arXiv:2603.02193v1 [cs.LG] 2 Mar 2026

符号等变循环推理模型

Richar d Freinschlag [1] Timo Bertram [1] Erich Kobler [1] Andreas Mayr [1] G¨unter Klambauer [1 2]

摘要

数独和 ARC-AGI 等推理问题对神经网络来说仍然颇具挑战。循环推理模型（RRMs）这一结构化问题求解架构系列，包括分层推理模型（HRM）和微型递归模型（TRM），为大语言模型提供了一种紧凑的替代方案，但目前只能通过代价高昂的数据增强来隐式处理符号对称性。我们提出符号等变循环推理模型（SE-RRMs），通过符号等变层在架构层面强制实现置换等变性，从而保证在符号或颜色置换下得到相同的解。SE-RRMs 在 9×9数独任务上优于以往的 RRM，并且仅通过在 9×9 上训练，就能泛化到更小的 4×4 和更大的16×16和 25×25实例，而现有 RRM 无法外推到这些实例。在 ARC-AGI-1 和 ARC-AGI-2 上，SE-RRMs 仅用 200 万个参数，便以大幅减少的数据增强实现了有竞争力的性能，表明显式编码对称性能够提升神经推理的鲁棒性和可扩展性。代码见 https://github.com/ml-jku/SE-RRM。

1. 引言

推理问题的难点。尽管深度神经网络和大语言模型取得了巨大成功，但它们在解决推理问题方面仍面临困难（Hazra 等人，2024年；Malek 等人，2025年；Seely 等人，2025年），例如本科水平的物理学（Xu 等人，2025年）、化学结构（Mirza 等人，2025年；Bartmann 等人，2026年），以及真实世界的临床情景（Kim 等人，2025年）。尽管研究界的大部分工作都聚焦于抽象

[1]ELLIS 林茨分部、LIT 人工智能实验室与机器学习研究所，林茨约翰内斯·开普勒大学，奥地利 [2]临床人工智能研究所，林茨约翰内斯·开普勒大学，奥地利。通信作者：G¨unter Klambauer <lastname@ml.jku.at>。

《 43rd 国际机器学习会议论文集，韩国首尔。PMLR 306，2026年。版权归作者所有，2026年。

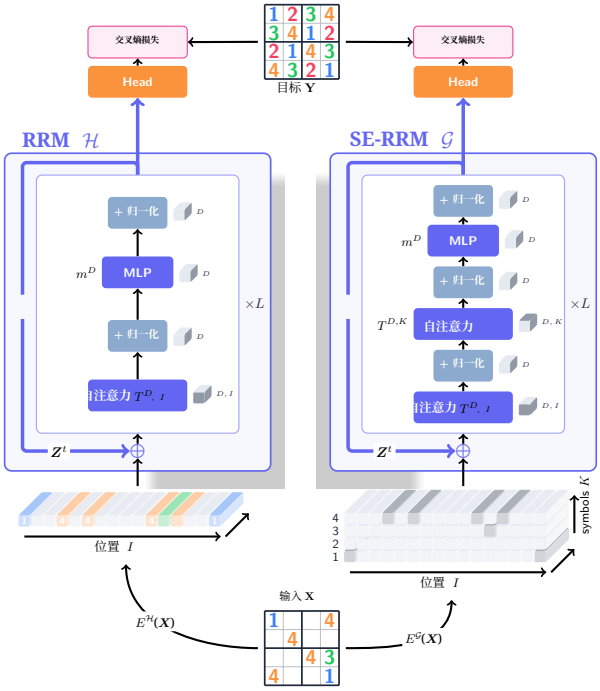

图 *1.* 比较用于求解 4×4 数独的 RRM（特别是 TRM）（Jolicoeur-Martineau，2025年）（左）与所提出的 SE-RRM（右）。对于 RRM，谜题在每个位置都通过符号特定嵌入进行编码，从而形成一个二维矩阵。相比之下，SE-RRM 引入第三维来关联位置和符号，并对所有符号使用相同的嵌入。SE-RRM 具有符号等变性，因为 Transformer 块中的自注意力 $T^{D,\,I}$ 和 $T^{D,\,K}$ 层依次作用于位置维度和符号维度，而归一化、加法和多层感知机层则逐 token 应用。

ARC-AGI 任务（Chollet，2019年；Chol- let 等人，2025年）、数独（Seely 等人，2025年；Long，2023）和迷宫问题（Wang 等人，2025年）等基准测试，可作为现实世界推理挑战的替代指标，涵盖规划、诊断、法律解释、因果分析、社会互动和风险评估。尽管此类基准测试无法充分体现不确定性、社会互动或规范性模糊的丰富内涵，但它们能够提取出许多实际问题的核心——约束满足与组合结构（Knuth，2015）。尤其是，数独为评估神经网络在硬约束和结构对称性下的推理能力提供了一个清晰的测试平台

。
符号求解器 组合优化和约束满足等结构化问题，传统上使用可满足性问题（SAT）、约束规划（CP）和混合整数优化（MIP）求解器等符号方法求解（Biere 等人，2021年；Zhang 等人，2023年；Bengio 等人，2021年；Kotary 等人，2021年）。这些方法依赖显式形式化和组合搜索，能够提供强有力的正确性和完备性保证；基于回溯的数独求解器会在解存在时找到解，否则证明问题无解。然而，符号求解器的计算开销可能很大：求解广义 $n^2 \times n^2$ 数独是 NP 困难问题（Yato 和 Seta，2003年），其最坏情况复杂度为 $O\big((n^2)^{n^4}\big)$。基于学习的方法可以通过学习求解器策略来加速搜索（Bengio 等人，2021年）；大语言模型也已得到探索，但若缺少任务专用支撑框架，它们在约束严格的问题上的表现仍然有限（Hazra 等人，2024年；Malek 等人，2025年；Seely 等人，2025年）。

用于解决推理问题的基于大语言模型的方法。大语言模型（LLM）（Peters 等人，2018年；Devlin 等人，2019年；Brown 等人，2020年）解决以自然语言表述的推理问题的能力有限（Wei 等人，2022b；a；Srivastava 等人，2023年），不过强化学习和大规模后训练可以提升这种能力（Guo 等人，2025年）。其推理能力通常通过 BIG-Bench（Srivastava 等人，2023年）等基于文本的基准测试进行评估，而这些测试远比结构化组合问题、约束满足问题或可满足性问题（SAT）简单。在更难的符号任务中，纯粹基于大语言模型的方法表现会急剧下降：前沿模型在 3-SAT 上表现吃力（Hazra 等人，2024年），在 ARC-AGI-2 上的得分则低于 5%（Chollet 等人，2025年）。不过，除了让大语言模型直接根据输入预测输出这一策略（Cole 和 Osman，2025年）之外，也可以利用大语言模型推断潜在函数，例如 Python 程序（Li 等人，2024年；Ellis 等人，2020年；Pourcel 等人，2025年；Mirchandani 等人，2023年）。后一种方式目前看起来更有前景，因为 ARC-AGI-2 上最好的成绩依赖于高度编排的大语言模型引导的符号处理流程，其中包含大量人工设计的组件，且计算成本高昂（Poetiq，2025年）。因此，尽管表现出色，基于大语言模型的方法仍只能有限地解决通用推理问题。

用于结构化问题的循环推理模型（RRMs）。近期研究提出了循环推理模型架构，尤其是分层推理模型（HRM）（Wang 等人，2025年）和微型递归模型（TRM）（Jolicoeur-Martineau，2025年），以应对数独、迷宫和 ARC-AGI 等离散推理基准测试。循环推理模型是一类旨在执行结构化、多步问题求解的神经网络架构，其方法是通过迭代优化逐步求解问题。从核心来看，循环推理模型由一个神经计算模块构成，该模块在不动点迭代方案中被反复应用。模型并非通过单次前向传播得出解，而是反复更新其循环状态，因此可以在中间步骤生成输出。每个模块接收两个主要输入：（i）结构化问题实例的嵌入；（ii）循环状态，用于保留先前迭代中的计算表征。该模块会生成更新后的循环状态，并可在任意迭代步由此导出或解码任务特定输出。训练过程中，代表任务解的固定监督信号会在不动点迭代中反复施加。每次施加监督信号后，梯度都会被截断。随后，在没有监督的情况下继续进行若干次循环模块迭代，以优化循环状态；之后，再将相同的监督信号施加于最近一次迭代步或若干步。模型参数通过在结构化问题求解任务的数据集上执行这一过程进行优化。迭代更新可以与学习得到的停止准则相结合，从而实现自适应计算时间，并动态控制每个任务实例的推理步数。

恒定输入循环神经网络。 在循环推理模型中使用恒定输入并非没有先例：早期 Jordan 型循环网络（Jordan，1986年）已采用恒定的“计划”向量来研究内部生成的时间动态。深度平衡模型（DEQs）（Bai 等人，2019年）同样依赖固定的外部输入，但它们不再采用循环形式，而是通过不动点求解器计算平衡态，并通过隐式微分进行优化，而非使用时间展开动态。正如 DEQ 相关文献所指出的，这一研究脉络可追溯至早期通过循环反向传播训练的平衡循环网络（Almeida，1987年；Pineda，1987年；Schmid-huber，2015）。自适应计算，尤其是在测试时的自适应计算，是 DEQ 的自然应用场景，也是循环推理模型的关键组成部分。

RRM 的优势。 RRM 是结构化推理领域一个前景广阔的方向，可与基于大语言模型的方法互为补充。作为一种循环神经网络，它们能够原生执行任意数量的顺序计算，这对于状态跟踪和解决多种推理问题是必需的（Merrill 等人，2025年）。如果大语言模型难以有效推理，那么它们抵御分布偏移的能力可能有限（Kim 等人，2025年）。鉴于 RRM 的训练明确侧重于推理任务，它们很可能在分布偏移下表现出更强的鲁棒性。这些特性在组合优化和调度等领域尤其重要，因为这些领域的问题规模和结构变化很大。尽管递归机制可能会导致 RRM 的推理延迟更高，

但它们通常参数效率较高，且只需有限的数据即可训练，无需大规模预训练或强化学习。此外，RRM 将推理深度与模型容量解耦，使计算量能够在推理时灵活扩展。其中一些优势已展现出实际应用前景，例如 TRM 在保险定价中的应用（Padayachy 等人，2026年）。

循环推理模型（RRM）的局限性。现有循环推理模型（RRM）的一个关键局限，是缺乏显式的符号等变性。在许多推理问题中，例如数独、ARC-AGI 和组合优化，符号可以互换，解应在符号置换下保持等变（Cohen 和 Welling，2016年；Bronstein 等人，2017年）。当前的 RRM 并未在架构层面编码这种对称性，而是依赖成本高昂的数据增强；这会增加样本复杂度，并可能妨碍模型泛化到未见过的符号配置。因此，RRM 可能会低效地学习等价符号之间冗余的区别，从而限制训练效率和鲁棒性。

贡献 a）我们提出符号等变循环推理模型（*SE-RRM*），这是一类新型循环推理架构，针对等价符号的置换具有显式等变性，例如数独中的数字或 ARC-AGI 问题中的颜色。

b) 我们证明，SE-RRM 能够减少对大量数据增强的依赖。在未通过大量数据增强处理符号等变性的情况下，与其他循环推理模型相比，SE-RRM 的表现显著更好，且可训练参数仅为其一小部分。**c)**我们表明，在架构层面实现符号等变性，使模型能够在推理过程中整合新符号，同时保持或提升循环推理的测试时扩展特性。

相关工作。置换等变性此前已在深度集合（Zaheer 等人，2018年）和图神经网络（Scarselli 等人，2008；Kipf 和 Welling，2017 年；Defferrard 等人，2016 年；Gilmer 等人，2017年）中得到探索，这些方法通常对图节点的置换具有不变性或等变性。此外，对于位置的置换等变性是 Transformer 架构引入的注意力操作的一项关键属性（Vaswani 等人，2017年）。Transformer 架构已应用于多维问题。例如，轴向注意力（Ho 等人，2019年）沿图像的行和列分别应用一个 Transformer 块，而交叉注意力（Huang 等人，2019年）则将同一原理应用于图像分割领域。MSA-Transformer（Rao 等人，2021年）在序列方向和残基方向分别采用一种注意力操作，用于不同蛋白质序列的多序列比对。

SE-RRM 通过轴向注意力实现了置换等变循环层；具体而言，它在问题的不同符号之间应用自注意力层。

2. 符号等变循环推理模型（SE-RRM）

2.1. 预备知识：原始 RRM

目标。我们的目标是为结构化问题求解任务（例如数独、迷宫或 ARC-AGI）训练一个机器学习模型。我们假设可以使用一个训练数据集，其中包含任务的表示及其对应的解，此外还有一个独立的评估数据集。这些数据集可能包含遵循同一套规则的任务（如数独），也可能包含受不同底层规则支配的任务（如 ARC-AGI，其中包含多种任务类型）。根据任务类型的复杂程度，通常需要多个任务实例才能推断出其支配规则。对于涉及多种任务类型的数据集（例如 ARC-AGI），通常会使用任务类型标识符来区分不同类型。

每个任务 $\boldsymbol{X}$ 都表示为一个符号元组 $x_i \in C$，即 $\boldsymbol{X} = (x_1, \ldots, x_I)$。假设任务定义在一个底层结构化领域上，该领域经展平后按固定顺序排列，从而得到元组表示。每个符号 x_i 表示该展平领域中第 i 个元素所关联的值。集合 C 由 K 问题求解任务的符号类型 ($K = |C|$) 组成（例如，数独中的数字以及一个掩码符号），通常也称为颜色集合。任务的解$\boldsymbol{Y} = (y_1, \ldots, y_I)$以类似方式表示为一个符号元组 $y_i \in C$，与 $\boldsymbol{X}$共享相同的索引和展平领域。根据任务的具体规则，C 中的某些符号可能不会出现在任务描述、解或两者中（例如，数独中的掩码符号只出现在输入中）。在任务数据集中，还可以为任务添加任务类型标识符$p \in \mathbb{N}$。这样便能指明哪些任务遵循相同的规则；训练任务集于是由配对 $(\boldsymbol{X}, p)$ 构成，而非单独的任务 $\boldsymbol{X}$。

RRM 块。如上所述，我们将 RRM 的主要神经计算块抽象为一个映射$\mathcal{H}$：$\mathbb{R}^{D\times I} \times \mathbb{R}^{D\times I} \mapsto \mathbb{R}^{D\times I}$。我们假设时间步 t 的循环状态记为 $\boldsymbol{Z}^t \in \mathbb{R}^{D\times I}$。结构化问题求解任务的嵌入由$E^{\mathcal{H}}(\boldsymbol{X})$提供。随后，RRM 块根据任务嵌入 $E^{\mathcal{H}}(\boldsymbol{X})$ 和前一循环状态 $\boldsymbol{Z}^t$计算后续循环状态 $\boldsymbol{Z}^{t+1}$：

$$\boldsymbol{Z}^{t+1} = \mathcal{H}(E^{\mathcal{H}}(\boldsymbol{X}), \boldsymbol{Z}^t), \tag{1}$$

任务嵌入。一种嵌入 $E^{\mathcal{C}}$：$C^I \mapsto \mathbb{R}^{D\times I}$，

它将符号任务描述映射到实向量空间，是处理任务所必需的。为构造这种嵌入，HRM 或 TRM 等方法会使用映射 $C \mapsto \mathbb{R}^D$，将任务字母表中的每个符号映射为一个 D 维实向量。这种符号级映射会在任务描述的所有位置上统一应用（在元组中逐元素应用）。所得矩阵$\boldsymbol{E} = E^{\mathcal{C}}(\boldsymbol{X})$在基于注意力的处理场景中可被视为词元序列 $\boldsymbol{e}_i \in \mathbb{R}^D$，即$\boldsymbol{E} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_I)$。由于嵌入 $\boldsymbol{E}$ 表示的是源自结构化领域的符号序列，因此通常有必要加入位置信息。位置嵌入（例如 RoPE、Su 等人（2024））可以添加到词元 $\boldsymbol{e}_i$ 中，使基于注意力的操作能够推理该领域内符号的相对位置及其关系。形式上，设 $E^{\mathcal{P}}$： $C^I \to \mathbb{R}^{D\times I}$ 表示一种位置嵌入映射，它为原始元组 X 中的每个位置 $i = 1, \ldots, I$ 分配一个 D 维向量。组合任务嵌入 $E^{\mathcal{H}}$ 可定义为

$$E^{\mathcal{H}} : C^I \mapsto \mathbb{R}^{D\times I}, \qquad (2)$$
$$E^{\mathcal{H}} := E^{\mathcal{C}} + E^{\mathcal{P}},$$

其中，元组 $\boldsymbol{X}$ 的符号级嵌入 $E^{\mathcal{C}}(\mathbf{X})$ 捕获 $\boldsymbol{X}$ 中各符号的内容，而位置嵌入 $E^{\mathcal{P}}(\mathbf{X})$ 则编码各自对应的位置。

当一个数据集包含多种不同的任务类型时，嵌入函数 $E^{\mathcal{H}}$ 会被扩展以编码任务类型信息，因此成为一个包含两个参数的函数，即任务实例 $\boldsymbol{X}$ 及其关联的任务类型 p。原则上，这种嵌入可以通过不同方式实现。在 HRM 和 TRM 中，任务类型由额外的位置 P 表示，其嵌入对应于一个学习得到的映射 $\mathbb{N} \to \mathbb{R}^{D\times P}$，该映射将任务类型 p 嵌入为一个矩阵 $\in \mathbb{R}^{D\times P}$。随后，通过用该任务类型嵌入增强原始任务表示，得到整体任务嵌入，结果为

$$E^{\mathcal{H}} : C^I \times \mathbb{N} \to \mathbb{R}^{D\times(I+P)}. \qquad (3)$$

解预测。 总体目标是预测任务的正确解，即预测表示该解的正确符号元组 $y_i \in C$。由于这定义了一种离散的正确/错误损失，循环推理模型对解元组中的每个位置采用可微代理损失。循环推理模型通常会为元组中每个位置上的每个符号输出预测概率。随后，可对每个位置独立应用分类交叉熵损失。

为获得符号的预测概率，循环推理模型通常将第 T 次迭代时的最终循环状态 $\boldsymbol{Z}^T \in \mathbb{R}^{D\times I}$ 映射为 logit 预测 $\hat{}$ $\boldsymbol{Y} \in \mathbb{R}^{K\times I}$。随后对这些对数几率应用 softmax 激活函数，或 stable-max 等类似函数，从而可以计算分类交叉熵损失，其计算依据是真实的训练解符号。对于从最终循环状态到对数几率的输出映射$\boldsymbol{O}$：$\mathbb{R}^{D\times I} \mapsto \mathbb{R}^{K\times I}$，可以沿循环状态的位置维度独立应用单个函数 $\mathbb{R}^D \to \mathbb{R}^K$（即在所有位置间共享）。通常，这通过使用简单的线性映射（由矩阵 $\in \mathbb{R}^{D\times K}$参数化）来实现。

总体 RRM 架构与深度监督。RRM架构的核心是应用公式（1）中描述的后续不动点运算，其中$\boldsymbol{Z}^0$ 保持固定，但取值为随机抽取，且相对于位置索引保持不变。这一原理存在多种变体：HRM 和 TRM 都可以视为循环推理模型的改编版本，因为它们将该模块中多个连续的迭代步骤组织成超块。连续的模块通常会省略任务嵌入，并将近期的先前循环状态添加到嵌入输入中，使某些不动点迭代步骤能够同时纳入紧邻的状态和更早的近期状态。

HRM 和 TRM 都采用的一项重要技术是深度监督。模型不会将最终预测的误差沿整个不动点迭代方案反向传播，而是根据选定的中间循环状态计算预测输出，即 $\boldsymbol{O}(\boldsymbol{Z}^t)$，并在时间 t 立即根据相应损失的梯度更新模型参数： $L(\boldsymbol{Y}^t, \boldsymbol{O}(\boldsymbol{Z}^t))$。梯度仅针对紧接在用于计算输出的循环状态之前的最后一个或多个迭代步骤计算。

RRM块的神经架构。 一个RRM 块$\mathcal{H}$ 实现不动点迭代的一步，由 L 个堆叠的Transformer风格层组成，这些层处理任务嵌入 $E^{\mathcal{H}}$（$\boldsymbol{X}$）以及循环状态 $\boldsymbol{Z}^t$，如公式（1）所示。在某些时间步，该架构的变体还可以额外纳入最近先前的循环状态 $\boldsymbol{Z}^{t-r}$ 或省略任务嵌入，这可以表示为

$$\boldsymbol{Z}^{t+1} = \mathcal{H}\Big(\mathbf{1}_{\text{emb}}\, E^{\mathcal{H}}(\boldsymbol{X}) + \mathbf{1}_{\text{prev}}\, \boldsymbol{Z}^{t-r},\, \boldsymbol{Z}^t\Big), \qquad (4)$$

其中 $\mathbf{1}_{\text{emb}}$ 与 $\mathbf{1}_{\text{prev}}$ 是指示变量，当包含嵌入或先前状态时等于1，省略时等于0。在RRM的变体中，将普通 $\mathcal{H}$ 块连接到超块的确切连线似乎并不关键，因为一些工作（ARC Prize Team，2025年；Liao& Poggio，2026年）表明，它或许可以由一叠Transformer层（对先前层具有捷径访问）替代。

RRM 块的第一层，我们称之为初始层 $\boldsymbol{H}_0$，被定义为所有包含的输入之和，即

$$\boldsymbol{H}_0 := \mathbf{1}_{\text{embed}}\, E^{\mathcal{H}}(\boldsymbol{X}) + \mathbf{1}_{\text{prev}}\, \boldsymbol{Z}^{t-r} + \boldsymbol{Z}^t. \qquad (5)$$

随后使用以下更新方案处理该初始层

$$\boldsymbol{H}'_l = \mathrm{Norm}\big(\boldsymbol{H}_l + T^{D,I}(\boldsymbol{H}_l)\big), \qquad \underline{\mathrm{D}} \times \underline{\mathrm{I}} \quad (6)$$

$$\boldsymbol{H}_{l+1} = \mathrm{Norm}\big(\boldsymbol{H}'_l + m^D(\boldsymbol{H}'_l)\big) \qquad \underline{\mathrm{D}} \times I. \quad (7)$$

用于 L 额外的步骤，生成块的输出 $\boldsymbol{Z}^{t+1} = \boldsymbol{H}_L$。每个 RRM 块主要由两类操作构成。第一类是自注意力层（Vaswani 等人，2017年），记作 $T^{D,\,I}$：$\mathbb{R}^{D\times I} \to \mathbb{R}^{D\times I}$，它沿着 $D \times I$ 矩阵的位置维度进行运算。第二类是逐位置多层感知机，具体而言是 SwiGLU 模块（Shazeer，2020年），记作 m^D：$\mathbb{R}^D \to \mathbb{R}^D$，它们分别独立作用于各个 I 位置。所采用的归一化操作是对 D 特征进行均方根归一化（Zhang 和 Sennrich，2019年），并且同样独立作用于所有 I 位置。

2.2. 循环推理模型的显著特性

循环推理模型的结构设计赋予其某些属性，这些属性可以通过形式化方式加以刻画，并在扩展中加以利用。在这些属性中，我们注意到它与位置等变性之间存在一种有趣的关系：如果没有位置嵌入，循环推理模型对输入位置的任意置换都具有等变性。以下命题将对此作出更精确的说明。

命题 2.1。 一个 *RRM* 块 $\mathcal{H}$ 不使用位置嵌入，如公式 (5) 至 (7) 所定义，在置换 π 下具有等变性：$[I] \mapsto [I]$输入位置$\in [I] := \{1, \ldots, I\}$，即

$$\Pi_2^\pi\Big(\mathcal{H}\Big(\mathbf{1}_{emb}\, E^{\mathcal{H}}(\boldsymbol{X}) + \mathbf{1}_{prev}\, \boldsymbol{Z}^{t-r},\, \boldsymbol{Z}^t\Big)\Big) =$$

$$= \mathcal{H}\Big(\mathbf{1}_{emb}\, E^{\mathcal{H}}(\Pi_2^\pi(\boldsymbol{X})) + \mathbf{1}_{prev}\, \Pi_2^\pi(\boldsymbol{Z}^{t-r}),\, \Pi_2^\pi(\boldsymbol{Z}^t)\Big),$$

其中

$$\Pi_a^\pi : \mathcal{S}^{d_1\times\cdots\times d_N} \to \mathcal{S}^{d_1\times\cdots\times d_N}$$

$$\big(\Pi_a^\pi(A)\big)_{i_1,\ldots,i_N} := A_{i_1,\ldots,i_{a-1},\pi(i_a),i_{a+1},\ldots,i_N}.$$

该命题很容易证明。它依赖于自注意力的等变性（Ilse 等人，2018年；Lee 等人，2019年）。此外，由于多层感知机和归一化操作独立应用于每个位置，它们自然满足关于位置置换的等变性。类似的论证也适用于残差加法。由于纯符号嵌入 $E^{\mathcal{C}}$ 同样独立应用于每个位置，可以通过归纳证明整个 RRM 块是位置等变的。我们甚至可以进一步将这一等变性扩展到不使用任何位置嵌入的整个RRM架构，考虑到 $\boldsymbol{Z}^0$ 相对于位置索引是常数，并假设 $\boldsymbol{O}$ 由沿位置维度独立应用的单一函数构造而成。

位置等变性 虽然等变性通常是一种理想属性，但对于严格的位置等变性而言并非如此。不同位置上的相同词元可能对正确结果具有截然不同的含义，正如在数独、迷宫和ARC-AGI中所观察到的那样。

超出训练基数的外推（位置和/或符号）。 尽管训练时使用固定数量的位置 I，该架构在测试时可以处理具有不同数量位置的输入 $J \neq I$，因此可能会实现一定程度的泛化（Kazemnejad 等人，2023； Anil 等人， 2022）。相比之下，它无法外推至未见符号：具有更大符号/颜色集合的问题 $|C'| > |C|$（包含 $C \subset C'$）无法处理，因为原始 RRM 通过将符号映射到 $E^{\mathcal{C}}$ 来编码符号，从而将其表示为高维特征空间中的特征。对未见符号的外推在少样本设置中可能尤其重要：如果测试时出现若干新符号，且新旧符号之间存在不同的关系。

2.3. 符号等变性：SE-RRM

我们建议在用于结构化问题求解任务的机器学习模型中采用符号等变性，因此在原始 RRM 架构中加入符号等变性，并将其称为 SE-RRM。SE-RRM 的构建核心思想如图 1 所示：SE-RRM 引入第三维，用于关联位置与符号。与原始 RRM 类似，SE-RRM 基于结构相同的不动点运算，但使用的张量形状不同。为清楚说明符号并更好地区分这两种方法，我们将 SE-RRM 的不动点运算记作 $\mathcal{G}$：

$$\boldsymbol{Z}^{t+1} = \mathcal{G}\Big(\mathbf{1}_{\mathrm{emb}}\, E^{\mathcal{G}}(\boldsymbol{X}) + \mathbf{1}_{\mathrm{prev}}\, \boldsymbol{Z}^{t-r},\, \boldsymbol{Z}^t\Big) \quad (8)$$

任务嵌入。 对于 SE-RRM，符号嵌入函数 $E^{\mathcal{C}}$ 经过修改，实现映射 $E^{\mathcal{C}}$： $C^I \mapsto \mathbb{R}^{D\times I\times K}$，其中 D 为特征数量， I 为位置数量， $K \equiv |C|$ 为符号总数。请注意，与原始 RRM 不同，原则上，我们对所有 I 个位置上的不同符号使用相同的嵌入$\mathbf{d} \in \mathbb{R}^D$。特殊符号是例外，例如掩码或未知词元，我们会为它们使用单独的嵌入（$\mathbf{s}_1, \ldots, \mathbf{s}_n$）。如果某个位置没有符号，则使用零嵌入（$\mathbf{0}$）。形式上，这可以视为一个实现下列映射的函数 h：

$$h : C^I \times C \times [I] \to \mathbb{R}^D$$

$$h(\boldsymbol{X}, c, i) = \begin{cases} \mathbf{d} \in \mathbb{R}^D, & \text{if } x_i = c \;\&\; c \in C_{\text{usual}} \\ \mathbf{s}_1 \in \mathbb{R}^D, & \text{if } x_i = c \;\&\; c = s_1 \\ \vdots & \\ \mathbf{s}_n \in \mathbb{R}^D, & \text{if } x_i = c \;\&\; c = s_n \\ \mathbf{0} \in \mathbb{R}^D, & \text{if } x_i \neq c \end{cases}$$

其中 $\boldsymbol{X}=(x_1,\ldots,x_I)$，$C=C_{\text{usual}}\cup\{s_1,\ldots,s_n\}$，其中$s_1,\ldots,s_n$表示特殊符号，而 C_{usual}包含其余所有（常见）符号。 h统一应用于所有I个位置以及所有 K个符号。因此，$\mathbf{d},\mathbf{s}_1,\ldots,\mathbf{s}_n$均为常向量。

将 $E^{\mathcal{H}}$ 扩展到 $E^{\mathcal{G}}$ 以编码任务类型信息的方式，SE-RRM与原始RRM有所不同。在原始RRM中，我们使用一个额外位置来编码任务类型；而对于SE-RRM，我们学习一个任务类型嵌入$\mathbb{N}\to\mathbb{R}^{D\times 1\times K}$。该嵌入沿位置维度广播，得到有效的任务类型嵌入 $E^{\mathcal{T}}$：$C^I\times\mathbb{N}\to\mathbb{R}^{D\times I\times K}$，并将其加到其他嵌入上，即 $E^{\mathcal{G}}:=E^{\mathcal{C}}+E^{\mathcal{P}}+E^{\mathcal{T}}$。我们将全部 K 个符号的任务类型嵌入初始化为同一向量。这种相同的初始化确保在学习开始时具有符号等变性，但允许针对特定任务类型发生偏离：对于这些特定任务类型，在单个任务层面上符号等变性并不合适（在这些情况下，符号等变性仅对整个任务类型成立）。

输出映射的适配。 在SE-RRM中，我们通过将特征维度线性映射到单个logit来生成预测，从而将其缩减为形状$(1,I,K)$或(I,K)，由此我们可以提取某个位置的预测符号，或计算分类损失。

SE-RRM 块的神经架构。 初始层$\boldsymbol{H}_0$类似于 RRM 块，其中使用 SE-RRM 嵌入和 SE-RRM 循环状态。不同于原始 RRM 中每个块仅包含一个 Transformer 风格层，SE-RRM 包含两个 Transformer 风格自注意力层：一个沿位置维度应用，另一个沿符号维度应用（Rao 等人，2021年；Tolstikhin 等人，2021年；Ho 等人，2019年；Huang 等人，2019年）。更新方案 $\mathcal{G}$ 如下：

$$\boldsymbol{H}'_l=\mathrm{Norm}\big(\boldsymbol{H}_l+T^{D,I}(\boldsymbol{H}_l)\big),\quad \underline{\mathrm{D}}\times\underline{\mathrm{I}}\times K \tag{9}$$

$$\boldsymbol{H}''_l=\mathrm{Norm}\big(\boldsymbol{H}'_l+T^{D,K}(\boldsymbol{H}'_l)\big),\quad \underline{\mathrm{D}}\times I\times\underline{\mathrm{K}} \tag{10}$$

$$\boldsymbol{H}_{l+1}=\mathrm{Norm}\big(\boldsymbol{H}''_l+m^{D}(\boldsymbol{H}''_l)\big),\quad \underline{\mathrm{D}}\times I\times K \tag{11}$$

经过 L 个额外步骤后，得到块输出 $\boldsymbol{Z}^{t+1}=\boldsymbol{H}_L$。每个方程右侧所写的表达式（例如 $\underline{\mathrm{D}}\times\underline{\mathrm{I}}\times K$）表示表示的维度，以及自注意力所应用的维度。$T^{D,\,I}$表示自注意力层沿位置维度应用，其中 token 维度为特征维度。$T^{D,\,K}$ 表示沿符号维度应用，其中 token 维度为特征维度。与原始 RRM 一样，RMSNorm（Zhang 和 Sennrich，2019年）用作归一化层，并统一应用于所有符号和位置。

至于原始 RRM，当在任务嵌入中省略位置编码和任务类型编码时，位置等变-性可以得到证明。

命题 2.2。 一个 *SE-RRM* 块 $\mathcal{G}$ 不含位置嵌入，定义见公式 (5) 和 (9) 至(11) 在输入位置的置换下具有等变性 π：$[I]\mapsto[I]\ \in[I]:=\{1,\ldots,I\}$，即

$$\Pi_2^{\pi}\Big(\mathcal{G}\Big(\mathbf{1}_{emb}\,E^{\mathcal{G}}(\boldsymbol{X})+\mathbf{1}_{prev}\,\boldsymbol{Z}^{t-r},\,\boldsymbol{Z}^{t}\Big)\Big)=$$
$$=\mathcal{G}\Big(\mathbf{1}_{emb}\,E^{\mathcal{G}}(\Pi_2^{\pi}(\boldsymbol{X}))+\mathbf{1}_{prev}\,\Pi_2^{\pi}(\boldsymbol{Z}^{t-r}),\,\Pi_2^{\pi}(\boldsymbol{Z}^{t})\Big),$$

其中 Π_a^{π} 的定义与定理2.1相同。

证明与定理2.1类似，唯一的区别是增加了一个额外维度，并在公式（10）中沿符号维度执行自注意力操作。公式（10）统一应用于所有位置，因此不会破坏位置等变性。由于所有其他操作在额外的符号维度上也都是统一的，位置等变性得以保留。

此外，SE-RRM 块还具有在符号置换下保持等变性的有益特性。

命题 2.3。 一个 *SE-RRM* 块 $\mathcal{G}$ 不含任务类型嵌入，定义见公式 (5) 和 (9) 至(11) 在输入符号的置换下具有等变性 ρ：$[K]\mapsto[K]\ \in[K]:=\{1,\ldots,K\}$，即

$$\Pi_3^{\rho}\Big(\mathcal{G}\Big(\mathbf{1}_{emb}\,E^{\mathcal{G}}(\boldsymbol{X})+\mathbf{1}_{prev}\,\boldsymbol{Z}^{t-r},\,\boldsymbol{Z}^{t}\Big)\Big)=$$
$$=\mathcal{G}\Big(\mathbf{1}_{emb}\,E^{\mathcal{G}}(\Pi_3^{\rho}(\boldsymbol{X}))+\mathbf{1}_{prev}\,\Pi_3^{\rho}(\boldsymbol{Z}^{t-r}),\,\Pi_3^{\rho}(\boldsymbol{Z}^{t})\Big),$$

其中 Π_a^{π} 定义为定理2.1中的形式。

证明与定理2.2类似，但根据应用于另一个维度进行了相应调整。通过适当的嵌入引入符号维度即可实现这种等变性。

计算复杂度。 Attention 的高效实现（Dao 等人，2022）的计算复杂度为 $O(I^2)$，内存复杂度为 $O(I)$。对于 SE-RRM 中的注意力机制，这分别变为$O(I^2K+K^2I)$ 和 $O(IK)$。因此，对于所考虑的问题 $I\gg K$，SE-RRM 会使计算复杂度和内存复杂度分别按因子 K线性增加。相反，在 $K\gg I$的领域中，计算复杂度由 $O(K^2I)$主导，这可能使 SE-RRM 不切实际。

3. 实验与结果

我们评估分层推理模型（分层推理模型）、微型递归模型（微型递归模型）以及我们提出的符号等变循环推理模型（SE-RRM），评估对象为

表 *1.* 不同网格大小下各方法的数独表现。我们报告完全求解率（*FSR*），即准确求解的谜题所占百分比，以及网格点准确率（*GPA*），即每种模式下正确预测的未填数独单元格所占百分比，并给出 95% 威尔逊评分置信区间。除 GPT-OSS（OpenAI 等人，2025年）外，所有模型均在 9×9 数独网格上训练。表现最佳的 RRM 以粗体标出。SE-RRM 能够泛化到更大的问题规模，而其他循环推理模型无法处理未见符号。

	4×4 迷你数独		9×9 数独		16×16 大型数独		25×25超大型数独	
模型	FSR	GPA	FSR	GPA	FSR	GPA	FSR	GPA
HRM	$0^{(0-1.32)}$	$29.19^{(27.45-30.95)}$	$63.53^{(63.38-63.67)}$	$86.11^{(86.10-86.12)}$	_[a]	_[a]	_[a]	_[a]
TRM	$0^{(0-1.32)}$	$45.88^{(43.96-47.79)}$	$71.94^{(71.80-72.08)}$	$89.80^{(89.79-89.81)}$	_[a]	_[a]	_[a]	_[a]
SE-RRM（我们的方法）	$\mathbf{95.46}^{(92.01-97.08)}$	$\mathbf{99.15}^{(98.67-99.41)}$	$\mathbf{93.73}^{(93.66-93.81)}$	$\mathbf{97.58}^{(97.54-97.63)}$	$0^{(0-1.75)}$	$\mathbf{51.95}^{(51.43-52.47)}$	$0^{(0-8.38)}$	$\mathbf{31.49}^{(30.22-32.75)}$
GPT-OSS-20B[b]	$100^{(98.68-100)}$	$100^{(99.85-100)}$	$21.67^{(13.12-33.62)}$	_[c]	_[d]	_[d]	_[d]	_[d]

[a]无法外推到此 规模。[b]基于大语言模型的系统在不使用特定任务的程序合成或符号搜索的情况下进行评估 h.
[c]该模型有时会重 拒绝生成任何输出，因此不计算 GPA。[d]模型拒绝提供解 s.

三类结构化问题求解任务：数独、ARC-AGI 和迷宫求解。所有循环推理模型都遵循 HRM（Wang 等人，2025年）中确立的同一总体训练协议：模型采用监督学习进行端到端训练，并在每个循环步骤施加深度监督；推理则通过以固定迭代次数反复应用同一循环块来完成。大语言模型基线在评估时不使用特定任务的程序合成、符号搜索或人工设计的编排，从而反映其作为普通神经求解器的性能。HRM 和 TRM 采用一种 Q 学习方法，作为深度监督的停止策略，使模型在训练期间可以应用少于最大深度监督步数的步骤，从而减少计算时间。相比之下，SE-RRM 使用一种更简单的随机机制：在每个深度监督步骤，除非已达到最大步数，否则训练时监督以概率 p（设为超参数）终止。在推理期间，所有模型——HRM、TRM 和 SE-RRM——都执行完整的最大深度监督步数，本节所有实验中该值均设为 16。

此外，我们使用 RoPE2d（Heo 等人，2024年）而非 RoPE 对空间位置进行编码 I。附录中提供了所有超参数的详细比较以及各种消融实验。在所有实验中，我们使用仅含 200 万个参数的 SE-RRM，其规模显著小于对比方法。

3.1. 数独

数据。我们使用 9×9Wang 等人提供的数独数据集Wang 等人 (2025年)（HRM）中的 1,000 个样本子集，为每个训练样本计算 1,000 种数据增强，与 HRM 和 TRM 类似，并使用包含 422,786 个样本的完整测试集进行验证。此外，我们还测试了对较小和较大数独谜题的零样本泛化。为此，我们考虑4×4、16×16和 25×25 谜题。对于 4×4 谜题，测试集包含全部 288 种独特组合[1],而我们使用 `py-sudoku` 创建了 216 个难度各异且具有唯一解的谜题，用于 16×16 数独。最后，我们从 42 个大小为 25×25 的样本中提取了谜题，这些谜题来自 `Sudoku-SMT-Solvers`[2]中整理的谜题。

方法。我们在网格大小不一的数独谜题上评估所有模型，范围从小型 4×4 实例到大型 25×25 谜题。所有循环推理模型（HRM、TRM 和 SE-RRM）均仅使用标准 9×9 数独进行训练，并在不进行微调的情况下，在其他网格大小上测试，以评估泛化和外推能力。性能通过两项指标衡量：完全求解率（*FSR*），定义为准确求解的谜题所占比例；以及网格点准确率（*GPA*），定义为正确预测的初始未填单元格所占比例。我们使用 Wilson 评分区间报告 95% 置信区间。

我们采用相应论文中的超参数，因此在推理时为 HRM、TRM 和 SE-RRM 执行 16 步深度监督。步数对性能的影响的消融实验见附录。

结果。定量评估结果汇总于表 1。在训练设置（9×9）下，HRM 和 TRM 均取得了出色的表现，优于 GPT-OSS-20B（OpenAI 等人，2025年）。GPT-OSS-20B 无法使用外部工具，尽管在预训练期间接触过类似数独的问题，却连标准的 9×9 谜题都无法可靠地求解。我们提出的 SE- RRM 模型大幅超越了所有基线（$> 11\%$ FSR 和 $> 7\%$ GPA），证明利用符号等变性进行推理更加有效。请注意，TRM（Jolicoeur-Martineau，2025年）报告称，使用 MLP-Mixer（Tolstikhin 等人，2021年）在数独上的准确率为 87.4%，但此版本仅支持固定数量的

[1]https://github.com/Black-Phoenix/ 4×4 数独数据集[2] https://github.com/liamjdavis/Sudoku-SMT-Solvers。请注意，求解器 Z3 求解器仅解出了 100 道谜题中的 42 道。

位置 因此无法外推到不同的网格 ds.

接下来，我们评估模型如何泛化到更简单的4×4 谜题。HRM 和 TRM 无法外推到 4×4迷你数独，因为 FSR 降至零，表明这两个模型都没有学到内在的数独规则。相比之下，GPT-OSS-20B 能够完美解决这些简单的逻辑谜题。值得注意的是，SE-RRM 能够将数独规则外推到 4×4 网格，并达到 95.46% 的 FSR，且几乎填对所有未填单元格（GPA 为 99.15%）。

最后，我们测试对更大网格（16×16 和25×25）的泛化，其中 I （位置数量）呈二次增长， K （符号数量）呈线性增长。HRM 和 TRM 在未经任何训练的情况下无法应用于更大的数独网格，因为额外的符号（“10”到“25”）需要全新的嵌入。尽管多次尝试，GPT-OSS-20B 并未给出任何有意义的输出。这一结果，连同在 9×9 网格上的表现，印证了先前的观察：大语言模型难以应对约束严格的符号推理任务。尽管经过训练的 SE-RRM 模型未能解出任何更大的数独，但 16×16 网格上 51.95% 的 GPA 以及 25×25网格上 31.49% 的 GPA 仍显著高于随机选择，表明我们训练的模型能够泛化到训练数据分布之外。

测试时扩展。 循环推理模型可以在推理时调整步数。对于数独这类 NP 困难问题，这能显著提升模型性能。表 2显示，SE-RRM 不仅在不同深度监督步数下都优于其他循环推理模型，而且总体表现也最佳。

表*2.* 9×9 数独上的测试时扩展。推理时，完全求解率（FSR，%）随深度监督步数的变化。

	推理步数							
模型	1	2	4	8	16	32	64	128
HRM	2.01	20.47	49.37	59.45	63.53	65.67	67.09	68.15
TRM	2.51	31.02	55.49	66.46	71.95	75.36	77.86	79.78
SE-RRM（我们的方法）	**16.05**	**62.06**	**77.31**	**87.38**	**93.73**	**96.82**	**98.22**	**98.84**

表*3.* ARC-AGI 基准测试上的 pass@2，以及迷宫上的完全求解率（FSR），并附有 95% Wilson 评分置信区间。HRM 和 TRM 的报告数值取自各自的论文。

	ARC-AGI-1	ARC-AGI-2	Maze
模型	pass@2		FSR
HRM	$40.3^{(35.56-45.13)}$	$5.0^{(2.31-10.48)}$	$74.5^{(71.66-77.15)}$
TRM	$44.6^{(39.71-49.40)}$	$\mathbf{7.8}^{(4.0-13.64)}$	$85.3^{(82.92-87.41)}$
SE-RRM（我们的方法）	$\mathbf{45.3}^{(40.44-50.15)}$	$7.1^{(3.42-12.61)}$	$\mathbf{88.8}^{(86.64-90.65)}$

3.2. ARC-AGI

数据。 在本节中，我们使用 ARC-AGI-1（Chollet， 2019年）和 ARC-AGI-2（Chollet 等人， 2025年）的官方训练集和评估集，其中包含几何谜题。每道谜题最多使用 10 种颜色，分布在一个至多包含 30×30 个位置的网格上，并通过 2 至 11 个少样本示例展示其底层任务。与 HRM 和 TRM 一样，我们还纳入了 ConceptARC 数据集中的 160 道谜题（ Moskvichev 等人， 2023），并将评估谜题中的所有少样本示例用于训练。因此，ARC-AGI-1 和 ARC-AGI-2 的训练集分别包含 960 道和 1,280 道谜题。最终模型在 ARC-AGI-1 官方的 400 道测试谜题和 ARC-AGI-2 挑战赛的 120 道测试谜题上进行评估。

方法。 我们比较 HRM、TRM 和 SE-RRM 官方报告的性能数值。与 HRM 和 TRM 所采用的大量颜色数据增强不同，由于 SE-RRM 具有符号等变性，我们对每个谜题仅使用 8 种二面体数据增强。此外，附录中还展示了若干消融实验，逐步将 SE-RRM 的超参数应用于 TRM。

结果。 SE-RRM 在 ARC-AGI-1 和 ARC-AGI-2 上的表现均优于 HRM，且结果与 TRM 相当。在表3中，我们报告了 HRM 和 TRM 原始论文中的结果，并与 SE-RRM 进行了比较。

3.3. 迷宫

数据。 最后，我们在基于网格的迷宫求解任务上评估这些模型。这类任务需要多步规划、全局约束满足以及一致的状态跟踪。所有模型均在由Wang 等人 (2025年) 定义的 Maze-hard 数据集上训练，该数据集包含 1,000 个训练迷宫和 1,000 个测试迷宫。每个迷宫都定义在一个 30×30 网格上，最短解路径长度为 110。请注意，训练 SE-RRM 时我们没有进行任何数据增强。

方法。 与之前的谜题不同，符号等变性不适用于此任务，因为墙壁与起点和终点并不等价。因此，我们允许每个符号使用不同的嵌入，从而打破符号等变性。由此得到的模型既不对符号的置换等变，也不对位置的置换等变。

结果。 与数独和 ARC-AGI 上的结果一致，SE-RRM 在迷宫任务上也取得了有竞争力的结果（表 3）。破坏符号等变性的 SE-RRM 表现略优于 TRM，且远优于 HRM。这可能是因为其输入嵌入的维度更高，或 SE-RRM 块中增加了一层 Transformer。无论如何，这表明 SE-RRM 可以成功应用于不需要符号等变性的问题。

4. 结论与局限性

总的来说，在这三个领域中，我们表明，符号等变循环推理模型为结构化问题求解任务提供了一种稳健且可扩展的替代方案，既可替代符号求解器，也可替代基于大语言模型的方法。我们还表明，SE-RRM 所需的数据增强显著更少。具体而言，我们在 ARC-AGI 上训练时，每个样本使用 8 次而非 1000 次数据增强。在数独任务中，尽管参与比较的模型都没有采用大量符号置换进行数据增强，SE-RRM 仍取得了更优结果。此外，我们的架构使循环推理模型能够适应新符号，从而外推到不同的任务形式。

局限性。 本文中，我们仅使用了参数量为 2M 的 SE-RRM。这远少于 TRM（7M）或 HRM（27M）。较少的参数量可以部分抵消训练和推理过程中增加的计算复杂度与内存复杂度。

致谢

我们感谢 Sebastian Lehner 提供的宝贵讨论。ELLIS 林茨分部、LIT 人工智能实验室和机器学习研究所均由上奥地利州联邦政府资助。我们感谢以下项目：FWF 双边人工智能（10.55776/COE12）、FWF AIRI FG 9-N（10.55776/FG9）、AI4GreenHeatingGrids（FFG-899943）、Stars4Waters（HORIZON-CL6-2021-CLIMATE-01-01）。我们还感谢 NXAI GmbH、奥迪股份公司、Silicon Austria Labs（SAL）、默克医疗保健有限合伙公司、GLS（滑铁卢大学）、¨TUV Holding GmbH、Software Competence Center Hagenberg GmbH、dSPACE GmbH、TRUMPF SE +Co. KG。

影响声明

本文旨在推动机器学习领域的发展。我们的工作可能会产生许多社会影响，但我们认为无需在此特别强调其中任何一项。

参考文献

Almeida, L. B. 一种适用于组合环境中带反馈的异步感知器的学习规则。载于*IEEE* 第一届国际神经网络会议，第 2 卷，第 609–618 页。IEEE，1987年。

Anil, C., Wu, Y., Andreassen, A., Lewkowycz, A., Misra, V., Ramasesh, V., Slone, A., Gur-Ari, G., Dyer, E., and Neyshabur, B. 探索大语言模型中的长度泛化。神经信息处理系统进展，35:38546–38556，2022年。

ARC Prize Team。HRM 背后的隐藏驱动因素 f在 arc-ag 上的表现 i. https://arcpr ize.org / 博客/hrm-analysis，2025年。访问日期：2025年1月27日。Bai, S.、Kolter, J. Z. 和 Koltun, V. 深度均衡模型。神经信息处理系统进展，32，2019年。

Bartmann, C.、Schimunek, J.、Ielanskyi, M.、Seidl, P.、Klambauer, G. 和 Luukkonen, S. MolecularIQ：通过在分子图上进行符号验证来刻画化学推理能力。国际学习表征会议（*ICLR*），2026年。

Bengio, Y.、Lodi, A. 和 Prouvost, A. 用于组合优化的机器学习：方法论概览。《欧洲运筹学杂志》，290(2)：405–421，2021年。

Biere, A.、Heule, M.、van Maaren, H. 和 Walsh, T.（编）。《可满足性手册——第二版》，第336卷，隶属于《人工智能与应用前沿》。IOS出版社，2021年。ISBN 978-1-64368-160-3。doi：10.3233/FAIA336。URLhttps://doi.org/10.3233/FAIA336。

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., 和 Van- dergheynst, P. 几何深度学习：超越欧几里得数据。*IEEE* 信号处理杂志, 34(4): 18–42, 2017年。

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shya m, P., Sastry, G., Askell, A., 等。语言模型是少样本学习者。神经信息处理系统进展, 33: 1877–1901, 2020年。

Chollet, F. 关于智能度量。*arXiv* 预印本 *arXiv:1911.01547*, 2019年。

Chollet, F., Knoop, M., Kamradt, G., Landers, B., 和 Pinkard, H. ARC-AGI-2：前沿人工智能推理系统面临的新挑战。*arXiv* 预印本 *arXiv:2505.11831*, 2025年。

Cohen, T. 和 Welling, M. 群等变卷积网络。载于国际机器学习会议，第 2990–2999 页。PMLR，2016年。

Cole, J. 和 Osman, M. 别把洗澡水和婴儿一起倒掉：为何以及如何将深度学习应用于 ARC。*arXiv* 预印本 *arXiv:2506.14276*，2025年。

Dao, T.、Fu, D. Y.、Ermon, S.、Rudra, A. 和R´e, C. FlashAttention：具有输入输出感知能力的快速、内存高效的精确注意力，2022年。

Defferrard, M.、Bresson, X. 和 Vandergheynst, P. 基于快速局部化谱滤波的图卷积神经网络

谱滤波。载于 Lee, D.、Sugiyama, M.、Luxburg, U.、Guyon, I. 和 Garnett, R.（编），神经信息处理系统进展，第 29 卷，2016年。

Devlin, J., Chang, M.-W., Lee, K. 和 Toutanova, K. Bert：用于语言理解的深度双向 Transformer 预训练。载于《*2019*年计算语言学协会北美分会人类语言技术会议论文集》，第*1*卷（长短论文），第4171–4186页，2019年。

Ellis, K., Wong, C., Nye, M., Sable-Meyer, M., Cary, L., Morales, L., Hewitt, L., Solar-Lezama, A. 和 Tenenbaum, J. B. DreamCoder：通过唤醒-睡眠贝叶斯程序学习，积累可泛化、可解释的知识。*arXiv*预印本 *arXiv:2006.08381*，2020年。

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. 和 Dahl, G. E. 用于量子化学的神经消息传递。载于 Precup, D. 和 Teh, Y. W.（编），第 *34* 届国际机器学习会议论文集，第70卷，第1263–1272页，2017年。Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X. 等。DeepSeek-R1通过强化学习激励大语言模型进行推理。《自然》，645(8081):633–638，2025年。Hazra, R., Venturato, G., Martires, P. Z. D. 和 De Raedt, L. 大语言模型能否推理？通过3-SAT进行表征。*arXiv*预印本 *arXiv:2408.07215*，2024年。Heo, B., Park, S., Han, D. 和 Yun, S. 用于视觉 Transformer 的旋转位置嵌入。载于欧洲计算机视觉会议，第289–305页。施普林格，2024年。Ho, J., Kalchbrenner, N., Weissenborn, D. 和 Salimans, T. 多维 Transformer 中的轴向注意力。*arXiv*预印本 *arXiv:1912.12180*，2019年。Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y. 和 Liu, W. CCNet：用于语义分割的交叉注意力。载于*IEEE/CVF*国际计算机视觉会议论文集，第603–612页，2019年。Ilse, M., Tomczak, J. 和 Welling, M. 基于注意力的深度多实例学习。载于国际机器学习会议，第2127–2136页。PMLR，2018年。Jolicoeur-Martineau, A. 少即是多：使用微型网络进行递归推理。*arXiv*预印本 *arXiv:2510.04871*，2025年。Jordan, M. 并行分布式处理方法。*ICS*报告，1986年。Kazemnejad, A., Padhi, I., Natesan Ramamurthy, K., Das, P. 和 Reddy, S. 位置编码的影响Transformer中的长度泛化。 神经信息处理系统进展，36:24892–24928，2023年。Kim, J.、Podlasek, A.、Shidara, K.、Liu, F.、Alaa, A.与Bernardo, D.。大语言模型在临床问题求解中因僵化推理而产生的局限性。《科学报告》，15(1):39426，2025年。Kipf, T. N.与Welling, M.。基于图卷积网络的半监督分类。载于国际学习表征会议，2017年。Knuth, D. E.《计算机程序设计艺术》：排序与查找。第*4*卷，分册*6*。可满足性。Addison-Wesley Professional，2015年。Kotary, J.、Fioretto, F.、Van Hentenryck, P.与Wilder, B.。端到端约束优化学习：综述。*arXiv*预印本 *arXiv:2103.16378*，2021年。Lee, J.、Lee, Y.、Kim, J.、Kosiorek, A.、Choi, S.与Teh, Y. W.。集合Transformer：基于注意力的置换不变神经网络框架。载于国际机器学习会议，第3744–3753页。PMLR，2019年。Li, W.-D.、Hu, K.、Larsen, C.、Wu, Y.、Alford, S.、Woo, C.、Dunn, S. M.、Tang, H.、Naim, M.、Nguyen, D.等。结合归纳与转导进行抽象推理。*arXiv*预印本 *arXiv:2411.02272*，2024年。Liao, Q.与Poggio, T.。简单递归模型：带跳跃连接的简化单状态推理。Research Square预印本，2026年。网址 https://www.researchsquare.com/article/rs-8492126/v1。第1版发布于2026年1月7日。Long, J.。大语言模型引导的思维树。*arXiv*预印本 *arXiv:2305.08291*，2023年。Malek, A.、Ge, J.、Lazic, N.、Jin, C.、Gy¨orgy, A.与Szepesv´ari, C.。前沿大语言模型仍难以应对简单推理任务。 *arXiv*预印本 *arXiv:2507.07313*，2025年。Merrill, W.、Petty, J.与Sabharwal, A.。状态空间模型中的状态错觉。 *arXiv*预印本 *arXiv:2404.08819*，2025年。Mirchandani, S.、Xia, F.、Florence, P.、Ichter, B.、Driess, D.、Arenas, M. G.、Rao, K.、Sadigh, D.与Zeng, A.。作为通用模式机器的大语言模型。*arXiv* 预印本 *arXiv:2307.04721*，2023 年。Mirza, A. 、Alampara, N.、Kunchapu, S.、R´ıos-Garc´ıa, M.、Emoekabu, B.、Krishnan, A.、Gupta, T.、Schilling-Wilhelmi, M.、Okereke, M.、Aneesh, A.等。用于对照化学家专业能力评估大语言模型化学知识与推理能力的框架。《自然·化学》，第1–8页，2025年。

Moskvichev, A.、Odouard, V. V.与Mitchell, M.。ConceptARC 基准：评估ARC 领域中的理解与泛化。*arXiv*预印本 *arXiv:2305.07141*，2023年。

OpenAI, Agarwal, S., Ahmad, L., Ai, J., Altman, S., Applebaum, A., Arbus, E., Arora, R. K., Bai, Y., Baker, B., Bao, H., Barak, B., Bennett, A., Bertao, T., Brett, N., Brevdo, E., Brockman, G., Bubeck, S., Chang,C., Chen, K.,Chen, M., Cheung, E., Clark, A., Cook, D., Dukhan, M., Dvo- rak, C., Fives, K., Fomenko, V., Garipov, T., Georgiev, K., Glaese, M., Gogineni, T., Goucher, A., Gross, L., Guzman, K. G., Hallman, J., Hehir, J., Heidecke, J., Hel- yar, A., Hu, H., Huet, R., Huh, J., Jain, S., Johnson, Z., Koch, C., Kofman, I., Kundel, D., Kwon, J., Kyrylov, V., Le, E. Y., Leclerc, G., Lennon, J. P., Lessans, S., Lezcano-Casado, M., Li, Y., Li, Z., Lin, J., Liss, J., Lily, Liu, Liu, J., Lu, K., Lu, C., Martinovic, Z., McCallum, L., McGrath, J., McKinney, S., McLaughlin, A., Mei, S., Mostovoy, S., Mu, T., Myles, G., Neitz, A., Nichol, A., Pachocki, J., Paino, A., Palmie, D., Pantuliano, A., Parascandolo, G., Park, J., Pathak, L., Paz, C., Peran, L., Pimenov, D., Pokrass, M., Proehl, E., Qiu, H., Raila, G., Raso, F., Ren, H., Richardson, K., Robinson, D., Rotsted, B., Salman, H., Sanjeev, S., Schwarzer, M., Sculley, D., Sikchi, H., Simon, K., Singhal, K., Song, Y., Stuckey, D., Sun, Z., Tillet, P., Toizer, S., Tsimpourlas, F., Vyas, N., Wallace, E., Wang, X., Wang, M., Watkins, O., Weil, K., Wendling, A., Whinnery, K., Whitney, C., Wong, H., Yang, L., Yang, Y., Yasunaga, M., Ying, K., Zaremba, W., Zhan, W., Zhang, C., Zhang, B., Zhang, E., and Zhao, S. gpt-oss-120b 与 gpt-oss-20b 模型卡。*arXiv* 预印本 *arXiv:2508.10925*，2025年。

Padayachy, K.、Richman, R. 和 W¨uthrich, M. V. Tab-TRM：用于表格数据保险定价的微型递归模型。*arXiv* 预印本 *arXiv:2601.07675*，2026年。

Peters, M. E.、Neumann, M.、Iyyer, M.、Gardner, M.、Clark, C.、Lee, K. 和 Zettlemoyer, L. 深度上下文化词表示。*arXiv* 预印本 *arXiv:1802.05365*，2018年。

Pineda, F. 将反向传播推广到循环神经网络和高阶神经网络。在神经信息处理系统中，1987年。

Poetiq。Poetiq 以一半的成本刷新了 ARC-AGI-2 的最先进水平。Poetiq AI 博客文章，2025年12月。网址：https://poetiq.ai/posts/arcagi_verified/。访问日期：2026-01-13。

Pourcel, J.、Colas, C. 和 Oudeyer, P.-Y. 面向进化式程序合成的自我改进语言模型：关于 arc-agi 的案例研究。*arXiv* 预印本 *arXiv:2507.14172*，2025年。

Rao, R. M., Liu, J., Verkuil, R., Meier, J., Canny, J., Abbeel, P., Sercu, T., and Rives, A. MSA Transformer. 收录于国际机器学习会议，第 8844–8856 页。PMLR，2021年。Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. 图神经网络模型。*IEEE* 神经网络汇刊，20(1):61–80，2008年。Schmidhuber, J. 神经网络中的深度学习：综述。神经网络，61:85–117，2015年。Seely, J., Imajuku, Y., Zhao, T., Cetin, E., and Jones, L. Sudoku-Bench：使用数独变体评估创造性推理。*arXiv* 预印本 *arXiv:2505.16135*，2025年。Shazeer, N. GLU变体改进了 Transformer。*arXiv* 预印本 *arXiv:2002.05202*，2020年。Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., 等。超越模仿游戏：量化并外推语言模型的能力。机器学习研究汇刊，2023年。Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., and Liu, Y. RoFormer：采用旋转位置嵌入的增强型 Transformer。神经计算，568:127063，2024年。Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., 等。MLP-Mixer：一种用于视觉的全多层感知机架构。神经信息处理系统进展，34:24261–24272，2021年。Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. 注意力就是你所需要的一切。神经信息处理系统进展，30，2017年。Wang, G., Li, J., Sun, Y., Chen, X., Liu, C., Wu, Y., Lu, M., Song, S., and Yadkori, Y. A. 分层推理模型。*arXiv* 预印本 *arXiv:2506.21734*，2025年。Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., 等。大语言模型涌现能力。*arXiv* 预印本 *arXiv:2206.07682*，2022a。Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., 等。思维链提示能激发大语言模型中的推理能力。神经信息处理系统进展，35: 24824–24837，2022b。Xu, X., Xu, Q., Xiao, T., Chen, T., Yan, Y., Zhang, J., Diao, S., Yang, C., and Wang, Y. UGPhysics：一项面向本科物理推理的综合基准测试，采用

大语言模型。*arXiv* 预印本 *arXiv:2502.00334*，2025年。

Yato, T. 和 Seta, T. 寻找另一个解的复杂性与完备性及其在谜题中的应用。电子、通信与计算机科学基础研究会学报，86(5):1052–1060，2003年。

Zaheer, M.、Kottur, S.、Ravanbakhsh, S.、Poczos, B.、Salakhutdinov, R. 和 Smola, A. 深度集合。*arXiv* 预印本，*arXiv:1703.06114*，2018年。

Zhang, B. 和 Sennrich, R. 均方根层归一化。神经信息处理系统进展，32，2019年。

Zhang, J.、Liu, C.、Li, X.、Zhen, H.-L.、Yuan, M.、Li, Y. 和 Yan, J. 通过机器学习求解混合整数规划的综述。神经计算，519:205–217，2023年。

A. 符号说明

表 *A1.* 循环推理模型（RRMs）、分层推理模型（HRMs）、微型递归模型（TRMs）和符号等变循环推理模型（SE-RRM）描述中使用的符号。

定义	符号	Type
标量和索引		
特征维度	D	$\mathbb{N}$
位置（网格点）数量	I	$\mathbb{N}$
符号 / 颜色数量	K	$\mathbb{N}$
循环迭代索引	t	$\mathbb{N}$
每个块的层数	L	$\mathbb{N}$
最终迭代步	T	$\mathbb{N}$
额外任务类型位置的数量（RRM）	P	$\mathbb{N}$
集合与空间		
符号／颜色字母表	C	有限集，$\lvert C\rvert = K$
输入与解空间	C^I	符号元组
任务与解		
输入任务	$\boldsymbol{X} = (x_1, \ldots, x_I)$	C^I
解任务	$\boldsymbol{Y} = (y_1, \ldots..., y_I)$	C^I
位置 i 处的符号	x_i, y_i	C
任务类型标识符	p	$\mathbb{N}$
嵌入与表示		
符号级嵌入（SE-RRM）	$E^{\mathcal{C}}$	$C^I \mapsto \mathbb{R}^{D\times I\times K}$
位置嵌入（SE-RRM）	$E^{\mathcal{P}}$	$C^I \to \mathbb{R}^{D\times I\times K}$
任务类型嵌入（SE-RRM）	$E^{\mathcal{T}}$	$C^I \times \mathbb{N} \to \mathbb{R}^{D\times 1\times K}$
组合任务嵌入（SE-RRM）	$E^{\mathcal{G}}$	$E^{\mathcal{C}} + E^{\mathcal{P}} + E^{\mathcal{T}}$
组合任务嵌入（RRM）	$E^{\mathcal{H}}$	$C^I \mapsto \mathbb{R}^{D\times I}$ 或 $C^I \times \mathbb{N} \to \mathbb{R}^{D\times(I+P)}$
SE-RRM符号嵌入函数	h	$C^I \times C \times [I] \to \mathbb{R}^D$
共享常规符号嵌入向量	$\mathbf{d}$	$\mathbb{R}^D$
特殊符号嵌入向量	$\mathbf{s}_1, \ldots, \mathbf{s}_n$	$\mathbb{R}^D$
HRM/TRM 任务嵌入矩阵	$\boldsymbol{E} = E^{\mathcal{H}}(\boldsymbol{X})$	$\mathbb{R}^{D\times I}$
SE-RRM 任务嵌入张量	$\boldsymbol{E}' = E^{\mathcal{G}}(\boldsymbol{X})$	$\mathbb{R}^{D\times I\times K}$
位置 i 处的词元嵌入	$\boldsymbol{e}_i$	$\mathbb{R}^D$
中间块状态	$\boldsymbol{H}_l, \boldsymbol{H}'_l, \boldsymbol{H}''_l$	$\mathbb{R}^{D\times I}$ 或 $\mathbb{R}^{D\times I\times K}$
第 t 步的循环状态	$\boldsymbol{Z}^t$	$\mathbb{R}^{D\times I}$ 或 $\mathbb{R}^{D\times I\times K}$
模型与块		
RRM 块 (HRM/TRM)	$\mathcal{H}$	$\mathbb{R}^{D\times I} \times \mathbb{R}^{D\times I} \mapsto \mathbb{R}^{D\times I}$
SE-RRM块	$\mathcal{G}$	$\mathbb{R}^{D\times I\times K} \times \mathbb{R}^{D\times I\times K} \mapsto \mathbb{R}^{D\times I\times K}$
嵌入包含指示器	$\mathbf{1}_{\text{emb}}$	$\{0, 1\}$
前一状态包含指示器	$\mathbf{1}_{\text{prev}}$	$\{0, 1\}$
神经算子		
位置自注意力	$T^{D,I}$	$\mathbb{R}^{D\times I} \mapsto \mathbb{R}^{D\times I}$
符号自注意力	$T^{D,K}$	$\mathbb{R}^{D\times K} \mapsto \mathbb{R}^{D\times K}$
特征多层感知机	m^D	$\mathbb{R}^D \mapsto \mathbb{R}^D$
归一化（均方根层归一化）	Norm	$\mathbb{R}^D \mapsto \mathbb{R}^D$
预测与损失		
输出映射（RRM）	$\boldsymbol{O}$	$\mathbb{R}^{D\times I} \mapsto \mathbb{R}^{K\times I}$
输出投影矩阵（RRM）	$\boldsymbol{W}$	$\mathbb{R}^{K\times D}$
输出投影矩阵（SE-RRM）	$\boldsymbol{W}$	$\mathbb{R}^{1\times D}$
第 t步的对数几率	$\hat{\boldsymbol{Y}}^t$	$\mathbb{R}^{K\times I}$ (RRM) 或 $\mathbb{R}^{I\times K}$ (SE-RRM)
分类交叉熵损失	$L(\hat{\boldsymbol{Y}}, \boldsymbol{Y})$	$\mathbb{R}_{\geq 0}$
对称算子		
位置置换	π	$\{1, \ldots, I\}$ 上的双射
符号/颜色置换	ρ	$\{1, \ldots, K\}$ 上的双射
张量上的置换算子	Π_a^π	对轴 a执行置换 π
任务类型组件		
任务类型嵌入位置（RRM）	$\boldsymbol{p}$	$\mathbb{R}^D$
任务类型嵌入矩阵（SE-RRM）	$\boldsymbol{P}$	$\mathbb{R}^{D\times K}$

B. 符号等变性

基于一个小型 4×4 Sud oku，我们通过图A1直观地解释符号等变性的概念。此外，我们还展示了一个简单的 ARC-AGI 示例，见图A2。

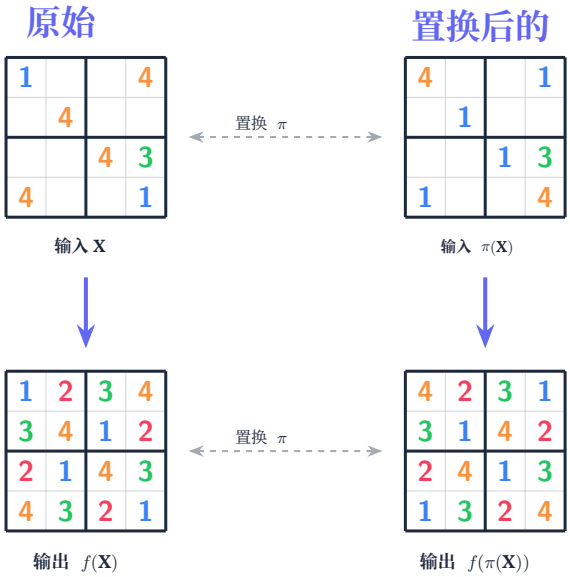

图 *A1*。符号置换下的模型等变性示意图。若对输入应用置换 π 后，输出也发生相应的置换，使得 $f(\pi(X)) = \pi(f(X))$，则模型 f 具有等变性。这里，置换 π 交换符号 1 和 4。与其他推理模型不同，SE-RRM 保证了置换等变性。

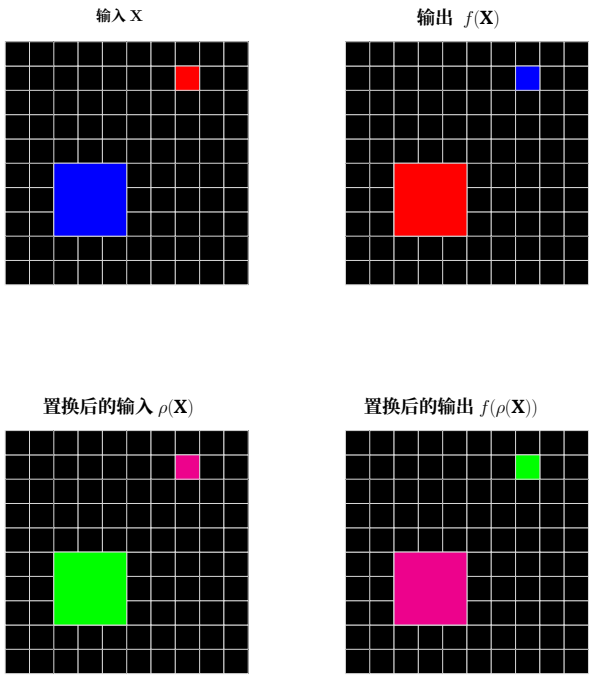

图 *A2*。 ARC-AGI 任务示例图：需要交换输入中方块的颜色。原始 RRM 需要依赖大量颜色增强，而 SE-RRM 则可以直接学习任务类型规则。

C. 架构与超参数 C.1. 数独

表 *A2.* 数独超参数比较

超参数	HRM	TRM	SE-RRM
深度监督步数	16	16	16
深度监督早停	Q 学习	Q 学习	随机（每步有 5% 的概率）
递归循环	$H=2, L=2$	$H=3, L=6$	$H=3, L=6$
批大小	384	768	272
训练轮数	50000	50000	10000
学习率	0.0001	0.0001	0.0005
权重衰减	1	1	1
学习率调度	预热 + 恒定学习率	预热 + 恒定学习率	预热 + 恒定学习率
隐藏层大小	512	512	256
参数	27M	7M	2M
Transformer 层	8	2	4（每个维度各 2 个）
位置编码	RoPE	RoPE	RoPE2D

C.2. ARC-AGI

表 *A3.* ARC-AGI 的超参数比较。所有模型在 ARC-AGI-1 和 ARC-AGI-2 上使用相同的超参数。

超参数	HRM	TRM	SE-RRM
深度监督步数	16	16	16
深度监督早停	Q 学习	Q 学习	随机（每步有 10% 的概率）
递归循环	$H=2, L=2$	$H=3, L=4$	$H=3, L=4$
批大小	768	768	272
训练轮数	100000	100000	100000
学习率	0.0001	0.0001	0.0005
权重衰减	0.1	0.1	0.1
学习率调度	预热 + 恒定学习率	预热 + 恒定学习率	预热 + 余弦衰减
谜题嵌入的学习率	0.01	0.01	0.01
谜题嵌入的权重衰减	0.1	0.1	0.3
隐藏层大小	512	512	256
参数	27M	7M	2M
Transformer 层	8	2	4（每个维度各 2 个）
位置编码	RoPE	RoPE	RoPE2D

C.3. 迷宫

表 *A4.* 迷宫的超参数比较

超参数	HRM	TRM	SE-RRM
深度监督步数	16	16	16
深度监督早停	Q 学习	Q 学习	随机（每步有 10% 的概率）
递归循环	$H=2, L=2$	$H=3, L=4$	$H=3, L=4$
批大小	384	768	64
训练轮数	20000	50000	10000
学习率	0.0001	0.0001	0.0005
权重衰减	1	1	1
学习率调度	预热 + 恒定学习率	预热 + 恒定学习率	预热 + 恒定学习率
隐藏层大小	512	512	256
参数	27M	7M	2M
Transformer 层	8	2	4（每个维度 2 个）
位置编码	RoPE	RoPE	RoPE2D

D. 其他实验结果

D.1. ARC-AGI-1 上的消融实验

在表 A5中，我们逐步将 TRM 转变为 SE-RRM，并展示其对 ARC-AGI-1 基准测试的影响。首先，我们将 Q 学习替换为深度监督中的随机停止机制（TRM + 随机停机）。然后，我们将 RoPE 改为 RoPE2D。有趣的是，这两项改动都明显降低了 TRM 的性能。进一步统一所有其余超参数后，这种影响更加显著。不过，这并不那么令人意外，因为 SE-RRM 的参数也明显更少。再加入二维嵌入并采用我们的架构后，我们便得到了 SE-RRM。

表 *A5*。 ARC-AGI-1 基准测试上的 pass@2，附带 95% 威尔逊评分置信区间。我们逐步调整 TRM 的超参数和架构，使其转变为 SE-RRM。

模型	ARC-AGI-1 (%)
TRM	$44.6^{(39.71-49.40)}$
TRM +随机停机	$34.8^{(30.25-39.54)}$
TRM + RoPE2d	$38.4^{(33.62-43.10)}$
TRM +随机停机 +RoPE2d	$33.1^{(28.57-37.75)}$
TRM +所有超参数均已匹配	$32.3^{(27.86-36.98)}$
TRM + 所有超参数均匹配 +等变性 =SE-RRM（我们的方法）	$\mathbf{45.3}^{(40.44-50.15)}$

D.2. 学习动态

在图 A3 中，我们比较了 HRM、SE-RRM 和 TRM 在不同数据集上训练期间，其在 ARC-AGI1 上的 Pass@2 准确率，以及在数独和迷宫上的完全求解率（FSR）的变化。

D.3. 数独和迷宫上的消融实验

我们进一步评估了数独和迷宫任务中不同超参数设置的效果。具体而言，我们分别采用随机停机和 Q 学习训练 TRM，使用 RoPE 和 RoPE2D 评估这两种模型，并尝试将 TRM 的超参数用于 SE-RRM，以及反过来将 SE-RRM 的超参数用于 TRM。需要注意的是，我们没有尝试在 SE-RRM 中使用 Q 学习，因为其编码的形状不同，无法直接实现。结果见表A6。SE-RRM 在不同配置下似乎都表现良好，而 TRM 有时则完全无法奏效。我们推测，TRM 的超参数是针对 Q 学习优化的，因此不太适用于其他设置。

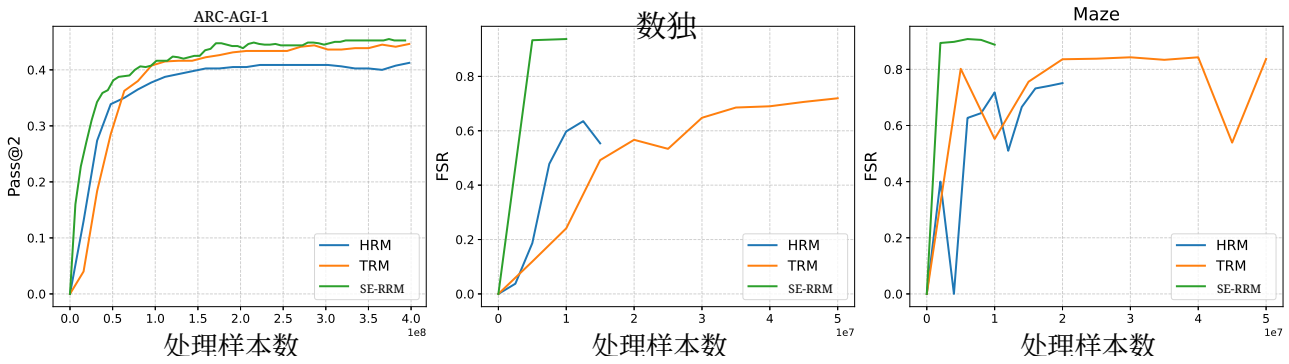

图 *A3.* HRM、TRM 和 SE-RRM 的学习曲线。纵轴展示 HRM、TRM 和 SE-RRM 在数独、ARC-AGI-1 和迷宫上的验证指标。横轴展示训练过程中处理的样本数。SE-RRM 的数据效率更高，因为它处理较少样本数就能达到更高的指标。

在表 A7 中，我们通过共享嵌入消融研究了强制等变性的影响。在强制等变性的情况下，SE-RRM 在数独任务上的表现最佳；不施加强制等变性时，表现略差。需要注意的是，即使每个数字的初始嵌入各不相同，模型仍可在训练过程中学会等变性。在迷宫任务中，各符号的语义明显不同。因此，强制施加等变性会导致性能下降。

表 *A6.* TRM 和 SE-RRM 在数独和迷宫上的消融实验。

模型	停止机制	位置编码	超参数	数独（%）	迷宫（%）
TRM	Q 学习	RoPE	TRM	$71.9^{(71.80-72.08)}$	$85.3^{(82.92-87.41)}$
TRM	随机	RoPE	TRM	$79.4^{(79.31-79.55)}$	$84.9^{(82.55-86.99)}$
TRM	Q 学习	RoPE2D	TRM	$76.5^{(76.34-76.59)}$	$85.4^{(83.08-87.45)}$
TRM	随机	RoPE2D	TRM	$0^{(0.00-0.00)}$	$59.3^{(56.23-62.30)}$
TRM	随机	RoPE2D	SE-RRM	$0^{(0.00-0.00)}$	$65.0^{(61.99-67.89)}$
SE-RRM	随机	RoPE2D	SE-RRM	$93.7^{(93.66-93.81)}$	$\mathbf{88.8}^{(86.64-90.65)}$
SE-RRM	随机	RoPE	SE-RRM	$\mathbf{95.4}^{(95.32-95.45)}$	$78.3^{(75.64-80.74)}$
SE-RRM	随机	RoPE	TRM	$68.4^{(68.30-68.58)}$	$80.1^{(77.51-82.46)}$

表 *A7*。 SE-RRM 在数独和迷宫上的结果，包括消融实验。第二列表示是否通过共享嵌入强制实现符号等变性。

模型	符号等变性	数独 FSR	Maze FSR
TRM	错误	$71.9^{(71.80-72.08)}$	$85.3^{(82.92-87.41)}$
HRM	错误	$63.5^{(63.38-63.67)}$	$74.5^{(71.66-77.15)}$
SE-RRM	True	$\mathbf{93.7}^{(93.66-93.81)}$	$0^{(0.00-0.38)}$
SE-RRM	否	$89.7^{(89.61-89.79)}$	$\mathbf{88.8}^{(86.64-90.65)}$