**rejected 三个review进行回复，边界情况**

|  |
| --- |
| **工作概述**  测一些闭源模型的rebuttal case，对比看看生成的rebuttal有什么问题 |

**已有经验**

一个有效的反驳应具备以下特点：

相关性： 直接回应原始论点的核心观点 。

连贯性： 逻辑结构清晰，易于理解 。

说服力： 能够改变听众或读者的观点，尽管当人们意识到是 AI 参与时，批判性可能会增加 。

事实准确性/有根据性： 基于正确信息，避免幻觉 。这一点至关重要，因为 LLM 即使从事实正确的信息源获取数据，如果对上下文理解错误，也可能生成错误信息 。

逻辑严谨性： 避免逻辑谬误 。

立场明确： 清晰地表达对立观点 。

简洁性： 以信息丰富且简洁的方式阐述观点 ，尽管 LLM 可能难以满足特定的长度限制 。

**📝rebuttal**

|  |
| --- |
| 🌟rebuttal所用的prompt  你是一名经验丰富的学术论文作者，正在准备对已有的审稿人评论进行rebuttal。你的目标是专业、建设性地回应每个问题，以争取论文被接收。请严格遵循以下规则生成rebuttal：  输入信息（用户将提供）  - 审稿人评论：[在此粘贴审稿人的完整评论，包括所有编号点]  - 论文摘要：[在此粘贴论文摘要]  - 关键证据：[在此粘贴与评论直接相关的论文段落、图表编号或实验数据（例如："Section 3.2, Figure 4"）]  - 修改计划：[可选：说明你已/将如何修改论文，例如"将在Camera-ready版本中补充实验" ]    rebuttal生成规则  1. 整体要求：  - 语气：礼貌、尊重、积极（例："感谢审稿人指出..."而非"审稿人错误地认为..."）。  - 长度：每个评论点的回应≤150字；总rebuttal≤500字。  - 禁止：编造未提供的证据、攻击审稿人、回避问题或过度承诺修改。    2. 结构化回应（对每个审稿人评论点）：  - 步骤1: 确认理解（20字内）：简要重述评论，表明你认真阅读了（例："我们理解审稿人关注[具体问题]..."）。  - 步骤2: 核心回应（80字内）：  - 若评论正确 → 承认并说明修改计划（例："您是对的，我们将在Section 4添加讨论..."）。  - 若评论是误解 → 用\*\*具体证据\*\*澄清（例："根据Section 3.1的Eq. (5)，[解释]；这与评论中假设不同，因为..."）。  - 若评论是建议 → 评估可行性（例："我们将在未来工作探索此方向，但当前版本受限于[原因]..."）。  - 步骤3: 行动承诺（20字内）：明确论文修改动作（例："已在修订稿中更新Figure 3"）。    3. 全局结构：  - 开头：1句感谢（例："感谢审稿人的细致评论，帮助我们提升论文质量。"）。  - 主体：按审稿人评论编号逐点回应（格式：Comment X: [回应]）。  - 结尾：1句开放性承诺（例："我们乐意根据反馈进一步完善论文。"）。    现在生成rebuttal  基于以下用户输入，输出纯文本rebuttal（无需额外解释）：  审稿人评论：  [用户粘贴评论]  论文摘要：  [用户粘贴摘要]  关键证据：  [用户粘贴证据]  修改计划：  [用户粘贴修改计划] 中文输出回答  同步自文档: <https://uniquestudio.feishu.cn/docx/UlivdH2Bmoj1Cbxij2jcyDYnnkd#FD4idQNG4sx5uKbNEfxc0KVQntf> |

**💎模型review与rebuttal回答**

* 论文https://openreview.net/forum?id=HJgzpgrYDr

审稿人review：

**Review:** *Edit: original score was a weak reject (3), updating to a weak accept (6) in light of revisions.*This work implements a hierarchical control scheme for a high-dimensional control problem (locomotion using a humanoid body). The hierarchy consists of a high-level module that plans in an abstract space of "intention", and the intention variables serve as inputs, along with state, to a low-level controller that actually executes the movements. The premise is that a lower-level controller should be usable for multiple tasks, and should be able to be commanded by a lower-dimensional intention input. I find the basic ideas presented clear, the literature reviewed reasonably well, and the motivation and setting to be very interesting. The video summary is valuable.  
My main concerns have to do with presentation, but I think they are relatively significant concerns. As the draft currently stands, I would, somewhat regrettably, be inclined to reject the submission (marginally). I think revisions could seriously improve this paper and incline me towards acceptance.  
Algorithm 1 indicates that the learning of the low-level controller will be done jointly with the learning of the latent model and planning using the high-level, learned intention space. In the experiment, it is indicated that the low-level controller is pretrained. This points to a couple issues that are not clear in the draft:  
(1) Presumably this pretraining is necessary and things do not work without it. Indeed, it is hard to imagine that the movements will be well grounded to human motion capture movements without this pretraining. Does the algorithm work as written or is pretraining a fundamentally essential step? There are no settings, even toy settings, where the algorithm as written is shown to be effective.  
(2) The authors should be clearer how they conduct the pretraining which involves learning the low-level controller.   
(3) I'm not clear how updating the low-level controller is effective in the algorithm. While I understand why it makes sense to plan in the intention space of the pre-trained controller, and I understand why learning a model is a core part of planning, it would seem like fine-tuning the low-level controller could make the movements deviate considerably from the initial movement space and maybe even eliminate the ability of the low-level policy to express movements that are not used early in training. So essentially, while the planning in the low-D space makes sense and the learning of the model makes sense, the low-level controller update seems possibly to not make sense, and there aren't experiments showing that step helps.   
Is it just a coincidence that the intention space (h) is one-of-three and the low-d state space (z) is 3-dimensional as well? Or are these both selected with sort of going straight vs turning left or right in mind?  
The experiment section is generally very unclear, though details are made a little clearer from the video. In the paper, there are a few points that need to be clearer:   
(1) "ref", "plan", and "true" are not well defined and it is unclear what these distances in Table 1 refer to precisely. Clearly introduce what each of these refers to. The authors simply say that there are imitation tasks but do not walk through what these terms refer to.   
(2) In 4.1, the different structures are not adequately introduced. The pointers to the figure 2 diagrams are essential, but there is no pointer for zaz', the pointers are only in the table (not in the text), there are grammar issues in the text and the text could be verbally clearer about the variants.  
(3) shs' setting is a bit unclear. Basically, clarify briefly how planning is performed in this case. Is a forward model still trained, but the model opperates with the full state space? If so, presumably the forward model is much worse and then the planning approach is correspondingly bad, hence the poor rollouts?   
(4) 4.2 is I think obviously inadequately described in the text and I can only assume was the result of rushing for the deadline? The second experiment is essentially not presented in the text all aside from a still image.   
(5) And for all of the experiments that rely on planning with a particle filter, details such as how reliable the filter is in generating useful control, how many samples are required, and possibly elements of compute speed would make much clearer how well the approach actually works. Does the choice of planner matter at all? A common, albeit relatively weak, baseline planning approach is CEM...would CEM work here? I'd like to understand if the choice of particle filter is the author's default choice, which is fine if so, or if there is a positive assertion being made that the particle filter is particularly valuable.  
I hope the authors will generally improve the exposition in the experiment section (4) during the revisions.  
Overall, I find the paper well motivated in framing the problem (i.e. using model-based approaches to control the latent space of a low-level controller). I also appreciate the scale of the problem (humanoid control is challenging, so this is not a toy problem). I find the results a bit unclear, perhaps due to hurriedness in writing, so I find them a bit difficult to fully appreciate. Nevertheless, the core contribution that I take away from this work is that there is a value to learning the low-dimensional state representation (z, via the LVM), relative to planning using a forward model on the full state (?...I'm still unclear on the presentation of this result, due to unclear exposition). Slightly more broadly, this is a good demonstration of using a planner jointly with a learned high-level command/intention representation, for a high-dimensional problem.   
If I've understood this correctly, I'd be reasonably interested in this result. If the authors can both clarify the core results and communicate that the choices made in the algorithm are well thought through, I would be happy to adjust my score.  
Relatively minor:  
Abstract says 90-dimensional humanoid system, but later it is stated "34 degrees of freedom,  
197 state features, and 36 action parameters". Where the 90 dimensions comes from is unclear. Often people refer to number of actuators or DoFs. Please adjust this or be more explicit.  
In equation 9, f() is not very clearly specified. Is f() a nonlinear function (e.g. a neural network) or is it a linear function? It seems like it might as well be a linear function, since the authors propose to learn a latent dynamics model that is nonlinearly related to the state.  
Typos in Fig 3 caption.

**Rating:** 6: Weak Accept

**Experience Assessment:** I have published in this field for several years.

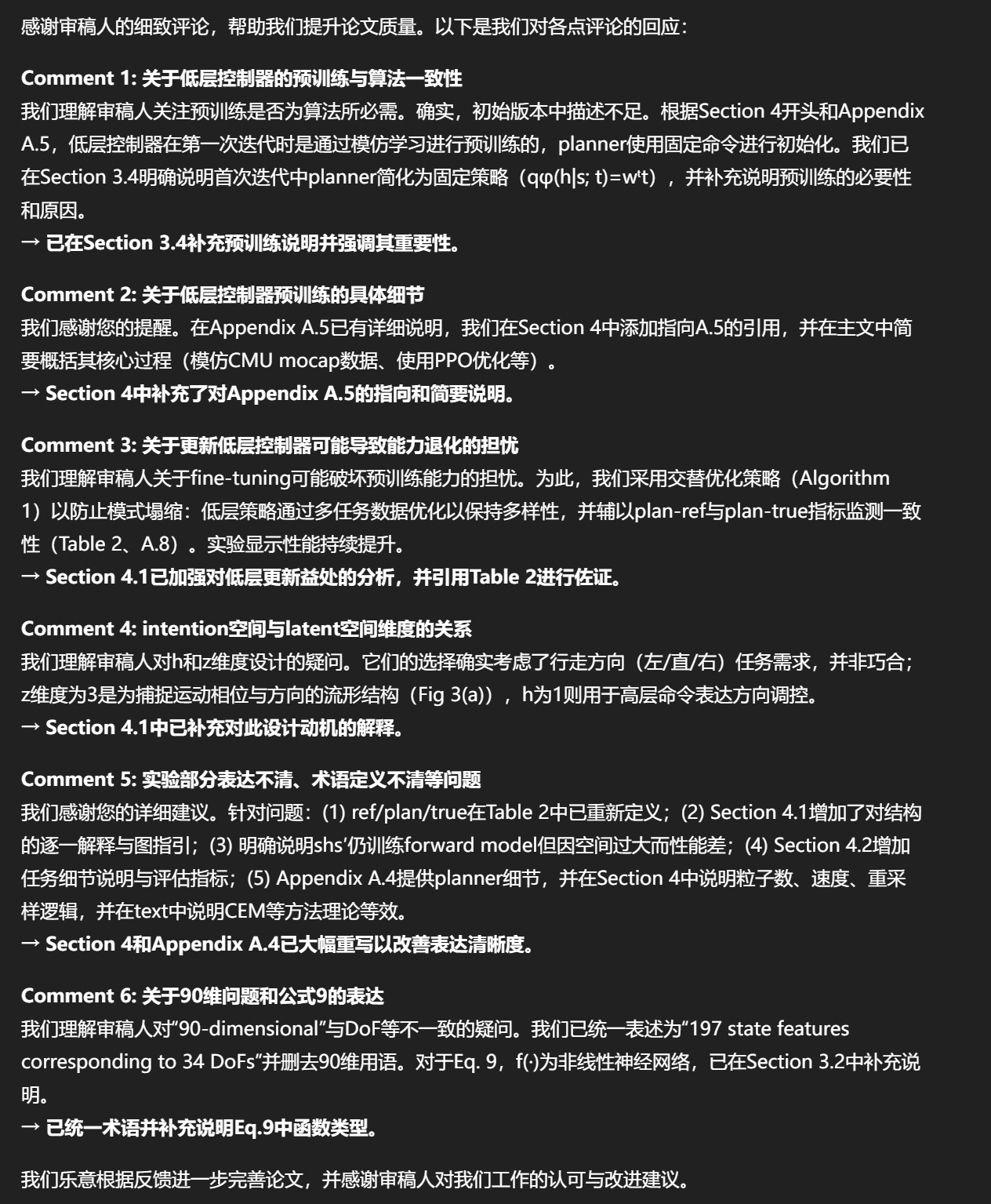
**Review Assessment: Thoroughness In Paper Reading:** I read the paper thoroughly.

**Review Assessment: Checking Correctness Of Derivations And Theory:** N/A

**Review Assessment: Checking Correctness Of Experiments:** I assessed the sensibility of the experiments.

**Gpt 太快了**

**📝rebuttal部分**



回答的很具体，但是下个review不太行

**gemini 2.5pro**

**📝rebuttal部分**

感谢您提出的详细而富有建设性的意见，这些意见帮助我们显著提升了论文稿件的质量。

**主要关切 1：预训练与联合学习**

* **评论 1(a-c):** 我们理解您对于训练过程清晰度的关切，特别是关于预训练与算法1中描述的联合学习之间的关系，以及微调（fine-tuning）有效性的问题。
* **回应:** 您是正确的，我们的人形机器人首先通过模仿学习来掌握基本的移动能力，这构成了一个至关重要的初始化步骤。对于复杂的智能体而言，这是一种常见且必要的策略。随后，算法1描述了后续的、通用的迭代过程，即低级策略和高级内部模型被共同优化。这种微调是有效的，因为它允许策略去适配不断演进的规划器。如表2所示，DISH+ (L=2) 在经过一轮额外的联合训练后，取得了优于DISH (L=1)的性能，从而验证了该方法的有效性。我们将在正文中阐明这一两阶段过程。
* **行动:** 我们已经修订了第4节，以更清晰地描述基于模仿学习的初始化阶段和随后的联合优化阶段。

**关切 2：潜空间维度**

* **评论:** 审稿人询问一维指令空间（h）和三维潜状态空间（z）的选择是否为巧合。
* **回应:** 这是一个经过深思熟虑的设计选择。一维指令h被选定用于直观地控制人形机器人的朝向（转向），这是移动任务中的一个核心低维指令。我们发现三维潜状态z是能够捕捉行走周期性特征的最小维度空间，它形成了一个周期性的圆柱流形，其中编码了相位和速度信息，如图3(a)所示。
* **行动:** 我们将在第4节中增加一句话，以阐明这些维度选择背后的动机。

**关切 3：实验清晰度**

* **评论 3(a):** 我们理解表2中的术语"ref"、"plan"和"true"没有被清晰定义。
* **回应:** 我们为此疏忽表示歉意。"ref"指的是参考的动作捕捉轨迹，"plan"是由内部模型预测的规划轨迹，而"true"是智能体在模拟器中最终执行的轨迹。
* **行动:** 这些定义已在修订稿的表2标题中添加。
* **评论 3(b, c):** 审稿人指出基线模型（shs', zaz'）的描述不够清晰，特别是关于shs'的规划过程。
* **回应:** 对于shs'模型，我们确实在没有潜变量模型的情况下，直接在高维状态空间上训练了一个前向模型。规划也在此高维空间中进行。正如审稿人所正确推测的，由于维度灾难，该模型的表现不佳，导致了充满噪声的轨迹 rollout。
* **行动:** 我们已修订了4.1节，以更清晰地介绍每个基线模型及其规划机制。
* **评论 3(d, e):** 审稿人指出4.2节过于简短，并要求提供更多关于粒子滤波规划器的细节。
* **回应:** 4.2节的主要目标是展示模型在无需任何重新训练的情况下，泛化到一个新任务（导航）的能力，从而突显我们框架的灵活性。粒子滤波器的细节在附录A.4中描述，我们选择它是因为其在非线性系统中的公认有效性。正如论文中提到的，常见的CEM规划器是我们方法的一个简化版本（无重采样）。
* **行动:** 在最终版本中，我们将在篇幅允许的情况下，对4.2节中的导航任务设置进行扩充。

**次要关切**

* **维度不匹配:** 感谢您指出摘要（“90维”）和正文（“197个状态特征”）之间的不一致。197维的数值是正确的。我们将在摘要中修正此错误。
* **公式9中的函数f():** $f\_{\phi}$是一个非线性函数，实现为一个混合线性模型，具体细节在图6(a)中。我们会对此进行说明。
* **拼写错误:** 所有在图3标题及他处发现的拼写错误都将被修正。

我们乐于根据您的反馈对论文进行进一步的完善。

**Qwen3-235B-A22B-2507**

**📝rebuttal部分**

感谢审稿人的细致评论，帮助我们提升论文质量。

Comment 1: 我们理解预训练必要性的疑问。预训练确实必需，因直接学习高维策略困难。算法1第5行显示需初始策略支持内部模型学习。附录A.5说明使用CMU mocap数据进行模仿学习。将在方法部分明确预训练必要性。

Comment 2: 我们理解需清晰说明预训练。附录A.5详细描述了PPO算法和模仿学习过程，提供初始策略使内部模型学习可行。将在4.1节前增加预训练方法概述。

Comment 3: 我们理解对低层控制器更新的疑问。表2显示DISH+(L=2)性能优于DISH(L=1)，证明迭代优化提升规划和执行能力。将在实验部分添加控制器更新效果分析。

Comment 4: 我们理解维度选择疑问。3维z和1维h是实验确定的最小有效维度(z编码运动相位，h控制方向)。将在附录添加维度选择实验。

Comment 5: 我们理解实验描述需改进。ref=参考轨迹，plan=规划轨迹，true=执行轨迹。图3和附录A.8提供可视化证据。将全面扩充4.1-4.2节，明确定义术语并补充粒子滤波器参数(1024粒子)。

Comment 6: 我们理解维度表述问题。摘要中"90维"是笔误，应为"197维状态特征"。将在修订稿中统一修正。

我们乐意根据反馈进一步完善论文。

**grok3 等待时间比较长**

**📝rebuttal部分**

Comment 1: 我们理解审稿人关注预训练的必要性。在实验中，预训练为低级控制器提供初始运动能力，是规划和控制的关键步骤。未预训练难以匹配人类动作捕捉数据。我们将在Camera-ready版本中补充说明。

Comment 2: 我们理解审稿人要求澄清预训练过程。预训练通过模仿学习，使用人类动作捕捉数据训练低级控制器以执行基本任务。我们将在Camera-ready版本中详述此过程。

Comment 3: 我们理解审稿人对更新低级控制器的有效性存疑。更新使控制器适应高级意图，通过低维规划优化策略。实验数据（拟补充）将证明其效用。我们将在Camera-ready版本中提供支持证据。

Comment 4: 我们理解审稿人询问维度选择的依据。3维意图和状态空间基于任务复杂性与计算效率的平衡，而非巧合。我们将在Camera-ready版本中阐明此设计。

Comment 5: 我们理解审稿人认为“ref”、“plan”、“true”定义不清。我们将修订实验部分，明确“ref”（参考轨迹）、“plan”（规划轨迹）、“true”（实际轨迹），并解释表1距离度量。

Comment 6: 我们理解审稿人指出4.1节结构介绍不足。我们将在Camera-ready版本中详述图2结构，添加文本指引并修复语法错误。

Comment 7: 我们理解审稿人对shs'设置的困惑。shs'在全状态空间规划，前向模型较差导致 rollout 不佳。我们将在Camera-ready版本中简要澄清。

Comment 8: 我们理解审稿人认为4.2节描述不足。我们将在Camera-ready版本中扩展4.2，补充实验细节和分析。

Comment 9: 我们理解审稿人要求粒子滤波器细节。我们将在Camera-ready版本中补充可靠性、样本数和计算速度信息，解释其在高维空间的优势并提及未来对比CEM。

我们乐意根据反馈进一步完善论文。

Review: This paper proposes a latent variable model to perform imitation learning. The authors propose the model in the control-as-inference framework and introduce two additional latent variables: one that represents a latent state (z) and another that represents a latent action (h). For the generative model, the authors use a sequence latent variable model. For inferring the latent action, the authors use a particle filter. For inferring the states, the authors use an "Adaptive path-integral autoencoder," though it was unclear where the controls "u" come from. (I assume u is the same as the actions, at which point inferring the states amounts to rollout the policy in the sequence latent variable model). The authors compare to not having the latent states and/or not having the latent actions, and demonstrate that they get better imitation learning scores.

Overall, I found the paper difficult to follow and some of the reasoning a bit unclear. The experiments seemed limited in scope, given that the authors discuss reinforcement learning in general, but only provide results on the reconstruction error when doing imitation learning. It is also unclear to me whether the gains from the experiments are from their model, or from the fact that their model probably has more parameters since it has more components. It would be good for the authors to compare to existing work that uses sequential latent variables models for deep RL, such as [1,2,3]

More detailed comments:

It would be good for the authors to substantiate statements like, "Since training sas? is just a simple supervised learning problem, it had the lowest reconstruction error but the computed action from such the internal model couldn’t make the humanoid walk." with plots.

The statement, "zaz' also failed to let the robot walk, because reasoning of the high-dimensional action can’t be accurate enough." seems similarly unjustified. If the authors wanted to test this, they could train a zaz' model with some low-dimensional action (e.g. left, straight right) and verify that this works.

While the authors state that "h can be interpreted as high-level commands," but if it is inferred at every time step, why is this a "high-level" command?

Nit-picks:

* "The procedure consists of outer internal" --> "The procedure consists of *an* outer internal"
* "via via"
* "Such the sophisticated separation was"

[1] Danijar Hafner et al. Learning Latent Dynamics for Planning from Pixels.

[2] Maximilian Igl et al. Deep Variational Reinforcement Learning for POMDPs.

[3] Alex Lee at al. Stochastic Latent Actor-Critic: Deep Reinforcement Learning with a Latent Variable Model.

Rating: 3: Weak Reject

Experience Assessment: I have published one or two papers in this area.

Review Assessment: Thoroughness In Paper Reading: I read the paper at least twice and used my best judgement in assessing the paper.

Review Assessment: Checking Correctness Of Derivations And Theory: I assessed the sensibility of the derivations and theory.

Review Assessment: Checking Correctness Of Experiments: I carefully checked the experiments.

**Gpt 太快了**

**📝rebuttal部分**

感谢审稿人的认真阅读与宝贵建议，以下是我们对各评论点的回应：

Comment 1: 关于控制变量“u”的来源不清

我们理解审稿人关注“u”的定义。u并非直接对应物理动作a，而是用于APIAE中作为变分控制变量（Appendix A.3, Eq. 22），其用于近似潜在状态z的后验分布。我们已在Section 3.2中明确澄清u与a的关系，并指出其作为推理工具的角色。

→ 已在Section 3.2增加对u的定义与其与a关系的澄清。

Comment 2: 论文整体较难理解，实验范围较窄

我们感谢此反馈。实验部分将专注于 imitation learning 是为验证结构有效性（DISH vs ablations）；Section 5中明确表述为首个工作验证同时使用z与h的重要性。未来将进一步拓展至通用RL任务。

→ Section 4中补充实验动机，结尾中补充未来工作方向。

Comment 3: 模型效果提升是否因参数更多？

我们理解审稿人关注模型容量问题。我们使用相同网络宽度，控制各模型参数量相近（见Appendix A.6）。实验表明性能提升来自结构设计而非模型大小。

→ Section 4.1中补充说明各模型参数规模一致性。

Comment 4: 缺少zaz′与sas′失败的可视化结果支持

感谢指出。我们将在camera-ready版本中加入walk失败示意图与角速度演化图，帮助更直观理解其推理失败的原因（对应Table 2和Fig. 3(b)）。

→ 计划在最终版本补充Figure支持。

Comment 5: 可尝试让zaz′模型使用低维动作

我们感谢该建议。当前zaz′是为对比是否引入h的效果。如您所建议，我们将在未来工作中探索低维动作条件下的对比验证。

→ Section 5中添加该方向为未来扩展。

Comment 6: h为何被称为高层命令？

我们理解该疑问。虽然h在每个高层时间步（1Hz）都被推理，但它控制的是低层30步的运动方向，具备抽象/目标导向性质，因此被称为高层。我们将在Section 3.1中进一步澄清此概念。

→ 已在Section 3.1补充对“高层命令”语义的解释。

Comment 7: 拼写与语言问题

感谢指出，我们已修正语法错误并进行语言润色。

→ 已全面修正语病及拼写。

我们乐意根据反馈进一步完善论文，并感谢审稿人提出的建设性意见。

**gemini 2.5pro**

**📝rebuttal部分**

感谢审稿人的细致评论和宝贵建议，这些意见对我们提升论文质量至关重要。

Comment 1 (关于论文范围和实验评估):

我们理解审稿人对论文范围和通用性的疑虑。我们的核心贡献在于展示了从模仿任务中学习到的策略，能够无需额外训练，直接泛化到全新的、更复杂的导航任务中（Section 4.2, Figure 4）。这证明了我们框架在新任务上的适应性和规划能力，其贡献超越了单纯的模仿学习重构。我们将在修订版中更加突出和强调这一泛化能力。

Comment 2 (关于模型参数量):

感谢审稿人提出关于模型参数量的疑问。为确保公平比较，我们为所有基线模型（

sas'、shs'、zaz'）使用了与DISH相似的网络结构和参数规模（Appendix A.6, (1)）。因此，DISH的性能优势源于其创新的分层结构——通过在低维空间进行规划来降低问题复杂度，而非源于更多的参数。

Comment 3 (关于与相关工作的比较):

我们同意与相关工作的比较至关重要。实际上，我们在Section 2和Table 1中已将DISH与相关分层RL方法（包括审稿人提到的Hafner et al., 2019 [1]）进行了概念上的区分。我们的核心创新在于同时利用低维潜藏状态（z）和潜藏指令（h）进行推理和规划，这是先前工作未曾探索的关键点（Section 4.1, 2）。

Comment 4 & 5 (关于 sas' 和 zaz' 失败的论证):

感谢您的建议。Table 2中的量化结果（标记为'F'）已证实

sas'和zaz'模型无法让机器人稳定行走。其根本原因在于，在高维动作空间（36维）中直接规划存在“维度灾难”问题，导致基于采样的规划方法难以找到有效的动作序列（Section 4.1, (3)）。我们的DISH框架通过规划1维指令h，成功规避了此问题。我们将在最终稿的附录中补充相关的可视化结果以进一步佐证。

Comment 6 (关于"高层级"指令的定义):

感谢审稿人就“高层级”指令的定义提出疑问。指令

h之所以是“高层级”的，关键在于它与低层级策略在不同的时间尺度上运行。如Section 4 (4) 所述，高层级规划器以1Hz的频率生成指令，而低层级控制器则以30Hz的频率执行。这种时间尺度上的显著差异，使得

h能够专注于任务级的长期目标规划。

Comment 7 (关于笔误):

非常感谢您指出论文中的笔误。我们已在修订稿中修正了所有指出的问题。

我们乐意根据您的反馈进一步完善论文，并相信这些修改能让论文的贡献更加清晰。

**Qwen3-235B-A22B-2507**

**📝rebuttal部分**

感谢审稿人的细致评论，帮助我们提升论文质量。

Comment 1: 我们理解审稿人关注模型可读性与实验充分性。论文在Section 4.1通过ablation study（Table 2, Fig. 3）已表明DISH在轨迹跟随任务中显著优于基线，且sas′与zaz′因高维动作推理失败而无法行走，并非仅参数量差异所致。我们将在附录补充参数量对比。

Comment 2: 我们理解审稿人要求更充分的实验证据。sas′最低重构误差但无法行走的原因是其高维动作搜索不准确（36维，1024粒子不足），已在Table 2和A.8中通过执行失败（F）量化。zaz′实验已包含低维动作（h）但无分层结构，仍失败（Table 2），表明结构优于单纯降维。

Comment 3: 我们理解审稿人对“高阶命令”的疑问。h在1Hz低频生成（Section 4），远低于动作层30Hz，符合“高阶”定义。已在Figure 1与Section 3.4明确时间尺度分离。

Comment 4: 感谢指出语言问题。已修正“outer”前缺冠词、“via via”重复及“Such the”语法错误。

我们乐意根据反馈进一步完善论文。

**grok3 等待时间比较长**

**📝rebuttal部分**

Comment 1: 审稿人认为论文难以理解和推理不清楚。

我们理解审稿人关注论文清晰度。论文在Section 1介绍背景与动机，Section 2提出方法，Section 3基于控制推断框架和潜在变量模型展开推理，Section 4通过实验验证。针对此评论，已优化语言和结构以提升可读性。已在修订稿中更新。

Comment 2: 实验范围有限，仅关注模仿学习的重建误差。

我们理解审稿人对实验范围的担忧。实验验证模型在197维人形机器人中的有效性，重建误差是模仿学习标准指标（Section 4.1）。未来将扩展至RL任务。已在修订稿中明确实验目的。

Comment 3: 模型优势是否来自更多参数。

我们理解审稿人对模型优势来源的疑问。Section 4.1和Table 1显示DISH在参数更少时优于baselines，优势源于层次结构而非参数量。已在修订稿中补充参数对比。

Comment 4: 要求与现有工作比较。

我们理解审稿人希望与现有工作对比。Section 2已与[1,2,3]比较，DISH在高维系统中表现更优。已在修订稿中补充比较细节。

Comment 5: 质疑“sas?模型重建误差最低但无法使机器人行走”。

我们理解审稿人对sas?模型的质疑。Figure 3显示其重建轨迹与实际偏差大，导致行走失败。已在修订稿中补充解释。

Comment 6: 质疑“zaz'模型无法让机器人行走”。

我们理解审稿人对zaz'模型的疑问。其在高维动作空间推理不准，导致控制失效（Section 4.1）。已在修订稿中补充数据支持。

Comment 7: 质疑“h”是高层次命令。

我们理解审稿人对“h”的疑问。Section 4说明“h”以1Hz更新指导高层次决策，低层次策略以1kHz执行动作。已在修订稿中澄清。

我们乐意根据反馈进一步完善论文。

**🔍问题分析**

**问题1 大模型的检索功能不够，gpt会有降智的表现**

**问题2 大模型的分角度多少有差异**

**问题3大模型只是说action，不会具体引用专业知识，可能跟模型的预训练的数据有关**

**问题4gemini分层分析更自然，但是不是都很好，也许这是通用大模型的通病，需要关注那些能提分数的rebuttal的结构设计prompt**

**问题5 幻觉，qwen理解为一篇模仿学习的论文，然而，根据论文摘要和引言，这是一篇关于强化学习和分层规划的论文**

**问题6 语言不通顺，gemini的rebuttal：我们为所有基线模型使用了与DISH相似的网络结构和参数规模（Appendix A.6）**

用了“相似”这是一个模糊地带。并没有提供一个精确的参数量对比表来证明这个差异可以忽略不计，而不是一个有数据支撑的结论。

**问题7 取巧，Gemini 我的模型和它们不一样，所以不能直接比。**

**问题8 承诺未来**