arXiv:2602.16702v1 [cs.CV] 18 Feb 2026

显著性感知　多路径思维：重新审视视觉语言推理

Mingjia Shi [1] **Yinhan He** [1] **Yaochen Zhu** [1] **Jundong Li** [1]

摘要

视觉语言模型（VLM）旨在联合利用视觉与文本模态进行推理。尽管为大语言模型（LLM）分配额外的推理时计算已被证明有效，但要在视觉语言模型中实现类似的扩展仍然具有挑战性。一项关键障碍在于，视觉输入通常仅在生成开始时提供一次，而文本推理（例如，早期视觉摘要）则以自回归方式生成，导致推理越来越受文本主导，并使早期视觉定位错误不断累积（图 1）。此外，推理过程中用于视觉定位的朴素引导往往粗糙且充满噪声，难以引导长文本中的推理。为应对这些挑战，我们提出了显著性感知原则（SAP）选择。SAP 作用于高层推理原则，而非词元级轨迹，使其能够在噪声反馈下稳定控制离散生成，同时允许后续推理步骤在需要重新定位时再次查阅视觉证据。此外，SAP 支持多路径推理，可并行探索多样的推理行为。SAP 与模型无关且无数据，无需额外训练。实证结果表明，在词元生成预算相当的情况下，SAP 达到了有竞争力的性能，尤其能减少物体幻觉；与思维链式长序列推理相比，它还带来了更稳定的推理和更低的响应延迟。

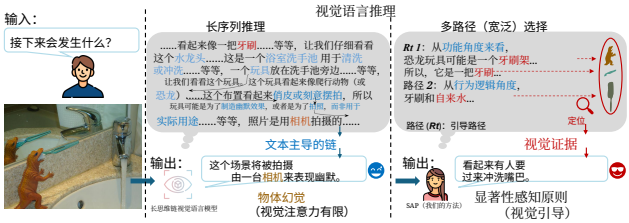


图 *1.* 视觉语言推理中的文本依赖。我们的 SAP 是一种多路径方法，它以视觉定位为指导，缓解了文本主导的 LongCoT 中的物体幻觉。

1. 引言

视觉-语言模型（VLMs）旨在通过联合处理视觉输入（例如图像或区域）和文本输入（例如指令或中间思考）来解决多模态推理任务（Liu 等人，2025）。近期进展在大语言模型（LLMs）领域表明，分配额外的推理时计算——例如生成更长的推理序列或探索多条推理路径——通常可以提升推理质量（Wei 等人，2022；Wang 等人，2022）。这种能力通常称为推理时扩展，已成为增强基于大语言模型的推理系统的核心机制（Liu 等人，2025）。然而，尽管人们对多模态思维链（COT）推理研究的兴趣日益浓厚（Wang 等人，2025c），视觉语言模型能否实现类似的推理时扩展仍是一个悬而未决的问题。

传统的推理时扩展依赖于大语言模型在较长的推理过程中迭代优化推理的能力；随着生成的推进，中间推理状态会被重新审视、修正和扩展（Wei 等人，2022；Yao 等人，2023年）。在多模态场景中，这不仅要求生成更长的文本推理链，还要求在整个推理过程中持续融入并重新评估视觉证据。与仅依赖语言的推理不同，后者的中间表示始终处于同一文本模态；视觉语言推理通常需要反复将文本推理状态与视觉定位对齐（Wang 等人，2025c）。因此，有效扩展的关键在于重新审视视觉信息，而不是仅在推理过程开始时生成一次视觉摘要（这可被视为外推得到的视觉信息，可能会累积误差）。

然而，由于自回归生成过程中文本与视觉模态之间存在系统性差异，要实现如此高质量的视觉定位并不容易。文本表示会在每个解码步骤中显式生成并更新，而视觉信息通常通过固定编码或有限的跨模态交互融入模型（Lu 等人，2019年）。随着

[1]弗吉尼亚大学。通信请联系：Jundong Li <jundong@virginia.edu>。

预印本。*2026*年*2*月*19*日。

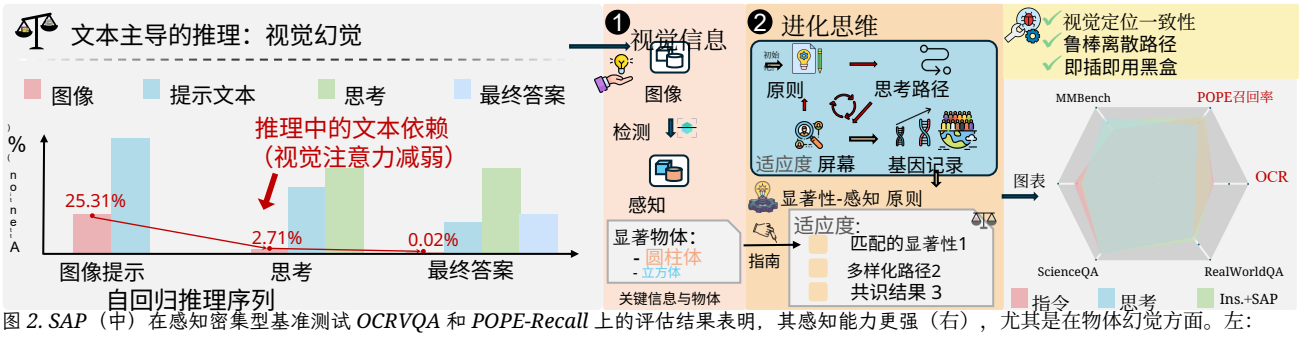


图 2. *SAP*（中）在感知密集型基准测试 *OCRVQA* 和 *POPE-Recall* 上的评估结果表明，其感知能力更强（右），尤其是在物体幻觉方面。左：Qwen3-VL-8B-Thinking 在 MS-COCO 上推理时的文本依赖。右：显著性感知原则（SAP）的选择流程，以及 SAP 与 Qwen3-VL-8B 系列模型在基准测试上的实证比较。

生成长度增加，这种差异会导致推理过程越来越受文本主导。此前的研究通过实验证明，这一现象与幻觉及带有偏见的视觉推理有关（Hendricks 等人，2018年；Rohrbach 等人，2018年）。虽然看起来只需在早期概括视觉内容，之后依靠文本即可（Wang 等人，2025c；Zhao 等人，2025），但这种概括本质上会造成信息损失：早期概括中产生的遗漏或误解，无法在后续推理步骤中得到纠正。因此，推理时扩展会放大早期的视觉定位错误，使推理路径逐渐偏离潜在的视觉证据（图2）。

缓解这一问题的一种通用方法是在推理过程中引入监督信号，促使模型关注利用不足的视觉模态（Zhao 等人，2024年）。然而，为视觉语言推理提供有效指导面临着根本性挑战。在多模态任务中，由于评估者和标注者具有主观性，监督信号往往反映出关于如何使用视觉证据的不一致且隐含的评估原则，从而导致反馈嘈杂且难以校准（Liu 等人，2025）。此外，推理时的推理过程通过离散的文本生成过程展开，而指导信号本质上较为粗略，难以将其传播至中间的多模态推理轨迹（Zhu 等人，2025）。综合来看，这些因素使得在多模态场景中实现稳定且通用的推理时优化尤为困难。

为应对这些挑战，我们针对视觉语言推理提出了显著性感知原则选择 （SAP）。SAP采用一种模型无关的黑盒测试时扩展表述，将视觉显著性作为模态感知引导信号，使信息丰富的视觉证据能够在整个推理过程中 得到持续且高质量的利用，而非依赖一次性的视觉摘要。通过作用于高层推理原则而非词元级轨迹，SAP缓解了早期视觉落地错误的累积，并防止长程推理漂移至纯文本状态。

此外，在离散原则空间中执行推理时优化，使SAP能够对粗粒度且依赖评估器的监督信号保持稳健，从而避免依赖细粒度轨迹评分或特定任务启发式。

实证结果表明，在相似的词元预算下，SAP 无需额外训练或新数据，便能达到与强大的专有视觉语言模型相当的性能。通过将推理时计算从单路径长思维链（LongCoT）（Chen 等人，2025）重新分配到多路径、原则引导的探索，SAP 实现了性能提升，而以往要获得这种提升，需要进行计算效率低下的更深层推理。此外，多路径推理支持并行执行：不同于后续词元严格依赖前序词元的 LongCoT 推理，SAP 可独立实例化和评估多条推理路径。这一特性支持在并行模型实例之间灵活地进行负载均衡，从而在大规模部署中实现更高的吞吐量和更低的延迟。贡献总结如下：

- 我们强调视觉语言模型倾向于以文本为主进行推理这一已知特性，并证明在长时间推理中，这种倾向也可能阻碍推理，因为它会限制单条推理路径在后续推理阶段重新审视视觉证据的能力。
- 我们提出显著性感知原则选择 （SAP），这是一种模型无关且无数据的方法，能够促进更好地利用视觉证据。
- 我们证明，SAP 在保持有竞争力的词元预算和性能的同时，实现了推理时间的扩展。通过并行评估多条推理路径，SAP 的响应延迟低于 LongCoT 式的顺序推理。

2. 问题表述

本文将推理时扩展表述为一个优化问题，即对推理阶段执行的推理路径进行优化。给定提示词 p 和语言模型 φ，推理路径 $\rho \in \mathcal{R}$规定了推理的

展开过程及其步骤，其中 $\mathcal{R}$ 表示可行推理路径的空间。路径 ρ 决定中间推理的结构和进展。推理时扩展的目标是确定一条能够最大化任务效用的推理路径：

$$\rho^\star = \arg\max_{\rho \in \mathcal{R}} \; f(\rho), \tag{1}$$

其中，简化后的 $f(\rho) \leftarrow f(\rho;\, p,\, \varphi)$ 用于评估 ρ 所诱导输出的效用。在此，推理时扩展并非仅指沿单一轨迹生成更长的推理链，而是指分配额外的推理时计算，以探索多条备选推理路径并筛选出质量更高的路径。随着推理预算增加，可以实例化并评估 $\mathcal{R}$ 中更多候选项，从而提高找到效用更高的推理路径的可能性。从这个意义上说，推理时扩展对应于对推理行为的搜索进行扩展，而不只是延长单条路径。

3. 方法

为实现上述目标，我们提出了显著性感知原则选择（SAP），这是一种用于视觉语言推理的推理时优化方法。SAP 由三个关键组件构成：（***i***）原则引导的推理生成，通过高层原则对推理行为进行参数化；（***ii***）演化式原则精炼，在序数型噪声反馈下执行基于种群的选择；以及（***iii***）显著性感知评估，为比较候选原则提供稳定的模态感知信号。SAP 方法的概览如图右侧所示。2。

3.1. 原则引导的推理生成 n

3.1.1. 不可处理的路径搜索

对任务效用进行直接优化 f，在推理路径空间 $\mathcal{R}$ 中进行，如公式（1）所示，以实现视觉语言模型的测试时扩展，由于理论和实践两方面的原因，这难以实现。首先，我们注意到，推理路径并非定长变量，而是语言模型 φ自回归生成所产生的涌现输出。随着推理时计算量的增加，推理路径会变得更长且愈发稀疏，使得直接采样和评估在实践中耗时更多。此外，该空间 $\mathcal{R}$ 本质上是离散且冗余的。对词元序列进行细微的局部修改，往往不会对最终推理结果产生有意义的影响；而语义上不同的推理行为之间，可能相隔着巨大的组合间隙。因此，路径空间中的信息分布稀疏：许多不同的路径会产生相似的最终输出，而高质量路径则十分罕见且难以识别。这种离散性和稀疏性使得难以高效地比较、排序或优化推理路径，从而使得对 $\mathcal{R}$ 进行直接优化并不现实。为了使这一问题适用于优化，我们使用一组高层次的推理原则对路径进行参数化。

3.1.2. 原则引导的推理

我们首先将 原则推理定义为一种推断形式，在这种形式中，推理过程由明确的高层原则引导。这些原则规定模型应当如何进行推理，而不是规定应生成的确切中间词元。与特定路径相比，推理原则为推理行为提供指导——例如，在生成过程中应如何重新审视视觉证据、应如何验证中间假设，或应如何平衡不同模态——但不会将推理限制为某一种特定的词元级实现。

例如，在视觉-语言推理任务中，一条原则可以指示模型每当形成中间文本结论时，都应明确地重新检查视觉证据，或者通过检查图像中所有相关物体的相对位置来验证空间关系。我们注意到，这样的原则 x 本身是一种简洁的文本描述，但它并不规定单一的推理轨迹。相反，在固定的语言模型下，同一原则 $x \in \mathcal{X}$ 可能诱导出多个具体的推理轨迹，这取决于模型如何在词元级实例化该原则。我们将每一种此类实例化表示为在给定原则 x下的推理路径 $\rho(x)$。这种参数化产生了等价的优化目标

$$x^\star = \arg\max_{x \in \mathcal{X}} \; f\big(\rho(x)\big), \tag{2}$$

从而将推理时优化从非结构化的推理路径转向更为紧凑的原则空间。

3.1.3. 原则选择

原则引导的推理将推理时控制从词元级轨迹转向高层推理指令。这一转变提出了一个核心问题：在推理过程中应如何选择并精炼推理原则？

在视觉语言模型中，原则选择具有挑战性，原因有二，且均与多模态推理密切相关。首先，在长程生成过程中，视觉证据越来越难以重新检视，因此有必要优先选择那些在整个推理过程中明确保持视觉定位的原则。此外，推理时可用的反馈本质上是离散且有序的，仅提供推理行为之间的相对比较（例如，偏好 A 而非 B），而不是将其强制转化为标量分数。

这些特性表明，原则选择应直接针对离散候选项进行，依靠相对比较而非绝对分数，并且保持稳健。

算法 1 初始化：采样原则和 τ 路径

要求： 图像 $\mathcal{I}$，提示词 p，#初始化. N，#路径 τ
$\mathcal{O} \leftarrow$ GROUND($\mathcal{I}$)，$x \leftarrow [\,]$
对 $i = 1$ **至** N **执行**
  $x_i \leftarrow \varphi_x(p, \mathcal{O})$ {采样一条原则}
  r_i、$s_i \leftarrow [\,]$、$[\,]$
  对于 $j = 1$ **到** τ **执行**
    $r_{i,j}$，$s_{i,j} \leftarrow \varphi_{rs}(p,\ \mathcal{I},\ x_i)$
    追加（r_i、$r_{i,j}$）；追加（s_i、$s_{i,j}$）
  结束循环
  追加(x, x_i)
结束循环
返回 x，$\{r_{i,j}, s_{i,j}\}_{i\in[N],\, j\in[\tau]}$

表 *1.* 符号说明：显著性感知原则选择 (SAP)。

p, x	提示词 [输入], 推理原则 [变量]
r, s	推理 [思考] 和摘要 [输出]
μ, λ, τ	#精英个体、#新个体、#路径（宽度） [超参数。]
φ, f	大语言模型、效用和适应度指标（r, s） [给定的]

面对含噪且不一致的反馈时也应如此。受这些要求的启发，我们采用基于种群的进化选择策略来优化推理原则。

3.2. 面向视觉语言推理的显著性感知原则演化

基于式（2）中的原则级形式化，我们将视觉语言模型的推理时扩展构建为一种针对推理原则的显著性引导进化优化问题。所提出的方法整合了三个关键组成部分：（***i***）将原则层面推理视为路径发现，（***ii***）将视觉显著性作为模态感知的对齐信号，以及（***iii***）将基于种群的进化搜索作为优化机制。这些组成部分共同构成了一种实用且稳健的方法，可在离散生成和含噪多模态反馈条件下实现推理时扩展。

3.2.1. 方法概述

给定输入图像 $\mathcal{I}$ 和提示词 p，我们首先提取一组显著视觉元素 $\mathcal{O} = \text{GROUND}(\mathcal{I})$，用于识别可能与推理相关的图像区域。这些显著元素作为评估时的模态感知参考，而非直接注入语言模型。基于提示词 p 和显著元素 $\mathcal{O}$的条件，我们初始化一个推理原则种群 $\{x_i\}_{i\in[\mu+\lambda]}$，其中每条原则都规定了如何利用视觉证据的高层指导。

对于每个原则 x_i，语言模型 φ 执行推理，生成多条路径的推理输出和摘要（r 和 s，（τ：#路径））

$$(r_{i,j}, s_{i,j}) \sim \varphi(r, s \mid x_i, p, \mathcal{I}), \quad j \in [\tau],$$

它们实例化了推理路径 $\rho(x_i)$，该路径由

算法 2 ($\mu + \lambda$) 演化：选择精英个体并重采样

要求： 原则 x、图像 $\mathcal{I}$、提示词 p、#路径 τ
$N \leftarrow |x|$ $\{N = \mu + \lambda\}$, $f \leftarrow [\,]$
对于 $i = 1$ **到** N **执行**
  $r_i, s_i \leftarrow$ 多路径推理（p，$\mathcal{I}$，x_i，τ）
  $\ell_i \leftarrow$ 离散评估(r_i, s_i)
  $f_i \leftarrow$ 将映射等级转换为分数(ℓ_i) {仅供内部使用}
  追加(f, f_i)
结束循环
idx $\leftarrow$ 取前 K 项($f, k = \mu$)
$x^{\mu} \leftarrow x$[索引]
$x^{\lambda} \leftarrow [\,]$
对于 $n = 1$ **到** $N - \mu$ **执行**
  $x_n \leftarrow \varphi_x$（p，x^{μ}）{来自精英个体的后代}
  追加（x^{λ}，x_n）
结束循环
$x^{+} \leftarrow$ 拼接（x^{μ}，x^{λ}）
返回 x^{+} 新的原则种群

原则 x_i。该设计确保视觉信息在整个推理过程中始终可用，同时允许不同原则诱导出不同的推理行为。

3.2.2. 显著性感知原则评估

每条原则都会根据一组离散标准进行评估，这些标准涵盖推理质量和视觉定位（详情见附录B）。具体而言，SAP 根据推理结果中的四种序数信号得出，如下所示：

- **共识匹配** c_i**：** 它衡量附录 B 中详述的代表性答案是否由 x_i引出并与种群层面多数意见一致，从而反映不同原则下全局答案的稳定性。

- **原则内多样性** d_i**：** 它反映同一原则下不同 τ 推理路径的差异程度，在鼓励受控探索的同时避免不稳定。

- **不确定性惩罚** u_i**：** 它衡量所选推理路径的置信度，并对系统性过度自信或含糊不清的行为予以惩罚。

- **证据有效性** e_i**：** 它评估推理中提及的视觉实体是否对应于 $\mathcal{O}$ 中有效的显著区域，在不向语言模型暴露显著性信息的情况下确保视觉定位的一致性，并避免使用过于简单的原则。

所有标准的取值均为 {低、中、高}，这体现了多模态监督本质上具有的顺序性。

3.2.3. 演化优化

上述离散标准合并为一个标量适应度：

$$f_i = \sum_{t\in\{c,d,e\}} w_t \mathrm{Score}(t_i) - w_u \mathrm{Pen}(u_i), \qquad (3)$$

其中，Score 和 Pen（代表惩罚）定义为序数映射（low，medium，high）$\to$(0，1，2)，$\{w_c, w_d, w_e, w_u\}$ 则是各项标准对应的权重。

根据适应度分数，我们使用（μ+λ）选择方案进行种群演化。在每次迭代中，保留排名前μ的原则作为精英（适应度最高的个体），同时通过以精英为条件进行采样，生成一组新的 λ 原则。将精英和新生成的原则合并，作为更新后的种群

$$\{x_i^+\}_{i\in[\mu+\lambda]} = \{x_m^\mu\}_{m\in[\mu]} \cup \{x_n^\lambda\}_{n\in[\lambda]}.$$

通过在显著性感知评估下迭代演化原则，所提出的方法将推理时计算资源分配给多样化推理行为的探索，同时逐步倾向于那些在长程推理过程中保持视觉依据的原则。这为式(2)中的原则级优化目标提供了一种基于种群的近似，从而实现视觉语言模型的稳健推理时扩展。详情见算法1和算法2，

3.3. 框架分析

在本部分中，我们将通过明确刻画进化搜索实现优化改进的概率，以及其所覆盖的有效推理原则的范围，从理论上量化 SAP 中推理时扩展的运作方式，具体而言，使用（μ，λ，T）进行刻画，其中T为迭代总次数。在采用精英选择的进化算法中，迭代过滤可确保种群适应度在各迭代代数中非递减，因此随着优化过程的推进，表现最佳的原则会单调改进。

定理 3.1 (通过$(\mu+\lambda)$ 演化实现优化与泛化)。 考虑原则空间 $\mathcal{X}$上的$(\mu+\lambda)$ 进化过程。令 F_t 表示迭代 t时种群中的最优适应度，令 T 表示迭代总次数。则以下性质成立：

1. (优化稳定性) 由于采用精英选择，最优适应度非递减： $F_{t+1} \geq F_t$, $\forall t$。此外，若每个后代样本实例化出适应度严格高于 F_t的原则的概率至少为 q_t ，则单次迭代中适应度提升的概率满足$\Pr(F_{t+1} > F_t)$ $\geq 1-(1-q_t)^\lambda$。

2. (泛化即覆盖) 假设存在一组有效原则 $\mathcal{X}_{\mathrm{good}} \subset \mathcal{X}$，且每个后代样本属于 $\mathcal{X}_{\mathrm{good}}$ 的概率至少为 $q > 0$ 。那么，进化过程在 T 次迭代内发现至少一个有效原则的概率下界为 $1-(1-q)^{\lambda T}$。

定理 3.1解释了推理时扩展如何在 SAP 中产生。优化保证确保了稳定性：精英保留通过在各迭代中保留表现最佳的原则来防止退化。同时，改进界表明，增大 λ 会提高每次迭代中发现严格更优原则的概率。在优化之外，泛化界刻画了不断增长的原则空间覆盖。由于每次迭代评估 λ 个新采样的原则，经过 T 次迭代后，搜索探索了数量级约为 λT的不同候选。随着 λ 或 T 增大，遇到当前种群之外的有效原则的概率单调增长。综合来看，这些结果表明，SAP 中的推理时扩展是通过分配额外计算来探索并在备选推理原则中进行选择而实现的，而不是延长单一推理轨迹。形式化证明见附录D。

3.3.1. 与长思维链的比较

SAP 进一步通过以并行且基于种群的方式运行，展现出相对于传统长思维链（CoT）推理的扩展优势。考虑固定的推理时 token 预算 L。在长思维链推理中，token 沿单一轨迹顺序生成，且基于 Transformer 的解码相对于序列长度产生二次注意力成本，即 $\mathcal{O}(L^2)$，同时还存在严格的顺序依赖，限制了解码步骤之间的并行。相比之下，SAP 将相同的 token 预算分配到多条更短的推理路径上，并以¯ ¯并行方式进行评估。设 $L=(\mu+\lambda)\tau\ell$，其中 ℓ 表示每条推理路径的平均长度。每条路径产生的注意力成本为 $\mathcal{O}(\bar{\ell}^2)$，总注意力成本为$\mathcal{O}\big((\mu+\lambda)\tau\,\bar{\ell}^2\big)$，当 $\bar{\ell} \ll L$ 时，该成本严格低于 $\mathcal{O}(L^2)$。此外，不同原则与路径上的推理路径相互独立，可以并行执行，从而避免长思维链固有的顺序瓶颈。因此，SAP 通过并行探索与选择来扩展推理时推理，而不是拉长单一推理链，从而在推理预算下提供计算效率更高的替代方案。

3.3.2. 无数据与无梯度扩展

SAP 是一种即插即用的方法，完全无需数据和梯度。进化搜索直接在推理原则空间中进行，利用预训练模型中已经编码的固定知识，无需引入额外的训练数据或外部监督。SAP 不会获取新知识，而是在推理过程中探索如何组织和应用现有知识。这与推理的本质高度契合：推理关注的是信息如何被组织、组合和验证，

而非学习了哪些新信息（Zhu 等人，2025）。因此，SAP 通过重新分配计算资源，在模型能力范围内搜索更优的推理行为，从而提升推理质量，同时避免数据收集或重新训练所带来的成本和潜在偏见。

4. 实证研究

4.1. 实验设置

我们在一系列多样化的多模态推理基准测试上评估 SAP。所有实验均采用纯推理设置，不更新任何参数。

模型。我们使用 *Qwen3-VL-8B* (Li等人，2026年）作为骨干视觉语言模型。所有推理、原则生成和聚合步骤均由同一个冻结模型完成。对于视觉定位，我们采用 *Segment Anything Model*（任意分割模型）(Kirillov等人，2023年），用于自动生成掩码，并由轻量级外部图像描述生成器生成对象标签。除非另有说明，模型只能观察原始图像以及简洁的文本视觉依据摘要，无法访问中间视觉裁剪图或区域级监督。为便于比较，公开报告的模型（xAI，2024年；Achiam等人，2023年；Anthropic，2024年；Team等人，2024年）列于表2中。

推理流程。 给定提示词 p 以及一张或多张图像 $\mathcal{I}$，SAP 维护一个包含 $\mu + \lambda$ 条推理原则的种群。对于每条原则 x_i，模型生成 τ条推理轨迹，并在一次单次 前向调用中完成。原则适应度使用式（3）中定义的离散效用函数计算，种群演化遵循（$\mu+\lambda$）选择方案。在最后一代之后，SAP 可选择执行一次聚合调用，以综合得出最终答案。

基准测试。 我们使用 VLMEvalKit（Duan etal.，2024年）在 16 个视觉语言基准上评估 SAP。各基准的详细信息见附录 C。为清晰起见，基准按其评估重点分为两组：推理 / 知识 / 通用 以及光学字符识别 / 文档 / 图表 / 现实世界。所有结果均在统一的推理时解码预算下报告。

超参数。 除非另有说明，我们使用$\mu + \lambda = 4$，并配合 $\mu = 2$、 $\lambda = 2$ 和 $\tau = 2$。离散适应度权重固定为 $w_c = w_d = w_e = w_u = 1$。进化代数总数 T默认设为 2，其扩展情况在效率消融实验中进一步考察。每张图像的已定位物体最大数量设有上限，以限制计算开销，且所有离散化阈值在各数据集上保持固定。

评估协议。 我们遵循各项基准测试的官方评估协议。对于精确答案数据集，我们采用精确字符串匹配。对于多项选择数据集，答案会映射到相应的选项标签。我们不进行数据集特定调优。所有报告的结果均基于完整的验证集或测试集划分计算平均值。在全文中，平均性能 （平均性能）指所有基准测试的平均得分，但不包括 MME，因为它采用不同的评分尺度。

基线方法。我们在相同的模型骨干网络和评估设置下，将 SAP 与标准推理策略进行比较。具体而言，我们考察了： （***i***）指令，直接根据输入图像和问题生成答案；（***ii***）思考，允许在生成过程中明确呈现中间推理轨迹；以及（***iii***）*SAP*，在推理时进行基于原则的进化推理。所有方法均使用相同的检查点和解码配置，且不进行任何额外微调。

4.2. 比较

SAP 在各项基准测试中取得了更高的平均性能，并且其推理时行为比 LongCoT 式扩展稳定得多。尽管思考模型在若干推理密集型基准测试（如 MMBench 和 SEEDBench）上提升了结果，但在对感知和定位敏感的任务上也出现了明显退步。值得注意的是，与 Instruct 基线相比，LongCoT 在 POPE召回率（83.9 → 79.6）、TextVQA（82.3 → 77.2）和 OCRVQA（63.2 → 44.1）上的表现明显下降，这表明更长的顺序推理可能会削弱视觉定位能力。

相比之下，SAP 在这些基准测试上的表现接近基线，性能下降始终控制在较小范围内（通常为 < 5 个百分点），同时在面向推理的任务上仍具竞争力。因此，在相同的推理时间预算下，SAP 的总体平均分高于 LongCoT。重要的是，这一提升并非来自引入额外知识，而是源于保留了在后续推理阶段重新查阅视觉证据的能力（例如更高的 POPE召回率），从而避免了 LongCoT 推理时扩展中观察到的鲁棒性—深度权衡。

4.3. 消融研究

所有模块消融实验均与默认设置以及相应的标准 *In-struct* 和 LongCoT思考基线进行评估和比较。

SAP 的模块。 表 3 表明，SAP 的核心组成部分，即演化和多样性，对于有效的推理时优化是必不可少的。当移除种群演化（EVO GEN=0）后，仅增加路径数量并不能提升性能，这表明没有选择机制的路径多样性是不够的。相反，禁用多路径推理（$\tau = 1$）会显著降

表2. 在16个视觉语言基准测试上对比更长与更宽的推理策略。SAP 在感知方面表现更佳，平均性能更高，同时保持稳定的视觉定位能力。根据评估重点，基准测试分为两组。除非另有说明，所有方法均使用相同的骨干网络和推理时解码预算。基线：Qwen3-VL-8B-Instruct；Qwen3-VL-8B-Thinking：LongCoT、顺序推理时扩展；SAP（本文方法）：在多个路径上进行并行推理时扩展。

	复现结果			报告结果（参考文献）				
基准测试（↑）	指令	思考	ins.+SAP（我们的方法）	Grok-1.5V	GPT-4V	Claude-S	Claude-O	Gemini-1.5
平均值（MME 除外）	75.4	75.9	**76.6**					
第一组：视觉与思维：感知 / 推理 / 常识								
POPE召回率（Li等人，2023b）	83.9	79.6↓	**89.9**↑	–	–	–	–	–
MME-p（Fu 等人，2025）	1744.7	1663.2↓	**1755.5**↑	–	–	–	–	–
MME-r (Fu 等人，2025)	656.7	**722.5**↑	689.9↑	–	–	–	–	–
MMBench(Liu 等人，2024a)	70.7	**87.0**↑	82.9↑	–	–	–	–	–
SEEDBench（Li等人，2023a）	68.5	**78.3**↑	77.3↑	–	–	–	–	–
MMMU（Yue 等人，2024年）	**62.4**	62.3↓	62.3↓	53.6	56.8	53.1	**59.4**	58.5
第二组。特殊案例：科学 / 文本 / 光学字符识别 / 文档 / 图表 / 现实世界								
ScienceQA(Saikh 等人,2022)	88.0	**92.5**↑	92.2↑	–	–	–	–	–
TextVQA（Singh 等人，2019年）	**82.3**	77.2↓	81.3↓	**78.1**	78.0	–	–	73.5
OCRVQA（Pham 等人，2025）	**63.2**	44.1↓	62.8↓	–	–	–	–	–
OCRBench（Liu 等人，2024b）	56.8	**65.9**↑	60.2↑	–	–	–	–	–
ChartQA（Masry 等人，2022）	**82.7**	77.2↓	78.1↓	76.1	78.5	81.1	80.8	**81.3**
DocVQA（Mathew 等人，2021）	**92.2**	91.6↓	90.6↓	85.6	88.4	**89.5**	89.3	86.5
InfoVQA（Mathew 等人，2022）	82.6	**84.1**↑	83.1↑	–	–	–	–	–
AI2D（Kembhavi 等人，2016）	79.2	**82.2**↑	80.1↑	88.3	78.2	**88.7**	88.1	80.3
BLINK（Fu 等人，2024b）	**68.2**	60.3↓	62.6↓	–	–	–	–	–
RealWorldQA (实)	70.3	72.8↑	**73.3**↑	**68.7**	61.4	51.9	49.8	67.5

表 3。 SAP 模块的消融实验。不应聚合有害原则。 演化和过滤经过验证都很重要。

设置	#生成式进化	#路径 τ	决策	平均性能
指令	0	1	–	75.4
无进化	0	τ	精英	73.4
无多路径	T	1	精英	73.0
完整 SAP	T	τ	精英	**76.6**
使用摘要器	T	τ	agg.	74.8

低性能，因为演化式优化退化为单轨迹优化。完整 SAP 配置，
它结合了演化式优化和精英选择的路径多样性偏好，取得了最佳平均性能。将精英选择替换为聚合会进一步降低性能，这表明过多的路径和原则实际上会适得其反，而有效的推理时推理需要进行明确的过滤。

进化超参数。以下结果表明，由于 SAP 不会引入新知识，其推理时扩展存在固有上界。增加精英个体数量 μ 并不能持续提升性能，这说明较高的存活率可能会保留次优原则，并阻碍种群的演化。总体而言，在推理时优化中，均衡探索比最大化存活率更有效。

λ #新生个体	2	2	4	4	6
μ #精英个体	2	4	2	4	6
平均性能	76.6	74.8	**76.8**	76.5	**76.8**

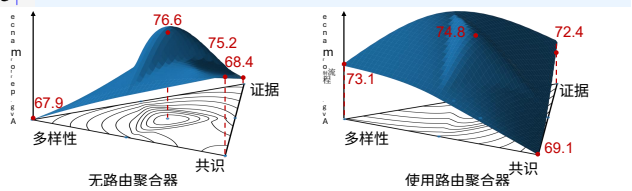


图3。适应度中奖励项权重的消融实验（损失设计）：聚合路径以生成最终答案，可降低多模态推理对共识的依赖。多样性、共识、证据分别指 w_d、w_c、w_e 的权重比例，其中 $\sum w = 1$，且不确定性惩罚始终为 1，$w_u = 1$。

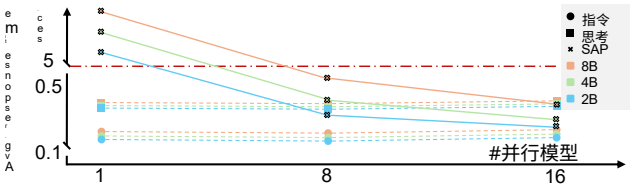


图4。SAP 支持通过并行实现推理时加速。

适应度权重。此消融实验考察离散适应度函数中不同权重组合的影响。我们调整第3.2.2节所定义标准的相对权重，即共识匹配、原则内多样性、证据一致性和不确定性惩罚，同时保持进化过程不变。该分析评估了原则选择中的不同归纳偏置如何影响推理时推理行为。图3显示，每个适应度组成部分的重要性取决于是否采用路径聚合。

在默认的非聚合设置中，提高共识权重（w_c）的配置表现更好，表明推理路径之间的明确一致对于稳定最终预测至关重要。启用额外的路径聚合以进行最终决策时，性能对原则内多样性的敏感度显著提高（w_d），因为聚合会隐式地促成一致，并将优化重点转向探索多样且富有信息的推理行为。在聚合和非聚合设置中，证据一致性始终是必要的补充信号。仅凭证据依据不足以带来出色表现；而将较高的证据一致性与较强的共识（不使用聚合）或较高的多样性（使用聚合）相结合，则能获得更稳定、可靠的推理结果。

响应时间。图4 表明，在单设备上，由于采用多路径生成和迭代式原则细化，SAP 的延迟高于 LongCoT 式推理。然而，在并行设置下，SAP 的响应时间显著更短。SAP 将推理分解为 T轮短路径生成和原则更新（默认每轮包含T=2 条路径），每轮中的所有路径彼此独立，可以并行执行。这种结构避免了 LongCoT 所需的长 token 级依赖和大量中间上下文同步，从而减少流水线停顿并提高 GPU利用率。通过调整迭代次数和路径数，SAP 在总 token 数和计算量上与 LongCoT 相当，同时借助并行实现更低的实际运行延迟。

模型规模。为考察模型规模的影响，我们还评估了 SAP 在 Qwen3-VL-Instruct-2B、-4B 和 -30B 上的平均性能。SAP 始终能够提升这些模型的性能。

尺寸	2B	4B	8B	30B
指令	60.4	75.4	75.4	77.3
+SAP	66.6	75.6	76.6	78.1

模型架构。为了进一步验证 SAP 的通用性，我们在其他视觉语言模型骨干网络上对其进行评估，包括 InternVL3.5-8B（Wang 等人，2025a）以及 DeepSeek-VL2（4.5B 混合专家）（Wu 等人，2024b）。在这两种模型上，SAP 均持续提升总体基准测试与推理导向基准测试上的实证性能，表明其收益并不依赖于特定架构或模型规模。这些结果表明，SAP 可作为适用于多种视觉语言模型的通用推理时优化策略。

模型	InternVL3 5	+SAP	DeepSeek-VL2	+SAP
MMMU	74.6	75.3	51.2	56.7
MMBench	83.4	84.6	79.5	79.8

5. 相关工作

多模态推理时扩展。 近期研究认为，降低对视觉特征的过度依赖可以提升视觉语言模型的鲁棒性（Wang 等人，2024年），尤其是通过缓解伪相关和幻觉（Zhao 等人，2025；Wang 等人，2025c）。我们补充了这一观点：在长序列推理过程中，若后续推理步骤需要重新获取视觉证据，过度抑制视觉特征可能有害。Lin 等人（Lin 等人，2025）研究了采用多模态思维链的推理时扩展，证明其相较纯文本推理具有优势，但代价是冗长的顺序生成。Wang 等人（Wang 等人，2025b）引入了一种视觉价值模型，用以引导推理时搜索朝向视觉定位的迭代代数。然而，现有方法主要沿单一推理轨迹扩展推理，要么延长思维链，要么执行顺序搜索。相比之下，我们的工作探索多路径推理，将计算分配给并行的原则级路径，并配合离散评估，从而能够在视觉证据中进行选择性重新落地，并提升稳定性与并行性。

进化提示词。 以往研究表明，通过提示词，大语言模型（LLM）能够表现出推理行为（Kojima 等人，2022）。研究发现，简单的指令级线索即可实现零样本推理，而（Pryzant 等人， 2023年）则通过梯度启发式搜索，将提示词视为可优化的物体。最近，（Wu 等人，2024a）综述了进化计算与大语言模型日益扩大的交叉领域，（vanStein &B¨ack，2024年）提出了 LLAM-EA，其中进化算子作用于由大语言模型生成的算法结构。从基于提示词的推理转向涉及大语言模型的进化框架，主要是在离线环境中应用演化。与此不同，我们研究的是推理时的进化推理，即结构化推理单元的种群不断演化。

6. 结论

在这项工作中，我们重新审视了视觉语言模型普遍倾向于文本主导的推理这一现象，并表明在长时程推理中，这种行为会产生负面影响：当后续推理步骤需要重新查阅视觉证据时，它会阻碍这一过程。基于这项分析，我们为视觉语言推理提出了一种新的推理时扩展范式：显著性感知原则选择 （SAP）。该方法无数据、与模型无关，并在高层推理原则而非词元级轨迹的层面上运行。SAP 通过支持多路径、原则引导的探索，并持续获取视觉证据，为顺序式 LongCoT 推理提供了一种有效且可并行化的替代方案。

影响声明

本文旨在推动机器学习领域的发展。我们的工作可能带来许多潜在的社会影响，但我们认为无需在此特别强调其中任何一项。进一步优化并行推理效率，或可减少不必要的推理开销并降低能耗。特别是，如果能为所有问题建立高效的原则库，就可以进一步减少各条路径上的并行推理消耗，从而进一步降低能耗。

参考文献

RealWorldQA 数据集。
https://huggingface.co/datasets/visheratin/realworldqa。

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. 等。GPT-4 技术报告。*arXiv* 预印本 *arXiv:2303.08774*，2023年。

Anthropic, A. Claude 3 模型系列：Opus、Sonnet、Haiku。*Claude-3* 模型卡，1(1):4，2024年。

Chen, Q., Qin, L., Liu, J., Peng, D., Guan, J., Wang, P., Hu, M., Zhou, Y., Gao, T., and Che, W. 迈向推理时代：推理大语言模型长思维链综述。*arXiv* 预印本 *arXiv:2503.09567*，2025。

Duan, H., Yang, J., Qiao, Y., Fang, X., Chen, L., Liu, Y., Dong, X., Zang, Y., Zhang, P., Wang, J., et al. VLMEvalKit：用于评估大型多模态模型的开源工具包。载于第*32*届 *ACM* 国际多媒体会议论文集，第11198–11201页，2024年。

Fu, C., Chen, P., Shen, Y., Qin, Y., Zhang, M., Lin, X., Yang, J., Zheng, X., Li, K., Sun, X., et al. MME：多模态大语言模型的综合评估基准测试。载于神经信息处理系统第三十九届年会数据集与基准测试专题，2025。

Fu, L., Kuang, Z., Song, J., Huang, M., Yang, B., Li, Y., Zhu, L., Luo, Q., Wang, X., Lu, H., 等。OCRBench v2：用于评估大型多模态模型视觉文本定位与推理能力的改进型基准。*arXiv* 预印本 *arXiv:2501.00321*，2024a。

Fu, X., Hu, Y., Li, B., Feng, Y., Wang, H., Lin, X., Roth, D., Smith, N. A., Ma, W.-C., and Krishna, R. Blink：多模态大语言模型能够看见，却无法感知。载于欧洲计算机视觉会议，第 148–166 页。施普林格，2024b。

Hendricks, L. A., Burns, K., Saenko, K., Darrell, T., and Rohrbach, A. 女性也玩单板滑雪：克服图像描述模型中的偏见。载于 欧洲计算机视觉会议论文集（*ECCV*），第 771–787 页，2018年。Kembhavi, A., Salvato, M., Kolve, E., Seo, M., Hajishirzi, H., and Farhadi, A. 一张图表胜过十几张图像。载于欧洲计算机视觉会议，第 235–251 页。施普林格，2016。

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., 等。任意分割。发表于*IEEE/CVF*国际计算机视觉会议论文集，第 4015–4026 页，2023年。

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y. 和 Iwasawa, Y. 大语言模型是零样本推理器。神经信息处理系统进展，35：22199–22213，2022。

Li, B., Wang, R., Wang, G., Ge, Y., Ge, Y. 和 Shan, Y. Seed-Bench：通过生成式理解对多模态大语言模型进行基准测试。*arXiv* 预印本 *arXiv:2307.16125*，2023a。

Li, M., Zhang, Y., Long, D., Chen, K., Song, S., Bai, S., Yang, Z., Xie, P., Yang, A., Liu, D. 等。Qwen3-vl-embedding 和 Qwen3-VL-Reranker：用于先进多模态检索与排序的统一框架。*arXiv* 预印本 *arXiv:2601.04720*，2026年。

Li, Y., Du, Y., Zhou, K., Wang, J., Zhao, W. X. 和 Wen, J.-R. 评估大型视觉语言模型中的物体幻觉。*arXiv* 预印本 *arXiv:2305.10355*，2023b。

Lin, Y., Wang, A., Chen, M., Liu, J., Liu, H., Su, J., 和 Xiao, X. 探究多模态思维链的推理时扩展：一项初步研究。*arXiv* 预印本 *arXiv:2502.11514*，2025。

Liu, Y., Duan, H., Zhang, Y., Li, B., Zhang, S., Zhao, W., Yuan, Y., Wang, J., He, C., Liu, Z., 等。MMBench：你的多模态模型是全能选手吗？载于欧洲计算机视觉会议，第 216–233 页。施普林格，2024a。

Liu, Y., Li, Z., Huang, M., Yang, B., Yu, W., Li, C., Yin, X.-C., Liu, C.-L., Jin, L., 和 Bai, X. OCRBench：大型多模态模型中光学字符识别的隐秘之谜。中国科学：信息科学，67(12):220102，2024b。

Liu, Z.、Wang, P.、Xu, R.、Ma, S.、Ruan, C.、Li, P.、Liu, Y. 与 Wu, Y. 通才奖励建模的推理时扩展。*arXiv* 预印本 *arXiv:2504.02495*，2025。

Lu, J.、Batra, D.、Parikh, D. 与 Lee, S. Vilbert：面向视觉与语言任务的任务无关视觉语言表征预训练。神经信息处理系统进展，32，2019年。

Masry, A., Do, X. L., Tan, J. Q., Joty, S. 和 Hoque, E. Chartqa：一个关于图表问答的基准，结合视觉推理与逻辑推理。载于计算语言学协会论文集：*ACL 2022*，第 2263–2279 页，2022。

Mathew, M., Karatzas, D. 和 Jawahar, C. Docvqa：一个用于文档图像视觉问答的数据集。载于*IEEE/CVF*冬季计算机视觉应用会议论文集，第 2200–2209 页，2021。

Mathew, M., Bagal, V., Tito, R., Karatzas, D., Valveny, E. 和 Jawahar, C. 信息图视觉问答。载于*IEEE/CVF*冬季计算机视觉应用会议论文集，第 1697–1706 页，2022。

Pham, H. Q., Nguyen, T. K.-B., Van Nguyen, Q., Tran, D. Q., Nguyen, N. H., Van Nguyen, K., 和 Nguyen, N. L.-T. Viocrvqa：用于通过理解图像中的越南语文本进行视觉问答的新型基准数据集和视觉阅读器。多媒体系统，31(2)：106，2025。

Pryzant, R., Iter, D., Li, J., Lee, Y. T., Zhu, C., 和 Zeng, M. 使用“梯度下降”和束搜索进行自动提示优化。*arXiv* 预印本 *arXiv:2305.03495*，2023年。

Rohrbach, A., Hendricks, L.A., Burns, K., Darrell, T., 和 Saenko, K. 图像描述生成中的物体幻觉。*arXiv* 预印本 *arXiv:1809.02156*，2018年。

Saikh, T., Ghosal, T., Mittal, A., Ekbal, A., and Bhattacharyya, P. Scienceqa：一种用于学术文章问答的新型资源。国际数字图书馆杂志, 23(3):289–301, 2022.

Singh, A., Natarajan, V., Shah, M., Jiang, Y., Chen, X., Batra, D., Parikh, D., and Rohrbach, M. 面向能够阅读的视觉问答模型。载于*IEEE/CVF*计算机视觉与模式识别会议论文集, pp. 8317–8326, 2019年.

Team, G., Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., Tanzer, G., Vincent, D., Pan, Z., Wang, S., et al. Gemini 1.5：解锁跨数百万词元上下文的多模态理解。*arXiv*预印本 *arXiv:2403.05530*, 2024年.

van Stein, N. 和 B¨ack, T. Llamea：一种用于自动生成元启发式算法的大语言模型进化算法。*IEEE*进化计算汇刊，2024年。

Wang, W., Gao, Z., Gu, L., Pu, H., Cui, L., Wei, X., Liu, Z., Jing, L., Ye, S., Shao, J., 等。Internvl3. 5：提升开源多模态模型在通用性、推理和效率方面的表现。*arXiv*预印本 *arXiv:2508.18265*，2025a。Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., 和 Zhou, D. 自洽性提升了语言模型中的思维链推理。*arXiv*预印本 *arXiv:2203.11171*，2022年。

Wang, X., Yang, Z., Li, L., Lu, H., Xu, Y., Lin, C.-C., Lin, K., Huang, F., and Wang, L. 利用视觉价值模型扩展推理时搜索，以提升视觉理解。载于*IEEE/CVF*国际计算机视觉会议论文集，第 1173–1184 页，2025b。

Wang, Y., Sun, P., Li, Y., Zhang, H., and Hu, D. 文本语义能否缓解声音对象分割偏好？载于欧洲计算机视觉会议，第 340–356 页。施普林格，2024年。

Wang, Y., Wu, S., Zhang, Y., Yan, S., Liu, Z., Luo, J., and Fei, H. 多模态思维链推理：一项全面综述。*arXiv* 预印本 *arXiv:2503.12605*，2025c。

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al. 思维链提示词能够激发大语言模型的推理能力。载于神经信息处理系统进展，35:24824–24837，2022。

Wu, X., Wu, S.-h., Wu, J., Feng, L., and Tan, K. C. 大语言模型时代的进化计算：综述与路线图。*IEEE*进化计算汇刊，2024a。

Wu, Z., Chen, X., Pan, Z., Liu, X., Liu, W., Dai, D., Gao, H., Ma, Y., Wu, C., Wang, B., 等。Deepseek-vl2：用于高级多模态理解的混合专家视觉语言模型。*arXiv* 预印本 *arXiv:2412.10302*，2024b。

xAI。Grok-1.5 视觉。https://x.ai，2024。技术报告。

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., 和 Narasimhan, K. 思维树：借助大语言模型进行审慎的问题求解。神经信息处理系统进展，36:11809–11822，2023年。

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超大规模跨学科多模态理解与推理基准。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 9556–9567 页，2024年。

Zhao, J., Wei, X. 和 Bo, L. R1-omni：采用强化学习的可解释的全方位多模态情绪识别。*arXiv* 预印本 *arXiv:2503.05379*，2025。

Zhao, L., Deng, Y., Zhang, W., andGu, Q. 通过基于图像的引导缓解大型视觉语言模型中的物体幻觉。*arXiv* 预印本 *arXiv:2402.08680*，2024年。

Zhu, H.、Hao, S.、Hu, Z.、Jiao, J.、Russell, S. 和 Tian, Y.《叠加推理：关于连续思维链的理论视角》。*arXiv* 预印本 *arXiv:2505.12514*，2025。

A. 更多讨论

局限性。 我们指出当前方法仍存在一些局限，有待进一步研究。SAP 只有在大规模或并行推理设置下才能发挥最大效用并快速响应，在这些设置中，其多路径设计可以摊销计算成本。此外，其性能取决于进化系统的提示模板和原则生成的质量，这需要进行谨慎设计。未来的工作可以探索用于指导原则生成本身的元级原则，从而实现对原则空间中多样性和随机性的更系统控制。

视觉模态在文本自回归生成过程中的作用逐渐减弱。如图2所示，在漫长的视觉语言推理过程中，随着生成不断推进，视觉信息的贡献往往会逐渐减弱。推理早期的步骤可能会积极依赖视觉证据，而后续步骤则越来越专注于文本空间。这一现象反映了模态之间的根本差异：文本表示会在每个解码步骤中显式生成并更新，而视觉输入通常仅通过固定编码或有限的跨模态交互间接融入。一旦视觉信息被压缩进早期的文本摘要，任何遗漏或误解都很难在后续推理步骤中得到纠正。因此，随着生成长度增加、推理时计算规模扩大，这些早期的接地错误反而会加剧，而非得到缓解。除了这种模态不平衡外，视觉语言模型的推理时优化还面临两项内在挑战。

挑战一：信号不一致且具有序数性。首先，多模态推理中的监督信号本质上是不一致且具有序数性的。对视觉语言推理输出的评估——无论由人还是模型提供——往往基于异质且隐含的标准，尤其是在如何使用视觉证据方面。对于相同的推理结果，不同的评估器可能会强调不同方面，例如视觉定位、逻辑完整性或简洁性。与此类不一致主要影响风格偏好的纯语言场景不同，在视觉语言推理中，这种不一致会直接影响视觉证据是否以及如何被纳入推理。因此，监督并未定义一个经过良好校准的单一标量目标。相反，自然可得的是一种序数信号：它表示一种偏序，指出哪些推理行为优于其他行为，而不是将异质判断强行转化为单一定量分数。

挑战二：稀疏且离散的搜索空间。 其次，推理时推理的搜索空间是离散且稀疏的。无论是令牌级推理轨迹，还是更高层次的推理原则，都是从组合空间中抽取的离散对象，缺乏平滑的局部结构。在视觉语言推理中，这一挑战会进一步加剧，因为不同的推理行为可能依赖于不同的视觉证据子集，使得候选项之间的局部比较难以明确定义。因此，依赖连续性假设或局部梯度信息的优化策略并不适用于这一场景。

综合来看，这些特性表明，视觉语言模型中的有效推理时优化应当：（i）直接作用于离散候选项；（ii）依赖相对比较，而非绝对标量监督；（iii）在含噪且不一致的反馈下保持稳健。这些要求促成了一种面向推理原则层面、基于种群且由选择驱动的优化范式，我们在方法中采用了这一范式。

B. I实现细节与代码报告：显著性感知原则选择 n

SAP 概述。 给定一个问题 p 以及一张或多张图像 $\mathcal{I} = \{I_k\}_{k=1}^{K}$，SAP 执行以下步骤：（i）定位；（ii）原则生成与进化选择；（iii）最终聚合。在整个过程中，大语言模型只观察原始图像 和一份简洁的文本视觉依据摘要；物体行和裁剪区域仅用于适应度评估，绝不会反馈给大语言模型。所有评估信号均被离散化为三个等级：低/中/高。

步骤 1：多图像视觉定位。对于每张图像 I_k，SAP 会调用视觉定位服务（例如基于SAM的AMG），获取一组已定位物体 $\mathcal{O}_k = \{o_{k,j}\}_{j=1}^{M_k}$，每个物体都带有一个标签 $\ell_{k,\ j}$（由外部图像描述生成器/标注器提供），以及可选的掩码/边界框/分数元数据。SAP 会将 $\{\mathcal{O}_k\}_{k=1}^{K}$ 转换为简洁的视觉依据摘要字符串 g，其中仅描述数量和图像尺寸（也可附带其他轻量统计数据）。SAP 还会构建一个标签映射，以便将任何物体引用规范化为 img#k obj#j($\ell_{k,\ j}$)。

步骤 2：初始化原则种群。SAP 通过一次大语言模型调用，在以（$p,\ g$）为条件的情况下采样初始原则种群 $\{x_i\}_{i=1}^{\mu+\lambda}$（如果实现需要，也可以选择性地加入图像，但并非必需）。每个 x_i 都是一段简短的文本

指令 ，用于指定一种推理策略（例如，“在回答前使用已定位物体进行验证”）。

步骤 3：每条原则只调用一次，进行多路径推理。 对于每次生成，SAP 都会针对每条原则 x_i 仅调用一次大语言模型。提供给大语言模型的内容包括（p、 g、 $\mathcal{I}$、 x_i），并要求其在同一原则下，在内部生成 τ条不同的推理路径：

$$\{(r_{i,j}, s_{i,j})\}_{j=1}^{\tau}.$$

每条路径都会生成一个候选最终答案、一份简短的理由列表、一个不确定性等级 $u_{i,\ j}$ ∈{低、中、高}，以及一份证据引用列表。大语言模型还会输出一个原则内多样性等级$d_i \in$ {低、中、高}，用于概括这些 τ路径之间的差异程度。

步骤 4：代表性答案与全局共识。 SAP 会为每条原则选出一条代表性路线（通常优先选择不确定性较低且答案非空的路线），并据此得到一个代表性答案 $y_i^\star$。随后，SAP 会对 $\{y_i^\star\}$ 进行多数投票，计算出全局共识答案 $\bar{y}$，并将获胜比例转换为离散的共识等级$c \in$ {低、中、高}。此外，SAP 还会为每条原则指定一个离散的共识匹配等级

$$c_i = \begin{cases} \texttt{high}, & y_i^\star = \bar{y}, \\ \texttt{median}, & y_i^* \sim_\varphi \bar{y} \\ \texttt{low}, & \text{otherwise}. \end{cases}$$

该设计强制规定，仅当各原则代表的答案与种群层面的一致意见相符时，这些原则才会获得奖励。 $y^* \sim_\varphi \bar{y}$ 表示 y^* 被判定为与 $\bar{y}$ 含义相同，由语言模型 φ进行判定。

步骤5：证据验证与离散化。 SAP使用标签映射，将每条被引用的证据项规范化为格式img#k obj#j($\ell_{k,\ j}$)。随后，它评估所引用的对象索引是否为有效的接地输出（有效引用与无效引用）。有效引用的比例被离散化为证据等级 $e_i \in$ {低，中， 高}，空证据视为低。该证据等级用作轻量级接地一致性信号；它从不要求大语言模型查看对象裁剪区域。

步骤6：离散等级适应度与选择。 对于每个原则 x_i，SAP根据四个离散信号计算标量适应度分数：共识匹配 c_i、原则内多样性 d_i、代表性不确定性 u_i （来自代表性路线）以及证据等级 e_i。在内部，SAP将低/中/高映射为序数分数，并用整数权重将它们组合：

$$f_i \leftarrow w_c \,\mathrm{Score}(c_i) + w_d \,\mathrm{Score}(d_i) + w_e \,\mathrm{Score}(e_i) - w_u \,\mathrm{Pen}(u_i),$$

其中 Score(low、medium、high) =(0、1、2)，Pen(low、medium、high) =(0、1、2)。SAP 将排名靠前的 μ 个原则选为精英个体，并在还有迭代代数时，要求大语言模型根据精英个体及其适应度提出 λ 个新原则，从而形成下一代 ($\mu + \lambda$) 种群。

步骤 7：最终聚合。最后一代结束后，SAP 通过聚合已评估的原则生成最终答案。可选地，SAP 会基于 (p、 g) 以及代表性摘要列表，再调用一次大语言模型，综合生成包含简要理由和离散不确定性等级的最终答案。如果聚合器失败，SAP 将回退到全局共识组中的代表性答案，并采用保守的不确定性等级。

B.1. 框架：原则进化推理

基于上述表述，我们通过基于种群的进化搜索来近似原则层面的优化目标，并将其称为原则进化推理 （PER）。PER 被提出作为一种通用的推理时扩展框架，通过依据效用函数 f 迭代改进推理原则种群。

整体过程包括初始化和迭代进化。如算法 1 所示，根据提示词 $\{x_i\}_{i\in[\mu+\lambda]}$ 对初始原则种群进行采样 p。对于每个原则 x_i，语言模型会生成多个推理轨迹和摘要，

$$(r_{i,j}, s_{i,j}) \sim \varphi(r, s \mid x_i, p), \quad j \in [\tau],$$

这些实例化了由该原则诱导的推理路径 $\rho(x_i)$。

给定这些 e 对于输出，每条原则都通过特定任务的效用指标进行评估，从而得出适应度分数

$$f_i = f(\{r_{i,j}, s_{i,j}\}),$$

用于估计相应推理路径的有效性。基于这些适应度分数，PER 按照算法中的描述进行种群演化。2。具体而言，采用 $(\mu + \lambda)$ 选择方案：排名前μ 条原则

$$\{x_m^{\mu}\}_{m\in[\mu]} = \mathrm{Top}_{\mu}\big(\{(x_i, f_i)\}_{i\in[\mu+\lambda]}\big)$$

被保留 为精英个体，同时通过以精英个体为条件进行采样，生成一组新的 λ原则 s,

$$x_n^{\lambda} \sim \varphi(x \mid \{x_m^{\mu}\}_{m\in[\mu]}, p), \quad n \in [\lambda].$$

精英个体和新生成的原则合并后，构成更新后的种群

$$\{x_i^{+}\}_{i\in[\mu+\lambda]} = \{x_m^{\mu}\}_{m\in[\mu]} \cup \{x_n^{\lambda}\}_{n\in[\lambda]}.$$

通过迭代应用种群演化，PER 为方程（2）中的优化问题提供了一种基于种群的近似方法，使模型能够在离散生成和噪声反馈下实现鲁棒的推理时扩展。

C. 相关基准测试

基准测试。我们在一组多样化的视觉语言基准上评估 SAP，涵盖鲁棒性、通用推理、科学理解、以 OCR 为中心的推理、文档分析和现实世界泛化。

POPE （Li等人，2023b）通过构建均衡的正样本和负样本，评估物体幻觉和存在性验证。它直接衡量模型能否将答案正确地建立在视觉证据之上，而非先验偏差。

MME （Fu等人，2025）通过区分感知能力和推理能力来评估多模态鲁棒性。其推理子集侧重于超越直接视觉识别的逻辑推断，因此适合用于评估推理时推理策略。

MMBench （Liu 等人，2024a）是一个用于通用多模态理解的大规模多项选择基准，涵盖物体识别、属性推理、空间关系和常识推断。它是衡量视觉语言模型整体能力的标准指标。

SEEDBench (Li等人，2023a) 评估感知、逻辑和常识任务中的通用视觉推理与理解，并着重考察模型在异构问题类型中的一致性。

MMMU (Yue 等人，2024年) 面向数学、物理、化学和工程学领域的大学水平多模态推理。许多问题涉及图表或多张图像，需要结构化推理和跨图像整合。

ScienceQA (Saikh 等人，2022) 评估结合视觉语境的科学问答，将图表与领域知识和多步推理相结合。

TextVQA (Singh 等人，2019年) 衡量场景文本理解能力，正确作答取决于能否读取并推理自然图像中嵌入的文本。

OCRVQA (Pham 等人，2025) 和 **OCRBench** (Liu 等人，2024b；Fu等人，2024a) 进一步强调复杂布局和严苛视觉条件下以 OCR 为中心的推理，重点关注文本定位的准确性和鲁棒性。

ChartQA （Masry 等人，2022）评估对图表和绘图的推理，要求基于可视化数据进行数值比较和解读。

DocVQA （Mathew 等人，2021）和 **InfoVQA** （Mathew 等人，2022）聚焦于文档理解，模型必须对文本内容密集的扫描文档、表格和表单进行推理。

AI2D （Kembhavi 等人，2016）测试科学教育场景中的图解理解，强调关系推理和结构推理，而非视觉组件。

BLINK （Fu 等人，2024b）评估多模态实体链接，要求将视觉实体与背景知识进行对齐。

RealWorldQA 使用真实生活图像和无约束问题评估开放世界鲁棒性，测试超越精选基准的泛化能力（rea）。
https://huggingface.co/datasets/visheratin/realworldqa

D. 算法分析

D.1. 符号与设定

令 $\mathcal{X}$ 表示推理原则的（离散）空间。（$\mu+\lambda$）进化过程维持一个种群

$$P_t = \{x_i^{(t)}\}_{i=1}^{N}, \qquad N = \mu + \lambda,$$

在迭代 $t \in \{0, 1, \ldots, T\}$。每个原则 $x \in \mathcal{X}$通过一次模型调用产生一组推理结果（轨迹和摘要），并被赋予一个适应度值 $f(x) \in \mathbb{R}$ 通过离散评估与序数映射（详见式（3））。将迭代 t 时种群中的最优适应度定义为

$$F_t = \max_{x \in P_t} f(x).$$

精英个体与后代。 在第 t 次迭代中，精英集合 $E_t \subseteq P_t$ 由 P_t 中按适应度排序的前 μ 个原则组成。后代多重集 $O_t = \{X_n^{(t)}\}_{n=1}^{\lambda}$ 通过从基于大语言模型的提议器所诱导的条件分布中采样生成：

$$X_n^{(t)} \sim \varphi_x(\cdot \mid p, E_t), \qquad n \in [\lambda],$$

下一代种群则由以下方式形成：

$$P_{t+1} = E_t \cup O_t.$$

我们强调，SAP 中的离散评估仅用于产生序数等级以及内部排序/选择信号；这里不假设进行连续置信度校准。

D.2. 关于采样与依赖性的假设

假设 D.1 （条件独立同分布后代采样）。 给定截至第 t 次迭代的历史记录 $\mathcal{H}_t$ （包括 p、所有过去的种群，因此也包括 E_t），后代样本 $X_1^{(t)}, \ldots, X_\lambda^{(t)}$ 根据 $\varphi_x(\cdot \mid p, E_t)$ 条件独立同分布。

当每个后代都是在相同的 D.1 条件下，通过独立调用提议器（或使用独立的解码随机性）生成时，(p, E_t)假设成立。若实现中引入了轻微依赖，则可在标准负依赖条件下扩展下文的概率界；为表述清晰并方便审阅，我们采用独立同分布形式。

D.3. 优化保证

D.3.1. 单调最优适应度（无退化）

定理 D.2 (在 $(\mu+\lambda)$ 选择下，最优适应度单调递增。 对于所有迭代 $t \geq 0$，最优适应度均非递减：

$$F_{t+1} \geq F_t.$$

证明。 根据定义， E_t 包含按 $f(\cdot)$ 排序的 P_t 中排名前 μ 的元素，因此至少包含一个在 P_t上使f 最大化的最大化点。因此

$$\max_{x \in E_t} f(x) = \max_{x \in P_t} f(x) = F_t.$$

由于 $P_{t+1} = E_t \cup O_t$ 和 $E_t \subseteq P_{t+1}$,

$$F_{t+1} = \max_{x \in P_{t+1}} f(x) \geq \max_{x \in E_t} f(x) = F_t.$$

□

D.3.2. 单步改进概率

虽然定理D.2 表明了稳定性（无回归），优化速度取决于该过程产生严格更优原则的可能性。

改进集合。 定义能够严格改进当前最佳结果的原则集合：

$$\mathcal{A}_t = \{x \in \mathcal{X} : \ f(x) > F_t\}.$$

假设 **D.3 假设**（每次迭代改进质量）。 存在 $q_t \in [0,1]$，使得在给定 $\mathcal{H}_t$ 的条件下，

$$\Pr\Big(X_n^{(t)} \in \mathcal{A}_t \mid \mathcal{H}_t\Big) \geq q_t \quad \text{for each } n \in [\lambda].$$

假设 D.3 不要求显式知道 $\mathcal{A}_t$；它只假定提议者分布在严格改进的原则上保留至少 q_t的概率质量。

定理 D.4（单步改进界）。 在假设 *D.1* 和 *D.3*下，迭代 t 处严格改进的概率满足

$$\Pr(F_{t+1} > F_t \mid \mathcal{H}_t) \geq 1 - (1 - q_t)^{\lambda}.$$

因此，

$$\Pr(F_{t+1} > F_t) \geq \mathbb{E}\big[\, 1 - (1 - q_t)^{\lambda} \,\big].$$

证明。 若至少有一个后代位于 $\mathcal{A}_t$ 中，则发生严格改进：

$$\{F_{t+1} > F_t\} \supseteq \bigcup_{n=1}^{\lambda} \{X_n^{(t)} \in \mathcal{A}_t\}.$$

等价地，未发生改进意味着没有任何后代命中 $\mathcal{A}_t$:

$$\{F_{t+1} = F_t\} \subseteq \bigcap_{n=1}^{\lambda} \{X_n^{(t)} \notin \mathcal{A}_t\}.$$

对 $\mathcal{H}_t$ 进行条件化，并利用条件独立性（假设 D.1），

$$\begin{aligned}
\Pr(F_{t+1} = F_t \mid \mathcal{H}_t) &\leq \Pr\left(\bigcap_{n=1}^{\lambda} \{X_n^{(t)} \notin \mathcal{A}_t\} \,\middle|\, \mathcal{H}_t \right) \\
&= \prod_{n=1}^{\lambda} \Pr(X_n^{(t)} \notin \mathcal{A}_t \mid \mathcal{H}_t) \\
&\leq (1 - q_t)^{\lambda},
\end{aligned}$$

其中最后一个不等式由假设 D.3得出。取补可得

$$\Pr(F_{t+1} > F_t \mid \mathcal{H}_t) = 1 - \Pr(F_{t+1} = F_t \mid \mathcal{H}_t) \geq 1 - (1 - q_t)^{\lambda}.$$

最后， 对 $\mathcal{H}$ 取期望并移除条件化，即可得到所述无条件界 $_t$. □

注释 D. 5（小-q_t 区域及 λ 的作用）. 当 q_t 较小时，改进概率满足

$$1 - (1 - q_t)^{\lambda} \approx \lambda q_t,$$

这表明，增加 λ 会近似线性地提高每次迭代中获得更优原则的概率。

D.3.3. 多次迭代： T 步内改进幅度的（宽松）界

为完整起见，我们给出一个标准的并集型下界，用于说明在 T 次迭代的时间范围内至少取得一次严格改进的概率。

推论 D.6 （在 T 次迭代内至少取得一次改进（下界））。 在假设 *D.1* 和 *D.3*下，并假设 $q_t \geq q_{\min}$ 对于 $t = 0, \ldots, T-1$,

$$\Pr(\exists\, t \in \{0, \ldots, T-1\} : F_{t+1} > F_t) \geq 1 - (1 - q_{\min})^{\lambda T}.$$

证明。 对每次迭代进行条件化，并应用定理 D.4：

$$\Pr(F_{t+1} = F_t \mid \mathcal{H}_t) \leq (1 - q_t)^{\lambda} \leq (1 - q_{\min})^{\lambda}.$$

对于未取得改进的概率上界而言，假设各次迭代之间具有条件独立性并非必要：通过迭代条件化，

$$\Pr(\text{no improvements in } T \text{ iterations}) \leq \prod_{t=0}^{T-1} (1 - q_{\min})^{\lambda} = (1 - q_{\min})^{\lambda T}.$$

取补集即可得到结论。 □

上述推论有意采取保守表述；将其纳入是为了说明随 λT 扩展的情况，并与下文的覆盖式论证保持一致。

D.4. 原理空间中的覆盖式泛化

D.4.1. 覆盖模型

我们采用基于覆盖的泛化概念：更大的推理预算会探索更多候选原则，从而提高发现有效原则的概率；这些原则能够在初始种群之外诱发可靠的推理行为。

定义 D.7 （有效原则集合）。 令 $\mathcal{X}_{\text{good}} \subseteq \mathcal{X}$ 表示感兴趣的任务分布下有效原则的（未知）集合（例如，根据 SAP 的离散评估，能够产生稳定、有依据且可靠输出的原则）。

假设 D.8 （逐样本覆盖下界）。 存在 $q_{\text{good}} \in [0, 1]$，使得对于每次迭代 t 和每个后代样本 $X_n^{(t)}$,

$$\Pr\left(X_n^{(t)} \in \mathcal{X}_{\text{good}} \mid \mathcal{H}_t\right) \geq q_{\text{good}}.$$

假设 D.8 指出，提议者保留了以非微小概率采样到有效原则的可能性。这是随机搜索中覆盖保证的标准要求。

D.4.2. 在 T 次迭代内的发现概率

命题 D.9 （覆盖随 λT 增加）。在假设 *D.1* 和 *D.8*下，在 T 次迭代内至少采样到一个有效原则的概率下界为

$$\Pr\left(\exists\, t \in \{0, \ldots, T-1\},\ \exists\, n \in [\lambda] \textit{ s.t. } X_n^{(t)} \in \mathcal{X}_{\text{good}}\right) \geq 1 - (1 - q_{\text{good}})^{\lambda T}.$$

证明。 定义在迭代 t 中未采样到任何有效原则这一事件：

$$\mathcal{E}_t := \bigcap_{n=1}^{\lambda} \{X_n^{(t)} \notin \mathcal{X}_{\text{good}}\}.$$

在 $\mathcal{H}_t$ 条件化，并利用条件独立性（假设 D.1），

$$\Pr(\mathcal{E}_t \mid \mathcal{H}_t) = \prod_{n=1}^{\lambda} \Pr(X_n^{(t)} \notin \mathcal{X}_{\text{good}} \mid \mathcal{H}_t) \leq (1 - q_{\text{good}})^{\lambda},$$

其中，不等式由假设 D.8得出。现在考虑在 T次迭代中均未发现的事件： $\bigcap_{t=0}^{T-1}\mathcal{E}_t$。通过迭代条件化，

$$\Pr\left(\bigcap_{t=0}^{T-1}\mathcal{E}_t\right)=\mathbb{E}\left[\prod_{t=0}^{T-1}\Pr(\mathcal{E}_t\mid\mathcal{H}_t)\right]\le\prod_{t=0}^{T-1}(1-q_{\text{good}})^{\lambda}=(1-q_{\text{good}})^{\lambda T}.$$

取补可得

$$\Pr\left(\bigcup_{t=0}^{T-1}\mathcal{E}_t^c\right)\ge 1-(1-q_{\text{good}})^{\lambda T},$$

这恰好是在 T次迭代内至少采样到一个有效原则的概率 s. □

注 D.10（ λ与 T中的扩展）。 界 $1-(1-q_{\text{good}})^{\lambda T}$关于 λ和 T均单调递增。当 q_{good}较小时，它满足

$$1-(1-q_{\text{good}})^{\lambda T}\approx\lambda T q_{\text{good}},$$

从而明确表明，增大探索宽度（λ）和/或迭代次数（T）可提升原则空间的覆盖。

D.5. 将优化与覆盖关联至推理时扩展

上述结果共同支持 SAP/PER 的推理时扩展解读：精英选择确保优化的稳定性（定理 D.2），而通过增大 λ和/或 T来增加推理预算，既能提高每次迭代实现严格改进的概率（定理 D.4），也能通过覆盖提高发现有效原则的概率（命题 D.9）。这些性质符合我们设定中扩展的操作性定义：分配额外计算资源，以探索并选择不同的推理原则和路径，而非延长单一固定轨迹。

E. 模板设计

系统提示模板设计。 SAP 的一个核心目标是防止推理时推理偏向纯文本启发式方法，并确保在较长的生成时域内，推理始终以视觉证据为依据。为实现这一目标，我们精心设计了一系列用户提示模板，并在 SAP 的所有阶段中共享这些模板，同时施加特定于各阶段的约束。不同的视觉语言模型应对应不同的模板。

动机。 我们发现，长时域视觉语言推理中反复出现两种失效模式。首先，视觉信息通常只在生成开始时提供一次，而文本推理状态（例如早期的视觉摘要）则会自回归地更新。随着生成推进，推理因此会越来越受文本主导，使早期的视觉定位错误随时间不断累积。其次，辅助文本信号——例如落地摘要或评估器反馈——往往存在噪声、不完整或带有启发性；然而，模型很容易过度信任这些信号，将其视为事实证据。这些因素共同导致模型过度依赖中间文本状态，而不是重新查阅原始图像，从而使推理时扩展不稳定。

设计原则。 为了解决这些问题，SAP 在所有用户提示词中严格执行证据层级。明确将原始图像指定为首要且最可靠的信息来源。文本落地摘要被设定为可选参考，可能缺失、不完整或不准确。其作用仅限于提示在图像中应该看哪里，而非应该相信什么。每当文本线索与视觉观察相冲突时，系统都会指示模型重新检查图像；若仍存在歧义，则采取保守回应。这一原则贯穿 SAP 的所有阶段，以遏制文本主导的自回归漂移。

模板概述。 SAP 使用四种用户提示模板，分别对应不同的推理时阶段：(*i*) 原则初始化，(*ii*) 原则演进，(*iii*) 原则引导的多路径推理，以及 (*iv*) 最终答案聚合。尽管作用各不相同，所有模板都会明确强调视觉主导性、不确定性感知，以及在噪声反馈下采取保守决策。

（1）原则初始化模板。原则初始化的目标是在生成具体答案之前，生成一组多样化的高层推理原则。提示词指示模型关注如何根据视觉证据进行推理，而不是编码特定任务的结论。

生成若干用于回答视觉问题的高层推理原则。图像是主要的信息来源。视觉依据摘要是可选的，且可能不完整或不准确。每条原则都应描述一种基于视觉证据进行推理的通用方式，而不是给出具体答案。避免假设无法直接从图像中核实的细节。

该模板鼓励推理策略的多样性，同时确保所有原则都以可观察到的视觉线索为依据。

(2) 原则演化模板。 在演化式优化过程中，原则会根据其观测到的效用进行更新。演化提示词会呈现先前的原则及其相对表现，同时明确避免过度拟合噪声反馈或短期反馈。

优化用于视觉问答的推理原则。图像仍然是最可靠的证据。适应度信号和摘要可能含有噪声，应谨慎使用。保留有效原则的优势，舍弃具有误导性的原则，并提出能够更好地促进依据视觉证据进行核验的新原则。不要假定高分原则总是正确的。

这种设计确保演化更新倾向于采用能够改善视觉定位的原则，而不是利用评估器信号中的偶然相关性。

(3) 原则引导的多路径推理模板。 给定一条固定原则，SAP 要求模型并行生成多条不同的推理路径。每条路径都必须独立地使用视觉证据论证其结论，并在证据薄弱时明确报告不确定性。

根据图像和下方唯一启用的推理原则回答视觉问题。在同一原则下生成多条不同的推理路径。
对于每条路径：- 基于可观察到的视觉证据进行推理。- 尽可能指出视觉线索。- 如果证据薄弱或含糊不清，请说明不确定性。视觉依据摘要仅供参考，也可能有误。如果不确定，请勿猜测。

通过设计，该模板会避免盲目依赖早期的文本摘要，并鼓励在不同推理路径中反复查看图像。

(4) 最终聚合模板。最终聚合阶段会综合不同原则下得出的候选答案。提示词强调，应依据视觉支持谨慎选择，而不是依赖文本一致性或多数投票。

综合多个候选答案，得出最终答案。将图像视为权威证据。候选答案可能彼此矛盾或有误。优先选择得到视觉证据更充分支持的答案。如果没有明确合理的答案，请说明不确定性。不要引入无法从图像中验证的新假设。

隐式模板约束。 除了阶段特定的指令外，SAP 还强制执行若干适用于所有提示模板的隐式约束。这些约束完全在提示词层面实现，无需引入额外监督即可塑造推理时行为。

证据　　层级块。 所有 SAP 提示词都明确将视觉证据置于任何文本信号之上　　ls.

[证据层级]
- 图像是最可靠的证据。- 文本摘要和反馈仅起辅助作用，且可能带有噪声。- 文本绝不能凌驾于视觉观察之上。- 若仍存在不确定性，请谨慎作答。

以定位为参考规则。 落地摘要严格视为弱参考。

[定位参考规则]- 定位摘要可能不完整或不准确。- 仅将其用于指引在图像中查看的位置。- 除非经视觉核实，否则不要假定定位属性成立。- 若出现冲突，以图像为准。

严格输出规则。 T 为支持可靠的评估与比较，SAP 强制执行严格的结构化输出约束 t.

[严格输出规则]- 严格遵循指定的 JSON 模式。 - 不要添加额外字段或自由格式的说明。 - 使用预定义的离散等级表示不确定性和评估结果。 - 如果缺少信息，请返回有效的占位符。

效果。 这些提示模板和共享约束共同构成统一的推理时控制机制，将计算资源重新分配到针对视觉证据的反复验证上。SAP 不引入新的知识、外部监督或特定任务启发式，而是重塑推理过程中现有模型容量的使用方式，从而在离散生成和嘈杂的多模态反馈下实现稳定且有效的推理时扩展。