arXiv:2602.01004v1 [cs.CV] 1 Feb 2026

SRVAU-R1：通过反思感知学习增强视频异常理解

Zihao Zhao [1] Shengting Cao [2] Muchao Ye [1]

摘要

多模态大语言模型（MLLMs）的推理能力取得了显著进展，并在视频异常理解（VAU）任务中展现出良好的应用前景。然而，现有基于多模态大语言模型的方法仍主要聚焦于对异常的表层描述，缺乏对异常行为的深入推理，例如明确的自我反思和自我纠正。为此，我们提出了用于视频异常理解的反思增强推理方法 **S**elf-**R**eflection-Enhanced Reasoning for**V**ideo**A**nomaly**U**nderstanding（SRVAU-R1），这是一种将反思融入多模态大语言模型推理的反思感知学习框架。具体而言，SRVAU-R1 首次引入了专为视频异常理解设计的反思导向思维链数据集，通过初始推理、自我反思和修订后的推理提供结构化监督。在此基础上，该方法提出了一种新颖的反思感知学习范式，结合监督微调与强化微调，以增强视频异常理解中的多模态推理能力。在多个视频异常基准上进行的大量实验表明，SRVAU-R1 始终优于现有方法，在时间异常定位准确率和推理质量方面均取得显著提升。[1]

1. 引言

从复杂、非结构化的视频数据中准确识别并深入理解交通违规、盗窃或突发危险事件等异常，仍是智能感知系统在公共安全、交通管理以及

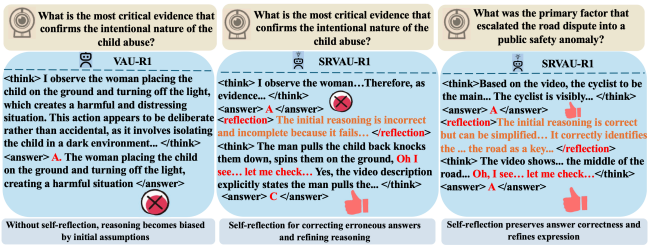


图*1.* 现有的视频异常理解方法（如 VAU-R1）仅提供表层分析，且缺乏对生成结果的反思。本文提出一种名为 SRVAU-R1 的新方法，在视频异常理解中引入显式自我反思，以纠正或完善初始推理。

城市治理（Ma 等人，2025年；Tang 等人，2024年；Liu 与 Ma，2019年）等应用中的核心挑战。传统方法主要将其形式化为视频异常检测任务，定位帧级或片段级异常时间片段（Chen 等人，2023年；Ding 等人，2024年；Gong 等人，2019年；Leng 等人，2024年；Zhou 等人，2023年；Tian 等人，2021年）。自引入多模态大语言模型（多模态大语言模型）（Ye 等人，2025年；Doshi 与 Yilmaz，2023年；Zanella 等人，2024年）以来，该任务已转变为视频异常理解（视频异常理解）（Zhu 等人，2025）任务，旨在为异常事件生成准确的自然语言描述。

局限性。 然而，现有基于多模态大语言模型的视频异常理解方法仍局限于对异常的表层描述（Du 等人，2024年；Sultani 等人，2018年；Yang 等人，2024年）。其突出局限在于缺乏对异常事件进行反思的能力，而这种能力对于视频异常理解至关重要，有助于处理异常的模糊性。例如，多模态大语言模型可能会将一段视频描述为“某人进入禁区”，但要判断这一事件是否确实异常，还需要结合时间和上下文线索进行反思，例如此人是否是维修工、是否正在应对紧急情况，或是否在恶意行事。

这一假设是合理的，因为近期关于多模态推理的研究表明，显式的自我反思和自我纠正机制对于提升模型推理的稳定性、一致性和深度至关重要（Madaan 等人，2023年；Kumar 等人，2024年；Gandhi 等人，2025年）。通过反思中间推理过程，

[1]美国爱荷华州爱荷华城爱荷华大学计算机科学系 [2]美国伊利诺伊州盖尔斯堡诺克斯学院计算机科学系。通讯作者：Muchao Ye<muchao-ye@uiowa.edu>。

预印本。*2026*年*2*月*3*日。[1]数据和代码将公开发布。

模型能够识别并修正错误或冗余步骤，从而形成因果可解释性更强、更连贯的推理链（Zelazo，2015）。因此，在进行视频异常理解时，必须使多模态大语言模型能够反思并再三思考，从而正确理解异常的成因、时间演变和上下文语义。因此，如何在视频异常理解的多模态推理中有效融入自我反思机制，仍是一个尚待解决且颇具挑战性的问题。

我们的方法：SRVAU-R1。由于缺乏高质量基准，以及缺乏支持多模态大语言模型进行视频异常理解细粒度时序推理的定制化学习范式，这一研究问题颇具挑战性。为应对这些挑战并引入**自**我**反**思增强的推理能力，应用于**视频异常理解**，我们提出了一种称为SRVAU-R1 的反思感知学习框架。具体而言，SRVAU-R1 首先包含一个反思感知数据构建流程，通过反思完善初始推理，并以修订后的推理扩充训练数据，如图1所示。基于此，SRVAU-R1 采用两阶段训练策略：阶段1在反思增强的训练数据上进行监督微调（SFT），使模型在学习异常推理的同时获得自我评估和自我纠正能力；阶段2则采用基于组相对策略优化（GRPO）（Guo 等人，2025年）的强化微调（RFT），引导模型生成简洁、连贯且具有认知意义的反思推理过程。具体而言，我们设计了一个复合奖励函数，联合优化推理正确性、反思质量和时间一致性。

我们的主要贡献可以总结为如下：

- **面向反思的视频异常理解数据构建流程**。在SRVAU-R1 中，我们提出了一种面向视频异常理解的面向反思的数据构建流程，并构建了首个专为反思增强型视频异常理解定制的思维链训练数据集。该数据集将每个样本明确结构化为初始推理、自我反思和修订后的推理，为在监督微调和强化学习中学习推理修正与自我评估提供了关键的监督信号。
- **面向视频异常理解中多模态大语言模型的反思感知学习框架**。SRVAU-R1 为视频异常理解中的多模态大语言模型引入了一种两阶段训练框架，将反思增强的监督微调与反思感知强化学习相结合。该框架使多模态大语言模型不仅能够提高异常推理准确率，还能在推理过程中诊断并纠正自身的推理错误，从而实现更稳健的视频异常理解。
- **全面的实证评估与分析**。我们在多个视频异常理解任务和数据集上进行了广泛的实验与消融研究。结果表明，SRVAU-R1在异常理解性能、时序定位准确率和推理稳定性方面始终优于现有方法，明确验证了显式自我反思机制对于复杂视频异常推理的有效性与必要性。

2. 相关工作

视频异常任务中的推理范式。随着多模态大语言模型的推出，视频异常的研究范式已逐渐从以检测为导向的表述（Yao 等人，2023年；Huang 等人，2024年；Liu 等人，2021年；Wu 等人，2020年；Zhou 等人，2023年；Joo 等人，2023年；Liu 等人，2018年；Lu 等人，2020年；Ye 等人，2019年）转向视频异常理解（Lv 和 Sun，2024年；Zhu 等人，2025）。现有的视频异常理解方法主要沿两条发展路线推进：（1）提示驱动方法（Yang 等人，2024），通过少样本提示或基于规则的模板引导多模态大语言模型分析异常，训练成本较低，但泛化能力有限；（2）学习驱动方法，通过微调（Zhang 等人，2024年）将异常语义和因果知识内化到模型参数中。与先前基于强化学习的视频异常理解方法（如 VAU-R1（Zhu 等人，2025）和 VAD-R1（Huang 等人，2025a））相比，SRVAU-R1 提供了一种基于监督微调和强化微调的新型学习流程，借助面向反思的数据构建流程，使视频异常理解中的多模态大语言模型能够进行反思感知推理。在 SRVAU-R1 下，多模态大语言模型将展现出带有反思的高级推理，并具有更高的推理质量。

用于多模态大语言模型推理的强化学习。近年来，GPT-o3、Gem- ini 2.5、Seed1.5-VL 和 DeepSeek-R1 等高能力多模态模型的出现，使强化学习成为多模态模型后训练阶段提升推理能力的关键技术（Bi 等人，2025年；Feng 等人，2025年；Huang 等人，2025b；Wang 等人，2024年；Zhou 等人，2025年；Guo 等人，2025年）。具体而言，GRPO 及其后续变体（Yue 等人，2025b；Yu 等人，2025年；Dou 等人，2025年；Liu 等人，2025a）凭借优化效率和推理稳定性，已成为训练大语言模型的主流强化学习方法。然而，基于强化学习的方法通常需要大量计算资源和大规模训练数据。为降低适配成本，一些工作将监督微调与强化学习相结合，利用思维链冷启动初始化（例如 Reason-RFT（Tan 等人，2025年）和 Vision-R1（Huang 等人，2025b）），或采用数据效率更高的视觉强化微调范式（例如 Visual-RFT（Liu 等人，2025b）），并进一步将这些策略扩展至视频异常理解方法（Zhu 等人，2025年；Huang 等人，2025a）。总体而言，我们的工作与 SRPO（Wan等人，2025年）秉持相同的愿景：需要在监督微调和强化学习阶段系统地建模自我反思与错误纠正。与 SRPO 相比，SRVAU-R1 包含

量身定制 了突出时间定位的奖励设计 on.

3. SRVAU-R1框架

本节将介绍 SRVAU-R1，这是一种面向 VAU 的新型两阶段端到端训练框架，其核心是将自我反思融入多模态推理，如图 2 所示。在第 3.1 节中，我们首先介绍反思感知数据构建流程，用于构建面向反思的数据集，使模型具备推理过程中进行自我反思的基本能力。在第 3.2 节中，我们将介绍 SRVAU-R1 的学习范式。

3.1. 反思感知数据构建

在 SRVAU-R1 中，我们首先需要准备反思感知数据来训练多模态大语言模型，以促使其在推理过程中进行反思。这一点至关重要，因为多模态大语言模型的预训练数据并未明确对此类行为进行建模。此类数据集的构建如图 2(a) 所示。

在该流程中，我们首先采用思维链（CoT）提示，从基础多模态模型（例如 `Qwen2.5-VL-3B`）中获取初始推理输出。随后，将初始推理结果与真实标注一同输入具有更强能力的多模态教师模型（例如 `Qwen3-VL-30B`），该模型通过两条互补路径生成显式反思：纠正错误的推理，以及细化正确但冗余的推理。这种双路径反思机制使模型能够同时学习错误诊断和推理优化。所使用的提示模板见附录中的第 D 节，部分生成样例见附录中的第E节。

形式化地，给定一个视频异常理解数据集 $D = \{(v,\ q)\}$，以及一个基于视频 v 的问题 q，基础模型 f 首先生成**初始推理**输出 a_1。随后，高容量多模态模型 f_{teach} 生成显式的**自我反思**信号 r，用于识别潜在的推理错误，例如证据误解、时间错位或缺失的因果联系。 f_{teach} 将提供修订后的推理 a_2。通过这一过程，我们在学习中为视频异常理解显式引入推理–反思–修订结构化监督。

3.2. 面向反思的学习范式

在 SRVAU-R1 的学习范式中，如第 3.2.1 节所述，首先利用 SFT，通过融入外部知识蒸馏，有效拓展多模态大语言模型的推理边界并增强其认知能力。随后，在第 3.2.2 节中，我们介绍了一种基于 GRPO 的反思感知 RFT 策略。通过引入精心设计的反思感知奖励机制，持续鼓励模型提升在强化学习过程中，其自我反思的有效性以及推理的时序一致性。这两个阶段共同促使 SRVAU-R1 形成稳健、可解释且以时间为依据的视频异常理解推理行为。

3.2.1. 使用 SFT 进行冷启动初始化

为了实现稳定且有效的反思感知视频异常理解，我们首先对 SRVAU-R1 策略模型执行面向反思的冷启动初始化。该阶段旨在强化学习之前向模型注入基本的自我反思与自我纠正能力，以防止仅使用强化学习所导致的潜在不稳定反思行为（Yue 等人，2025a）。

具体而言，给定高质量的反思增强训练数据集 $D_{\text{reflect}} = \{(v,\ q,\ a_1,\ r,\ a_2)\}$（来自第 3.1 节），我们提出一种面向反思的 SFT 策略，将反思知识显式注入策略模型（Shah 等人，2025年）。在该反思的条件下，策略模型 f 被优化以生成修订后的推理轨迹和最终答案 a_2。作为一项 SFT 任务，冷启动阶段的训练目标是常用的负对数似然：

$$\mathcal{L}_{\text{cold-start}} = -\mathbb{E}_{(q,a_1,r,a_2)\sim\mathcal{D}}\left[\sum_{t=1}^{T}\log \pi_{\text{initial}}\Big(a_2 \mid q, a_1, \langle\text{reflection}\rangle r \langle/\text{reflection}\rangle\Big)\right], \tag{1}$$

其中， r 表示显式反思内容， a_2 表示反思修正后的推理过程及其对应的标准答案。 π_{initial} 是 f的策略（模型参数）。

这种面向反思的冷启动初始化有两个互补的目的：（1）通过在反思指导下显式学习修订初始推理，策略模型逐渐具备识别推理错误、时间不一致和因果结构缺失的能力，而这些能力对于稳健的视频异常理解至关重要。（2）经反思增强的冷启动为模型提供了结构化的推理先验，从而有效避免模型在后续强化微调过程中产生不稳定或肤浅的反思行为。因此，反思逐渐内化为模型决策过程的一部分，而非仅仅发挥事后纠正的作用。通过这一初始化，SRVAU-R1 从通用多模态理解平稳过渡到反思感知的视频异常推理，为后续的反思感知的 RFT 阶段奠定了坚实基础。

3.2.2. 反思感知的 RFT

背景。 在强化微调阶段，我们采用GRPO作为核心优化框架。具体而言，给定输入查询q，旧策略 π_{old} 会采样一组包含 G 个候选回复-

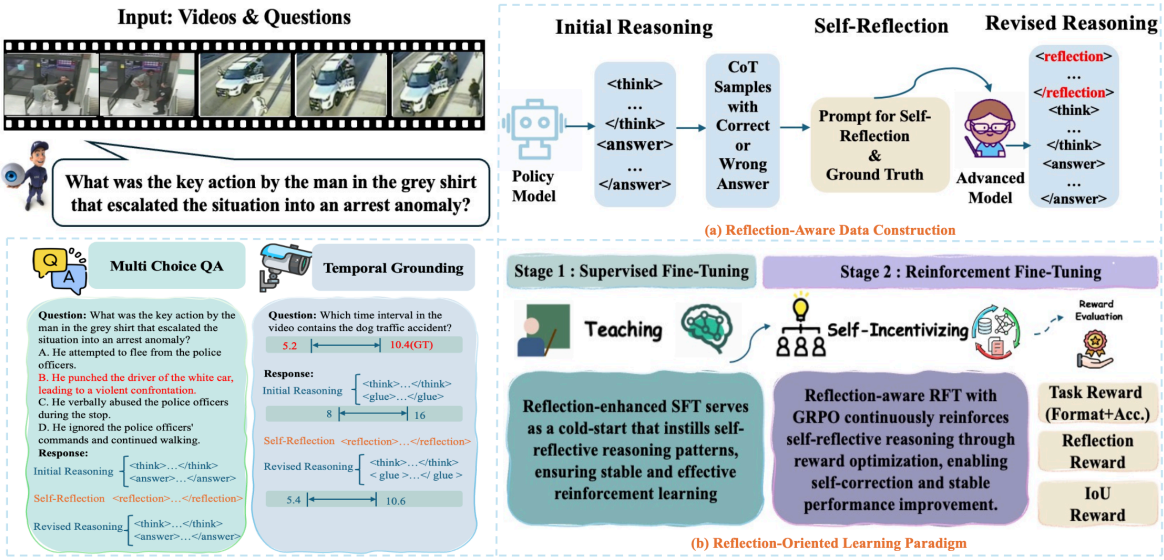


图 2. SRVAU-R1框架概览。首先，SRVAU-R1包含（a）反思感知数据构建流程：在先进模型的指导下生成初始思维链候选项和明确的自我反思标注。SRVAU-R1可灵活应用于视频异常理解任务，包括多项选择问答和时序异常定位，这些任务需要结构化推理与反思。最后，SRVAU-R1还包含一种新颖的（b）学习范式，由反思增强监督微调和强化微调组成。在强化学习过程中，通过任务奖励、反思奖励和时间交并比奖励优化模型参数。

复 $\{o_1, o_2, \ldots, o_G\}$。对于每个候选回复 o_i，根据预定义的奖励函数计算任务特定奖励 r_i。随后，GRPO在组内对这些奖励进行归一化，以计算每个候选回复的相对优势： $A_i = \frac{r_i - \text{均值}(\{r_1, r_2, \ldots, r_G\})}{\text{标准差}(\{r_1, r_2, \ldots, r_G\})}$。基于估计的优势，GRPO优化目标定义为：$\mathcal{L}_{\text{GRPO}}(\theta) = \mathbb{E}_{q,\ \{o_i\}\sim\pi_{\text{old}}} \left[\frac{1}{G} \sum_{i=1}^{G} \min\left(\mathcal{J}_i, \widehat{\mathcal{J}}_i\right) - \beta \mathbb{D}_i \right]$，其中 $\mathcal{J}_i = \rho_i(\theta) A_i$, $\mathcal{J}_i = \text{clip}(\rho_i(\theta)\ 1-\epsilon,\ 1+\epsilon) A_i$，以及 $\mathbb{D}_i = D_{\text{KL}}(\pi_\theta \| \pi_{\text{ref}})$，其中 ϵ 为用于稳定策略更新的裁剪阈值，β 为KL散度正则化项的系数，促使学习到的策略保持接近冻结的参考策略 π_{ref}。

SRVAU-R1 中的整体奖励设计。我们的强化学习阶段基于 SRPO（Wan等人，2025年）框架。由于 GRPO 中原有的奖励设计主要关注答案正确性和格式遵循，缺乏反思等关键的复杂多模态推理能力，SRPO 通过加入反思感知奖励改进了 GRPO。然而，SRPO 的奖励机制并不适用于视频理解任务，仅针对静态输入任务。对于视频异常理解（VAU）而言，这一局限尤为明显，因为模型必须基于模糊的视觉证据和不断演变的因果关系，在较长的时间跨度上进行推理。为解决这一局限，我们在 SRPO 奖励机制中加入了时间交并比奖励项。在这一奖励机制下，这项改进既优化最终输出的正确性和自我评估能力，也提升了对时间输入的理解能力贯穿整个推理过程。对于每个候选回答，总奖励定义为

$$R_{\text{total}} = \alpha R_{\text{task}} + \beta R_{\text{reflection}} + \gamma R_{\text{tIoU}}, \tag{2}$$

其中，R_{task} 表示任务级奖励，$R_{\text{reflection}}$ 衡量自我反思的质量和有效性，R_{tIoU} 评估时序定位准确率。系数 α、β和 γ 用于平衡各组成部分的贡献。

任务奖励。任务奖励用于确保模型输出的基本正确性和结构一致性。它由格式奖励和准确性奖励组成：

$$R_{\text{task}} = R_{\text{format}} + R_{\text{accuracy}}. \tag{3}$$

格式奖励 R_{format} 鼓励模型遵循标准化推理结构（例如，使用⟨think⟩ ⋯ ⟨/think⟩ 标签），从而确保生成的推理过程具有一致性和可解析性。如果输出格式正确，我们将 $R_{\text{format}} = 0.5$；否则将$R_{\text{format}} = 0$。准确性奖励 R_{accuracy} 评估模型的初始推理输出是否与真实标签相匹配。如果第一次预测正确，我们将 $R_{\text{accuracy}} = 0.5$；否则将 $R_{\text{accuracy}} = 0$。这一设计确保模型始终与任务目标保持一致，并避免仅依赖事后反思。

反思奖励。反思奖励明确评估模型通过自我评估与修正来改进其推理的有效程度。其定义为：

$$R_{\text{reflection}} = I_{\text{eff}} + I_{\text{ref}} + \alpha f_{\text{len}}(L_{\text{response}}), \tag{4}$$

其中 $I_{\text{ref}} = 0.25$ 表示反思片段使用 <reflect> 标签的格式正确，$I_{\text{ref}} = 0$ 表示格式不正确；$f_{\text{len}}(L_{\text{response}})$是反思简洁性正则项，用于促进简洁而信息丰富的反思：$f_{\text{len}}(L_{\text{response}}) = \exp\left(-\frac{|L_{\text{response}} - T_{\text{target}}|}{T_{\max} - T_{\text{target}}}\right)$，该项对过于冗长的反思施加软惩罚并稳定梯度行为；以及

$$I_{\text{eff}} = \begin{cases} +0.25, & \text{C} \to \text{C}, \\ +0.5, & \text{W} \to \text{C}, \\ 0, & \text{W} \to \text{W}, \\ -0.25, & \text{C} \to \text{W}, \end{cases} \tag{5}$$

该指标衡量反思对改善推理结果的实际有效性。其中，C 和 W 分别表示应用反思前后预测是否正确（Correct）或错误（Wrong）。这一公式明确区分了有效反思与表面化或有害反思，避免模型仅通过冗长表达或形式上的结构合规来利用反思机制。

时间交并比奖励。为了为时间定位提供细粒度监督，我们引入了一种基于交并比（Intersection-over-Union, IoU）的时间重叠奖励。该奖励并非作为独立的定位目标，而是作为一种辅助信号，使时间推理结果与真实异常时间段保持一致。具体而言，时间奖励定义为

$$R_{\text{tIoU}} = \begin{cases} 1, & \text{normal and correct}, \\ \text{IoU}\big([s_1, s_2], [s_1^*, s_2^*]\big), & \text{anomaly and correct}, \\ 0, & \text{otherwise}. \end{cases} \tag{6}$$

3.3. SRVAU-R1 中的顿悟时刻

我们观察到，SRVAU-R1 在视频异常推理过程中展现出明确且一致的“顿悟时刻”（Guo 等人，2025年）（我们将在第 4.4 节通过定性结果加以说明）。面对时间线索模糊、异常边界不清或涉及多步因果推断的复杂场景时，模型的初始推理往往基于不完整或局部证据形成。随后，在反思信号的触发下，模型会主动重新审视视频上下文，并重新评估先前的结论，从而修正推理轨迹（Pan 等人，2025年）。这一行为表明，模型能够进行自我反思：识别初始推理中的证据遗漏或因果不一致，并在第二阶段推理过程中迅速修正判断，重新定位异常的关键时间窗口，或修正核心因果假设。

4. 实验

4.1. 实验设置

训练数据。基于 VAU-Bench（Zhu 等人，2025年），我们从三个公开视频异常数据集中选取样本作为原始数据来源，包括 UCF-Crime、ECVA 和 MSAD，并按照第 3.1 节所述的数据构建流程，对这些数据进行系统筛选和重组。该过程生成了一个整合反思增强标注的训练数据集，包含 4,602 个视频样本，涵盖 19 个主要异常类别，包括暴力行为、事故、危险活动，以及现实场景中常见的突发异常事件。生成的数据集在视频时长、异常模式和语义复杂度方面具有显著多样性，为反思感知的异常推理提供了有力支持。关于数据集构建和标注过程的更多详情，见附录中的第 A 节。

指标与基线。我们在 VAU-Bench 上评估我们的模型（Zhu 等人，2025年）。遵循其设置，对于问答（QA）任务，我们采用多项选择准确率作为评估指标；对于时间异常定位，我们报告时间平均交并比（mIoU）以及不同交并比阈值下的召回率（R@0.3、R@0.5 和 R@0.7），以全面评估定位性能；对于生成的推理和解释性输出，我们使用 VAU-Eval（Zhu 等人，2025年），并采用 DeepSeek-V3（Liu 等人，2024年）作为评判模型，从五个维度评估每个回答：分类正确性、关键概念对齐、流畅度、信息丰富度和事实一致性。每个维度均按 10 分制评分，从而实现对推理质量的细粒度评估。遵循（Zhu 等人，2025年），我们使用来自 Qwen-VL 和 InternVL 系列的冻结模型作为基线。我们还将我们的结果与 VAU-R1 和 HolmesVAU（Zhang 等人，2025年）进行比较，以评估我们学习范式的有效性。

实现设置。我们的主要实验采用 Qwen2.5-VL-3B-Instruct（Bai 等人，2025年）模型作为基础模型 f，并采用 Qwen3-VL-30B-Instruct（Yang 等人，2025年）作为教师模型 f_{teach}。初始推理、自我反思和修订后的推理所用的提示模板见附录第D 节。自我反思冷启动监督微调阶段和后续的强化学习阶段均使用 4 块 NVIDIA RTX PRO 6000 GPU 进行训练。所有实验均采用全参数微调，不使用参数高效适配器或 LoRA 模块。推理时，对于基于 Qwen-VL 的模型，以帧每秒 1 帧的速率采样视频帧。对于作为教师模型使用的基于 InternVL 的模型，则均匀采样每个视频的 16 帧作为输入。统一的推理协议确保了不同模型系列之间比较的公平性。

表 *1.* MSAD 和 UCF-Crime 数据集上的问答准确率与 VAU-Eval 结果对比。受篇幅限制，ECVA 上的对比结果见附录第 C 节。

数据集	模型	问答准确率		VAU-Eval					
		准确性$_{\text{w/o 思考}}$	准确性$_{\text{w/ 思考}}$	分类↑	知识记忆↑	FLU↑	INF↑	事实准确性↑	总计↑
MSAD	InternVL2.5-2B	76.67	72.08	6.84	6.23	8.55	6.64	6.64	34.90
	Qwen2.5-VL-7B	84.58	83.33	6.75	6.41	9.27	7.74	6.92	37.08
	InternVL2.5-8B-MPO	82.50	84.17	6.83	6.33	8.32	6.37	6.86	34.72
	Qwen2-VL-2B	77.08	72.50	5.94	5.43	8.77	6.29	5.90	32.25
	Qwen2.5-VL-3B	85.83	82.50	5.77	5.24	9.02	6.74	5.70	32.47
	HolmesVAU	85.00	86.25	3.73	2.72	6.82	3.55	3.33	20.15
	VAU-R1	88.33	87.08	5.97	5.49	9.05	6.84	6.03	33.38
	SRVAU-R1	**89.58** (↑ 3.75)	**91.25** (↑ 8.75)	7.65	6.98	9.78	7.08	7.43	**38.30**
UCF-Crime	InternVL2.5-2B	84.86	68.13	4.40	3.08	8.09	5.69	3.47	24.74
	Qwen2.5-VL-7B	92.03	89.64	4.80	3.73	8.95	7.05	4.25	28.78
	InternVL2.5-8B-MPO	89.64	90.44	3.79	3.20	8.23	5.77	3.48	24.47
	Qwen2-VL-2B	87.25	83.67	3.47	2.48	7.75	4.49	2.82	21.02
	Qwen2.5-VL-3B	91.63	83.27	4.31	2.88	8.70	5.95	3.27	25.10
	HolmesVAU	86.45	85.66	3.05	1.97	6.30	3.08	2.39	16.79
	VAU-R1	92.03	91.63	4.42	2.98	8.71	5.89	3.39	25.49
	SRVAU-R1	**92.82** (↑ 1.19)	**96.81** (↑ 13.54)	7.22	6.76	9.16	6.56	6.96	**36.66**

4.2. 对比结果

问答引导推理评估。 如表1所示，实验结果表明，SRVAU-R1 在所有评估数据集上均持续且显著地优于开源通用多模态基线模型。在 MSAD 和 UCF-Crime 上，SRVAU-R1 的多项选择问答准确率以及 VAU-Eval 五个维度的汇总得分（Total）均达到最佳，明显优于广泛使用的多模态大语言模型基线和 VAU 方法。这一发现表明，观察到的性能提升并非源于模型容量增加或单纯的监督微调，而是得益于对反思感知训练机制的有效整合。

此外，我们观察到，与关闭推理时（即 $\text{Acc}_{\text{w/o think}}$）相比，启用显式推理生成时（即 $\text{Acc}_{\text{w/think}}$），基线模型的表现往往更差。这一现象表明，朴素的思维链提示可能会引入推理幻觉，进而降低最终答案的准确性。相比之下，SRVAU-R1 在 SFT 和 RFT 阶段均对自我反思进行显式建模，从而有效缓解了这一问题。总体而言，这些结果有力地支持了我们的核心主张：在监督微调和强化微调过程中显式增强模型的自我反思能力，能够显著提升视频异常理解中的高层语义推理、决策准确率和推理稳定性，使模型从单纯的异常识别转向更深入的异常理解。

时序异常定位评估。除了基于问答的异常推理之外，我们还进一步评估

表*2*。MSAD 和 UCF-Crime 数据集上的时序异常定位性能对比结果。

数据集	模型	mIoU	R@0.3	R@0.5	R@0.7
MSAD	Qwen2-VL-2B	0.00	0.00	0.00	0.00
	Qwen2.5-VL-7B	17.57	26.67	11.67	3.33
	Qwen2.5-VL-3B	11.05	15.42	5.35	1.67
	SRVAU-R1	**20.40** (↑ 9.35)	30.00	19.17	8.33
ECVA	Qwen2-VL-2B	0.17	0.30	0.00	0.00
	Qwen2.5-VL-7B	5.71	7.96	4.73	2.99
	Qwen2.5-VL-3B	6.35	7.21	1.99	0.50
	SRVAU-R1	**44.42** (↑ 38.07)	63.68	42.29	24.13

模型在时序异常定位（TAG）上的表现，以评估其在分布外（OOD）设置下的时序泛化能力。值得注意的是，在这一实验设置中，模型仅在 ECVA 数据集上训练，而 MSAD 则完全保留为 OOD 测试集，从而避免模型直接拟合测试数据分布。如表 2所示，SRVAU-R1 在 MSAD 和 ECVA 上的表现均持续优于所有开源通用多模态基线，在不同交并比阈值下的平均交并比和召回率均取得最佳结果。尤其是在 OOD 的 MSAD 数据集上，SRVAU-R1 的平均交并比达到 20.40，比 Qwen2.5-VL-3B 高出 9.35 个点，展现出在未见数据分布下强大的时序定位能力。在 ECVA 上，SRVAU-R1 的优势更加显著，其平均交并比达到 44.42，超过最强基线 38.07 个点，并且在 R@0.7 等更严格的定位标准下仍保持明显领先。这些结果表明，反思增强推理不仅有助于精确建模语义异常

表3. 反思数据对视频异常理解的影响。

方法	UCF-Crime	MSAD
Qwen2.5-VL-3B	83.27	82.50
+GRPO（无反思数据）	91.63	87.08
SRVAU-R1（本文）	**96.81**	**91.25**

表4. 分离反思感知监督微调与反思感知强化学习的影响。

方法	UCF-Crime	MSAD
Qwen2.5-VL-3B	83.27	82.50
仅反思监督微调	96.02	90.83
仅使用反思强化学习	93.23	88.48
SRVAU-R1（我们的方法）	**96.81**	**91.25**

表5。 反思数据规模对 SRVAU-R1 效果的影响。

方法	强化学习数据规模	UCF-Crime	MSAD
GRPO	1.5K	89.24	85.41
GRPO	3k	91.63	87.08
SRVAU-R1	1.5k	94.42	87.50
SRVAU-R1（我们的方法）	3k	**96.81**	**91.25**

表6. 反思范式（两步思考与思考–反思–再思考）及教师模型对反思感知推理的影响。

方法	UCF-Crime	MSAD
两步思考	89.64	85.41
SRVAU-R1（InternVL3-38B 教师模型）	95.62	88.33
SRVAU-R1（我们的方法）	**96.81**	**91.25**

边界，还能提升异常定位中的全局时序一致性。综合来看，这些发现表明，在强化微调过程中引入反思感知机制，显著增强了模型在未见数据分布下进行时序异常定位的泛化能力。与仅依赖监督微调或直接推理的基线相比，反思增强推理赋予模型在时序决策过程中自我纠正的能力，从而使其在复杂的跨分布视频异常理解场景中更加稳定且鲁棒。

4.3. 消融研究

我们进一步探究了 SRVAU-R1 中反思增强数据构建、训练阶段设计及反思推理范式本身对推理质量的影响。为保证普适性，所有消融实验均在默认设置下进行，使用 Qwen2.5-VL- 3B 作为 f 问答数据集中的模型。**反思数据的影响（无反思数据 vs. SRVAU-R1）**。如表 3所示，我们首先移除反思增强数据构建过程，仅使用原始异常标注和标准思维链推理来训练模型。结果表明，缺少明确的反思监督会导致时序定位准确率、推理连贯性和预测稳定性显著下降。这一发现说明，仅凭传统推理信号不足以应对复杂的视频异常理解。反思增强数据在引导模型进行自我纠正和因果推理方面发挥着关键作用，使模型能够修正错误的推理路径，并使预测更好地契合高层次异常语义。

分离反思 SFT 与反思 RL 的作用。如表4所示，为了区分反思机制在 SFT 和强化微调中的各自贡献，我们评估了以下两种消融设置：（1）在 SFT 期间使用反思增强数据，但不采用反思感知强化学习；（2）在强化学习期间引入反思感知奖励，但不使用反思增强 SFT 初始化。结果表明，仅使用反思增强 SFT 在一定程度上能够提升推理质量，但仍不足以应对复杂的时间决策和分布外（OOD）场景。相反，仅在强化学习期间引入反思感知奖励，无法可靠地促使模型形成稳定的反思行为，导致训练不稳定且收敛速度较慢。相比之下，SRVAU-R1 受益于反思增强监督微调阶段提供的结构良好的冷启动初始化，随后通过反思感知强化学习持续强化反思行为。这种两阶段协同效应的表现始终优于两种单阶段方案，表明有效反思学习关键取决于 SFT 与强化学习之间的协同。

反思数据规模的影响（部分反思数据集与 SRVAU-R1 的比较）。如表 5所示，我们通过使用从反思数据集中随机抽取的子集训练 SRVAU-R1，进一步考察了反思数据规模的影响。随着反思数据量减少，模型性能逐渐下降。尽管如此，即使在反思数据大幅减少的情况下，SRVAU-R1 仍优于完全不具备反思机制的基线。这一结果表明，反思增强训练具有较高的数据效率，同时也凸显了充足的反思数据对于充分发挥其潜力至关重要。

来自其他教师模型的反思数据（替代教师模型与 SRVAU-R1）。如表 6所示，为验证观察到的提升并非源于特定教师模型，我们使用

不同的多模态大模型生成反思数据，并据此重新训练 SRVAU-R1。具体而言，我们使用 InternVL3-38B f_{teach}来构建反思感知数据 D_{reflect}，用于 SRVAU-R1 的监督微调和强化微调。尽管不同教师模型的语言风格和描述粒度存在差异，SRVAU-R1 仍表现出稳定的性能提升。这表明，性能提升主要源自反思结构本身，而非与特定教师模型相关的蒸馏效应，进一步证明了所提方法的鲁棒性和通用性。

反思范式的有效性（两步思考与思考–反思–再思考）。最后，为了验证所提出的“思考–反思–再思考”推理范式的必要性，我们将 SRVAU-R1 与一种基于 GRPO 的两步思考基线进行比较。在这一替代设置中，模型连续生成两个 <思考> 和</思考> 推理步骤，并使用任务奖励和基于一致性的关系奖励进行训练，而不引入显式的反思标记。表 6 中的结果表明，与普通 GRPO 相比，两步思考未能带来稳定的提升，仅在少数数据集上取得了微小增益。相比之下，SRVAU-R1 通过显式反思并纠正先前的推理，显著提升了异常推理准确率和时间定位性能。这一比较清楚地表明，单纯增加推理步骤的数量无法替代显式的自我反思机制。反思在引导模型识别并修正有缺陷的推理轨迹方面发挥着不可或缺的作用，从而使视频异常理解更加可靠且易于解释。

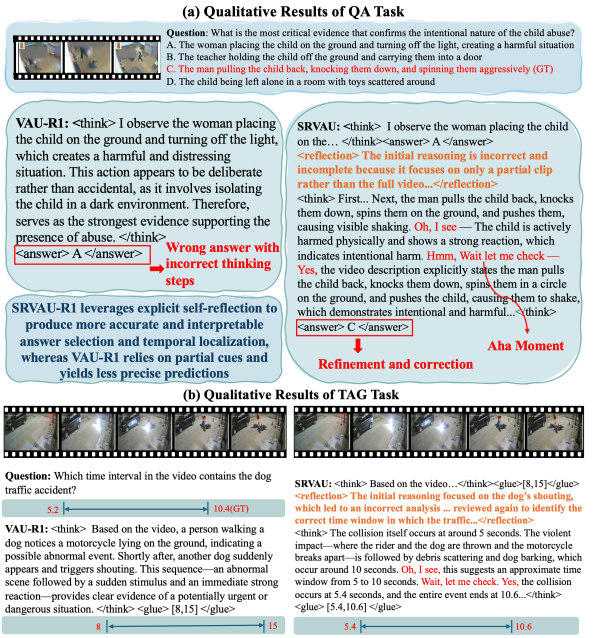


图*3.* **(a) 问答**和**(b) 时间异常定位**任务的定性结果。所有真实答案和异常区间均以红色突出显示。VAU-R1 和 SRVAU-R1 在相同的思维链提示下进行推理。SRVAU-R1 利用显式自我反思，生成更准确且更易解释的答案选择和时间定位；而 VAU-R1 依赖不完整的线索，因此预测不够精准。

4.4. 定性结果与案例研究

我们展示了两个不同视频异常理解任务的定性案例研究，并在相同的思维链提示下比较 VAU-R1 与我们的 SRVAU-R1。在图 3(a) 所示的问答任务中，VAU-R1 在初始推理中过度依赖局部视觉线索，因此对观察到的行为作出了良性的解读。相比之下，SRVAU-R1 在反思阶段明确意识到其初始推理并不完整。通过重新审视完整的视频上下文，它正确地将男子的攻击行为识别为故意虐待的决定性证据。这一修正伴随着一个清晰的“顿悟时刻”：模型放弃了先前的假设，并从根本上修正了推理轨迹。在图 3(b) 所示的时间异常定位任务中，VAU-R1 给出的异常区间较为粗略，时间定位不准确，且缺乏清晰的因果解释。相比之下，SRVAU-R1 在反思过程中意识到，其初始分析过分强调了非关键线索。随后，它重新审视视频的时间结构，并将异常区间准确定位在 5.4s–10.6s。这两个相互独立的案例研究共同表明 SRVAU-R1 通过明确的自我反思，使模型能够自我纠正有缺陷的初始推理，从而在语义理解和时间定位方面实现更准确、稳定且可解释的推理表现。

5. 结论

本文提出了 SRVAU-R1，这是一种面向视频异常理解中多模态大语言模型的反思感知数据构建与学习框架。通过这一流程，我们能够构建高质量的反思增强数据集，以初始推理、显式自我反思和修订后的推理形式提供结构化监督。在此基础上，通过结合反思增强的监督微调与反思感知强化学习，SRVAU-R1 不仅提高了推理准确性，还使模型能够在推理过程中反思并纠正自身的推理错误，从而形成更加鲁棒的推理行为。在多个视频异常基准上开展的大量实验表明，SRVAU-R1 能够持续提升异常理解性能和推理稳定性，验证了反思感知训练策略对于复杂视频推理任务的有效性。

影响声明

本研究提出了 SRVAU-R1，这是一种面向视频异常理解的反思感知学习框架，旨在提升多模态大语言模型在复杂视频场景中的推理稳定性。具备更可靠异常理解能力的模型有望支持公共监控、异常行为预警和事件分析等安全相关应用，从而辅助人类决策过程。同时，视频异常场景可能涉及暴力、危险或混乱行为。我们严格遵守既定的伦理准则，本研究使用的所有数据均来自公开数据集，并仅依照原始发布者的使用条款用于学术研究。总体而言，SRVAU-R1 旨在推动视频异常理解研究朝着实现更可靠、可解释的推理方向发展。未来工作应继续关注鲁棒性、公平性和隐私问题，以确保该技术能够在现实场景中得到负责任的部署。

参考文献

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年。Bi, J., Liang, S., Zhou, X., Liu, P., Guo, J., Tang, Y., Song, L., Huang, C., Vosoughi, A., Sun, G., 等。为什么推理很重要？多模态推理进展综述（v1）。*arXiv* 预印本 *arXiv:2504.03151*，2025年。Chen, Y., Liu, Z., Zhang, B., Fok, W., Qi, X., 和 Wu, Y.-C。MGFN：用于弱监督视频异常检测的幅度对比式全局浏览与局部聚焦网络。载于*AAAI*人工智能会议论文集，第37卷，第387–395页，2023年。Ding, D., Wang, L., Zhu, L., Gedeon, T., 和 Koniusz, P。用于多模态特征融合的可学习图算子扩展。*arXiv* 预印本 *arXiv:2410.01506*，2024年。Doshi, K. 和 Yilmaz, Y。迈向可解释视频异常检测。载于*IEEE/CVF*冬季计算机视觉应用会议论文集，第2655–2664页，2023年。Dou, Z., Zhao, Q., Wan, Z., Zhang, D., Wang, W., Raiyan, T., Chen, B., Pan, Q., Ouyang, Y., Gao, Z., 等。先规划后行动：面向大语言模型推理的高级规划引导强化学习。*arXiv* 预印本 *arXiv:2510.01833*，2025年。

Du, H., Nan, G., Qian, J., Wu, W., Deng, W., Mu, H., Chen, Z., Mao, P., Tao, X. 和 Liu, J. 探索是什么、为什么以及如何：一个用于理解视频异常因果关系的多维基准。*arXiv* 预印本 *arXiv:2412.07183*，2024年。

Feng, K., Gong, K., Li, B., Guo, Z., Wang, Y., Peng, T., Wu, J., Zhang, X., Wang, B., 和 Yue, X. Video-r1：在多模态大语言模型中强化视频推理。*arXiv* 预印本 *arXiv:2503.21776*，2025年。

Gandhi, K., Chakravarthy, A., Singh, A., Lile, N., 和 Goodman, N. D. 使推理者能够自我提升的认知行为，或高效之星的四个习惯。*arXiv* 预印本 *arXiv:2503.01307*，2025年。

Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., and Hengel, A. v. d. 记忆正常性以检测异常：用于无监督异常检测的记忆增强深度自编码器。载于*IEEE/CVF*计算机视觉国际会议论文集，第1705–1714页，2019年。Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al. Deepseek-r1：通过强化学习激励大语言模型的推理能力。*arXiv*预印本 *arXiv:2501.12948*，2025年。Huang, C., Wen, J., Liu, C., and Liu, Y. 用于视频异常检测的长短期动态原型对齐学习。载于第三十三届国际人工智能联合会议论文集，第866–874页，2024年。Huang, C., Wang, B., Wen, J., Liu, C., Wang, W., Shen, L., and Cao, X. VAD-R1：通过感知到认知思维链实现视频异常推理。*arXiv*预印本 *arXiv:2505.19877*，2025a。Huang, W., Jia, B., Zhai, Z., Cao, S., Ye, Z., Zhao, F., Xu, Z., Hu, Y., and Lin, S. Vision-R1：激励多模态大语言模型的推理能力。*arXiv*预印本 *arXiv:2503.06749*，2025b。Joo, H. K., Vo, K., Yamazaki, K., and Le, N. CLIP-TSA：用于弱监督视频异常检测的CLIP辅助时序自注意力。载于*2023*年*IEEE*国际图像处理会议（*ICIP*），第3230–3234页。IEEE，2023年。Kumar, A., Zhuang, V., Agarwal, R., Su, Y., Co-Reyes, J. D., Singh, A., Baumli, K., Iqbal, S., Bishop, C., Roelofs, R., et al. 通过强化学习训练语言模型进行自我纠错。*arXiv*预印本 *arXiv:2409.12917*，2024年。Leng, J., Wu, Z., Tan, M., Liu, Y., Gan, J., Chen, H., and Gao, X. 超越欧几里得：用于弱监督视频暴力检测的双空间表示学习。神经信息处理系统进展，37：17373–17397，2024年。

Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., et al. Deepseek-v3技术报告。*arXiv*预印本 *arXiv:2412.19437*，2024年。

Liu, K. 和 Ma, H. 探索监控视频异常检测中的背景偏差。载于第*27*届*ACM*国际多媒体会议论文集，第1490–1499页，2019年。

Liu, W.、Luo, W.、Lian, D. 和 Gao, S. 用于异常检测的未来帧预测——一种新的基线。载于*IEEE*计算机视觉与模式识别会议论文集，第6536–6545页，2018年。

Liu, Z., Nie, Y., Long, C., Zhang, Q., 和 Li, G. 一种通过记忆增强光流重建和光流引导帧预测实现的视频异常检测混合框架。载于*IEEE/CVF*计算机视觉国际会议论文集，第 13588–13597 页，2021年。Liu, Z., Chen, C., Li, W., Qi, P., Pang, T., Du, C., Lee, W. S., 和 Lin, M. 理解类R1-Zero训练：一种批判性视角。*arXiv* 预印本 *arXiv:2503.20783*，2025a。Liu, Z., Sun, Z., Zang, Y., Dong, X., Cao, Y., Duan, H., Lin, D., 和 Wang, J. Visual-RFT：视觉强化微调。*arXiv* 预印本 *arXiv:2503.01785*，2025b。Lu, Y., Yu, F., Reddy, M. K. K., 和 Wang, Y. 少样本场景自适应异常检测。载于欧洲计算机视觉会议，第 125–141 页。施普林格，2020年。Lv, H. 和 Sun, Q. 通过大语言模型进行视频异常检测与解释。*arXiv* 预印本 *arXiv:2401.05702*，2024年。Ma, J., Wang, J., Luo, J., Yu, P., 和 Zhou, G. Sherlock：通过全局-局部空间敏感大语言模型实现多场景视频异常事件提取与定位。载于*2025*年 *ACM Web Conference* 论文集，第 4004–4013 页，2025年。Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Alon, U., Dziri, N., Prabhumoye, S., Yang, Y. 等。Self-refine：通过自我反馈进行迭代优化。神经信息处理系统进展，36:46534–46594，2023年。Pan, K., Wu, Y., Bu, W., Shen, K., Li, J., Wang, Y., Li, Y., Tang, S., Xiao, J., Wu, F. 等。通过强化学习激发顿悟时刻：推进协作式视觉理解与生成。*arXiv* 预印本 *arXiv:2506.01480*，2025年。Shah, D. J., Rushton, P., Singla, S., Parmar, M., Smith, K., Vanjani, Y., Vaswani, A., Chaluvaraju, A., Hojel, A., Ma, A. 等。重新思考预训练中的反思。预印本，2025年。

Sultani, W.、Chen, C. 和 Shah, M. 监控视频中的真实世界异常检测。载于*IEEE*计算机视觉与模式识别会议论文集，第6479–6488页，2018年。

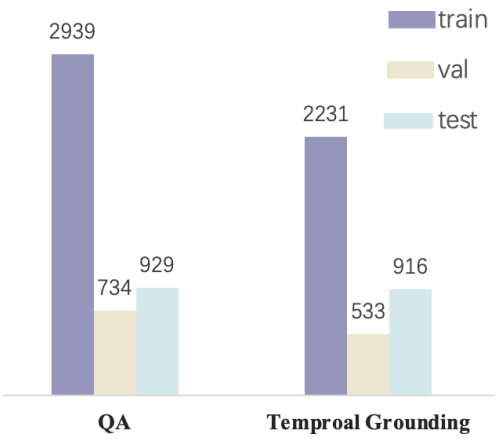
Tan, H.、Ji, Y.、Hao, X.、Lin, M.、Wang, P.、Wang, Z. 和 Zhang, S. Reason-RFT：用于视觉推理的强化微调。*arXiv*预印本 *arXiv:2503.20752*，2025年。

Tang, J., Lu, H., Wu, R., Xu, X., Ma, K., Fang, C., Guo, B., Lu, J., Chen, Q., 和 Chen, Y. Hawk：学习理解开放世界视频异常。神经信息处理系统进展，37:139751– 139785，2024年。Tian, Y., Pang, G., Chen, Y., Singh, R., Verjans, J. W., 和 Carneiro, G. 基于鲁棒时间特征幅值学习的弱监督视频异常检测。发表于*IEEE/CVF*国际计算机视觉大会论文集，第4975–4986页，2021年。Wan, Z., Dou, Z., Liu, C., Zhang, Y., Cui, D., Zhao, Q., Shen, H., Xiong, J., Xin, Y., Jiang, Y., 等。SRPO：通过反思感知强化学习增强多模态大语言模型推理。*arXiv*预印本 *arXiv:2506.01713*，2025年。Wang, W., Chen, Z., Wang, W., Cao, Y., Liu, Y., Gao, Z., Zhu, J., Zhu, X., Lu, L., Qiao, Y., 等。通过混合偏好优化增强多模态大语言模型的推理能力。*arXiv*预印本 *arXiv:2411.10442*，2024年。Wu, P., Liu, J., Shi, Y., Sun, Y., Shao, F., Wu, Z., 和 Yang, Z. 不仅要看，也要听：弱监督下的多模态暴力检测学习。发表于欧洲计算机视觉会议，第322–339页。施普林格，2020年。Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., 等。Qwen3技术报告。*arXiv*预印本 *arXiv:2505.09388*，2025年。Yang, Y., Lee, K., Dariush, B., Cao, Y., 和 Lo, S.-Y. 遵循规则：利用大语言模型进行视频异常检测推理。发表于欧洲计算机视觉会议，第304–322页。施普林格，2024年。Yao, Y., Wang, X., Xu, M., Pu, Z., Wang, Y., Atkins, E., 和 Crandall, D. J. Dota：驾驶视频中交通异常的无监督检测。*IEEE*模式分析与机器智能汇刊，45:444–459，2023年。

Ye, M., Peng, X., Gan, W., Wu, W., 和 Qiao, Y. Anopcn：通过深度预测编码网络进行视频异常检测。发表于第*27*届*ACM*国际多媒体会议论文集，第1805–1813页，2019年。

Ye, M., Liu, W. 和 He, P. Vera：通过视觉-语言模型的语言化学习实现可解释视频异常检测。载于计算机视觉与模式识别会议论文集，第8679–8688页，2025年。Yu, Q., Zhang, Z., Zhu, R., Yuan, Y., Zuo, X., Yue, Y., Dai, W., Fan, T., Liu, G., Liu, L. 等。Dapo：大规模开源大语言模型强化学习系统。*arXiv*预印本 *arXiv:2503.14476*，2025年。Yue, Y., Chen, Z., Lu, R., Zhao, A., Wang, Z., Song, S. 和 Huang, G. 强化学习真的能在基础模型之外激励大语言模型的推理能力吗？*arXiv*预印本 *arXiv:2504.13837*，2025a。Yue, Y., Yuan, Y., Yu, Q., Zuo, X., Zhu, R., Xu, W., Chen, J., Wang, C., Fan, T., Du, Z. 等。Vapo：面向高级推理任务的高效可靠强化学习。*arXiv*预印本 *arXiv:2504.05118*，2025b。Zanella, L., Menapace, W., Mancini, M., Wang, Y. 和 Ricci, E. 利用大语言模型实现无需训练的视频异常检测。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第18527–18536页，2024年。Zelazo, P. D. 执行功能：反思、迭代再加工、复杂性与发育中的大脑。发育评论，38:55–68，2015年。Zhang, H., Xu, X., Wang, X., Zuo, J., Han, C., Huang, X., Gao, C., Wang, Y. 和 Sang, N. Holmes-vad：通过多模态大语言模型实现无偏且可解释的视频异常检测。*arXiv*预印本 *arXiv:2406.12235*，2024年。Zhang, H., Xu, X., Wang, X., Zuo, J., Huang, X., Gao, C., Zhang, S., Yu, L. 和 Sang, N. Holmes-vau：迈向任意粒度的长期视频异常理解。载于 计算机视觉与模式识别会议论文集，第13843–13853页，2025年。Zhou, H., Yu, J. 和 Yang, W. 具有不确定性调节的双记忆单元用于弱监督视频异常检测。载于 *AAAI*人工智能会议论文集，第37卷，第3769–3777页，2023年。Zhou, H., Li, X., Wang, R., Cheng, M., Zhou, T. 和 Hsieh, C.-J. 非监督微调模型在视觉推理中“R1-zero的”顿悟时刻”：基于2B非监督微调模型的研究。*arXiv*预印本 *arXiv:2503.05132*，2025年。Zhu, L., Wang, L., Raj, A., Gedeon, T. 和 Chen, C. 推进视频异常检测：简明综述与新数据集。神经信息处理系统进展，37:89943–89977，2024年。

Zhu, L., Chen, Q., Shen, X. 和 Cun, X. Vau-r1：通过强化微调推进视频异常理解。*arXiv* 预印本 *arXiv:2505.23504*，2025年。



（*a*）跨任务的数据集划分统计。



（*b*）问答任务问题与答案选项的词云。

图 *4*。**数据集构成与语义分布概览**。（a）问答和时间定位任务的数据集划分统计，包括训练集、验证集和测试集。（b）根据训练数据中的反思文本构建的词云可视化，突出显示基于反思的异常推理过程中学到的主要语义线索。

A. 训练数据集

我们的训练数据取自三个成熟的视频异常数据集：UCF Crime、ECVA 和 MSAD，均由 VAU-R1（Zhu 等人，2025年）提供。这些数据集在异常类别、语义复杂度和场景多样性方面互为补充，共同为视频异常理解与推理提供了广泛覆盖。数据集的详细统计信息见图 4。

UCF-Crime 是视频异常检测与理解领域中应用最广泛的大规模真实世界数据集之一，主要采集自监控场景。它涵盖了多种高风险异常事件，包括暴力行为、抢劫、打斗和交通事故。该数据集的特点是视频时长较长、异常片段稀疏且背景复杂。由于异常通常仅占每段视频中很短的一部分，UCF-Crime 对时间建模和长程依赖推理提出了重大挑战。

ECVA </style'>更侧重于异常事件的语义结构和因果关系。它包含具有明确行为模式和可解释线索的异常视频。与面向检测的数据集相比，ECVA 更关注异常发生的过程和逻辑结构，为异常理解和推理任务提供了更合适的语义基础，尤其适用于分析异常行为的触发条件和后果。

MSAD (Zhu 等人，2024年）涵盖了多样化环境中的异常事件，包括室内、室外和公共场所。该数据集在异常类型、场景分布和视觉外观方面表现出高度多样性。MSAD 中的异常通常更加细粒度且语义模糊，给模型泛化和跨场景推理带来了更大挑战。因此，MSAD 尤其适用于评估分布外（OOD）场景下异常理解的稳定性。

反思增强数据构建。 为构建反思增强训练数据，我们采用统一的“初始推理 → 自我反思 → 推理修订”范式，明确建模视频异常理解中从错误判断到自我纠正的完整推理轨迹。首先，我们使用基础多模态大语言模型 Qwen2.5-VL-3B，为每个视频—问题对生成初始推理过程和初始预测。在此阶段，我们不会刻意引入反思提示，使模型在缺乏自我反思能力的情况下展现其自然推理行为。这一设计为后续反思提供了真实且具有挑战性的起点。接下来，我们基于初始推理输出及相应的真实事实标注，引入高性能多模态教师模型 Qwen3-VL-30B，对初始推理进行系统性自我反思。反思过程侧重以下方面：（1）是否遗漏了关键证据，或推理是否过度依赖局部视觉片段。（2）是否存在时间错位或异常时间定位错误。（3）

表 7. ECVA 数据集上的性能比较。

数据集	模型	问答准确率		VAU-Eval					
		准确率$_{w/o\ 思考}$	准确率$_{w/\ 思考}$	CLS↑	KM↑	FLU↑	INF↑	FAC↑	总计↑
ECVA	InternVL2.5-2B	78.84	58.84	2.86	2.78	7.57	4.62	3.03	20.86
	Qwen2.5-VL-7B	83.02	86.98	3.70	3.67	8.64	6.40	4.04	26.45
	InternVL2.5-8B-MPO	90.00	83.72	3.40	3.31	7.87	4.48	3.47	22.53
	Qwen2-VL-2B	86.98	83.95	2.41	2.36	7.81	3.81	2.57	18.96
	Qwen2.5-VL-3B	85.58	75.81	2.21	2.58	8.33	5.02	2.75	20.89
	VAU-R1	89.53	86.51	1.45	2.24	8.05	4.32	2.39	18.45
	SRVAU-R1	**90.18** (↑0.65)	**92.22** (↑5.71)	2.86	3.78	8.96	5.45	3.05	24.1

推理链是否存在因果断裂或逻辑跳跃。（4）初始结论是否与整体视频语义一致。反思以明确的自然语言形式给出，着重识别和分析初始推理中可能存在的问题，而不是直接呈现正确答案。在反思反馈的指导下，教师模型进一步生成修订后的推理过程和最终正确答案。修订后的推理明确展示反思如何影响推理路径的调整，例如重新检查关键时间片段、纳入先前忽略的行为线索，或纠正有误的因果假设。由此得到逻辑连贯且以证据为依据的异常解释。通过这一流程，视频异常理解中的推理失败与纠正过程被明确纳入监督信号。这使模型能够系统地学习如何发现错误、错误产生的原因以及如何纠正错误，为后续反思感知强化学习阶段提供高质量的初始化。

B. 超参数

在 SRVAU-R1 的强化学习阶段，我们采用组相对策略优化（GRPO）作为优化框架。展开批次大小和训练批次大小均设为 128；对于每个输入查询，我们采样 $G = 4$ 个候选响应，以进行组内相对优势估计，采样温度固定为 1.0。我们使用 AdamW 作为优化器，学习率为 2×10^{-5}，所有实验均采用 bfloat16（bf16）混合精度训练。在策略更新期间，我们结合使用 GRPO 的组归一化优势估计和 KL散度正则化项，以约束当前策略与冻结参考模型之间的分布偏移。为提高计算效率，我们启用梯度检查点和 FlashAttention，以降低内存消耗。所有模型均采用全参数微调，不使用 LoRA 或其他参数高效微调模块。对于视频输入，我们以每秒 1 帧的速度均匀采样帧。在奖励计算期间，我们对每组中的多个奖励组成项进行归一化，包括任务正确性、反思质量和时间交并比，以减少奖励尺度失衡并稳定多目标策略优化。

C. 复杂视频场景下的进一步评估结果（ECVA）

表7 报告了不同多模态模型在 ECVA 数据集上的问答准确率和 VAU-Eval 推理得分。与 MSAD 和 UCF-Crime 相比，ECVA 更具挑战性，这是因为其视频时长更长、视角变化更频繁，且异常多样性更高。这些因素对时间理解和因果推理提出了更高要求，导致各模型的整体表现均有所下降。在问答准确率方面，SRVAU-R1 在启用思考和不启用思考两种设置下均取得最佳结果，表明反思感知训练能够提升复杂场景中的最终决策能力。在 VAU-Eval 指标方面，SRVAU-R1 的流畅度（FLU）和信息丰富度（INF）得分相对较高，说明其生成的推理对关键证据的覆盖更全面，逻辑连贯性也更强。然而，与其在 MSAD 和 UCF-Crime 上的表现相比，SRVAU-R1 在 ECVA 上的整体推理得分仍受到限制。尤其是关键匹配（KM）方面的提升有限，这凸显了长程依赖以及包含多个共现异常的场景所带来的难度。总体而言，ECVA 上的结果表明，反思增强训练能够持续提升复杂视频中的决策准确率和推理质量，但仍不足以全面应对时间和语义复杂度较高情境下的细粒度推理挑战。若要进一步提升，可能需要更有针对性的反思监督或更细粒度的奖励设计。

D. 提示模板

初始推理提示

系统提示：
你是一名精准可靠的视频异常理解人工智能助手。你必须严格遵守以下格式
和推理规则：

1. 首先，在 <think>...</think> 中生成简洁、近似人类的推理过程。
2. 推理必须严格以视频中可观察到的动作和事件为依据。不要猜测，也不要引入视觉证据无法支持的信息。
3. 将最终答案放在 <answer>...</answer> 中输出。
4. 答案必须是一个大写字母（A/B/C/D）。

用户：
请分析给定的视频并回答以下选择题。**问题：** {data ["Question"]}**选项：**

- **A.** {数据 ["选项 1"]}
- **B.** {数据 ["选项 2"]}
- **C.** {数据 ["选项 3"]}
- **D.** {数据 ["选项 4"]}

反思感知数据构建提示

系统提示：
你是一位一丝不苟的多模态推理编辑器。你必须通过自我反思来改进模型的推理，并生成一份经过修订、有依据且与正确选项相符的思维链。**用户：**
你正在对先前为一道基于视频的选择题生成的推理过程进行自我反思。你的目标是：

1. 批判初始推理；以及
2. 生成更清晰、更可靠且支持正确选项的修订后推理。

反思规则：

- **如果初始答案不 匹配正确选项：**
  - 解释为什么所选的不正确选项无法得到视频的支持。– 解释哪些关键证据支持正确选项。
- **如果初始答案匹配 正确选项：**
  - 解释如何让推理更清晰、结构更完善，或更有按时间顺序排列的证据支撑。

保持反思简短，并聚焦于改进推理结构。**最终输出要求：**

- 生成**最终思考**：
- 在 <think>...</think>中撰写一段新的、改进后的推理。

自我反思监督微调提示

系统提示：
你是一个善于思考的多模态推理模型，经过训练，能够逐步分析视频并通过自我反思改进推理。**用户：**

给定一个关于视频的问题以及四个答案选项（A、B、C、D）。你的目标是进行两轮推理：

1. **初始推理：** 分析视频并生成初始思维链（CoT）推理，随后给出初始答案。

2. **反思与修订：** 反思你的初始推理，找出潜在问题或需要改进之处，然后生成修订后的推理和最终答案。

推理指南：

- 所有推理都必须严格以视频中可观察到的内容为依据。不要猜测或臆造。
- 按照视频中事件发生的时间顺序进行推理。
- 推理结构应为：发生了什么 →这意味着什么 → 与选项进行比较。
- 推理应清晰、简洁、准确且条理分明。

自我反思强化微调提示

系统提示：
你是一个用于视频理解的多模态推理模型。你的推理必须严格依据视频中可观察到的证据，避免幻觉或缺乏依据的推测。

用户：
你将会收到 一个关于视频的问题和四个答案选项（A–D）。你的任务是分两轮进行推理 ：

1. 给出**初步** 推理和**初步** 答案。
2. 反思你的初始推理，如有必要，请予以修订，并提供<начало id='1'>最终推理和<начало id='3'>最终答案。

输出要求（必须严格遵守）：
输出必须依次包含以下五个部分：

- 思考（<think>...</think>）
- 答案（<answer>...</answer>）
- 反思（<reflection>...</reflection>）

推理约束：

- 所有推理都应以视频中可见的事件和动作作为依据。
- 按照事件发生的时间顺序进行推理。
- 推理应清晰、客观且简洁。
- 不要引入视频中没有依据的细节。

E. 生成样本分析

示例：初始预测错误时的反思增强纠正

输入视频（采样帧）：



问题： 视频中支持纵火这一结论的最关键行为线索是什么？**选项：**

A. 停车场内火势迅速蔓延。**B. 一名男子身上着火，正逃离现场。** C. 停车场内存在易燃材料。 D. 该区域缺少消防设备。

初始思考（错误）：
<思考> 在该视频的背景下，支持纵火结论的最关键行为线索是停车场中的火势迅速蔓延。这表明这是蓄意行为，而非意外起火，因为火势在没有明显起因的情况下迅速蔓延。

</思考>
<答案>答 **<答案>**
由于推理不完整且受假设驱动，答案错误。

自我反思：
<反思> 初始推理错误地聚焦于火势迅速蔓延，却没有考虑到那个着火后逃跑的男子；这是支持纵火判断的更直接的行为线索。专家纳入了这一细节，使推理更加全面。下次分析时，请确保考虑所有关键行为线索。
<反思>

</think> 视频描绘了两起蓄意纵火事件。首先，有人在停车场纵火，火焰随即迅速蔓延至整个区域。 哦，我明白了，不久后，画面中一名男子身上着火，并从停车场跑开，这表明其行为可能是蓄意或可疑的。 等等，让我确认一下——没错，这一系列事件在极短时间内发生，火势迅速蔓延，而该男子的举动则构成了有力的行为证据，表明可能发生了犯罪行为。

</think>
<反思> B **<反思>**
通过明确的自我反思进行完善和纠正。

示例：初始预测正确时的反思增强式细化

输入视频（采样帧）：



问题： 穿黄色衬衫的男子做出了什么关键动作，使事态升级为逮捕异常？**选项：**

A. 他试图逃离警察。**B. 他殴打了白色汽车的司机，引发了暴力冲突。** C. 他在拦停过程中辱骂警察。D. 他无视警察的命令，继续行走。

初始思考（正确）：

<思考> 视频中，可以看到穿黄色衬衫的男子走近一辆白色警车……这一行为表明他试图逃避抓捕或逃跑，这是试图避免被逮捕时常见的做法。其他选项与所描述的行为不符：- 选项 A：没有……。- 选项 C：该男子……车辆。- 选项 D：该男子没有……车辆。因此，最合乎逻辑的结论是，该男子上警车这一行为是使事态升级为逮捕异常的关键动作

</think>
<答案> B **<答案>**
答案正确，但推理冗长，且缺乏时间连贯性。

自我反思：
<反思> 初始推理正确地指出，男子进入警车是一个关键动作。不过，推理可以更加简洁，也缺少时间背景以及对实际事件顺序的清晰呈现。 **<反思>**

</think> 视频开始时，一名穿黄色衬衫的男子正在过马路。一辆白色汽车驶近并拦住他，两名警察下车。哦，我明白了，随后，穿黄色衬衫的男子走向驾驶座并殴打司机，之后被其中一名警察按倒在地。等等，让我确认一下——没错，坐在驾驶座上的警察下车后，两名警察都用枪指着躺在地上的男子。这一过程跨越了多个片段，清楚地展现了事态如何从一次例行拦停升级为暴力冲突并最终导致逮捕。

</think>
<反思> B **<反思>**
经过反思后，模型在提升语言连贯性和时间顺序的同时，保留了正确的推理。