SimpleVLN：迈向简洁而前沿的视觉-语言导航

Lingyi Hong[1,*,†], **Weiren Cai**[2,*], **Jiayu Chen**[1,*], **Xinyu Zhou**[2,†,‡],
Bailin He[1], **Yunfei Guo**[1], **Qiao Sun**[3], **Boyu Ma**[2],
Jianfei Yang[2], **Wenqiang Zhang**[1]

[1] 复旦大学, [2] 南洋理工大学, [3] 联想研究院

项目主页

honglyhly@gmail.com, xinuzprf@gmail.com

摘要

视觉-语言导航（Vision-and-Language Navigation，VLN）要求具身智能体将语言与视觉观测对齐，并做出长时程决策。现有方法通常借助任务专用架构应对上述需求。随着多模态大语言模型（Multimodal Large Language Models，MLLM）能力不断增强，一个重要问题仍有待回答：**在引入任务专用架构之前，我们是否真正释放了原生 *MLLM* 在 *VLN* 上的潜力？**本文提出 **SimpleVLN**，一个保持模型架构不变的框架，从导航形式化、数据构建与优化三个方面展开研究。在形式化层面，SimpleVLN 将单轮决策输入与固定预算的全局–局部记忆相结合，在不让上下文无界增长的前提下高效利用导航历史；动作块（Action Chunk）则使模型每次调用即可高效输出多个原子动作。在数据构建方面，本文将专家示范与依赖策略的 DAgger 经验相结合，并设计可扩展的合成流程，为强化学习生成大规模任务。在优化层面，渐进式 IL–DAgger–RL 流程从动作模仿逐步推进到任务级结果优化。全面的实验验证了上述三个方面的有效性。本文在多种 MLLM 骨干网络和模型规模上实例化 SimpleVLN，展示了其通用性。SimpleVLN 在 R2R-CE 和 RxR-CE 划分上分别取得 76.7% 和 80.1% 的成功率，显著优于现有方法。结果表明，无需任务专用架构设计，即可有效释放原生 MLLM 的导航能力。

1 引言

视觉-语言导航（Vision-and-Language Navigation，VLN）要求具身智能体在未见过的环境中遵循自然语言指令 [1, 8]，综合了视觉理解、语言对齐、空间推理与长时程决策。现有方法通过面向感知与空间推理的任务专用设计 [22]、基于图的建模 [3] 以及分层或双系统决策 [16] 不断取得进展。这些架构行之有效，而多模态大语言模型（MLLM）同样具备强大的通用视觉理解、推理与指令遵循能力 [2]。然而，增强型 MLLM 系统的性能同时反映了骨干网络与任务专用组件的贡献，原生 MLLM 自身的能力因此并不清楚。由此引出一个尚未被充分探讨的问题：*在引入任务专用架构之前，我们是否真正释放了原生 MLLM 在 VLN 上的潜力？*为回答这一问题，本文研究在不修改架构的情况下，原生 MLLM 在 VLN 上能被推进到何种程度。

保持架构不变，仍需使 MLLM 适应序列化导航。本文从形式化、数据构建与优化三个方面研究并构建 **SimpleVLN**（图 1）。第一，在形式化层面，本文将单轮决策输入与固定预算的全局–局部记忆相结合，在高效利用导航历史的同时避免上下文无界增长；动作分块（Action Chunking）通过每次模型调用预测多个原子动作，进一步实现高效的动作输出。第二，在数据构建方面，本文将专家示范与通过 DAgger [10] 收集的依赖策略的经验相结合，并设计可扩展的任务合成流程，为强化学习生成大规模导航任务。第三，在优化层面，本文研究渐进式 IL–DAgger–RL 训练流程：模仿学习（IL）通过动作模仿初始化策略，DAgger 使其适应策略自身诱导的状态，强化学习（RL）则直接优化任务级结果，逐步提升导航精度。

*同等贡献。 † 共同项目负责人。 ‡ 通讯作者。

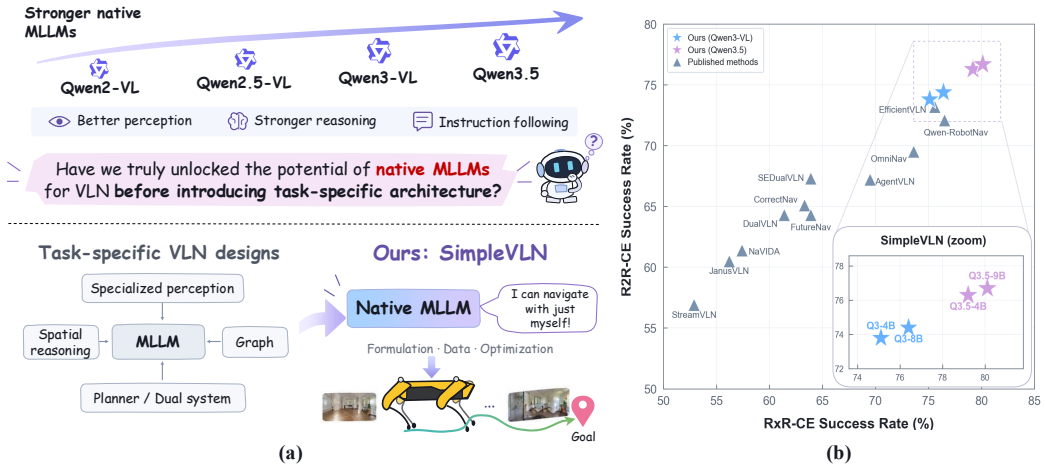


图 1: **SimpleVLN 的动机与性能对比**。(a) 在 VLN 中使用原生 MLLM 的动机。(b) 在 R2R-CE 和 RxR-CE 上与现有方法的对比。

全面的实验验证了上述三个方面的有效性。本文在多种 MLLM 骨干网络和模型规模上实例化 SimpleVLN，展示了其通用性。SimpleVLN 在 R2R-CE 和 RxR-CE 的 *val-unseen* 划分上分别取得 76.7% 和 80.1% 的 SR，显著优于现有方法。这些结果表明，无需任务专用架构设计，即可有效释放原生 MLLM 的导航能力。

本文贡献有三：(1) 提出 SimpleVLN，一个保持架构不变的框架，借助单轮决策输入、受管理的视觉历史以及动作块预测，使原生 MLLM 适应 VLN，无需任务专用的架构模块。(2) 利用专家示范、依赖策略的 DAgger 经验以及可扩展的 RL 任务合成构建训练数据，并设计渐进式 IL–DAgger–RL 流程以提升导航精度。(3) 在多个骨干网络系列和模型规模上实例化 SimpleVLN，在 R2R-CE 和 RxR-CE 上取得优异性能，大量实验验证了其形式化、数据构建与优化设计的有效性。

2 相关工作

视觉-语言导航。视觉-语言导航（VLN）要求智能体遵循以视觉观测为依据的自然语言指令 [1, 8]。早期 VLN 在预定义的离散视点上运行，VLN-CE [7] 则将导航扩展到使用低层动作的连续环境中。

面向 VLN 的多模态大语言模型。近期的 VLN 方法利用多模态大语言模型（MLLM）进行视觉–语言理解 [25, 4, 24]，同时引入流式上下文 [17]、3D 几何编码器 [22]、双系统架构 [16, 6, 18, 12] 或显式思维链推理 [9, 14] 来应对长时程导航。与之不同，SimpleVLN 无需修改架构，直接适配预训练 MLLM，保持与现有 MLLM 推理框架的兼容性。

导航训练范式。模仿学习利用专家轨迹训练导航策略 [17, 28]，但在策略执行时会受到分布偏移的影响。DAgger [10] 通过在模型诱导的状态上提供监督缓解这一问题 [17, 22]，而 RL [28, 5] 则直接优化任务级结果。SimpleVLN 将三者组合为渐进式 IL-DAgger-RL 流程，从专家监督过渡到任务完成度优化。

3 SimpleVLN

3.1 概述

SimpleVLN 无需修改架构，直接将现成的 MLLM [13, 2] 适配到 VLN。本文从三个角度研究原生 MLLM 导航：导航形式化、数据构建，以及一种通过环境交互从动作模仿转向结果驱动学习的渐进式优化策略。

如图 2所示，给定自然语言指令 $\mathcal{I}$，在每个原子环境步 k，智能体观测到一张以自我为中心的 RGB 图像 x_k，并从如下动作空间中执行动作 u_k：

$$\mathcal{A} = \{\textsc{Forward}(0.25\,\text{m}), \textsc{TurnLeft}(15^\circ), \textsc{TurnRight}(15^\circ), \textsc{Stop}\}.$$

当执行 Stop 或达到原子动作预算 $S_{\max}$ 时，回合终止；若智能体停在距目标 3 m 以内，则判定为成功 [7]。

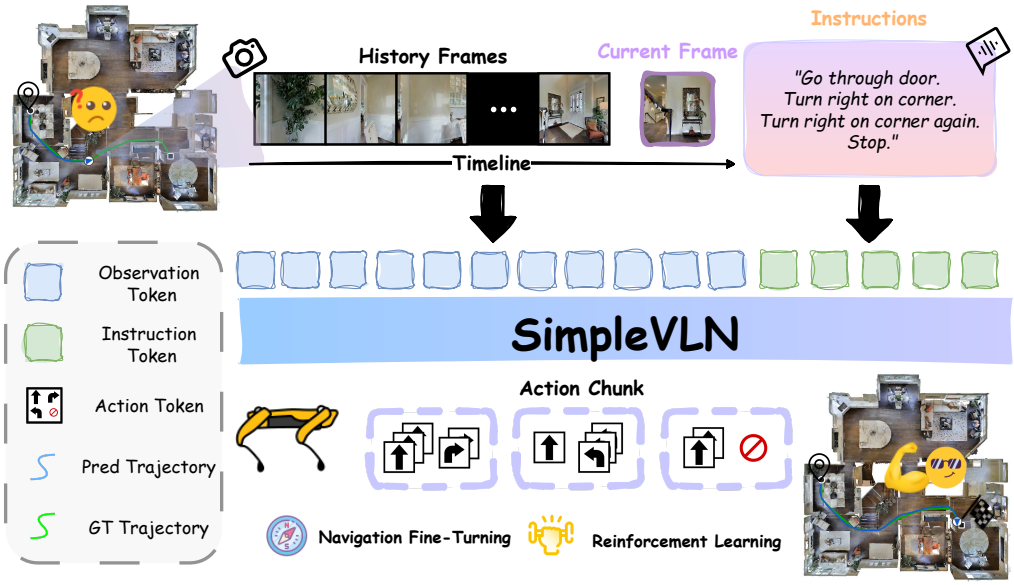


图 2: **SimpleVLN 概览**。在每次决策时，采样的历史帧、当前观测和指令被输入 MLLM，由其直接预测用于导航的动作块。

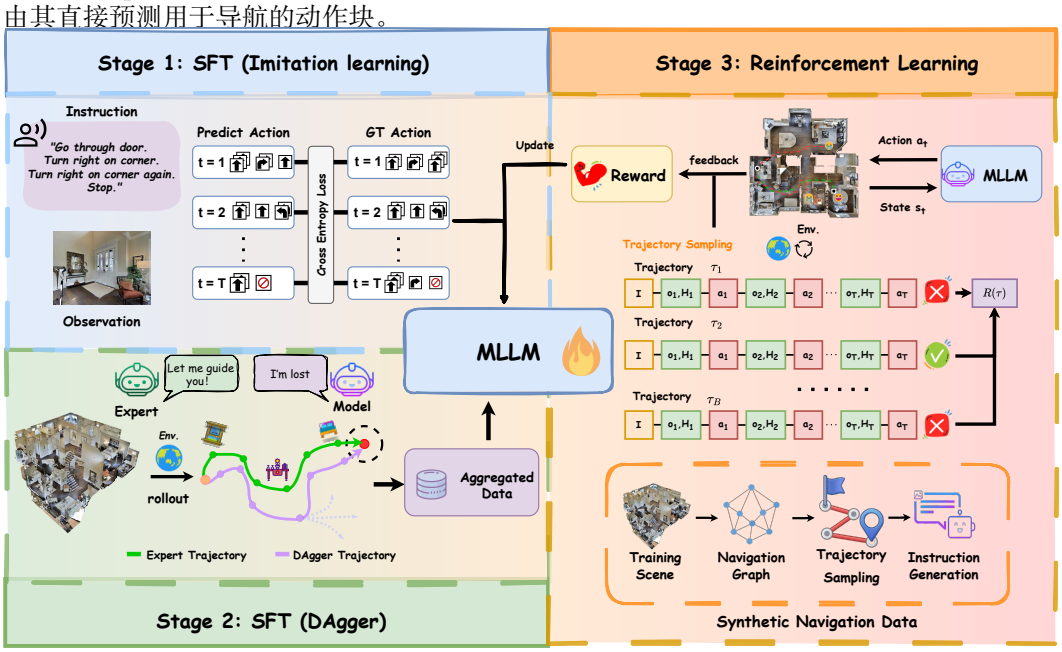


图 3: **SimpleVLN 的训练数据流程**。SimpleVLN 首先从专家示范中学习，然后通过 DAgger 收集并聚合纠正性轨迹。在 RL 阶段，它由采样的起点–终点路径和生成的指令合成导航回合，再通过环境交互与结果反馈优化策略。

3.2 决策级导航形式化

本文围绕模型决策步而非原子环境步对导航进行形式化。在每次决策时，SimpleVLN 由指令、当前观测和采样历史重建一个有界上下文，并利用原生 MLLM 预测导航动作。下文首先介绍决策上下文的构建，然后介绍输出动作表示。

单轮决策输入。在每次模型决策时，本文构建一个独立的输入，而不是将最新观测和预测动作追加到之前的对话中。令 $t = 1, 2, \ldots$ 为模型决策的索引，n_t 为决策 t 之前已执行的原子动作数，且 $n_1 = 0$。于是当前观测为 $o_t = x_{n_t}$。输入由指令 $\mathcal{I}$、当前观测 o_t 和采样的视觉历史 H_t 组成；不包含之前的对话轮次和过去的动作。记由此得到的视觉上下文为 $s_t = (o_t, H_t)$。

全局–局部历史选择。在决策 t 之前，原子步观测历史 $(x_0, \ldots, x_{n_t-1})$ 会随环境交互不断增长。保留全部观测会使输入膨胀，而仅保留近期帧又会丢失较早的路线上下文。因此，本文在两个互补的时间尺度上表示历史：局部历史保持近期的连续性，全局历史则稀疏覆盖较早的原子步观测。令 L_t 包含紧邻 o_t 之前的至多 N 个观测，G_t 包含从更早历史中均匀采样的至多 M 个观测：

$$\begin{aligned} L_t &= (x_k)_{k=\max(0,n_t-N)}^{n_t-1}, \qquad m_t = \min\{M, \max(0, n_t - N)\}, \\ G_t &= (x_{i_j})_{j=1}^{m_t}, \qquad H_t = G_t \,\|\, L_t. \end{aligned} \tag{1}$$

其中，$\|$ 表示序列拼接。当 $m_t > 0$ 时，采样索引满足 $0 \le i_1 < \cdots < i_{m_t} < \max(0, n_t - N)$，从而保持时间顺序。当可用的较早观测少于 M 个时，全部保留。当前观测 o_t 单独提供。因此，每次决策至多使用 $N + M + 1$ 个视觉输入，与回合长度无关。

动作分块。在输出端，一次模型决策可以指定多个原子动作。本文不再每次决策只预测一个原子动作，而是沿用 NaVIDA [30] 的做法，预测动作块 a_t，即一个有序的原子动作序列。令 K_t 为该动作块中的原子动作数，则 $a_t = (u_{n_t}, \ldots, u_{n_t+K_t-1})$，导航策略预测 $a_t \sim \pi_\theta\big(\cdot \mid \mathcal{I}, s_t\big) = \pi_\theta\big(\cdot \mid \mathcal{I}, o_t, H_t\big)$，其中 π_θ 表示原生 MLLM 导航策略。环境依次执行 a_t 中的动作，每执行一个动作产生一个观测，这些观测被保留用于后续的历史构建。动作块执行完毕后再次查询 MLLM，除非 STOP 或原子动作预算使回合终止。

3.3 导航数据构建

本文为后训练的每个阶段构建训练数据。专家示范提供动作监督，DAgger 则补充模型执行过程中遇到的依赖策略的状态。在 RL 阶段，复用监督学习所用的回合既不能扩展任务分布，也无法提供真正的新交互经验。因此，本文在训练场景中合成新的导航任务。

专家示范覆盖。本文汇集 R2R [1]、RxR [8] 和 ScaleVLN [15] 的轨迹，以扩大专家行为的覆盖范围。为匹配第 3.2节的决策级推理接口，每条轨迹被转换为监督样本：以 $(\mathcal{I}, s_t)$ 为输入、对应的专家动作块 $a_t^\star$ 为目标，其中 $s_t = (o_t, H_t)$。

基于 DAgger 的依赖策略经验。扩大专家示范并不能消除专家轨迹与当前策略访问状态之间的分布差异。因此，本文采用基于偏离度的 DAgger [10] 策略：将当前策略轨迹与真值（Ground Truth，GT）轨迹比较，偏离较大时选用专家轨迹，否则保留策略轨迹。所选轨迹被转换为决策级监督样本，并与示范数据聚合。

面向 RL 的任务合成。由于 R2R、RxR 和 ScaleVLN 的训练回合已用于监督学习，本文在这些数据集的训练场景中合成新的 RL 任务，而非复用已标注回合，并排除 val-unseen 场景。对每个数据源，从导航图中采样位于不同房间的起点–终点对，终点与带语义标注的物体相关联。一个 MLLM 根据参考路径上的路点生成导航指令，描述转向、房间、显著物体以及以自我为中心的视图。参考路径仅用于构建任务，并不作为动作目标。

3.4 渐进式优化

图 3总结了 SimpleVLN 的三阶段优化。首先，模仿学习利用专家示范学习导航动作。其次，基于 DAgger 的微调使策略适应其自身执行过程中遇到的状态，同时保留动作级监督。最后，强化学习在合成回合上通过环境交互优化任务级结果，将学习信号从示范动作转向结果反馈。

基于专家动作的模仿学习。令 $\mathcal{D}$ 为由元组 $(\mathcal{I}, s_t, a_t^\star)$ 构成的监督数据集，其中 $a_t^\star$ 为决策状态 s_t 对应的目标动作块。本文最小化完整目标动作块的负对数似然：$\mathcal{L}_{\mathrm{SFT}} = -\mathbb{E}_{(\mathcal{I}, s_t, a_t^\star)\sim\mathcal{D}}\left[\log \pi_\theta\big(a_t^\star \mid \mathcal{I}, s_t\big)\right]$。该目标使策略学会复现示范的动作块，从专家监督中学习。

依赖策略的监督适应。接下来，本文在第 3.3节所述示范数据与 DAgger 数据的聚合上训练。目标 $\mathcal{L}_{\mathrm{SFT}}$ 保持不变，而 $\mathcal{D}$ 额外包含由所选依赖策略轨迹得到的目标动作块 $a_t^\star$。

基于强化学习的结果优化。由于导航最终以任务完成度而非与特定动作序列的一致性来评估，本文在合成导航任务上优化任务级结果。利用第 3.3节所述的合成回合，策略与环境交互并依次预测决策级动作块：

$$a_t \sim \pi_\theta\big(\cdot \mid \mathcal{I}, s_t\big), \qquad \tau = \big((s_t, a_t)\big)_{t=1}^{T}. \tag{2}$$

其中，T 为回合终止前实际的模型级决策次数，区别于原子动作预算 $S_{\max}$。本文使用二值任务级奖励，使任务完成成为显式目标，而不要求与参考动作序列一致。成功判定遵循

表 1: 在 R2R-CE 和 RxR-CE *val-unseen* 划分上**与最先进方法的比较**。† 和 * 分别表示 RL 之前和之后的 SimpleVLN。Pano.：全景；Odo.：里程计；Depth：深度；S.RGB：单目 RGB。

方法	规模	观测				R2R-CE val-unseen				RxR-CE val-unseen			
		Pano.	Odo.	Depth	S.RGB	NE↓	OS↑	SR↑	SPL↑	NE↓	SR↑	SPL↑	nDTW↑
NaVid [25]	7B				✓	5.47	49.1	37.4	35.9	8.41	23.8	21.2	—
Uni-NaVid [24]	7B				✓	5.58	53.3	47.0	42.7	6.24	48.7	40.9	—
NaVILA [4]	7B				✓	5.22	62.5	54.0	49.0	6.77	49.3	44.0	58.8
ActiveVLN [28]	3B				✓	5.31	58.3	50.1	43.7	5.84	50.7	41.2	58.1
StreamVLN [17]	7B				✓	4.98	64.2	56.9	51.9	6.22	52.9	46.0	61.9
NavFoM [23]	7B				✓	5.01	64.9	56.2	51.2	5.51	57.4	49.4	60.2
JanusVLN [22]	7B				✓	4.78	65.2	60.5	56.8	6.06	56.2	47.5	62.1
NaVIDA [30]	3B				✓	4.32	69.5	61.4	54.7	5.23	57.4	49.6	67.0
CorrectNav [21]	7B				✓	4.24	67.5	65.1	62.3	4.09	69.3	63.3	75.2
DualVLN [16]	7B				✓	4.05	70.7	64.3	58.5	4.58	61.4	51.8	70.0
FutureNav [27]	4B				✓	4.24	70.2	65.4	61.3	4.60	60.9	52.8	68.0
FutureNav [27]	8B				✓	4.29	69.2	64.3	60.3	4.26	63.9	54.8	69.1
AgentVLN [18]	3B			✓	✓	3.88	73.5	67.2	64.7	3.92	69.5	61.3	74.6
OmniNav [19]	3B	✓				3.74	74.6	69.5	66.1	3.77	73.6	62.0	—
SEDualVLN [6]	7B				✓	3.75	73.7	67.3	62.5	4.12	63.9	52.4	72.8
Qwen-RobotNav [26]	4B				✓	4.22	73.6	66.9	60.5	4.15	71.3	61.5	68.6
Qwen-RobotNav [26]	8B				✓	4.36	72.7	65.7	59.6	4.16	73.4	63.5	69.9
Qwen-RobotNav [26]	4B	✓				3.80	77.2	69.5	63.6	3.80	75.2	65.0	71.9
Qwen-RobotNav [26]	8B	✓				3.53	78.5	72.1	66.6	3.58	76.5	65.7	72.5
Efficient-VLN [29]	8B				✓	3.29	78.2	73.2	65.0	3.18	75.6	60.0	-
SimpleVLN† (Qwen2.5-VL)	3B				✓	5.70	63.3	51.6	44.9	5.20	50.3	42.0	57.6
SimpleVLN† (Qwen2.5-VL)	7B				✓	4.87	68.4	59.1	52.0	5.31	58.9	49.2	62.9
SimpleVLN† (Qwen3-VL)	4B				✓	4.26	74.3	65.4	58.8	4.79	63.4	54.1	66.5
SimpleVLN* (Qwen3-VL)	4B				✓	3.39	79.4	73.8	66.7	3.29	75.1	65.3	71.5
SimpleVLN† (Qwen3-VL)	8B				✓	4.40	73.3	64.2	57.0	4.73	64.2	54.3	66.2
SimpleVLN* (Qwen3-VL)	8B				✓	2.93	79.2	74.4	69.5	3.11	76.4	66.4	77.4
SimpleVLN† (Qwen3.5)	4B				✓	3.76	77.5	67.1	59.2	4.32	65.6	55.3	67.8
SimpleVLN* (Qwen3.5)	4B				✓	2.79	80.1	76.3	72.3	2.61	79.2	69.7	80.3
SimpleVLN† (Qwen3.5)	9B				✓	3.59	76.9	69.2	61.2	4.48	64.0	54.2	67.0
SimpleVLN* (Qwen3.5)	**9B**				✓	**2.63**	**84.1**	**76.7**	**72.6**	**2.51**	**80.1**	**69.9**	**79.8**

表 2: **与最先进方法在推理延迟和导航性能上的比较**。延迟在 R2R-CE 上使用同一块 NVIDIA A100 GPU 测量，仅统计一次动作块决策的模型执行时间，不含预处理、环境执行及其他非模型开销。SimpleVLN 以 Qwen3-VL-4B 为骨干网络。

方法	规模	延迟 (ms)↓	R2R-CE				RxR-CE			
			NE↓	OS↑	SR↑	SPL↑	NE↓	SR↑	SPL↑	nDTW↑
StreamVLN [17]	7B	300	4.98	64.2	56.9	51.9	6.22	52.9	46.0	61.9
JanusVLN [22]	7B	906	4.78	65.2	60.5	56.8	6.06	56.2	47.5	62.1
DualVLN [16]	7B	683	4.05	70.7	64.3	58.5	4.58	61.4	51.8	70.0
FutureNav [27]	4B	1141	4.24	70.2	65.4	61.3	4.60	60.9	52.8	68.0
SimpleVLN	**4B**	**293**	**3.39**	**79.4**	**73.8**	**66.7**	**3.29**	**75.1**	**65.3**	**71.5**

第 3.1节的停止与距离准则：$R(\tau) = \mathbb{I}(\text{success}) \in \{0, 1\}$. 本文在合成回合上，对由策略–环境交互诱导的轨迹分布 $p_\theta(\tau)$ 最大化期望任务奖励：$\max_\theta \ \mathcal{J}_{\text{RL}}(\theta) = \mathbb{E}_{\tau \sim p_\theta}[R(\tau)]$. 与模仿学习不同，该目标不要求与特定参考动作序列一致：任何成功完成任务的轨迹都会得到奖励。因此，训练从专家动作监督出发，经由依赖策略的监督适应，最终过渡到回合级任务完成反馈。

4 实验

4.1 实验设置

基准与指标。本文在 Matterport3D 场景中基于 Habitat [7]，于 R2R-CE [1] 和 RxR-CE [8] 的 val-unseen 划分上评估 SimpleVLN，报告 NE、OS、SR、SPL 和 nDTW，并以 SR 为主要指标。

实现细节。本文使用 Qwen3-VL（4B/8B）[2] 和 Qwen3.5（4B/9B）[13]。IL 阶段在 R2R [1] 和 RxR [8] 上进行 3 个 epoch 的 SFT，冻结视觉编码器，峰值学习率为 2×10^{-5}。随后在收集的 DAgger 数据与 R2R/RxR 上训练 1 个 epoch，再在它们与 ScaleVLN [15] 的混合数

表 3: **导航接口与记忆设计**。使用在 R2R 上微调的 Qwen3-VL 4B，对动作分块、单轮导航和观测采样进行消融。SimpleVLN 的默认设计以灰底标出。

动作格式	NE↓	OS↑	SR↑
原子动作	5.2	54.5	45.6
动作块	**5.5**	**61.2**	**53.6**

(a) **动作分块**。*原子动作*每次决策预测一个原子动作；*动作块*每次决策预测多个原子动作并按顺序执行，若回合终止则提前停止。

对话格式	NE↓	OS↑	SR↑
多轮	5.9	55.2	35.7
单轮	**5.5**	**61.2**	**53.6**

(b) **单轮导航**。*多轮*在对话中累积观测与动作；*单轮*由指令、当前观测和固定预算的采样历史重建每次输入。

采样方案	NE↓	OS↑	SR↑
Chunk	6.1	55.2	46.3
Primitive (all)	5.8	57.5	51.9
Primitive (8)	**5.5**	**61.2**	**53.6**

(c) **观测采样**。*Chunk* 采样动作块边界；*Primitive (all)* 保留全部原子步观测；*Primitive (8)* 均匀采样八帧全局历史，不含局部帧与当前帧。

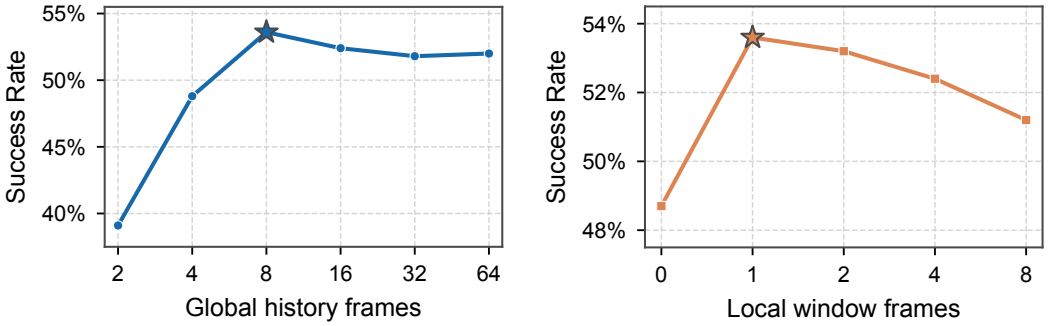


图 4: **记忆预算**。全局与局部记忆大小对 R2R-CE *val-unseen* SR 的影响。左图在 $N = 1$ 时改变 M，右图在 $M = 8$ 时改变 N；星号表示默认配置。

据上训练 2 个 epoch。主结果的 RL 使用 DAPO [20] 训练 40 步，批大小为 256，每个提示 16 次 rollout，学习率 1×10^{-6}，KL 系数为 0，裁剪比率为 0.30/0.20。

4.2 与最先进方法的比较

主要结果。表 1将 SimpleVLN 与已有方法进行比较，并报告其在多个骨干网络系列和模型规模上的结果，包括 RL 前的 Qwen2.5-VL (3B/7B) 变体。SimpleVLN 仅使用单目 RGB 观测和原生 MLLM 架构，即在 R2R-CE 上达到 76.7 SR、在 RxR-CE 上达到 80.1 SR，优于现有方法。RL 在所有评估的骨干网络上都稳定提升了 SR；以 Qwen3-VL 4B 为例，R2R-CE 上的 SR 从 65.4 提升至 73.8，RxR-CE 上从 63.4 提升至 75.1。表 2显示，SimpleVLN 每次动作块决策耗时 293 ms，与 StreamVLN 相当，并快于其余对比方法。

4.3 形式化

本节研究第 3.2节提出的导航形式化。除非另有说明，这些实验均使用在 R2R 上微调的 Qwen3-VL 4B，并在 R2R-CE *val-unseen* 划分上评估。

动作分块。表 3a评估以动作分块替代原子动作预测的效果。动作分块将 SR 从 45.6 提升至 53.6，提高 8.0 个点，说明它不仅降低了模型调用频率，还是一种有效的输出表示。

单轮决策输入。表 3b以重建的单轮输入加采样视觉历史替代累积的多轮对话，使 SR 从 35.7 提升至 53.6。尽管多轮对话可以保留更完整的交互历史，但持续追加观测和动作会产生过长的上下文，当前 MLLM 难以有效利用其进行导航。

观测采样。表 3c考察应保留哪些观测。仅采样动作块边界得到 46.3 SR，保留全部原子步观测得到 51.9，均匀采样八帧全局历史得到 53.6。这八帧预算仅针对全局历史，不包括局部历史和当前观测。结果表明，紧凑、均匀采样的全局历史优于保留每一帧。

记忆预算。图 4检验增大全局或局部历史预算能否提升导航精度。在 $N = 1$ 时，SR 在 $M = 8$ 处达到峰值，随后随全局预算增大而下降，说明过多的全局历史会引入冗余或相关性较低的观测，稀释有用的上下文。在 $M = 8$ 时，保留一个局部观测的 SR 最高，表明更多的邻近帧提供的信息大多冗余。因此本文采用 $M = 8$、$N = 1$：至多八帧全局历史和一帧局部历史，外加当前观测。

表 4: 在 R2R-CE 和 RxR-CE *val-unseen* 划分上的**示范数据与 DAgger 数据消融**。

R2R	RxR	ScaleVLN	DAgger	R2R-CE				RxR-CE			
				NE↓	OS↑	SR↑	SPL↑	NE↓	SR↑	SPL↑	nDTW↑
✓				5.53	61.2	53.6	48.1	6.45	50.3	42.0	57.6
✓	✓			5.28	63.4	56.7	51.3	5.75	54.5	46.7	62.4
✓	✓	✓		4.90	68.4	57.9	50.9	5.27	55.3	48.1	63.6
✓	✓	✓	✓	**4.26**	**74.3**	**65.4**	**58.8**	**4.79**	**63.4**	**54.1**	**66.5**

表 5: 在 R2R-CE 和 RxR-CE *val-unseen* 划分上的**合成 RL 数据消融**。数据集列表示用于合成 RL 回合的训练场景来源。

R2R	RxR	ScaleVLN	R2R-CE				RxR-CE			
			NE↓	OS↑	SR↑	SPL↑	NE↓	SR↑	SPL↑	nDTW↑
	RL 前基线		4.26	74.3	65.4	58.8	4.79	63.4	54.1	66.5
✓			3.91	74.0	68.5	63.2	4.31	65.8	55.3	68.3
	✓		3.76	71.8	67.5	59.3	3.98	68.5	59.5	68.8
		✓	4.17	70.9	66.6	61.3	4.34	66.3	56.2	67.3
✓	✓		3.75	74.1	69.4	63.8	3.77	69.7	60.8	68.1
✓	✓	✓	**3.74**	**73.1**	**70.1**	**65.6**	**3.82**	**71.0**	**62.1**	**68.4**

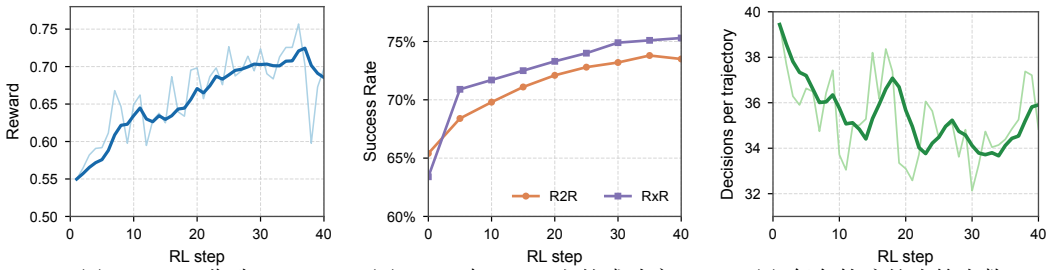
4.4 数据构建

本节研究第 3.3节所构建训练数据的作用。本小节所有实验均使用 Qwen3-VL 4B，并报告 R2R-CE 和 RxR-CE *val-unseen* 划分上的结果。

专家示范覆盖。表 4检验扩大专家示范覆盖范围能否改善导航。在 R2R 基础上加入 RxR，R2R-CE 上的 SR 从 53.6 提升至 56.7，RxR-CE 上从 50.3 提升至 54.5。再加入 ScaleVLN，SR 分别进一步提升至 57.9 和 55.3。结果表明，组合 R2R 与 RxR 带来明显的域内增益，而 ScaleVLN 进一步提升了对未见环境的泛化能力。

基于 DAgger 的依赖策略经验。接下来评估依赖策略的数据能否在组合专家示范的基础上带来提升。在表 4中，加入 DAgger 带来最大增益：R2R-CE 上提升 7.5 个 SR 点，RxR-CE 上提升 8.1 个点，分别达到 65.4 和 63.4。这说明用学习到的策略所收集的轨迹补充静态示范具有重要价值。

面向 RL 的任务重合成。表 5评估从 R2R、RxR 和 ScaleVLN 训练场景合成的 RL 任务。这些名称指代用于合成的场景池，而非复用的已标注回合。所有变体均从同一个 IL–DAgger 检查点出发，使用 GRPO [11]，批大小为 32，每个提示八次 rollout。每个场景池都使 SR 超过 RL 前基线，而三者混合效果最好，在 R2R-CE 上达到 70.1，在 RxR-CE 上达到 71.0。



(a) Rollout 奖励。 (b) R2R 与 RxR 上的成功率。 (c) 每条轨迹的决策次数。

图 5: **RL 训练动态**。总体而言，rollout 奖励与评估成功率上升，而每条轨迹的平均动作块决策次数下降。

4.5 优化

本节研究 RL 配置如何影响导航精度，以及模型的导航行为在训练过程中如何演变。表 6使用 Qwen3-VL 4B，以及由 R2R、RxR 和 ScaleVLN 训练场景构建的合成回合混合数据。

奖励设计。在批大小 32、每个提示八次 rollout 的 GRPO 设置下，本文比较 SPL 奖励与二值奖励。两者对失败回合均给予零奖励；成功回合分别获得 SPL 分数或奖励 1。二值奖励

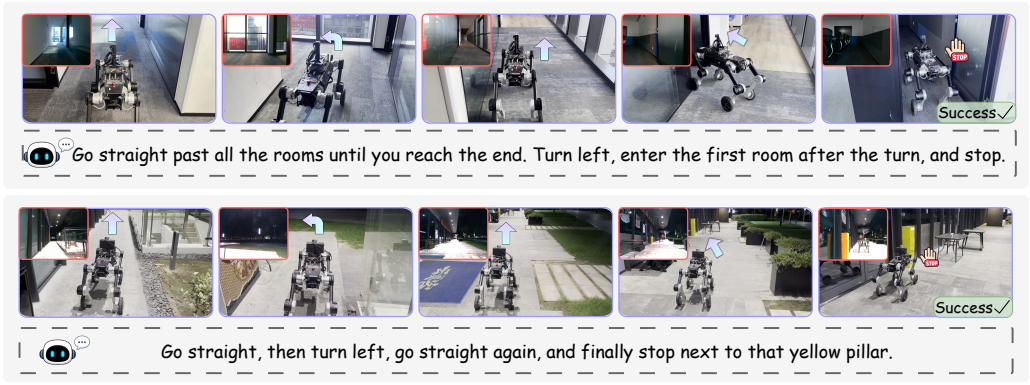
表 6: 使用 Qwen3-VL 4B 在 R2R-CE 和 RxR-CE *val-unseen* 划分上的**奖励设计与 RL 优化**。SPL 奖励对失败回合为零，对成功回合等于真值路径长度与实际执行路径长度之比；成功奖励为二值。

RL 配置	R2R-CE				RxR-CE			
	NE↓	OS↑	SR↑	SPL↑	NE↓	SR↑	SPL↑	nDTW↑
RL 前基线	4.26	74.3	65.4	58.8	4.79	63.4	54.1	66.5
GRPO + SPL(τ)	3.80	73.9	68.9	63.1	4.13	67.7	58.3	67.3
GRPO + $\mathbb{I}$(success)	3.74	73.1	70.1	65.6	3.82	71.0	62.1	68.4
DAPO + $\mathbb{I}$(success)	**3.45**	**76.6**	**71.5**	**65.3**	**3.39**	**73.7**	**64.0**	**69.4**

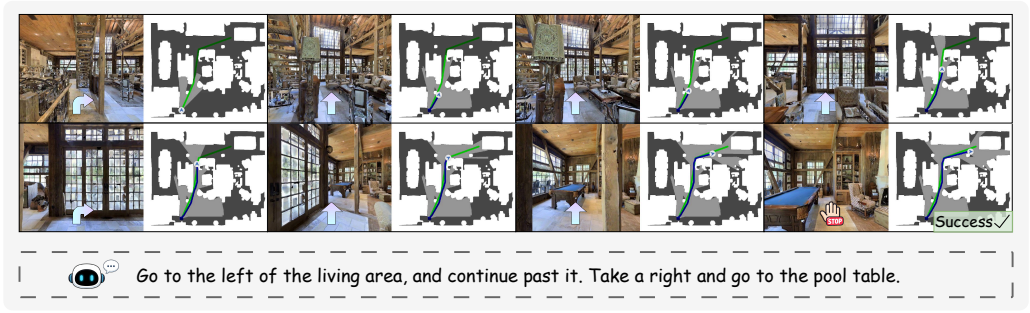
使 R2R-CE 上的 SR 从 68.9 提升至 70.1，RxR-CE 上从 67.7 提升至 71.0，支持采用简单的任务完成反馈。

RL 配置。在固定成功奖励的情况下，DAPO [20] 达到 71.5/73.7 SR，而 GRPO 为 70.1/71.0。两组消融均使用批大小 32，而第 4.1节的主结果配置使用 256。结果支持所评估的 DAPO 配置，但本文并不将增益完全归因于优化算法本身。

RL 训练动态。图 5显示，尽管存在波动，rollout 奖励与评估 SR 总体呈上升趋势，而每条轨迹的平均动作块决策次数逐渐下降。这表明 RL 在提升导航性能的同时也提高了决策效率。



(a) Qualitative results of SimpleVLN in the real world.



(b) Qualitative results of SimpleVLN on Simulation.

图 6: **SimpleVLN 在真实与仿真环境中的表现**。 SimpleVLN 能遵循多样的指令并到达指定目标；箭头表示导航方向。

4.6 真实世界实验

本文进一步使用 Daystar 四足机器人在真实环境中评估 SimpleVLN。模型在机载 NVIDIA Jetson Thor GPU 上运行。机器人配备一台 Intel RealSense D450 RGB-D 相机用于以自我为中心的感知，本文仅使用 RGB 作为观测。图 6显示，SimpleVLN 能成功遵循包含前进、转向、进入房间以及依据地标停止的多步指令，并在较长轨迹中保持连贯的指令遵循行为。

5 结论

本文提出 SimpleVLN，一个使原生 MLLM 适应 VLN 的框架。SimpleVLN 将单轮决策输入、全局–局部记忆与动作分块相结合，并使用由专家示范、依赖策略的 DAgger 经验和合成 RL 任务构建的训练数据。渐进式 IL–DAgger–RL 流程进一步将学习目标从动作模仿转向任务级结果优化。SimpleVLN 取得了优异的导航性能，表明无需任务专用架构设计，即可有效释放原生 MLLM 的导航能力。

参考文献

[1] Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian Reid, Stephen Gould, and Anton Van Den Hengel. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pages 3674–3683. IEEE, 2018.

[2] Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, et al. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*, 2025.

[3] Shizhe Chen, Pierre-Louis Guhur, Makarand Tapaswi, Cordelia Schmid, and Ivan Laptev. Think global, act local: Dual-scale graph transformer for vision-and-language navigation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16516–16526. IEEE, 2022.

[4] An-Chieh Cheng, Yandong Ji, Zhaojing Yang, Zaitian Gongye, Xueyan Zou, Jan Kautz, Erdem Bıyık, Hongxu Yin, Sifei Liu, and Xiaolong Wang. Navila: Legged robot vision-language-action model for navigation. *arXiv preprint arXiv:2412.04453*, 2024.

[5] Yue Hu, Avery Xi, Qixin Xiao, Seth Isaacson, Henry X Liu, Ram Vasudevan, and Maani Ghaffari. Longnav-r1: Horizon-adaptive multi-turn rl for long-horizon vla navigation. *arXiv preprint arXiv:2602.12351*, 2026.

[6] Jingzhi Huang, Junkai Huang, Wenxuan Song, Haoyang Yang, Hailong Huang, Haoang Li, and Yi Wang. Sedualvln: A spatially-enhanced dual-system for vision-language navigation. *arXiv preprint arXiv:2605.17249*, 2026.

[7] Jacob Krantz, Erik Wijmans, Arjun Majumdar, Dhruv Batra, and Stefan Lee. Beyond the nav-graph: Vision-and-language navigation in continuous environments. In *European Conference on Computer Vision*, pages 104–120. Springer, 2020.

[8] Alexander Ku, Peter Anderson, Roma Patel, Eugene Ie, and Jason Baldridge. Room-across-room: Multilingual vision-and-language navigation with dense spatiotemporal grounding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4392–4412, 2020.

[9] Bingqian Lin, Yunshuang Nie, Ziming Wei, Jiaqi Chen, Shikui Ma, Jianhua Han, Hang Xu, Xiaojun Chang, and Xiaodan Liang. Navcot: Boosting llm-based vision-and-language navigation via learning disentangled reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):5945–5957, 2025.

[10] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

[11] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y.K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[12] InternNav Team. InternVLA-N1: An open dual-system navigation foundation model with learned latent plans, 2025.

[13] Qwen Team et al. Qwen3.5: Towards native multimodal agents, february 2026. *URL https://qwen.ai/blog*, 2026.

[14] Shuo Wang, Yongcai Wang, Wanting Li, Xudong Cai, Yucheng Wang, Maiyue Chen, Zhizhong Su, Deying Li, and Zhaoxin Fan. Aux-think: Exploring reasoning strategies for data-efficient vision-language navigation. *Advances in Neural Information Processing Systems*, 38:30892–30916, 2026.

[15] Zun Wang, Jialu Li, Yicong Hong, Yi Wang, Qi Wu, Mohit Bansal, Stephen Gould, Hao Tan, and Yu Qiao. Scaling data generation in vision-and-language navigation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11975–11986. IEEE, 2023.

[16] Meng Wei, Chenyang Wan, Peng Peng, Xiqian Yu, Yuqiang Yang, Delin Feng, Wenzhe Cai, Chenming Zhu, Tai Wang, Jiangmiao Pang, et al. Ground slow, move fast: A dual-system foundation model for generalizable vision-language navigation. In *International Conference on Learning Representations*, volume 2026, pages 12380–12396, 2026.

[17] Meng Wei, Chenyang Wan, Xiqian Yu, Tai Wang, Yuqiang Yang, Xiaohan Mao, Chenming Zhu, Wenzhe Cai, Hanqing Wang, Yilun Chen, et al. Streamvln: Streaming vision-and-language navigation via slowfast context modeling. *arXiv preprint arXiv:2507.05240*, 2025.

[18] Zihao Xin, Wentong Li, Yixuan Jiang, Ziyuan Huang, Bin Wang, Piji Li, Jianke Zhu, Jie Qin, and Shengjun Huang. Agentvln: Towards agentic vision-and-language navigation. In *European Conference on Computer Vision*, pages 114–132. Springer, 2026.

[19] Xinda Xue, Junjun Hu, Minghua Luo, Shichao Xie, Jintao Chen, Zixun Xie, Quan Kuichen, Wei Guo, Zedong Chu, Mu Xu, et al. Omninav: A unified framework for prospective exploration and visual-language navigation. 2026:106216–106231, 2026.

[20] Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, et al. Dapo: An open-source llm reinforcement learning system at scale. *Advances in Neural Information Processing Systems*, 38:113222–113244, 2026.

[21] Zhuoyuan Yu, Yuxing Long, Zihan Yang, Chengyan Zeng, Hongwei Fan, Jiyao Zhang, and Hao Dong. Correctnav: Self-correction flywheel empowers vision-language-action navigation model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 18737–18745, 2026.

[22] Shuang Zeng, Dekang Qi, Xinyuan Chang, Feng Xiong, Shichao Xie, Xiaolong Wu, Shiyi Liang, Mu Xu, and Xing Wei. Janusvln: Decoupling semantics and spatiality with dual implicit memory for vision-language navigation. In *International Conference on Learning Representations*, volume 2026, pages 33001–33026, 2026.

[23] Jiazhao Zhang, Anqi Li, Yunpeng Qi, Minghan Li, Jiahang Liu, Shaoan Wang, Haoran Liu, Gengze Zhou, Yuze Wu, Xingxing Li, et al. Embodied navigation foundation model. In *International Conference on Learning Representations*, volume 2026, pages 127293–127322, 2026.

[24] Jiazhao Zhang, Kunyu Wang, Shaoan Wang, Minghan Li, Haoran Liu, Songlin Wei, Zhongyuan Wang, Zhizheng Zhang, and He Wang. Uni-navid: A video-based vision-language-action model for unifying embodied navigation tasks. *arXiv preprint arXiv:2412.06224*, 2024.

[25] Jiazhao Zhang, Kunyu Wang, Rongtao Xu, Gengze Zhou, Yicong Hong, Xiaomeng Fang, Qi Wu, Zhizheng Zhang, and He Wang. Navid: Video-based vlm plans the next step for vision-and-language navigation. *arXiv preprint arXiv:2402.15852*, 2024.

[26] Jiazhao Zhang, Gengze Zhou, Hale Yin, Yiyang Huang, Zixing Lei, Qihang Peng, Haoqi Yuan, Jie Zhang, Xudong Guo, Xiaoyue Chen, et al. Qwen-robotnav technical report: A scalable navigation model designed for an agentic navigation system. *arXiv preprint arXiv:2606.18112*, 2026.

[27] Lingfeng Zhang, Zeying Gong, Xiaoshuai Hao, Haoxiang Fu, Qiang Zhang, Mingliang Zhou, Hangjun Ye, Xiaojun Liang, Junwei Liang, and Wenbo Ding. Futurenav: Unified world-action modeling for vision-and-language navigation. *arXiv preprint arXiv:2606.30367*, 2026.

[28] Zekai Zhang, Weiye Zhu, Hewei Pan, Xiangchen Wang, Rongtao Xu, Xing Sun, and Feng Zheng. Activevln: Towards active exploration via multi-turn rl in vision-and-language navigation. *arXiv preprint arXiv:2509.12618*, 2025.

[29] Duo Zheng, Shijia Huang, Yanyang Li, and Liwei Wang. Efficient-vln: A simple yet strong baseline for efficient vision-language navigation. *arXiv preprint arXiv:2512.10310*, 2025.

[30] Weiye Zhu, Zekai Zhang, Xiangchen Wang, Hewei Pan, Teng Wang, Tiantian Geng, Rongtao Xu, and Feng Zheng. Navida: Vision-language navigation with inverse dynamics augmentation. *arXiv preprint arXiv:2601.18188*, 2026.