. Candidate generation and value scoring use batched computation. Increasing K improves search success with sublinear latency growth over the tested range; gains diminish from $K = 3$ to $K = 4$.

range, quadrupling the candidate count increases measured latency by 29.8%, consistent with the amortization of inference overhead through batched candidate generation and scoring.

6 Conclusion

We presented FG-WVAM, the first World-Value-Action Model for language-goal aerial object navigation. Future Target Grounding jointly predicts future observations and target heatmaps, while target-weighted generation strengthens target information in the predicted representations. The Future Value Expert learns from dense stage-aware supervision to evaluate predicted visual and grounding representations, enabling value-guided selection for action generation within the joint model. Experiments on UAV-ON and AerialDoJo show SR improvements of 12.3 and 10.2 percentage points over the strongest evaluated aerial WAM baseline, respectively, together with improved OSR and SPL and reduced DTS.

AI USE STATEMENT

We used AI-assisted writing tools to improve the clarity, grammar, and readability of the manuscript. All AI-assisted edits were reviewed by the authors, who take full responsibility for the final content of the paper.

REFERENCES

Hongzhe Bi, Hengkai Tan, Shenghao Xie, Zeyuan Wang, Shuhe Huang, Haitian Liu, Ruowen Zhao, Yao Feng, Chendong Xiang, Yinze Rong, Hongyan Zhao, Hanyu Liu, Zhizhong Su, Lei Ma, Hang Su, and Jun Zhu. Motus: A unified latent action world model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35101–35113, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Bi_Motus_A_Unified_Latent_Action_World_Model_CVPR_2026_paper.html.

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. 2026.

Tongtong Feng, Xin Wang, Haoran Hou, Ren Wang, Weiran Wang, Shaokai Zhu, Ziqi Jia, Hao Wang, Yu-Wei Zhan, Zongyuan Wu, Jinghao Cui, and Wenwu Zhu. AerialDojo-200K: A large-scale benchmark suite for open-world aerial object-goal search. *arXiv preprint*, 2026.

Weiqi Gai, Yuman Gao, Yuan Zhou, Yufan Xie, Zhiyang Liu, Yuze Wu, Xin Zhou, Fei Gao, and Zhijun Meng. USS-Nav: Unified spatio-semantic scene graph for lightweight UAV zero-shot object navigation. 2026.

Yunpeng Gao, Chenhui Li, Zhongrui You, Junli Liu, Zhen Li, Pengan Chen, Qizhi Chen, Zhonghan Tang, Liansheng Wang, Penghui Yang, et al. OpenFly: A comprehensive platform for aerial vision-language navigation. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=OKm3w71ymP.

Yue Liao, Pengfei Zhou, Siyuan Huang, Donglin Yang, Shengcong Chen, Yuxin Jiang, Yue Hu, Jingbin Cai, Si Liu, Jianlan Luo, Liliang Chen, Shuicheng Yan, Maoqing Yao, and Guanghui Ren. Genie Envisioner: A unified world foundation platform for robotic manipulation. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://openreview.net/pdf/90e39e28b6fa7a15dda5ba3cbff8972ce0618c2d.pdf.

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. *arXiv preprint arXiv:2303.05499*, 2023.

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*, pp. 13023–13029, 2025. doi: 10.1145/3746027.3758251.

Zichen Yan, Yuchen Hou, Shenao Wang, Yichao Gao, Rui Huang, and Lin Zhao. AION: Aerial indoor object-goal navigation using dual-policy reinforcement learning. 2026.

Weirui Ye, Xinhao Chen, Zhengkai Jiang, Haozhi Xue, Zhiyuan Zhang, Yansong Li, Jiangmiao Pang, and Jifeng Dai. World action models are zero-shot policies. *arXiv preprint arXiv:2602.15922*, 2026.

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. 2026a.

Weichen Zhang, Peizhi Tang, Xin Zeng, Fanhang Man, Shiquan Yu, Zichao Dai, Baining Zhao, Hongjin Chen, Yu Shang, Wei Wu, Chen Gao, Xinlei Chen, Xin Wang, Yong Li, and Wenwu Zhu. Aerial world model for long-horizon visual generation and navigation in 3d space. 2026b.

Xinhong Zhang, Qiyuan Zhu, Yubo Huang, Haolin Chen, Runqing Wang, Yuhao Mo, Zhongxin Chen, Yu Hu, Xinjiang Wang, Jian Sun, and Gang Wang. AeroAct: Action-centered world-action models for language-conditioned quadrotor flight. *arXiv preprint arXiv:2607.14997*, 2026c. doi: 10.48550/arXiv.2607.14997. URL https://arxiv.org/abs/2607.14997.

Baining Zhao, Jiacheng Xu, Weicheng Feng, Xin Zhang, Zhaolu Wang, Haoyang Wang, Shilong Ji, Ziyou Wang, Jianjie Fang, Zhiheng Zheng, Weichen Zhang, Yu Shang, Wei Wu, Chen Gao, Xinlei Chen, and Yong Li. WorldVLN: Autoregressive world action model for aerial vision-language navigation. 2026.

Shengtao Zheng, Kai Li, Weichen Zhang, Yu Meng, Chen Gao, Xinlei Chen, Yong Li, and Xiao-Ping Zhang. WorldFly: A world-model-based vision-language-action model for UAV navigation. 2026.

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*, pp. 21616–21626, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Zhou_Memory-Augmented_Scene_Understanding_and_Exploration_for_Open-World_Aerial_Object-Goal_Navigation_CVPR_2026_paper.html.

A Training and Inference Details

A.1 Implementation Details

Table 6 summarizes the training and inference settings, together with the available corpus-level dataset statistics. The dataset counts describe stored trajectory records and their loader partitions, rather than the number of closed-loop evaluation trials.

Table 6: Training, inference, and dataset configuration.

Parameter	Value	Parameter	Value
Training			
Optimizer	AdamW	LR schedule	Cosine
Video learning rate	10^{-5}	Joint learning rate	3×10^{-5}
Video warmup	5,000 steps	Joint warmup	500 steps
Video training	30k steps	Joint training	30k steps
Batch size	8	Weight decay	10^{-5}
Training precision	BF16	Training GPUs	$4\times$ A100 80GB
Grounding teacher	Grounding DINO-base (frozen; training only)		
Inference			
Camera views	4	RGB resolution	256×256
Prediction horizon	10 steps	Default candidates	$K = 3$
Tested K	1, 2, 3, 4	Candidate processing	Batched
Action chunk	2 actions	Executed prefix	1 action
Replanning	Every step	Selection space	Latent
RGB / heatmap decoding	Optional	DINO at inference	Not used

Dataset records

Corpus	All episodes	Train	Corpus test
UAV-ON	8,872	7,872	1000
AerialDoJo	6,003	5,191	812

Architecture. We use a multi-view adaptation of LTX-Video-2B with 28 transformer layers, a hidden size of 2048, and 32 attention heads of dimension 64. The video VAE is `AutoencoderKLLTXVideo`, and the text encoder is T5-v1.1-XXL. Both encoders are frozen. The four cameras provide front, left, right, and downward views.

Optimization and supervision. In addition to the settings in Table 6, AdamW uses betas $(0.9, 0.95)$ and epsilon 10^{-8}, with a gradient-clipping threshold of 1.0. Grounding training uses a learning rate of 10^{-5} and 5k warmup steps. Value pretraining runs for 90k steps at a learning rate of 5×10^{-5}, updating only the Future Value Expert. Successful expert trajectories and unsuccessful rollouts are sampled at a 1:1 ratio for value supervision. Action behavior cloning uses successful trajectories only.

Annotation settings. The frozen Grounding DINO-base teacher runs in FP32 with an input longest side of 512 pixels. Its prompt is the target category name followed by a period; the policy uses the full language goal. We retain boxes with confidence at least 0.3, apply NMS with an IoU threshold of 0.5, and keep at most three boxes per view and frame. Teacher heatmaps have resolution 32×32. For normalized box width w_b and height h_b, the Gaussian standard deviations are $\sigma_x = \max(0.35w_b, 1/32)$ and $\sigma_y = \max(0.35h_b, 1/32)$. A heatmap annotation is valid if any view contains a retained target box. The value-stage boundary requires target detections in two consecutive frames and is assigned to the first frame of the pair. Trajectories without such a pair receive exploration labels throughout. These stage boundaries are constructed offline.

Inference details. We use `FlowMatchEulerDiscreteScheduler`. Candidates have independent Gaussian future-noise initializations and share the observation history, pose history, lan-

guage goal, and flow schedule. The discrete action head runs alongside the world and value branches, with logits read at the final integration step. Grounding DINO is not used at inference; future RGB and heatmap decoding are optional.

Closed-loop evaluation. For each benchmark and evaluation split, we conduct 300 episodes per run with seeds 42, 43, and 44. An episode terminates on stop, a simulator terminal collision or failure, or a limit of 150 decision steps. These counts refer to closed-loop evaluation trials, whereas the dataset counts in Table 6 refer to stored trajectory records.

A.2 Training Dynamics

Figure 5 shows training-loss trajectories for three configurations. Value-only training exhibits a rapid initial loss decrease followed by smaller fluctuations. During joint training, action and value losses decrease while video loss remains comparatively stable. The paired-rollout configuration is shown as an additional training setting with a different value objective.



Figure 5: **Training-loss trajectories.** Solid lines show a 1k-step moving mean; pale traces show subsampled raw observations. **A (upper left):** Value-only flow-matching loss with frozen video, shown through 80k steps. **B (right):** Joint video, action, and value training over 30k steps. Video and action losses are logged directly; the unweighted value loss is recovered at matching steps as $(L_{\text{total}} - L_{\text{video}} - L_{\text{action}})/0.3$ and plotted on a logarithmic axis. **C (lower left):** An additional paired-rollout configuration with frozen video. Its value loss combines regression and pairwise ranking, and its action loss is success-only cross-entropy. Panels show separate runs with different objectives; their loss magnitudes are not directly comparable.

A.3 Flow Matching for Visual Futures

The video VAE maps the future multi-view sequence to $\mathbf{Z}_0^F = E_{\text{vae}}(\mathbf{o}_{t+1:t+T_f}^{1:V})$. For $\tau \in [0,1]$ and $\boldsymbol{\epsilon}^F \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the noisy latent and target velocity are

$$\mathbf{Z}_\tau^F = (1-\tau)\mathbf{Z}_0^F + \tau\boldsymbol{\epsilon}^F, \qquad \mathbf{u}_F^* = \boldsymbol{\epsilon}^F - \mathbf{Z}_0^F. \tag{7}$$

The visual expert predicts $\widehat{\mathbf{u}}_F = F_\theta^F(\mathbf{Z}_\tau^F, \mathcal{H}_t, \mathbf{e}_g, \tau)$. Visual tokens retain their camera identity through within-view attention and exchange information through cross-view attention. Equation 3 applies spatial target weights to this velocity-regression objective.

A.4 TEACHER HEATMAPS

For future step s and camera v, the frozen teacher returns boxes and confidences $\mathcal{B}_s^v = G_{\text{tea}}(\mathbf{o}_s^v, g) = \{(\mathbf{b}_j, c_j)\}_j$. We place a scale-adaptive Gaussian at each box center and take the maximum response:

$$H_s^{v,*}(x,y) = \max_j c_j \exp\left[-\frac{1}{2}\left(\frac{(x-\mu_j^x)^2}{(\sigma_j^x)^2} + \frac{(y-\mu_j^y)^2}{(\sigma_j^y)^2}\right)\right]. \tag{8}$$

The Goal Decoder produces $\widehat{H}_s^v$ from predicted grounding tokens. Normalizing target and predicted heatmaps jointly over view and location yields P_s^* and $\widehat{P}_s$ for Equation 2. Steps without a reliable annotation are masked out by m_s. The teacher heatmap is downsampled to the visual-latent resolution to obtain $\widetilde{H}^*$ for target-weighted generation. All teacher annotations are constructed from training data; the teacher is absent at inference.

A.5 STAGE-AWARE VALUE TARGETS

Trajectory supervision. We annotate successful expert trajectories and unsuccessful rollouts collected in training scenes. The annotations use trajectory observations, UAV poses, target locations, and offline grounding evidence. Successful trajectories provide examples of target discovery and approach, while unsuccessful rollouts provide examples of search behavior that does not complete the task. These annotations supply process-level targets for the Future Value Expert; target locations and future ground-truth observations are not policy inputs at inference.

Exploration score. Before reliable target grounding, $e_t \in [0,1]$ summarizes local changes between consecutive steps, including motion and changes in target evidence. It is a normalized exploration annotation, not a discounted time-to-discovery target. Its contribution is capped at $\beta = 0.25$ by Equation 4. This local score can fluctuate during exploration and does not require every exploratory step to reveal the target.

Approach progress. The first reliable grounding establishes t_g and the reference distance d_{t_g}. With UAV position $\mathbf{p}_t^{\text{pos}}$ and target position $\mathbf{x}_g$, $d_t = \|\mathbf{p}_t^{\text{pos}} - \mathbf{x}_g\|_2$. The normalized approach score is

$$p_t^{\text{app}} = \left[1 - \frac{d_t}{\max(d_{t_g}, \varepsilon_d)}\right]_0^1. \tag{9}$$

This measures distance reduction relative to discovery. The reference remains fixed after discovery, so retreat can reduce the score; temporary loss of visibility does not reset the stage. Equation 4 maps this progress to $[0.25, 1]$. The stage offset provides a baseline at discovery without introducing a separate stopping bonus.

Boundary interpretation. For $t_g > 0$, the approach label at discovery is at least β, while the preceding exploration label is at most β; hence $y_{t_g}^* - y_{t_g-1}^* \geq 0$. The intervals share a boundary, so a strictly positive jump is not required. These bounds describe annotation targets, not calibrated success probabilities or hard constraints on model predictions. The model predicts values from visual-future and grounding latents, with the expert-training objective specified in Appendix A.6.

A.6 VALUE AND ACTION EXPERT TRAINING

The Future Value Expert has its own QKV projections, self-attention, and feed-forward layers. It cross-attends to the visual and grounding states at corresponding world-model layers. Given trajectory targets $\mathbf{y}^*$ and Gaussian noise $\boldsymbol{\epsilon}^V$, we form

$$\mathbf{Y}_\tau = (1-\tau)\mathbf{y}^* + \tau\boldsymbol{\epsilon}^V, \qquad \mathbf{u}_V^* = \boldsymbol{\epsilon}^V - \mathbf{y}^*, \tag{10}$$

and optimize

$$\mathcal{L}_{\text{value}} = \mathbb{E}\left[\left\| V_\psi(\mathbf{Y}_\tau, \widehat{\mathbf{Z}}^F, \widehat{\mathbf{Z}}^G, \mathcal{H}_t, \mathbf{e}_g, \tau) - \mathbf{u}_V^* \right\|_2^2\right]. \tag{11}$$

The visual and grounding conditioning states are generated latent futures. Trajectory-value input tokens, labeled "Current Value" in earlier architecture figures, are noisy value tokens for flow matching; they do not define a separate current-frame reranking model.

For candidate k at layer l, the Action Expert reads $\mathbf{C}^{k,l} = [\widehat{\mathbf{Z}}^{F,k,l}; \widehat{\mathbf{Z}}^{G,k,l}; \widehat{\mathbf{Z}}^{V,k,l}]$, where $\widehat{\mathbf{Z}}^{V,k,l}$ denotes the value expert's hidden states. Its cross-attention update is

$$\mathbf{A}^{k,l+1} = \mathrm{FFN}_A\big(\mathbf{A}^{k,l} + \mathrm{CrossAttn}_A(\mathbf{A}^{k,l}, \mathbf{C}^{k,l})\big)\,. \tag{12}$$

The action decoder gives $\boldsymbol{\pi}_{t+s} = \mathrm{Softmax}(\mathbf{W}_A \mathbf{A}_{t+s})$, and behavior cloning minimizes

$$\mathcal{L}_{\text{action}} = -\frac{1}{T_a} \sum_{s=0}^{T_a-1} \log \pi_{t+s}(a^*_{t+s}). \tag{13}$$

The resulting action chunk is associated with the same generated future that the value expert scores. These losses are combined in Equation 6.

A.7 Value-Guided Candidate Selection

At each decision step, the model generates K candidate latent futures conditioned on the same observation history and language goal. The candidates are packed along the batch dimension and processed in parallel. For candidate k, the Future Value Expert produces a scalar score

$$S_k = V_\psi\Big(\widehat{\mathbf{Z}}^{F,k}, \widehat{\mathbf{Z}}^{G,k}, \mathcal{H}_t, \mathbf{e}_g\Big)\,. \tag{14}$$

The policy selects

$$k^* = \arg\max_{1 \le k \le K} S_k, \tag{15}$$

and the Action Expert uses the visual, grounding, and value representations associated with candidate k^* to generate the action chunk. The default setting is $K = 3$, and Section 5.7 evaluates $K \in \{1, 2, 3, 4\}$. Candidate generation, value prediction, and selection are performed once per decision step; no iterative candidate refinement or resampling is used.

B Value Error and Candidate Selection under Partial Observability

We give a conditional analysis of how future-prediction errors affect value-guided candidate selection. The analysis does not assume full state observability or that predicted latents constitute a sufficient belief state. It applies to a fixed candidate set at one decision and does not assert global policy optimality.

History-conditioned setup. Let h_t contain the full observation, pose, and action history together with the language goal g. Fix K generated candidates. Candidate k specifies an executed action prefix α_k; all prefixes are evaluated under the same subsequent policy π. Let $P_k(F, G \mid h_t)$ be the true joint distribution of future scene information F and target information G under execution of α_k followed by π, and let $\widehat{P}_k$ be its model approximation. These are distributions over task-relevant future quantities, not pointwise distances between arbitrary latent coordinates. Hidden states and future observations are marginalized conditional on h_t. The finite history window and latent representations used by the model are not assumed sufficient for predicting these distributions.

Let $u_g(F, G) \in [0, 1]$ be a bounded task-progress readout, and define $J_k = \mathbb{E}_{P_k}[u_g]$ and $\widehat{J}_k = \mathbb{E}_{\widehat{P}_k}[u_g]$. This readout is an analytical surrogate for the stage-aware supervision; it need not fully determine eventual success. Let S_k be the actual scalar ranking score. For analysis, suppose a common strictly increasing calibration function c maps these scores to the scale of u_g, and write $\widehat{v}_k = c(S_k)$. Calibration preserves ranking and is not an additional module required by the algorithm. Its approximation error is retained explicitly below; the analysis does not establish that a small error is achieved by training.

Proposition 1 (Prediction-to-value error propagation). For each candidate, define

$$\epsilon_{F,k} = \mathrm{TV}(P_k^F, \widehat{P}_k^F), \tag{16}$$

$$\epsilon_{G,k} = \mathbb{E}_{F\sim P_k^F}[\mathrm{TV}(P_k^{G|F}, \widehat{P}_k^{G|F})], \tag{17}$$

$$\epsilon_{V,k} = |\widehat{v}_k - \widehat{J}_k|. \tag{18}$$

Conditioning on h_t is suppressed for readability. The model conditional distribution is taken to be defined on the support of P_k^F. Then

$$|\widehat{v}_k - J_k| \le \epsilon_{V,k} + \epsilon_{F,k} + \epsilon_{G,k}. \tag{19}$$

Proof. For probability densities or mass functions with respect to suitable dominating measures, add and subtract $p_k^F \widehat{p}_k^{G|F}$ inside the joint L_1 distance. Integrating gives

$$\mathrm{TV}(P_k, \widehat{P}_k) \le \mathrm{TV}(P_k^F, \widehat{P}_k^F) + \mathbb{E}_{P_k^F}[\mathrm{TV}(P_k^{G|F}, \widehat{P}_k^{G|F})]. \tag{20}$$

Since $0 \le u_g \le 1$, the bounded-expectation inequality yields $|J_k - \widehat{J}_k| \le \mathrm{TV}(P_k, \widehat{P}_k)$. Combining this with $|\widehat{v}_k - \widehat{J}_k| = \epsilon_{V,k}$ proves Equation 19.

Proposition 2 (Conditional improvement over a candidate baseline). Let Q_k be the probability of eventual benchmark success under h_t, executed prefix α_k, and continuation policy π. Define the surrogate-to-success discrepancy $\epsilon_{A,k} = |J_k - Q_k|$. Suppose

$$\max_{k\le K}(\epsilon_{V,k} + \epsilon_{F,k} + \epsilon_{G,k} + \epsilon_{A,k}) \le \epsilon. \tag{21}$$

Let $\widehat{k} = \arg\max_k S_k = \arg\max_k \widehat{v}_k$ and let k_0 be any baseline choice in the same candidate set. Then

$$Q_{\widehat{k}} - Q_{k_0} \ge \max_{k\le K} Q_k - Q_{k_0} - 2\epsilon. \tag{22}$$

Consequently, selection improves conditional success probability whenever the available improvement $\max_k Q_k - Q_{k_0}$ exceeds 2ϵ.

Proof. Proposition 1 and the definition of $\epsilon_{A,k}$ imply $|\widehat{v}_k - Q_k| \le \epsilon$ for all candidates. Let $k^* = \arg\max_k Q_k$. Since $\widehat{k}$ maximizes $\widehat{v}_k$,

$$Q_{k^*} \le \widehat{v}_{k^*} + \epsilon \le \widehat{v}_{\widehat{k}} + \epsilon \le Q_{\widehat{k}} + 2\epsilon. \tag{23}$$

Rearranging and subtracting Q_{k_0} proves the claim.

Interpretation and scope. Equation 19 separates world-future prediction, target prediction, and value-readout errors. Holding the remaining terms fixed, reducing prediction errors tightens the bound on value error. Equation 22 explains when this accuracy can translate into improved candidate selection. Both results retain partial observability through conditioning on history; a generated future is not treated as an observed environment state. The analysis also requires evaluating the same candidate action that is eventually executed.

Future Target Grounding and target-weighted generation are motivated by reducing task-relevant prediction errors, but empirical heatmap KL and image-quality metrics do not establish the full distributional bounds assumed here. In particular, teacher heatmaps on annotated frames do not specify the complete conditional law of visibility, discovery, and target location. Nor do the losses establish a small calibration error or a small surrogate-to-success discrepancy. Thus, these propositions are sufficient-condition analyses, not claims that every prediction improvement monotonically increases success. They concern a single decision under a fixed continuation policy; candidate coverage and changes to the entire closed-loop policy require additional analysis. Improvements in navigation success and computational efficiency are established by our experiments.