arXiv:2602.15460v1 [cs.LG] 17 Feb 2026

多模态大语言模型在简单视觉规划任务中推理的分布外泛化

Yannic Neuhaus [1] **Nicolas Flammarion** [2] **Matthias Hein** [1] **Francesco Croce** [3]

摘要

将推理融入大型语言模型和大型视觉语言模型，最近显著提升了它们的能力。然而，推理模型的泛化能力仍缺乏明确界定，也尚未得到充分理解。本文提出一个评估框架，严格考察思维链（CoT）方法在简单规划任务上的泛化表现。具体而言，我们考虑一种基于网格的导航任务：模型获得一张地图，并须输出一系列移动指令，引导玩家从起始位置抵达目标，同时避开障碍物。该任务及其数据具有多样性，因此我们可以使用不同的输入表示形式（视觉和文本）及思维链推理策略对模型变体进行微调，并在分布内（ID）和分布外（OOD）测试条件下对其进行系统评估。实验表明，尽管思维链推理提升了所有表示形式的分布内泛化能力，但在排除与分布内数据简单匹配的情况后，大多数情形下的分布外泛化能力（例如，泛化到更大的地图）仍十分有限。值得注意的是，我们发现，结合多种文本格式的推理轨迹能够实现最佳且非平凡的分布外泛化。最后，纯文本模型的表现始终优于使用基于图像输入的模型，其中也包括一种近期提出的、依赖潜空间推理的方法。

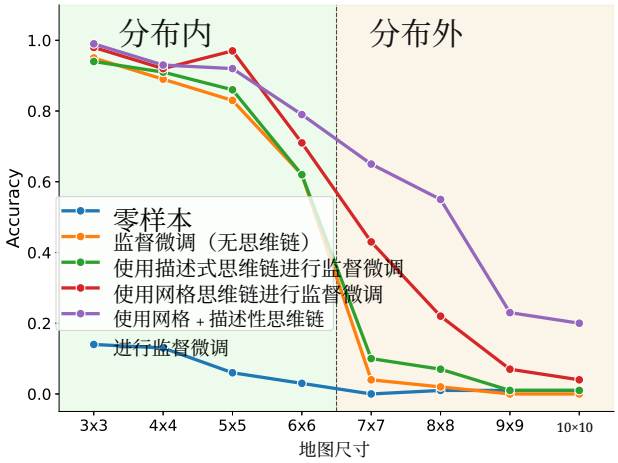


图*1.* **思维链（CoT）格式会影响分布外（OOD）泛化。** 我们在 FROZENLAKE 数据集上对 Qwen2.5-VL-7B-Instruct 进行微调，数据集包含大小最大为6×6的地图，采用网格格式。未使用 CoT 训练的模型，或使用描述格式 CoT 训练的模型，都无法泛化到分布外（更大的地图）；而将网格与描述推理步骤结合起来，则在地图大小达到 10×10 之前都能取得有意义的结果。这表明，推理的泛化能力会受到思维链轨迹格式的影响。

1. 引言

思维链（CoT）推理已被证明是提升大型语言模型（大语言模型）解决复杂任务能力的有力工具（Wei 等人，2022）。在实践中，测试时提示（“让我们一步一步地思考”）（Kojima 等人，2022）或专门设计的微调方案（Ho 等人，2023）都能促使大语言模型生成中间推理步骤，从而帮助得出正确的解，而非直接给出答案。这种方法还有助于提升可解释性，因为用户可以检查回答背后的推理过程以及潜在的失效点。近期，思维链范式已扩展至多模态模型，以增强其对视觉输入进行推理的能力（Huang 等人，2025；Yang 等人，2025）。尽管思维链已广泛取得成功并得到应用，但人们仍根本不了解思维链为何有效，以及推理型大语言模型学到了什么。例如，近期研究表明，当前的推理能力可能主要反映数据的统计特性，而非真正的算法学习；当输入偏离训练分布时，性能会下降，这为此提供了佐证（Stechly 等人，2024；Zhao 等人，2025）。

在这项工作中，我们希望系统地研究

[1]蒂宾根大学人工智能中心 – 蒂宾根大学 [2]EPFL [3]ELLIS 芬兰研究所 – 阿尔托大学。通信作者：Yannic Neuhaus <yannic.neuhaus@uni-tuebingen.de>。

预印本。*2026*年*2*月*18*日。

Image



Description

```
This is a 3x3 map.
The player is at: row 3, column 2;
The hole(s) are at: Row 2, Column 2;
Row 3, Column 3;
The goal is at: Row 1, Column 2.
```

Table

```
|       | Col 1 | Col 2 | Col 3 |
| Row 1 |   _   |   *   |   _   |
| Row 2 |   _   |   #   |   _   |
| Row 3 |   _   |   @   |   #   |
```

Grid

```
Grid:
.  G  .
.  X  .
.  P  X
```

图 2. **迷宫表示**。 在 Wu等人 （2024）的工作中，规划任务通过迷宫的多种表示形式加以介绍：包括 图像、文本 描述 和 表格。我们还引入了一种基于 ASCII 的 网格 表示形式，用更少的词元对地图进行编码。

大语言模型在受控设置下的分布内（ID），尤其是分布外（OOD）泛化行为，并在一项简单但具有挑战性的任务中进行研究，以便区分输入表示、推理和思维链格式等因素的影响。然而，使用常见基准测试进行如此全面的评估并不容易（Lu等人，2022；Yue等人，2024），因为这些任务并未明确区分分布内与分布外，也没有显而易见的、可供学习的真实算法。因此，我们首先基于 FROZENLAKE 数据集构建一个受控评估环境（Wu等人，2024）。这项任务要求进行视觉规划，找到正确路径以抵达宝藏，同时避免掉入湖泊（见图2中的示意图）：即便是小地图，对最先进的大语言模型而言也颇具挑战性（Wu等人，2024），因此这是一个相关的测试平台。此外，地图通常以图像形式呈现，但也可以用纯文本格式表示（图2），推理步骤亦然（通过沿着通往目标的路径逐步移动来表示）。至关重要的是，任务复杂度可以通过地图尺寸、起点与目标之间的距离、湖泊数量等轻松控制，从而使我们能够区分分布内和分布外数据。

随后，我们训练了多种标准模型（无思维链）和推理模型。这些模型共享相同的训练地图，但采用不同格式对其进行编码。我们使用四种输入格式： 图像、描- 述、表格和 网格（后三种为基于文本的格式）；而推理轨迹则可以采用其中一种基于文本的格式，或由多种格式组合而成（图 3）。总体而言，在分布外数据上，即地图尺寸更大时，模型性能会大幅下降；尤其是当起点和目标位置之间的距离大于训练数据中的距离时，下降更为明显。尽管思维链推理能够全面提升性能，但并未解决分布外泛化的局限性。值得注意的是，我们发现，唯一能够在规模达到 10×10地图时仍表现出非平凡性能的模型，使用了我们结合 网格/表格 + 描述 的思维链（见图 1），揭示了格式与泛化之间有趣的联系。即使在分布内测试地图上，这些模型也取得了最佳结果，优于近期依赖连续空间推理的复杂方法（Yang 等人，2025）以及专门的纯视觉模型（Xu 等人，2025）。我们还在固定地图尺寸和最优解长度的条件下，进一步分析了泛化与起点-目标距离之间的关系，证实了分布偏移的困难性以及采用恰当思维链格式所带来的益处。

总体而言，我们的研究表明，推理的分布外泛化仍然受到严重限制，并且似乎更多地依赖模式匹配，而非算法学习。然而，我们的结果也表明，特定的（多模态）数据格式可能带来更强的性能，从而为未来的研究开辟新的方向。我们引入的基准测试将成为衡量和提升未来大语言模型泛化能力的关键工具；其数据具有灵活性，因此随着更强大模型的出现，还可以加入更加复杂的层次。

2. 相关工作

推理型大语言模型在各种任务上取得了显著成果（Chen 等人，2025），理解其运行机制与局限性是当前活跃的研究领域（Shojaee 等人，2025；Mirzadeh 等人，2025；Wang 等人，2025）。特别是，Stechly 等人（2024）研究了最先进的大语言模型在文本规划领域和简单合成任务中的性能。在他们的研究中，思维链示例被加入上下文，而非用于微调。他们观察到，在各项任务中，思维链提示带来的提升有限，且仅限于这些提示与目标测试示例非常相似的情况。因此，Stechly 等人（2024）认为，模型无法通过此类示例学会通用的算法程序。Zhao 等人（2025）推出了 DATAALCHEMY，这是一个可控环境，通过简单的符号输入和变换对语言任务进行抽象。借此，他们能够控制从头训练的 Transformer 和微调后的大语言模型在不同分布偏移下的分布外泛化表现。思维链推理在分布内数据上效果显著，但在中等程度的分布偏移下性能便迅速下降，这再次表明，模型学会的只是识别训练集中的模式，而非逻辑结构。在本研究中，我们拓展了这条关于思维链推理分布外泛化的研究路线：我们的任务对人类而言微不足道，却对最先进的大语言模型颇具挑战；这些任务既适合从头训练，也适合微调，并且可以扩展到任意难度。与 DATAALCHEMY 等先前的基准测试不同，我们的数据集能够让我们研究输入与思维链格式之间的相互作用，涵盖文本和图像表示。近期，多模态推理模型领域已开始关注这一问题（Zhou 等人，2025）。这表明，某些设置下可以实现显著的分布外泛化。

	Image	Descriptive	Table	Grid
Step 1		This is a 3x3 map. The direct path to the goal is moving UP. However, there is a hole UP, so I cannot move UP. As the direct path is blocked, I will move LEFT instead.	| | Col 1 | Col 2 | Col 3 | | Row 1 | _ | * | _ | | Row 2 | _ | # | _ | | Row 3 | @ | _ | # |	Grid: . G . . X . P . X
Step 2		The direct path to the goal is moving UP or RIGHT. Thus, I will move UP.	| | Col 1 | Col 2 | Col 3 | | Row 1 | _ | * | _ | | Row 2 | @ | # | _ | | Row 3 | _ | _ | # |	Grid: . G . P X . . . X
Step 3		The direct path to the goal is moving UP or RIGHT. However, there is a hole RIGHT, so I cannot move RIGHT. Thus, I will move UP.	| | Col 1 | Col 2 | Col 3 | | Row 1 | @ | * | _ | | Row 2 | _ | # | _ | | Row 3 | _ | _ | # |	Grid: P G . . X . . . X
Step 4		The direct path to the goal is moving RIGHT. Thus, I will move RIGHT.	| | Col 1 | Col 2 | Col 3 | | Row 1 | _ | @ | _ | | Row 2 | _ | # | _ | | Row 3 | _ | _ | # |	Grid: . P . . X . . . X

图*3.* **不同格式的推理轨迹**。 我们展示了一个示例推理轨迹，从第一步到最终解，涵盖了多种格式。 Xu 等人(2025) 将迷宫表示序列用作图像，而我们则将对应步骤生成纯文本描述、表格和网格。在描述格式中，我们在每一步讨论下一步应如何进行；对于其他格式，我们则展示下一次移动后地图的可视化表示。

3. 用于评测推理模型泛化能力的受控环境

3.1. 任务

我们研究 Wu 等人（2024）提出的一项空间规划任务，名为 FROZENLAKE。在这项任务中，模型必须引导玩家穿过一个完全可观测的迷宫。输入包括对游戏规则和目标的文本描述，以及迷宫的基于网格的地图；其中一个格子里有玩家，另一个格子里有宝藏，其余格子则为空或有障碍。模型的输出是一个移动序列（“上”、“下”、“左”、“右”），应引导玩家在不掉进洞里的情况下到达宝藏。这种设定为我们的分析提供了若干优势：输入地图既可以轻松表示为图像，也可以表示为文本；而这一简单规划问题的解具有清晰的视觉和描述性结构。此外，任务难度可以通过*(i)* 地图大小和*(ii)*起点与宝藏之间的距离进行调整：这些可控的难度因素使我们能够系统地考察不同微调方法的分布外泛化。最后，尽管这项任务对人类玩家而言很简单，即使是最先进的——多模态大语言模型若不经微调，难以完成该任务（Wu等人，2024），例如，GPT-4o的准确率为46%（Hurst等人，2024），Gemini的准确率为29%（Gemini团队，2023）（在小型 3×3、4×4、 5×5和 6×6 地图上的平均准确率），见表格9。

3.2. 迷宫表示

在Wu等人（2024）中，规划任务采用多种迷宫表示方式：图像，非结构化文本描述以及 Markdown 格式的 ASCII表格。我们还引入了一种基于 ASCII 的网格表示，它同样具有类似表格的结构，但编码地图所需的词元更少。如图2所示，这四种迷宫表示方式使我们能够将相同的数据同时用于纯文本和多模态大语言模型。

3.3. 推理轨迹表示

任务的简单性使得通往正确解答序列的推理过程十分直接，因为每一步都对应着沿正确路径前往宝藏的一步（关于推理轨迹构造的更多细节，请参见附录。A）。这一点，加上数据的多样性，使得我们能够生成多种不同格式的思维链

r不同格式的推理轨迹（见图。 3):

• 图像： Xu 等人 (2025) 使用纯视觉模型，其推理轨迹由图像组成，这些图像展示了每次移动后的当前迷宫状态。虽然我们在实验中没有直接采用这种格式，因为标准（多模态）大语言模型无法在推理过程中处理图像，但它或许有助于未来将视觉元素整合到思维链的研究。

• 描述：这是对每一步推理过程的叙述，讨论哪些移动可以让玩家更接近宝藏、通往宝藏的路径是否被湖泊阻挡、哪些步骤可行，最后决定下一步该怎么走。

• 表格 和网格：对于通往宝藏路径上的每一步，我们使用前文讨论的两种基于文本的表示，来呈现下一步移动后的当前地图。

• 表格 +描述、网格 + 描述：我们将表格 或网格推理轨迹与描述结合起来。因此，思维链首先会以叙述的方式描述推理过程，说明下一步该怎么走；然后给出执行所选移动后的地图表格 或网格表示。这些格式包含更多信息，但代价是内容更长（因此推理时间的生成成本也更高，见表格5）。

3.4. 衡量分布外泛化

我们从多个维度衡量模型的分布外泛化能力，这些维度以不同方式控制任务的难度。

• **地图尺寸：** 首先，通过使用一系列不同地图尺寸进行训练，并在训练期间未见过的尺寸上进行测试，我们可以评估模型是否能泛化到更大的地图。请注意，较大的地图可以包含与较小地图相同的迷宫，只需用空单元格或障碍填充即可。

• **起点-目标距离：** 接下来，我们考虑起点与目标之间的 L_∞ 距离，即玩家坐标（s_1，s_2）与宝藏坐标（g_1，g_2）之间的距离

$$d_\infty = \max\{|s_1 - g_1|, |s_2 - g_2|\}.$$

例如，在 3×3 图 2中的示例里，该距离为 2。一般而言，尺寸为 $n \times n$ 的地图有 $d_\infty \leq n-1$。因此，我们可以测试模型对起点-目标距离大于训练时所见距离的地图的分布外泛化能力。

• **最优解长度：** 最后，我们可以通过 A* 算法计算最优路径的长度，并分析模型在最优解路径比训练集中的路径更长的地图上的表现。

这些因素彼此相关（例如，地图越大，起点-目标距离通常越远，解也越长），但它们捕捉的是不同类型的分布偏移，因此使我们能够研究大语言模型的不同方面。

4. 实验

4.1. 设置

数据。训练时，我们使用 Yang 等人（2025）介绍的数据集，其中包含 1,000 个数据点（100 3×3、200 4×4、 300 5×5 和 400 6×6 地图）。因此，训练集仅包含距离 $d_\infty \leq 5$。对于每个训练样例，我们都生成第 3 节所述所有格式的输入和推理轨迹。分布内测试集（3×3、 4×4、 5×5 和 6×6 地图）也与 Yang 等人（2025）中的完全相同，而分布外测试集则由我们使用 Gym库（Brockman 等人，2016）按照附录 A 中的说明生成。对于相应的分布外测试集，我们针对每种地图尺寸（或嵌入式地图情形中的距离）各生成 200 张地图。在评估过程中，我们从模型输出中提取移动序列，并在环境中模拟该序列。如果满足以下条件，则该序列正确：(i) 玩家没有进入有障碍的格子；(ii) 终点位置位于宝藏处。代码和数据可在 http://github.com/YanNeu/frozen_ood 获取。

模型。 我们采用 Qwen2.5-VL-7B- Instruct（Bai 等人，2025）作为基础模型，并遵循 Yang 等人（2025）的设置，包括其提示词和规则规范。所有模型均通过监督微调进行适配，训练 10 轮，详情见附录。A 。对于每种输入格式，我们测试 Qwen2.5-VL-7B-Instruct 的零样本性能，以及其在不使用推理轨迹（无思维链）的情况下微调后的版本。随后，我们训练模型，使输入和思维链采用相同或不同的格式（生成模型的完整列表见表格1）。

4.2. 分布内性能

首先，为了测试各模型的分布内泛化能力，我们测量它们在与训练集（从 3×3 到 6×6）尺寸相同的地图上的性能，并将结果报告在表格1中。我们发现，各种输入格式下的零样本性能仍然较低，这证实了即使是小地图，对通用大语言模型而言也是一项具有挑战性的任务；而且，基础模型的训练集中不包含 FROZENLAKE数据集，因此它非常适合用于测试分布外行为。在不同输入格式下，任何形式的思维链推理都比仅针对答案进行微调（无思维链模型）能取得更高的准确率。图像输入（无论是否使用思维链）的表现都始终逊于纯文本输入，这凸显了当前多模态大语言模型在充分利用视觉输入方面的局限性。在其他格式中，输入与思维链表示的所有组合都能有效提升性能，其中表格/网格 输入与表格/网格 +描述思维链结合时，平均准确率最高（91%）。这类组合推理轨迹尤其

表*1.* **数据格式对分布内和分布外泛化的影响**。 我们报告了零样本和微调后的大语言模型在不同输入格式与思维链轨迹（如有）格式组合下，于分布内和分布外地图尺寸上的准确率。对于分布外地图，我们区分了起点与目标位置之间的距离为 $d_\infty \geq 6$ 的测试集，即训练期间未见过此类距离。采用混合格式（网格/表格 + 描述）的思维链在分布内和分布外均取得了最佳性能；即使训练时最多使用 6×6 张地图，在 10×10 张地图上仍保持了不俗的准确率。

输入格式	CoT格式	分布内测试地图（$d_\infty \leq 5$）					分布外地图（随机 d_∞）					分布外地图（$d_\infty \geq 6$）				
		3x3	4x4	5x5	6x6	Avg	7x7	8x8	9x9	10×10	Avg	7x7	8x8	9x9	10×10	Avg
图像	零样本	0.14	0.05	0.03	0.02	0.06	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00
	无思维链	0.89	0.84	0.72	0.52	0.74	0.51	0.37	0.23	0.12	0.31	0.02	0.00	0.00	0.00	0.01
	描述	0.95	0.87	0.76	0.62	0.80	0.47	0.35	0.21	0.14	0.29	0.07	0.03	0.00	0.01	0.03
描述	零样本	0.41	0.25	0.23	0.13	0.26	0.12	0.14	0.10	0.10	0.12	0.01	0.06	0.04	0.03	0.03
	无思维链	0.92	0.91	0.86	0.64	0.83	0.65	0.55	0.49	0.41	0.52	0.19	0.08	0.04	0.02	0.08
	描述	0.94	0.92	0.90	0.67	0.86	0.67	0.63	**0.55**	**0.46**	0.58	0.07	0.02	0.01	0.01	0.02
	Grid	0.96	0.89	0.90	0.70	0.86	0.59	0.47	0.28	0.12	0.36	0.35	0.17	0.06	0.01	0.15
	表格	**0.99**	0.90	0.93	0.69	0.88	0.64	0.45	0.33	0.16	0.40	0.28	0.16	0.07	0.02	0.13
	网格 +描述	0.98	0.93	0.93	0.69	0.88	0.70	0.51	0.41	0.26	0.47	0.44	0.28	0.11	0.07	0.23
	表格 + 描述	0.93	0.93	0.93	0.71	0.88	0.66	0.52	0.40	0.32	0.47	0.31	0.21	0.12	0.09	0.18
表格	零样本	0.43	0.24	0.20	0.08	0.24	0.09	0.08	0.04	0.04	0.06	0.04	0.03	0.01	0.01	0.02
	无思维链	0.94	0.89	0.83	0.59	0.81	0.56	0.41	0.35	0.27	0.40	0.02	0.01	0.00	0.00	0.01
	描述	0.96	**0.94**	0.83	0.69	0.85	0.57	0.39	0.34	0.23	0.39	0.04	0.03	0.00	0.00	0.02
	表格	**0.99**	0.90	0.93	0.72	0.89	0.73	0.51	0.35	0.19	0.45	0.38	0.24	0.07	0.03	0.18
	表格 + 描述	0.97	0.93	0.95	0.78	**0.91**	0.81	0.57	0.41	0.27	0.51	0.48	0.32	0.12	0.06	0.25
Grid	零样本	0.14	0.13	0.06	0.03	0.09	0.02	0.01	0.00	0.02	0.01	0.00	0.01	0.01	0.01	0.01
	无思维链	0.95	0.89	0.83	0.62	0.82	0.55	0.39	0.34	0.26	0.38	0.04	0.02	0.00	0.00	0.01
	描述	0.94	0.91	0.86	0.62	0.83	0.53	0.41	0.35	0.25	0.39	0.10	0.07	0.01	0.01	0.05
	Grid	0.98	0.92	**0.97**	0.71	0.90	0.74	0.51	0.39	0.24	0.47	0.43	0.22	0.07	0.04	0.19
	网格 +描述	**0.99**	0.93	0.92	**0.79**	**0.91**	**0.83**	**0.68**	0.49	0.35	**0.59**	**0.65**	**0.55**	**0.23**	**0.20**	**0.41**

与分布外泛化相关，接下来我们将对此展开讨论。

4.3. 分布外泛化

地图尺寸。 我们的第一个分布外测试集包含训练期间未见过的更大地图尺寸，即 7×7、 8×8、9×9和 10×10。与分布内数据类似，这些地图通过随机采样玩家和宝藏的初始位置生成，不控制起点与目标之间的距离（随机 d_∞ 结果见表 1）。这意味着起点、目标和解可以放入尺寸小于或等于 6×6的地图中，然后嵌入更大的地图。这些示例在地图尺寸方面属于分布外，但在起点-目标距离方面并非如此，如表 2所示。与分布内情形类似，基于文本输入的模型表现显著优于基于图像的模型。输入和思维链表示采用相同格式，相较于无思维链能够提升表现，尤其是 描述 格式的平均准确率达到 58%。同样，网格 输入搭配多模态 网格 + 描述思维链（59%）的表现略胜一筹。图 5 展示了该模型生成的思维链示例：即使面对尺寸为8×8 的分布外地图，它也会先用自然语言推理下一步移动，然后正确生成执行该移动后更新的地图。

起点-目标距离。如前所述，在之前的测试集中，我们不限制 d_∞，因此地图尺寸较大的样本分布外集合中的 d_∞ 可能仍然较小，且在训练期间出现过。然后，我们为分布外地图尺寸生成新的测试集，并强制满足 $d_\infty \geq 6$，即玩家与宝藏之间的距离大到无法容纳在任何一张训练地图中（表1右栏）。在这种更具挑战性的分布偏移下，大多数模型在 7×7 地图上的准确率就已降至 10% 以下，并且在更大的地图上迅速接近 0%。尤其是，输入和思维链均采用描述格式的模型，在随机 d_∞的分布外地图上取得了第二好的结果，但在 7×7 地图上的准确率骤降至 7%。这表明，我们的评估设置通过将 d_∞ 与地图尺寸一同控制，能够揭示泛化能力的严重局限。令人惊讶的是，对于所有基于文本的输入，将表格或网格与描述式思维链结合起来都能取得最佳结果。尤其是，采用网格输入、网格 + 和描述式思维链的模型，在 10×10 地图上的准确率仍有 20%，在具有 $d_\infty \geq 6$的分布外测试集上平均准确率达到 41%。这些结果表明，所有模型的分布外泛化通常仍面临挑战，但选择合适的数据格式可以带来改进。我们推测，在每一步生成当前地图的“视觉”表示（如表格或网格），有助于追踪地图导航的进度；而自然语言推理（描述式思维链部分）则更适合模型详细阐述下一步移动。网格格式也更紧凑，不含额外的

表格 2. **平均起点—目标距离**。 我们报告了每种地图尺寸下起点和目标位置之间的平均距离 d_∞。对于分布外地图 $d_\infty \geq 6$，我们确保地图尺寸和起点-目标距离相较于训练数据均属于分布外。

	分布内测试地图（$d_\infty \leq 5$）				分布外地图（随机 d_∞）				分布外地图（$d_\infty \geq 6$）			
	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10×10	7x7	8x8	9x9	10×10
均值 d_∞	1.50	1.79	2.16	2.75	3.27	3.66	4.01	4.47	6.00	6.35	6.72	7.00

表 3. **起点-终点距离（d_∞）固定时的泛化 10×10 地图尺寸**。我们比较了训练地图尺寸固定为 10×10 且 $d_\infty < 6$ 时，不同输入与思维链格式的表现；测试地图尺寸为 10×10，且 $d_\infty \in \{2, \ldots, 5\}$（分布内）或 $d_\infty \in \{6, \ldots, 9\}$（分布外）。使用网格输入以及网格或网格 + 描述思维链格式可取得最佳结果，而无思维链或采用其他格式思维链的模型，准确率在 $d_\infty = 6$ 时就已降至接近 0%。

输入格式	CoT 格式	分布内测试地图 ($d_\infty \leq 5$)					分布外地图 ($d_\infty \geq 6$)				
		2	3	4	5	Avg	6	7	8	9	Avg
图像	无思维链	0.70	0.46	0.21	0.03	0.35	0.03	0.01	0.00	0.00	0.01
	描述	0.69	0.41	0.16	0.06	0.33	0.01	0.00	0.00	0.00	0.00
描述	无思维链	0.82	0.75	0.57	0.34	0.62	0.01	0.01	0.00	0.00	0.01
	描述	**0.89**	0.80	0.68	0.41	0.69	0.00	0.00	0.00	0.00	0.00
	Grid	0.82	0.76	0.67	0.49	0.69	0.34	0.20	0.10	0.04	0.17
	网格 + 描述	0.82	0.73	0.61	0.47	0.66	0.36	0.25	0.17	**0.14**	0.23
表格	无思维链	0.88	0.77	0.54	0.16	0.59	0.01	0.00	0.00	0.00	0.00
	描述	0.89	0.76	0.47	0.09	0.55	0.00	0.01	0.01	0.00	0.00
Grid	无思维链	0.86	0.81	0.55	0.12	0.58	0.00	0.01	0.00	0.00	0.00
	描述	0.84	0.78	0.45	0.10	0.54	0.00	0.01	0.00	0.00	0.00
	Grid	**0.89**	**0.84**	**0.77**	**0.62**	**0.78**	0.46	**0.34**	0.17	0.08	0.26
	网格 + 描述	0.86	**0.84**	0.74	0.61	0.76	**0.50**	0.33	**0.21**	0.10	**0.29**

结构表格（参见图2 和表5），并且可能使模型更好地泛化到更大的地图。

固定地图尺寸下的起点-目标距离。为充分区分地图尺寸与 d_∞的影响，我们将所有训练图像（尺寸最大为 6×6）随机嵌入到 10×10 地图中，这不会改变正确路径，但会固定地图尺寸。此外，我们生成了一个新的测试集，其中包含 10×10 张地图，每个 $d_\infty \in \{2, \ldots, 9\}$值对应 200 个样本。此设置使我们能够单独分析起点-目标距离方面的泛化，而无需模型学习或泛化到不同尺寸的地图。表格3展示了不同数据格式下的分布内（$d_\infty \leq 5$）和分布外（$d_\infty \geq 6$）准确率（由于在较大地图上计算成本极高，我们省略了基于表格的思维链）。与之前的设置类似，图像输入的结果不如基于文本的输入。此外，对于无思维链和描述思维链，准确率在$d_\infty = 6$时就已降至接近 0%，这已经是最小的分布偏移，表明模型通常并未学会该任务的算法解法。使用网格输入，并采用网格或网格 +描述思维链格式，可获得最佳的分布内准确率（76–78%）和分布外准确率（26–29%），显著优于基线方法。基于网格的思维链格式甚至能提升采用描述输入时的性能，使其达到较好的分布外准确率。这与之前的实验结果一致，表明网格格式非常适用，即使测试时遇到较大的分布偏移，也能取得不俗的性能。

最优解长度。我们研究各种大语言模型分布外泛化性能的下一个维度是最优解 l的长度。在图4 （上图）中，我们展示了训练集和测试集中解决每张地图的最短路径长度分布，并汇总了不同地图尺寸下的统计数据。正如预期，将较大的分布外地图纳入测试集会使分布偏移朝较长解的方向移动。具体而言，对于训练集 $l \leq 14$，而测试集中的长度最高可达18。在图4下图中，我们展示了采用网格输入表示的模型准确率与测试地图解长度之间的关系。对于无思维链或采用描述思维链的大语言模型，当 $l \geq 9$时，准确率降至接近零，此时可用的训练样例非常少。相反，采用网格+描述思维链时，准确率直到$l = 13$仍保持在不低的水平，但之后便无法维持。作为进一步的消融实验，我们移除训练集中包含 $l > 10$ 的样例（总计 10 张地图）后重新训练该模型：该模型（图4中的虚线）表现与原模型相近。总体而言，泛化到解更长的地图同样极具挑战，但即使在这种情况下，我们的思维链表示也有助于提升准确率。所有输入格式的结果见图7。

4.4. 补充分析

与 Mirage 的比较。Mirage（Yang 等人，2025）通过潜在推理将图像整合到思维链中，其中

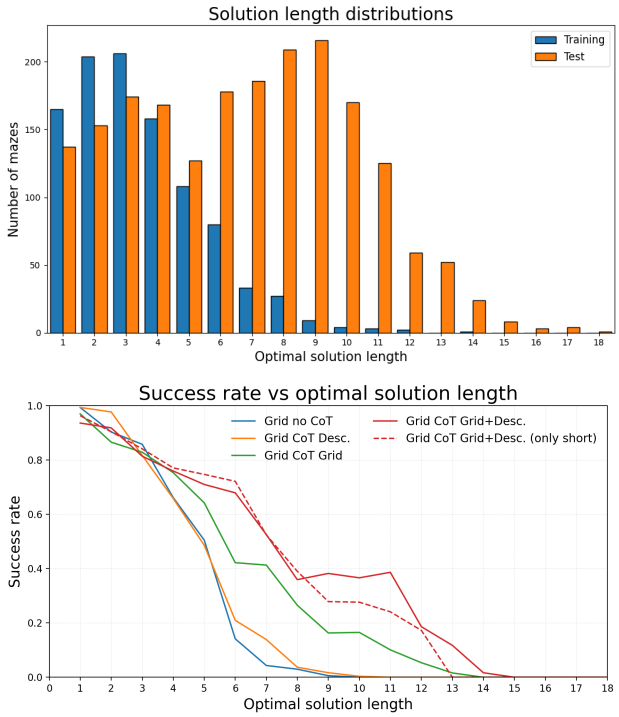


图*4.* **关于最优解长度的分布外泛化。上：** 我们展示训练地图和测试地图上最优（最短）解路径长度的分布（跨地图尺寸汇总）。**下：** 我们展示使用网格 作为输入表示、采用不同思维链格式的模型的成功率。即使仅使用最优解长度为 ≤ 10 （虚线）的训练地图，网格 +描述思维链在解长度为 11 和 12 的地图上也能取得非平凡的成功率。

LLM 可以在连续空间中生成词元，也就是说，不受词表中词元的限制。具体而言，系统会生成描绘通往目标的正确路径的辅助图像，而 Mirage 学会在给出基于文本的解之前，先生成辅助图像的编码版本。Yang 等人（2025）报告称，依赖这种连续推理和监督微调（SFT）的 Mirage Direct，在使用图像形式的 FrozenLake 地图时，表现优于仅使用解进行训练的 SFT（Direct SFT）以及使用基于文本的思维链进行训练的 SFT（CoT SFT）。在表格 4 中，我们报告了 Yang 等人（2025）的结果，并将其与我们使用图像输入格式、且不使用思维链或使用描述思维链训练的模型进行比较。令人惊讶的是，我们不使用思维链的模型已优于 Mirage Direct 和 Direct SFT，尽管两者使用了相同的数据：我们推测这是由于我们选择了更合适的训练超参数。此外，采用我们描述思维链的模型取得了最佳结果（平均准确率为 80 %），而 Yang 等人（2025）报告称，使用他们的推理轨迹时，这一版本表现极差（47%）：我们推测，我们的思维链表述比 Yang 等人（2025）的表述更简洁，因此带来了这一提升。最后，我们重新训练了 Mirage Direct，分别采用原始设置，以及在训练集中随机打乱辅助图像的设置：

表4. **与 Mirage 的比较。** 我们将 Yang 等人 (2025) 的模型分布内性能与我们的模型进行比较（训练 20 轮，* 仅训练 10 轮）。Mirage 的连续推理并未优于未使用思维链数据训练的模型。

模型	**分布内测试地图**（$d_\infty \leq 5$） 3x3	4x4	5x5	6x6	平均值
DirectSFT(Yang 等人，2025)	0.88	0.81	0.73	0.47	0.72
CoT SFT (Yang 等人，2025)	0.68	0.53	0.35	0.31	0.47
MirageDirect(Yang 等人，2025)	0.93	0.83	0.76	0.51	0.76
Mirage 直接（重新训练）	0.91	0.82	0.79	0.51	0.76
Mirage 直接打乱	0.92	0.85	0.82	0.5	0.77
图像，无思维链（我们的方法）	0.91	0.85	0.76	0.59	0.78
图像，描述性思维链（我们的方法） *	0.95	0.87	0.76	0.62	0.80

这两个模型的性能均与 Yang 等人 (2025) 报告的 Mirage Direct 性能相近，这表明他们的连续推理方法在此任务上并无优势。附录中的表格 6 还显示，Mirage 框架并未提升分布外性能。

与其他基线方法的比较。附录 9中的表格 B报告了其他基线方法在分布内地图上的结果。Li 等人 (2025) 使用比我们更大的训练集对 Anole（Chern 等人，2024）进行了 40 轮微调。他们的方法 MVoT 也在推理轨迹中加入了地图图像，准确率达到 86%，低于我们表格1中的多个模型。此外，VPFT（Xu 等人，2025）使用专门模型 LVM-7B（Bai 等人，2024）取得了 75.4% 的准确率（精确匹配指标），该模型仅处理图像输入和输出。VPRL（Xu 等人，2025）通过进一步使用强化学习（RL）进行微调，将准确率提高到 91.6%。仅通过对通用模型 Qwen2.5-VL-7B-Instruct 进行监督微调，我们便达到了相当的水平，采用的方案是网格/表格 +描述思维链（91%）。由于 Xu 等人 (2025) 的模型无法获取，我们无法测试其分布外泛化。这证实了与近期方法相比，我们的模型在分布内性能方面的有效性，而不仅仅体现在分布外泛化上。

推理长度。表格5 报告了表格1中推理模型给出正确解答的回复平均词元数。正如预期，结合多种格式的思维链输出最长，但推理长度并不一定与更好的表现相关。最佳结果来自网格输入与网格+描述思维链，且输出相对简短，这证实了紧凑的网格表示方式的有效性。

训练轮数的影响。为了测试延长训练的影响，我们采用不同的数据格式，并分别在有无思维链的设置下，对大语言模型进行微调，训练 20 轮和 30 轮（默认训练 10 轮）。详细结果见图6（见附录）。在分布内测试集上训练 20 轮后，以及在分布外测试集上训练 30 轮后，性能提升都很有限，但总体性能和模型排名

通往目标的直接路径是向上或向右移动。因此，我将向上移动。网格:

.....X...X....X.X.......
G..X.....X.....P....X.
....

通往目标的直接路径是向上或向右移动。因此，我将向上移动。G

rid:
.....X...X....X.X.......
G..X...P.X.X.....

通往目标的直接路径是向上或向右移动。因此，我将向上移动。网格:

.....X...X....X.X.....
P G..X.....X.X.....

通往目标的直接路径是向右移动。
Thus, 我将向右移动。
网格:
.....X . .
.X.... X .
.....
..... X .
..... P .
. X X
.....
..X.....

最终答案: \\boxed{上 , UP, UP, 右}

图*5*。 **我们的模型在分布外地图上的思维链推理示例**。 我们展示了一个示例，其中模型在网格输入和网格 + 描述思维链上训练后所生成的推理轨迹。即使面对分布外地图，模型也会先用自然语言推理下一步移动，然后生成执行该移动后的地图。

基本保持不变。

5. 讨论与结论

讨论。 我们的结果证实了先前研究（Stechly 等人，2024；Zhao 等人，2025）的观察：思维链推理虽然有助于大语言模型泛化到分布内测试样本，但在面对哪怕很小的分布偏移时也常常失效。这表明模型并未学会任务的算法解法，而是学到某种形式的模式识别和记忆。不过，与现有研究不同，我们证明，通过谨慎选择输入和推理的格式，可以显著提升分布外性能——

表 *5*。 **输出长度**。 我们报告了各模型正确解答（包括推理过程）的平均长度（以词元计）。

输入格式	CoT 格式	分布内地图	分布外地图（随机 d_∞）	分布外地图 ($d_\infty \geq 6$)
图像	零样本	328	562	564
	无思维链	10	11	19
	描述	65	69	192
描述	零样本	383	732	870
	无思维链	10	13	20
	描述	69	89	141
	Grid	160	586	1317
	表格	437	1311	2990
	网格 + 描述	213	771	1615
	表格 +描述	471	1657	3123
表格	零样本	463	893	937
	无思维链	10	11	19
	描述	69	75	180
	表格	428	1517	2925
	表格 +描述	490	1909	3904
Grid	零样本	785	1387	1338
	无思维链	10	11	19
	描述	67	77	157
	Grid	158	619	1194
	网格 + 描述	220	831	1459

有趣的是，我们的大量实验表明，丰富的推理步骤结合自然语言讨论与当前地图的结构化表示，能够提升模型应对各种测试时分布偏移（更大的地图、更远的起点和目标位置、更长的解）时的准确率，而标准方法在这些情况下会失效。

展望。 我们将本次评估视为比较和开发推理模型方法的起点，这些模型能够真正学会任务，并具备强大的分布外泛化能力。我们所引入的环境在数据格式和分布偏移方面具有灵活性，因此可以在一项人类轻易就能完成的任务上研究各种方法。例如，探索强化学习与不同数据格式的交互，或许会带来更多有趣的发现。尽管 图像 输入的效果不如基于文本的输入，但多模态模型正在迅速进步，可能是一个很有前景的方向，或可将视觉输入整合进思维链。此外，测试连续空间推理（Hao 等人，2024）是顺理成章的下一步，因为已有研究表明它能够编码并行搜索（Zhu 等人，2025），尽管这也面临训练方面的挑战。最后，任务的简单性或许使我们能够从更具理论性的角度，研究基于 Transformer 的模型如何解决该任务（或其某种变体）。

局限性。 遵循 Yang 等人 (2025)，我们将 Qwen2.5-VL-7B-Instruct 作为基础模型，因为它适合在学术预算范围内开展大量实验，

并使我们能够单独研究分布外性能。其他模型与各种数据格式的交互方式可能略有不同，这是一个值得未来研究探讨的有趣方向。

影响声明

本文旨在推动机器学习领域的发展。我们的工作可能会带来许多社会影响，但我们认为没有哪一项需要在此特别强调。

参考文献

Anthropic。推出新一代 Claude，2024 年。网址 https://www.anthropic.com/news/claude-3-family。

Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., 等。Qwen2. 5-vl 技术报告。*arXiv* 预印本 *arXiv:2502.13923*，2025。Bai, Y., Geng, X., Mangalam, K., Bar, A., Yuille, A. L., Darrell, T., Malik, J. 和 Efros, A. A. 序列模型训练使大型视觉模型能够实现可扩展学习。发表于*CVPR*，2024。Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J. 和 Zaremba, W. Openai gym，2016。网址 https://arxiv.org/abs/1606.01540。Chen, Q., Qin, L., Liu, J., Peng, D., Guan, J., Wang, P., Hu, M., Zhou, Y., Gao, T. 和 Che, W. 迈向推理时代：关于用于推理的大型语言模型长思维链的综述。*arXiv* 预印本 *arXiv:2503.09567*，2025。Chern, E., Su, J., Ma, Y. 和 Liu, P. Anole：一种用于交错图文生成的开放式、自回归原生大型多模态模型。*arXiv* 预印本 *arXiv:2407.06135*，2024。Gemini团队。Gemini：一系列能力强大的多模态模型。*arXiv* 预印本 *arXiv:2312.11805*，2023。Hao, S., Sukhbaatar, S., Su, D., Li, X., Hu, Z., Weston, J. 和 Tian, Y. 训练大型语言模型在连续潜在空间中进行推理。*arXiv* 预印本 *arXiv:2412.06769*，2024。Ho, N., Schmid, L. 和 Yun, S.-Y. 大型语言模型是推理教师。发表于*ACL*，2023。Huang, W., Jia, B., Zhai, Z., Cao, S., Ye, Z., Zhao, F., Xu, Z., Hu, Y. 和 Lin, S. Vision-r1：激发多模态大型语言模型的推理能力。*arXiv*

Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., et al. GPT-4o 系统卡。*arXiv* 预印本 *arXiv:2410.21276*，2024。Kojima, T., Gu, S. S., Reid, M., Matsuo, Y. 和 Iwasawa, Y. 大型语言模型是零样本推理器。发表于*NeurIPS*，2022。Li, C., Wu, W., Zhang, H., Xia, Y., Mao, S., Dong, L., Vuli´c, I. 和 Wei, F. 在空间中边推理边想象：多模态思维可视化。*arXiv* 预印本 *arXiv:2501.07542*，2025。Lu, P., Mishra, S., Xia, T., Qiu, L., Chang, K.-W., Zhu, S.-C., Tafjord, O., Clark, P. 和 Kalyan, A. 学会解释：通过思维链进行多模态推理，以回答科学问题。发表于*NeurIPS*，2022。Mirzadeh, S. I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S. 和 Farajtabar, M. GSM-symbolic：理解大型语言模型数学推理的局限性。发表于*ICLR*，2025。Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S. 和 Farajtabar, M. 思考的幻象：通过问题复杂度的视角理解推理模型的优势与局限性。发表于*NeurIPS*，2025。Stechly, K., Valmeekam, K. 和 Kambhampati, S. 无思维链？规划中的思维链分析。发表于*NeurIPS*，2024。von Werra, L., Belkada, Y., Tunstall, L., Beeching, E., Thrush, T., Lambert, N., Huang, S., Rasul, K. 和 Gal- louédec, Q. TRL：Transformer 强化学习，2020。网址：https://github.com/huggingface/trl。Wang, Y., Chang, F.-C. 和 Wu, P.-Y. 使用 Gevrey 类研究 Transformer 分布外鲁棒性的理论框架。*arXiv* 预印本 *arXiv:2504.12991*，2025。Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D. 等。思维链提示能够激发大型语言模型的推理能力。*NeurIPS*，2022。Wu, Q., Zhao, H., Saxon, M., Bui, T., Wang, W. Y., Zhang, Y. 和 Chang, S. 面向视觉语言模型空间规划任务的视觉感知与推理双重挑战评估。*arXiv* 预印本 *arXiv:2407.01863*，2024。Xu, Y., Li, C., Zhou, H., Wan, X., Zhang, C., Korhonen, A. 和 Vuli´c, I. 视觉规划：只用图像来思考。*arXiv* 预印本 *arXiv:2505.11409*，2025。

Yang, Z., Yu, X., Chen, D., Shen, M. 和 Gan, C. 机器心理意象：利用潜在视觉词元增强多模态推理。*arXiv* 预印本 *arXiv:2506.17218*，2025。

Yue, X.、Ni, Y.、Zhang, K.、Zheng, T.、Liu, R.、Zhang, G.、Stevens, S.、Jiang, D.、Ren, W.、Sun, Y. 等。Mmmu：面向专家级通用人工智能的大规模多学科多模态理解与推理基准测试。载于*CVPR*，2024 年。

Zhao, C.、Tan, Z.、Ma, P.、Li, D.、Jiang, B.、Wang, Y.、Yang, Y. 和 Liu, H. 大语言模型的思维链推理是海市蜃楼吗？数据分布视角。*arXiv* 预印本 *arXiv:2508.01191*，2025 年。

Zhou, Y.、Tu, H.、Wang, Z.、Wang, Z.、Muennighoff, N.、Nie, F.、Choi, Y.、Zou, J.、Deng, C.、Yan, S. 等。当可视化是推理的第一步：MIRA，一个视觉思维链基准测试。*arXiv* 预印本 *arXiv:2511.02779*，2025 年。

Zhu, H., Hao, S., Hu, Z., Jiao, J., Russell, S. 和 Tian, Y. 叠加式推理：关于连续思维链的理论视角。发表于*NeurIPS*，2025。

A. 数据与实验细节

数据生成。 对于分布外地图，我们以 0.1 的障碍概率随机生成地图。每种地图尺寸生成 200 张地图。玩家和目标随机放置；当起点和目标之间的距离有限制时，则使用拒绝采样。同样，为了将小地图嵌入 10×10 尺寸的地图中，我们随机生成一张10×10 地图，移除玩家和宝藏，然后将小地图随机放置在大地图中。

推理轨迹构建。 为了解序列的每一步生成描述性推理轨迹，我们首先确定玩家朝向宝藏的总体方向，该方向可以包含四个可能方向中的一个或两个。在 3×3示例中，该方向为“UP”。接下来，我们检查该方向是否被障碍阻挡。若未被阻挡，则选择此移动；否则，我们遍历其余方向：首先是另一个指向宝藏的方向（若存在），然后是不指向宝藏的方向。我们强制要求推理轨迹中所选方向始终与训练点的目标动作序列中的下一步移动相对应，以确保推理与解之间的一致性。

训练设置。 训练数据的统计信息见表格 7。为对模型进行微调，我们使用 trl（von Werra 等人，2020）软件包中的 SFTTrainer，并采用表格 8 中给出的超参数。所有模型均在两块 A100 GPU 上训练。

B. 附加实验

在分布外地图上与 Mirage 的扩展比较。为补充第4.2节的结果，表格6 进一步报告了 Mirage 模型（Yang 等人，2025）经我们重新训练后的分布外性能（原始模型不可用）。与分布内性能类似，Mirage 的连续推理方法相较于我们使用图像输入的简单监督微调并无收益，无论有无描述思维链。

回复长度的详细结果。表格10 报告了不同大语言模型在不同地图尺寸下平均回复长度（推理与解答）的明细。

其他方法的结果。表格9 报告了其他基线模型的结果。我们看到，Gemini-1.0-Pro-Vision（Gemini团队，2023）、Claude-3（Anthropic，2024）和 GPT-4o（Hurst 等人，2024）等前沿大语言模型在分布内测试地图上的零样本准确率较低（最佳的是 GPT-4o，准确率为 46%；结果取自（Wu等人，2024））。此外，我们报告了 Li 等人（2025）提出的三种微调变体： 直接、思维链和*MVoT*。这些方法使用 Anole-7B（Chern 等人，2024）作为基础模型，该模型能够以自回归方式生成交错文本与图像。直接 对应于

表6. **与 Mirage 的比较**。 我们还报告了表 4 中模型的分布外性能。Mirage 的潜在推理方法并未比不使用思维链的简单监督微调带来更好的分布外泛化。

模型	分布内测试地图（$d_\infty \leq 5$）					分布外地图（随机 d_∞）					分布外地图（$d_\infty \geq 6$）				
	3×3	4×4	5×5	6x6	Avg	7x7	8×8	9×9	10×10	平均值	7x7	8×8	9×9	10×10	平均值
Direct SFT（Yang 等人，2025）	0.88	0.81	0.73	0.47	0.72	-	-	-	-	-	-	-	-	-	-
CoT SFT（Yang 等人，2025）	0.68	0.53	0.35	0.31	0.47	-	-	-	-	-	-	-	-	-	-
Mirage Direct（Yang 等人，2025）	0.93	0.83	0.76	0.51	0.76	-	-	-	-	-	-	-	-	-	-
Mirage Direct（重新训练）	0.91	0.82	0.79	0.51	0.76	0.45	0.35	0.22	0.15	0.29	0.01	0.01	0.01	0.00	0.00
Mirage 直接打乱	0.92	0.85	0.82	0.5	0.77	0.46	0.41	0.26	0.14	0.31	0.05	0.06	0.02	0.00	0.03
图像，无思维链（我们的方法）	0.91	0.85	0.76	0.59	0.78	0.46	0.41	0.23	0.16	0.31	0.04	0.01	0.00	0.00	0.01
图像，描述性思维链（我们的方法）	0.95	0.87	0.76	0.62	0.80	0.47	0.35	0.21	0.14	0.29	0.07	0.03	0.00	0.01	0.03

表7。 **训练数据统计**。 我们报告平均 d_∞ 以及训练集上真实解序列的平均长度。

	3x3	4x4	5x5	6x6	总计
样本数量	100	200	300	400	1000
平均 d_∞	1.40	1.80	2.23	2.72	2.26
平均解长度	2.03	2.65	3.58	4.18	3.48

表 8。 **监督微调超参数**。 这些值适用于所有基于文本的模型。图像模型使用与 Yang 等人 （2025）

超参数	**值**
训练轮数	10
每_设备_训练_批次_大小	1
梯度_累积_步数	1
预热_步数	10
学习率	1e-5
权重_衰减	0.01
优化器	AdamW
bf16	True

不使用推理轨迹的监督微调，准确率为 78%。对于思维链，模型在包含坐标和文本描述的环境布局的推理轨迹上进行微调。与我们的实验不同，他们未能通过此类轨迹提升表现（准确率为 64%）。*MVoT*还在推理轨迹中加入了网格图像，准确率达到 86%。请注意，这些方法分别在更大的训练集上进行了 40 轮微调。最后，VPFT（Xu 等人，2025）使用专用模型 LVM-7B（Bai 等人，2024）取得了 75.4% 的准确率（精确匹配指标）；该模型仅将图像作为输入和输出。VPRL（Xu 等人，2025）通过进一步采用强化学习（RL）进行微调，将准确率提升至 91.6%。这与我们仅通过监督微调通用型 Qwen2.5-VL-7B-Instruct 所取得的结果相近，使用的是网格/表格 +描述思维链（91%），见表1。由于Xu 等人（2025）的模型尚未公开，我们无法测试其分布外泛化能力。

表 9。 **其他基线方法**。 我们报告了其他基线方法模型和视觉推理方法在分布内任务上的性能。结果取自此前报告这些结果的研究。

模型	分布内测试地图（$d_\infty \leq 5$）				
	3x3	4x4	5x5	6x6	Avg
Gemini-1.0(Xu 等人, 2025)	0.31	0.26	0.15	0.06	0.20
Claude-3 (Xu 等人, 2025)	0.52	0.33	0.16	0.15	0.29
GPT-4o (Xu 等人, 2025)	0.68	0.58	0.35	0.24	0.46
Anole Direct(Li 等人,2025)	0.83	0.80	0.75	0.75	0.78
Anole 思维链（Li 等人，2025）	0.94	0.72	0.50	0.39	0.64
MVoT（Li 等人，2025）	0.86	0.84	0.84	0.89	0.86
VPFT（Xu 等人，2025）	0.92	0.83	0.67	0.58	0.75
VPRL（Xu 等人，2025）	0.98	0.96	0.91	0.82	0.92

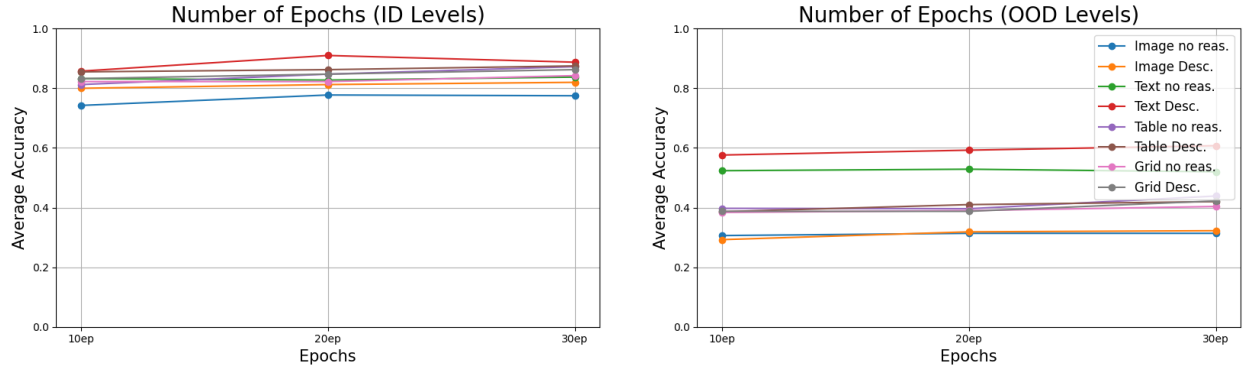


图*6*。 **训练轮数**。 我们将部分模型的微调训练轮数延长至20轮和30轮，并在分布内和分布外（随机 d_∞）测试地图上对其进行评估。然而，整体性能和模型排名基本保持不变。

表10。 **输出长度**。 我们报告了针对每种地图尺寸，各种微调模型的正确解答（包括推理）的平均长度（以词元计）。

输入格式	CoT格式	分布内测试地图 ($d_\infty \leq 5$) 3×3	4×4	5x5	6x6	Avg	分布外地图 (随机 d_∞) 7x7	8x8	9x9	10×10	Avg	分布外地图 ($d_\infty \geq 6$) 7x7	8x8	9x9	10×10	Avg
图像	零样本	210	429	309	363	328	450	590	662	547	562	524	457	774	502	564
	无思维链	9	10	10	11	10	12	11	10	9	11	19	-	-	-	19
	描述	52	59	64	86	65	81	74	65	54	69	196	217	-	162	192
描述	零样本	222	264	505	541	383	588	836	832	675	732	848	791	1010	833	870
	无思维链	9	10	11	12	10	13	13	13	13	13	21	19	18	22	20
	描述	52	62	73	87	69	86	90	91	90	89	156	127	122	157	141
	Grid	54	98	176	310	160	472	554	666	652	586	850	1023	1281	2116	1317
	表格	155	272	497	824	437	1136	1307	1534	1268	1311	2071	2534	2967	4387	2990
	网格 +描述	97	148	236	369	213	623	621	772	1069	771	1000	1371	1676	2413	1615
	表格 + 描述	188	320	527	848	471	1213	1451	1760	2205	1657	2238	2613	3289	4353	3123
表格	零样本	227	384	701	540	463	910	684	951	1027	893	848	918	886	1097	937
	无思维链	9	10	11	12	10	12	12	11	11	11	20	19	-	-	19
	描述	52	64	68	91	69	82	80	76	64	75	171	189	-	-	180
	表格	152	266	482	810	428	1242	1436	1687	1701	1517	2166	2652	3383	3498	2925
	表格 + 描述	192	324	544	899	490	1265	1633	1982	2755	1909	2333	2979	3813	6489	3904
Grid	零样本	355	668	1030	1086	785	1209	1314	1459	1567	1387	1067	1421	1523	1340	1338
	无思维链	9	10	11	12	10	12	11	11	11	11	19	18	-	-	19
	描述	52	61	72	83	67	81	80	79	71	77	161	145	171	150	157
	Grid	53	98	179	300	158	457	614	659	744	619	871	1081	1251	1572	1194
	网格 +描述	98	151	239	393	220	582	751	882	1109	831	1025	1298	1632	1880	1459

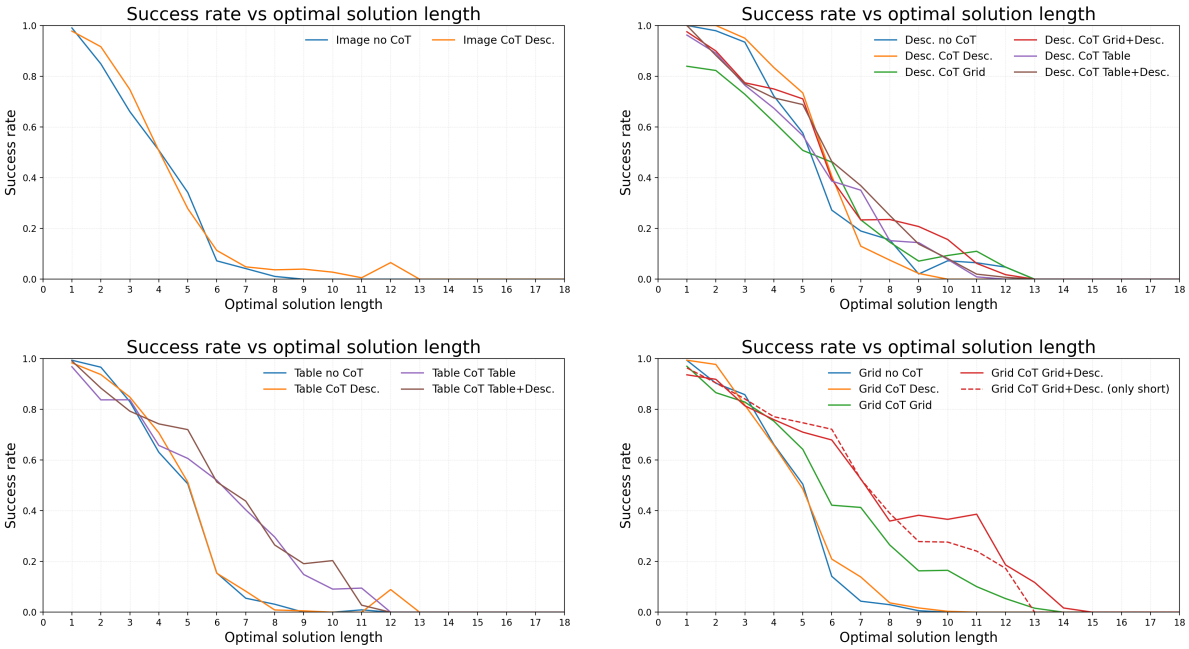


图 7. **分布外泛化** **关于最优解长度**。 我们展示了使用不同输入和 Co 的模型成功率 T
对解的表示 长度，补充图4中网格格式的结果。