Spec-o3: A Tool-Augmented Vision-Language Agent for Rare Celestial Object Candidate Vetting via Automated Spectral Inspection

Anonymous ACL submission

Abstract

Due to the limited generalization and interpretability of deep learning classifiers, The final vetting of rare celestial object candidates still relies on expert visual inspection—a manually intensive process. In this process, astronomers leverage specialized tools to analyze spectra and construct reliable catalogs. However, this practice has become the primary bottleneck, as it is fundamentally incapable of scaling with the data deluge from modern spectroscopic surveys. To bridge this gap, we propose Spec-o3, a tool-augmented vision-language agent that performs astronomer-aligned spectral inspection via interleaved multimodal chain-of-thought reasoning. Spec-o3 is trained with a two-stage post-training recipe: cold-start supervised fine-tuning on expert inspection trajectories followed by outcome-based reinforcement learning on rare-type verification tasks. Evaluated on five rare-object identification tasks from LAMOST, Spec-o3 establishes a new State-of-the-Art, boosting the macro-F1 score from 28.3 to 76.5 with a 7B parameter base model and outperforming both proprietary VLMs and specialized deep models. Crucially, the agent demonstrates strong generalization to unseen inspection tasks across survey shifts (from LAMOST to SDSS/DESI). Expert evaluations confirm that its reasoning traces are coherent and physically consistent, supporting transparent and trustworthy decision-making. Data and models are available. Code will be released.

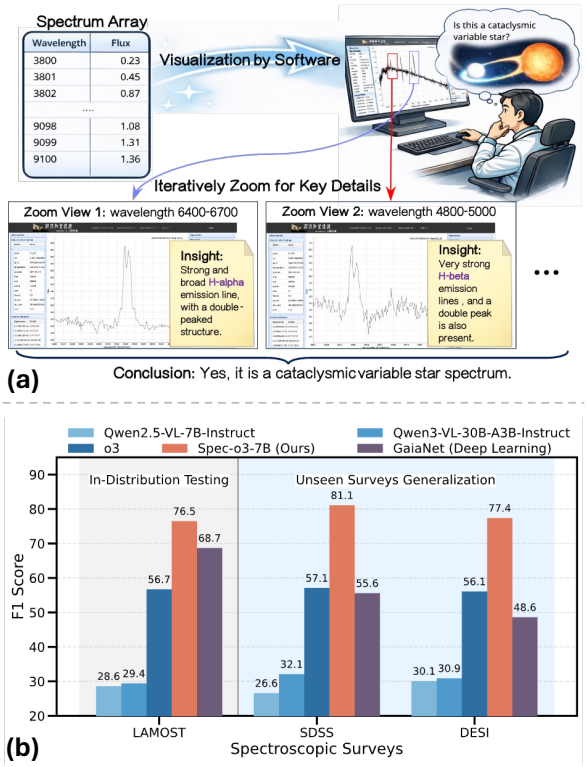

Figure 1: **(a) Astronomers' visual inspection workflow.** Astronomers visualize the raw numerical arrays to assess global morphology, then iteratively zoom into specific wavelength regions to examine fine-grained features for the final decision. **(b) Performance comparison.** Spec-o3 achieves state-of-the-art performance and good generalization across all datasets.

1 Introduction

Detecting rare celestial objects and establishing catalogs for them is one of the core objectives of many large-scale spectroscopic surveys (York et al., 2000; Cui et al., 2012; Aghamousa et al., 2016), as these long-tail objects are instrumental in deepening and refining our understanding of astrophysical theories. In practice, developing such catalogs generally involves a two-stage process that combines automated candidate screening with expert vetting (Tan et al., 2025; Kong and Luo, 2021; Inight et al., 2025). Deep learning algorithms can scale the screening stage over massive spectral archives (Lanusse et al., 2023), but the final vetting stage still relies on manual visual inspection to filter subtle contaminants and instrumental artifacts and ensure catalog reliability (Kim et al., 2024; Borra, 2015; Lan et al., 2023). However, the vetting stage is difficult to scale. For instance, building a cataclysmic variable catalog from LAMOST survey required experts to visually inspect about

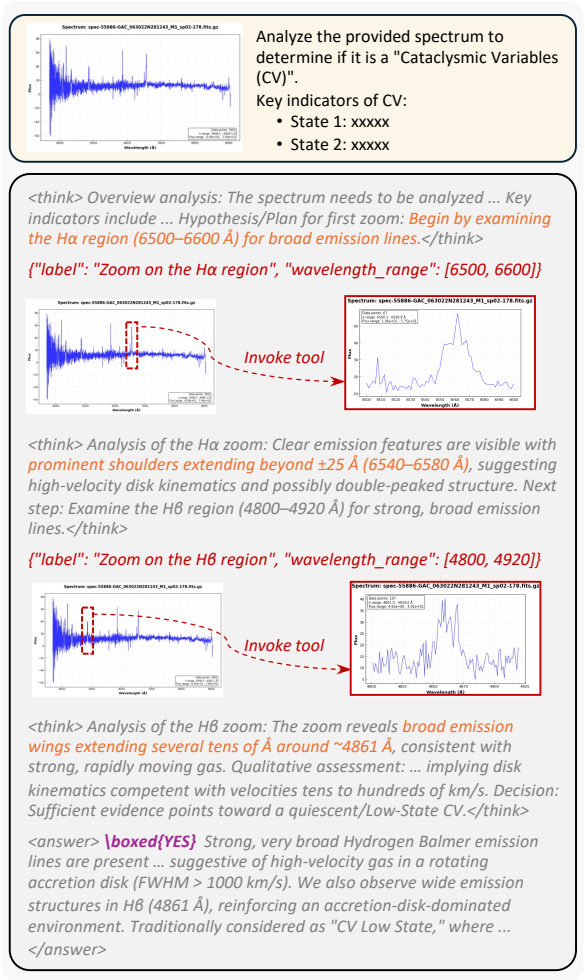

Figure 2: An illustration of Spec-o3's Interleaved Multimodal Chain-of-Thought. The agent iteratively alternates between textual reasoning (<think>...</think>) and fine-grained visual evidence from tool-rendered zoomed spectrum plots. Red JSON shows the tool calls. The final decision is in <answer>...</answer>.

170k candidates screened from roughly 10 million archived spectra, ultimately confirming only 323 objects (Sun et al., 2021). With the exponential data growth of next-generation surveys, candidate volumes will continue to surge (Fulmer et al., 2023; Vincent et al., 2023; Li et al., 2024b), making manual inspection a major bottleneck in modern astronomy (Fluke et al., 2020, 2023).

A key reason why manual inspection is still required is that deep learning models typically produce opaque probability scores and exhibit limited out-of-distribution generalization, which undermines expert trust (Lieu, 2025; Wetzel et al., 2025). Although post-hoc methods such as Grad-CAM (Selvaraju et al., 2017), Integrated Gradients (Sundararajan et al., 2017), LIME (Ribeiro et al., 2016), and SHAP (Lundberg and Lee, 2017) offer some interpretability, they mainly produce coarse feature attributions. Such noisy explanations often cannot be reliably mapped onto astrophysical structures (Stoppa et al., 2025; Lieu, 2025), rendering them insufficient to substitute for expert inspection.

We observe that real-world inspection is a "think-with-spectral-image" process as illustrated in Figure 1(a). Astronomers typically rely on software to visualize spectra stored as numerical arrays, whereby they form an initial judgment from the global morphology and subsequently iteratively zoom into task-relevant wavelength regions to verify local details. Finally, they make a vetting decision on whether to include the candidate in the catalog. The software details are provided in Appendix A. This raises a key question: **Can we design an expert-trusted and highly generalized vetting agent to inspect spectra like Astronomers?**

Aligning models with the iterative expert workflow may improve acceptance of automated vetting. Vision-Language Models (VLMs) have shown promise in generating expert-level explanations for transient imaging classification (Stoppa et al., 2025). Building on this insight, a natural direction is to augment VLMs with a spectral visualization tool (see Appendix B for details), so that they can inspect spectra like Astronomers in the vetting stage. However, even advanced think-with-image models such as o3 (OpenAI, 2025c) perform poorly on this task (Figure 1(b)), failing to consistently distinguish subtle spectral-shape differences. While task-specific fine-tuning is a natural remedy, it is often impractical because it requires a large amount of fully annotated expert trajectory data.

To overcome these limitations, we introduce **Spec-o3**, a tool-augmented agent for rare celestial object candidate vetting via automated spectral inspection. Spec-o3 follows an Interleaved Multimodal Chain-of-Thought (iMCoT) trajectory (Figure 2), alternating between textual reasoning and fine-grained visual evidence from tool-rendered zoomed spectral views before producing a final vetting decision. It adopts a two-stage post-training strategy that cold-starts with a small set of expert trajectories and then scales via outcome-based reinforcement learning.

Our main contributions are as follows:

- We propose Spec-o3, the first tool-augmented VLM-based agent that inspects spectra like Astronomers for rare celestial object vetting.
- We propose a two-stage post-training ap-

proach for spectral inspection. We curate a high-quality astronomy iMCoT dataset to provide a reliable cold-start Supervised Fine-tuning (SFT). Outcome-based reinforcement learning (RL) then improves tool usage and think-with-spectral-image inspection.

- We build a standardized benchmark for rare-object candidate vetting based on public LAMOST, SDSS, and DESI spectra, and we demonstrate strong generalization and reliability of Spec-o3.

2 Related Works

2.1 Cataloging Rare Celestial Objects

Research on rare celestial object cataloging can be broadly grouped into screening and vetting. Machine learning and deep learning are now standard for screening and can retrieve rare-object candidates at archive scale (He et al., 2024; Zhang et al., 2025a; Inight et al., 2025; Tan et al., 2025; Fang et al., 2025). Yet vetting still relies on expert visual inspection as the final quality-control gate, filtering artifacts and false positives while providing high-fidelity labels for downstream use (Lan et al., 2023; Alexander et al., 2023; Rojas et al., 2023). Prior works therefore focus on software and interfaces that streamline manual inspection to help experts surface key evidence and standardize vetting workflows (Juneau et al., 2024; Landriau et al., 2025), such as MARZ (Hinton et al., 2016), ASERA (Yuan et al., 2016), SpectrumVA (Li et al., 2024b) and Prospect (Ratajczak et al., 2025; Juneau et al., 2025). However, the literature still lacks end-to-end methods that automate expert vetting with an auditable, workflow-aligned inspection process.

2.2 Multimodal Large Language Models

Recent years have seen rapid advances in multimodal large language models (MLLMs). Early approaches typically paired pretrained vision encoders with LLMs using lightweight adapters or projection modules, enabling basic cross-modal alignment and simple multimodal reasoning (Liu et al., 2023; Li et al., 2023; Bai et al., 2023; Liu et al., 2024; Chen et al., 2024; Tu et al., 2025). As research progressed, models such as Qwen2.5-VL (Bai et al., 2025b), LLaVA-OneVision (Li et al., 2024a) and InternVL3 (Zhu et al., 2025) scaled up both data and model capacity, leading to substantial improvements on several tasks. More recent OmniMLLM systems (Fu et al., 2024; Hong et al., 2025a; Li et al., 2025; Zhao et al., 2025a) further extend the range of supported modalities, jointly processing images, video, speech, and other signals within a unified framework. Nevertheless, these models are still largely used as passive perception engines that map multimodal inputs to textual responses, and seldom engage in domain-specific, tool-augmented analysis of scientific data.

2.3 Think with Image

The think-with-image paradigm, presented first by o3 (OpenAI, 2025c), has inspired a series of open-source efforts to reproduce and extend tool-augmented visual reasoning. GRIT (Fan et al., 2025) and Pixel Reasoner (Wang et al., 2025) explicitly integrate visual information to achieve a more precise visual focus. VLM-R^3 (Jiang et al., 2025), Chain-of-Focus (Zhang et al., 2025b) and Mini-o3 (Lai et al., 2025) employ two-stage training pipelines to enhance tool-use capabilities. DeepEyes (Zheng et al., 2025) explores the use of reinforcement learning alone to enhance the model's ability to think with images. To improve generality, PyVision (Zhao et al., 2025b) and Thyme (Zhang et al., 2025c) introduce the execution of programmatic code into the visual reasoning loop, allowing flexible manipulation of visual operations. DeepEyes-v2 (Hong et al., 2025b) further extends the tool set by searching, enabling models to retrieve external knowledge. Yet their core interaction is pixel-level image editing, whereas spectral inspection requires repeatedly re-rendering local wavelength views from numerical spectra, which prevents a direct transfer.

3 Method

3.1 Spec-o3

Spec-o3 is a tool-augmented multimodal agent that performs "think-with-spectral-image" inspection. It is built upon Qwen2.5-VL (Bai et al., 2025b) and trained through a two-stage post-training procedure, combining cold-start (Section 3.2) with agentic reinforcement learning (Section 3.3).

As illustrated in Figure 3, Spec-o3 follows an iMCoT trajectory, alternating between textual reasoning and tool-generated spectral images. Spec-o3 takes as input a text prompt T_0 and an initial visualization I_0 rendered from the original spectral array. Here, T_0 includes a discriminative query about whether the spectrum matches the target category, along with an expert-written summary of its

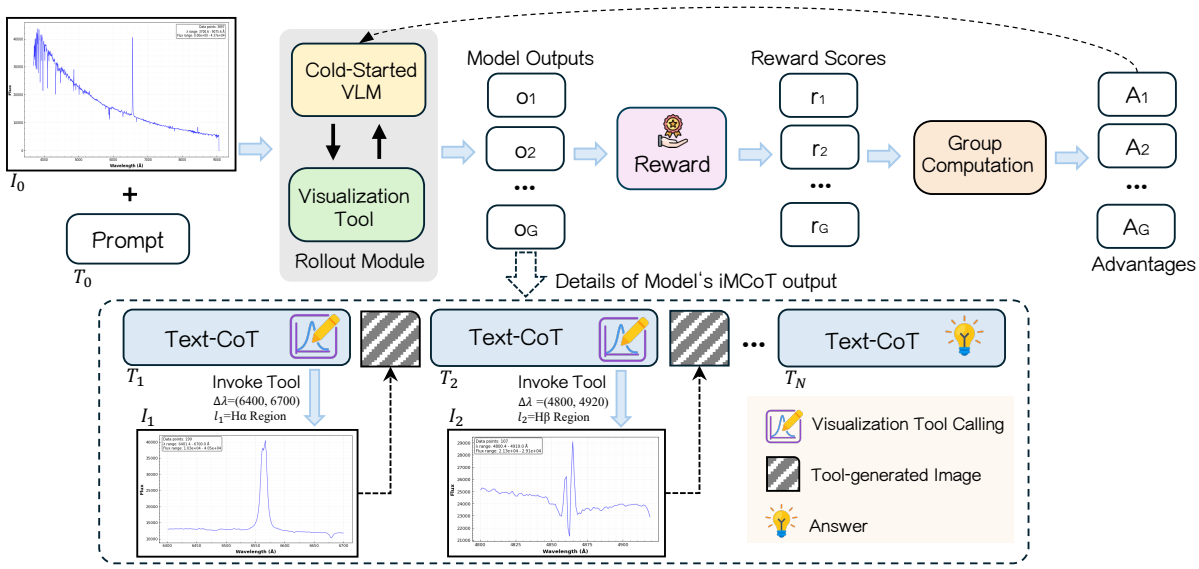

Figure 3: Overview of the Spec-o3 framework. Given a prompt T_0 and an initial view I_0, the VLM generates an iMCoT trajectory in which text reasoning blocks T_n are interleaved with tool-generated images parameterized by wavelength interval $\Delta\lambda$ and optional label l_n, until the final text output T_N is produced. The VLM is cold-start initialized and optimized with GRPO.

diagnostic features and common contaminants.

To formalize the iterative reasoning process, we define the state s_t of iMCoT at step t as follows:

$$s_t = \{(I_k, T_k)\}_{k=0}^{t} = \{I_{\le t}, T_{\le t}\}, \qquad (1)$$

where $I_{\le t} = \{I_0, I_1, \ldots, I_t\}$ denotes the set of rendered spectral visualizations observed so far, and $T_{\le t} = \{T_0, T_1, \ldots, T_t\}$ represents the corresponding textual tokens. Given the current state s_t, the action $a_t \sim \pi_\theta(a \mid s_t)$ is drawn from the policy π_θ. Specifically, Spec-o3 autonomously determines whether to directly output a final answer or to use the spectral visualization tool $Tool_t$ to acquire more fine-grained evidence from wavelength regions of interest. Here, $Tool_t$ denotes the tool invocation at step t, whose input is a wavelength interval $\Delta\lambda_t = (\lambda_t^{\min}, \lambda_t^{\max})$ accompanied by an optional textual label l_t for the queried diagnostic region. Upon successful execution, the tool returns a localized re-rendering I_{t+1} restricted to $\Delta\lambda_t$, which is appended to the state as the next observation. This interaction repeats until the model outputs the final text block T_N containing the answer, or reaches a preset limit on tool invocations. The resulting iMCoT trajectory can be written as

$$\tau = \big(T_0, I_0, T_1, I_1, T_2, I_2, \ldots, T_N\big). \qquad (2)$$

3.2 Cold Start

To initiate the cold start phase, we sample spectra for the five tasks, including Cataclysmic Variables (CV), Carbon Stars (CS), S-type Stars (SS), M-type Giants (MG), and White Dwarfs (WD), from the corresponding LAMOST official catalogs[1]. We additionally sample spectra from several confusing types selected by astronomers based on domain experience. After filtering all sampled spectra by Signal-to-Noise Ratio (SNR) > 10, we obtain an initial pool of approximately 4k spectra for trajectory construction.

Astronomers first formulate simplified inspection guidelines for each task, specifying key spectral features and typical contamination patterns. We then prompt GPT-5 (OpenAI, 2025b) with these guidelines and ground-truth labels, enabling it to invoke spectral visualization tool to generate initial reasoning trajectories. Astronomers first screen these drafts and discard trajectories that are evidently implausible. For each remaining draft, one astronomer performs a focused revision by correcting tool arguments and rewriting the accompanying analysis to ensure that each claim is supported by the rendered evidence. Two additional astronomers then independently audit the revised trajectory against the same guidelines. If either auditor requests changes, the trajectory is returned for revision and re-audited until all three experts approve it. The expert-approved trajectories are then standardized by GPT-5 into iMCoT-formatted reasoning traces optimized for learning. After rewrit-

[1] https://www.lamost.org/dr11/v2.0/catalogue

ing, the three astronomers jointly perform a final acceptance vote on each CoT sequence to confirm its fidelity to the validated tool actions and evidence statements. CoT sequences that fail this final vote are discarded. After this process, we compile a cold start dataset of approximately 1k high-quality trajectories.

Subsequently, we performed SFT on Qwen2.5-VL (Bai et al., 2025b). To prevent the model from memorizing tool outputs, we applied a token-level loss mask to the tool returns. This strategy encourages the model to learn tool invocation and interpretation methodologies rather than memorizing specific visualization results.

3.3 Agentic Reinforcement Learning

The cold start phase injects basic expert priors and enables stable tool use. However, performance is still constrained by the scarcity of full expert trajectories, especially in noisy and heavily contaminated cases. We therefore apply outcome-based reinforcement learning after cold start, using relatively abundant label-only data to further improve performance.

Optimization As illustrated in Figure 3, we utilize Group Relative Policy Optimization (GRPO) (Shao et al., 2024) for reinforcement learning. Consistent with the cold start phase, to prevent tool outputs from interfering with the training objective, we apply a token-wise loss mask to all response tokens of tool, effectively excluding them from the loss computation.

Reward Design In the agentic reinforcement learning phase, we use a rule-based outcome reward to evaluate each trajectory, following Guo et al. (2025). The reward prioritizes final prediction accuracy and enforces output-format constraints. Since tool use is already reliable after cold start, we do not add an explicit tool-usage reward term as in Zheng et al. (2025). Given a reasoning trajectory τ, the reward function is defined as:

$$r(\tau) = \begin{cases} 1, & \text{if } o_{\text{pred}} = o_{\text{gold}} \ \wedge \ f_{\text{fmt}}(y), \\ 1-\alpha, & \text{if } o_{\text{pred}} = o_{\text{gold}} \ \wedge \ \neg f_{\text{fmt}}(y), \\ 0, & \text{if } o_{\text{pred}} \neq o_{\text{gold}} \ \wedge \ f_{\text{fmt}}(y), \\ -\alpha, & \text{if } o_{\text{pred}} \neq o_{\text{gold}} \ \wedge \ \neg f_{\text{fmt}}(y), \end{cases} \tag{3}$$

where o_{pred} and o_{gold} are the predicted and ground-truth labels, $f_{\text{fmt}}(y)$ indicates whether the response y satisfies the required formatting constraints, and α controls the penalty for format violations.

4 Experiments

4.1 SpecVI-Bench

Most existing spectral datasets are designed for the candidate screening stage and therefore use stratified sampling to construct negative examples (Tan et al., 2025; He et al., 2024; Tan et al., 2022). These negatives do not reflect the vetting stage, in which contaminants can closely resemble the true targets. To evaluate and train models under this high-confusion vetting setting, we construct SpecVI-Bench.

SpecVI-Bench comprises the same five rare-object categories as in the cold-start phase, each formulated as an independent inspection task. For each task, positive samples are taken from the corresponding official catalogs. To approximate the inspection stage, we construct hard negatives using a rejection-sampling procedure. Specifically, for each task we train a weak classifier and then sample spectra from the full LAMOST archive, retaining only sources that are not listed in the corresponding official catalogs and whose predicted positive probability exceeds 0.8. On average, the rejection-sampling acceptance rate is 1.74%. We finally create approximately balanced training splits and deliberately imbalanced test splits to reflect the rarity of true targets in practice. To avoid contamination between training demonstrations and evaluation, we ensure that spectra used for cold-start trajectory construction are excluded from the SpecVI-Bench test split. The detailed statistics are provided in the Appendix C.

4.2 Experimental Setup

Datasets We evaluate on three datasets. SpecVI-Bench is the main benchmark. We further test generalization on (i) a Cross-Survey set from SDSS and DESI, which tests robustness to survey-specific instrumental differences, and (ii) a Cross-Task set with unseen target categories, which evaluates transfer to new inspection tasks. Introduction of the surveys is provided in Appendix D, and detailed dataset statistics are reported in Appendix E.

Baselines We benchmark Spec-o3 against three distinct categories of baselines: (1) Specialist Deep Learning Models, where we adapted and fine-tuned CarbonNet (He et al., 2024), GaiaNet (Ye et al., 2025) (task-specific architectures), and AstroCLIP (Parker et al., 2024) (a large-scale pre-trained spectral foundation model) on the SpecVI-Bench train-

Table 1: **Main results on SpecVI-Bench** across five tasks: Cataclysmic Variables (CV), Carbon Stars (CS), S-type Stars (SS), M-type Giants (MG), and White Dwarfs (WD). We report Accuracy (Acc) and F1 for each task and their macro-average. All Qwen-VL baselines use the Instruct variants. Spec-o3 is trained from Qwen2.5-VL. Best and second-best results are in **bold** and underlined, respectively.

Model	CV		CS		SS		MG		WD		Average	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Specialist Deep Learning Models												
CarbonNet (He et al., 2024)	92.7	76.6	95.2	**87.5**	85.7	64.4	83.1	52.0	61.8	40.8	83.7	64.3
AstroCLIP (Parker et al., 2024)	92.4	74.7	94.0	84.1	88.7	62.4	80.5	52.3	80.9	48.8	87.3	64.5
GaiaNet (Ye et al., 2025)	91.1	67.2	**95.4**	87.1	89.4	70.3	85.5	51.8	**83.8**	48.2	**89.0**	64.9
Proprietary VLMs												
GPT-4.1 (OpenAI, 2025a)	57.9	27.7	51.9	31.7	60.2	36.1	20.6	29.2	52.3	24.4	48.6	29.8
o3 (OpenAI, 2025c)	88.8	57.1	87.1	53.1	88.1	53.3	81.7	60.0	81.9	37.8	85.5	52.3
Open-source VLMs												
Qwen2.5-VL-3B (Bai et al., 2025b)	30.6	25.8	25.3	20.0	19.0	26.9	17.8	28.7	48.2	28.6	28.2	26.1
Qwen2.5-VL-7B (Bai et al., 2025b)	28.4	25.4	59.8	31.5	19.4	27.3	19.7	29.0	25.8	28.1	30.6	28.3
Qwen3-VL-8B (Bai et al., 2025a)	39.7	33.6	29.8	27.0	29.4	28.4	22.9	27.3	52.9	19.7	34.9	27.2
Qwen3-VL-30B-A3B (Bai et al., 2025a)	22.9	31.3	19.8	29.5	17.5	26.3	35.1	27.4	65.3	34.4	32.1	29.8
Ours												
Spec-o3-3B	92.0	80.7	88.0	75.9	92.6	84.0	**93.0**	82.7	78.5	43.2	88.8	73.3
Δ vs Qwen2.5-VL-3B	+61.4	+54.9	+62.7	+55.9	+73.6	+57.1	+75.2	+54.0	+30.3	+14.6	+60.6	+47.2
Spec-o3-7B	**93.1**	**81.0**	92.5	80.2	**94.2**	**84.5**	90.6	**83.4**	73.2	**53.6**	88.7	**76.5**
Δ vs Qwen2.5-VL-7B	+64.7	+55.6	+32.7	+48.7	+74.8	+57.2	+70.9	+54.4	+47.4	+25.5	+58.1	+48.2

ing split; (2) Proprietary VLMs, including GPT-4.1 (OpenAI, 2025a) and o3 (OpenAI, 2025c); (3) Open-Source VLMs, covering Qwen2.5-VL (Bai et al., 2025b) (base models for Spec-o3) and Qwen3-VL (Bai et al., 2025a). All VLM baselines utilize the same spectral visualization tool and prompt as Spec-o3 for fair comparison.

Metrics We report Acc and positive-class F1 for each task, with macro-averages across tasks. Tests are imbalanced, so F1 is primary and Acc is complementary. For VLMs, we require the final decision to be `\boxed{YES}` or `\boxed{NO}`, and we compute metrics by exact matching of this output.

Training Details We use Qwen2.5-VL-3B and Qwen2.5-VL-7B as our base models. The training phase is conducted using 8 × NVIDIA H100 GPUs. The RL stages employ GRPO framework with 8 rollouts per question, limiting the maximum number of tool calls to 8 per trajectory. Further analysis of the training dynamics and convergence behavior is provided in Appendix F.

4.3 Main Results

Table 1 reports the in-distribution results on the SpecVI-Bench. Spec-o3 achieves state-of-the-art macro-average performance across five tasks. Notably, Spec-o3-7B attains a macro-average F1 score of 76.5%, which surpasses the proprietary o3 model (52.3%) and its base model Qwen2.5-VL-7B (28.3%) by a substantial margin. Our two-stage post-training strategy delivers pronounced performance gains, elevating the macro-average F1 score from 26.1% to 73.3% (+47.2%) for the 3B model and from 28.3% to 76.5% (+48.2%) for the 7B model. These substantial improvements underscore the effectiveness of our framework in bridging the reasoning gap for spectral visual inspection.

Spec-o3 also outperforms specialist deep learning baselines fine-tuned on training split of SpecVI-Bench. While GaiaNet attains a similar average accuracy (89.0%), its macro-average F1 (64.9%) is notably lower than Spec-o3-7B (76.5%), indicating a better precision-recall balance on imbalanced dataset. These results position Spec-o3 as a reliable autonomous agent for the spectral inspection, with the additional benefit of interpretable trajectories.

4.4 Generalization Evaluation

Cross-survey Generalization We evaluate zero-shot performance on the Cross-Survey set from SDSS and DESI without any adaptation (Table 2). Specialist deep learning models show strong

Table 2: **Performance on unseen surveys**. We report F1 for CV, CS, and MG because SS and WD do not have sufficient samples. The LAMOST column is the training-survey reference, averaged over the same task subset on SpecVI-Bench. Best and second-best results are in **bold** and underlined.

Model	SDSS				DESI				LAMOST
	CV	CS	MG	Avg	CV	CS	MG	Avg	Avg
Specialist Deep Learning Models									
CarbonNet (He et al., 2024)	49.8	66.5	46.9	54.4	49.6	64.1	58.2	57.3	72.0
AstroCLIP (Parker et al., 2024)	53.9	64.1	46.4	54.8	48.3	53.9	50.8	51.0	70.4
GaiaNet (Ye et al., 2025)	58.5	67.1	41.3	55.6	48.3	53.8	43.7	48.6	68.7
Proprietary VLMs									
GPT-4.1 (OpenAI, 2025a)	29.4	33.4	28.8	30.5	28.9	34.4	31.8	31.7	29.5
o3 (OpenAI, 2025c)	59.8	54.3	57.1	57.1	60.2	51.4	56.8	56.1	56.7
Open-source VLMs									
Qwen2.5-VL-3B (Bai et al., 2025b)	22.6	25.3	25.4	24.4	23.3	30.8	23.1	25.7	24.8
Qwen2.5-VL-7B (Bai et al., 2025b)	26.7	25.9	27.2	26.6	25.9	33.7	30.6	30.1	28.6
Qwen3-VL-8B (Bai et al., 2025a)	30.9	26.9	29.8	29.2	33.8	23.9	30.4	29.4	29.3
Qwen3-VL-30B-A3B (Bai et al., 2025a)	32.1	34.4	29.8	32.1	31.5	31.1	30.2	30.9	29.4
Ours									
Spec-o3-3B	79.3	76.4	76.3	77.3	76.1	70.1	74.5	73.6	79.8
Spec-o3-7B	**84.9**	**79.8**	**78.5**	**81.1**	**82.6**	**72.8**	**76.7**	**77.4**	**81.5**

Table 3: **Performance on the Cross-Task testset** with unseen types (O, B, A), with F1 reported. Deep learning baselines are omitted because they lack this cross-task transfer capability, and best/second-best are in **bold** and underlined.

Model	O	B	A	Average
GPT-4.1	32.2	34.1	33.0	33.1
o3	57.6	64.2	60.8	60.9
Qwen2.5-VL-3B	26.0	28.1	27.2	27.1
Qwen2.5-VL-7B	29.4	31.8	30.3	30.5
Qwen3-VL-8B	30.7	32.2	30.8	31.2
Qwen3-VL-30B-A3B	32.0	33.7	32.7	32.8
Spec-o3-3B	73.5	72.6	77.1	74.4
Spec-o3-7B	**77.0**	**73.1**	**79.1**	**76.4**

degradation under these instrumental shifts. From the LAMOST reference average, their average F1 drops by -17.6% to -13.1% on SDSS and -20.1% to -14.7% on DESI, which suggests reliance on survey-specific artifacts. Spec-o3-7B remains stable with 81.1% on SDSS and 77.4% on DESI, which is close to its in-distribution reference of 81.5% and well above o3 at 56.1%. These results suggest that Spec-o3 relies on diagnostic spectral evidence that transfers across telescope projects.

Cross-task Generalization We evaluate zero-shot performance on three unseen inspection tasks (O-, B-, and A-type spectra), where specialist deep learning models are inapplicable. Despite the significant visual disparity between these new tasks and the training tasks, Spec-o3 demonstrates remarkable transferability. Spec-o3-7B achieves an average F1 of 76.4%, which is +15.5% over o3 and +45.9% over Qwen2.5-VL-7B. Spec-o3-3B reaches 74.4%, which is +13.5% over o3 and +47.3% over Qwen2.5-VL-3B. These results confirm that Spec-o3 has learned a generic, tool-assisted inspection paradigm that identifies diagnostic visual evidence defined in instructions, rather than merely memorizing specific task distributions.

4.5 Reliability Judgement

To assess the reliability of Spec-o3's explanations, we evaluate the reasoning trajectories with both human experts and automated LLM Judges. We randomly sampled 100 reasoning trajectories (50 each from Spec-o3-3B and Spec-o3-7B). Six astronomers with spectroscopy expertise rated each trajectory for coherence and physical consistency on a discrete 0-5 scale (see Appendix G for comprehensive evaluation details). In parallel, we used four proprietary models (GPT-5, Gemini-2.5-Pro, Claude-4-Sonnet, and Grok-4) as LLM Judges to score the identical trajectories.

Figure 4(a) compares the score distributions and Cumulative Distribution Function (CDF) from human experts and LLM Judges, showing close agreement and a strong concentration of high scores, which suggests that Spec-o3 typically produces coherent and physically consistent explanations. Figure 4(b) further reports Spearman correlations between the human scores and each LLM Judge,

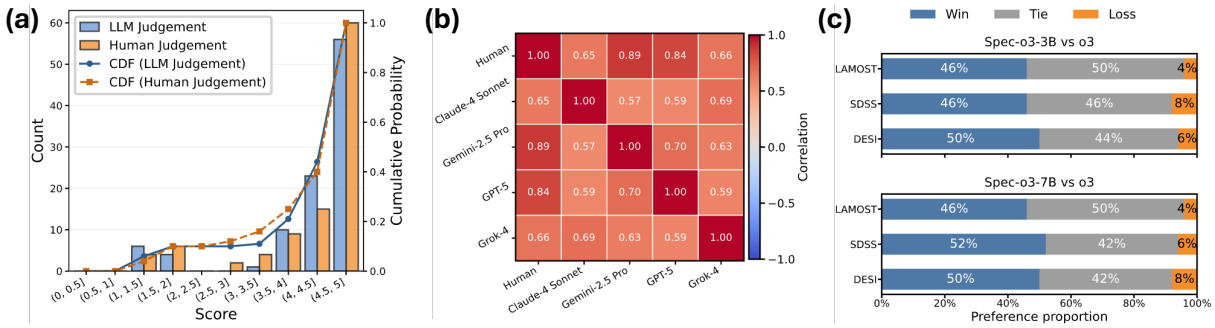

Figure 4: **(a) Quality distribution.** Score distributions and Cumulative Distribution Functions are compared between human and LLM judges to evaluate reasoning trajectories. **(b) Rating consistency.** The heatmap displays Spearman correlation coefficients between human experts and four LLM judges. **(c) Pairwise preference.** Human expert preferences on explanation quality are compared between Spec-o3 and o3 across three datasets.

Table 4: Ablation analysis of training stages and tool availability on SpecVI-Bench. Scores denote the average F1 across five tasks.

#	SFT	RL	Tool	3B-Based	7B-Based
0	✓	✓	✓	73.3	76.5
1	✗	✓	✓	35.7 (-37.6)	40.5 (-36.0)
2	✓	✗	✓	33.1 (-40.2)	41.6 (-34.9)
3	✗	✗	✓	26.1 (-47.2)	28.3 (-48.2)
4	✓	✓	✗	43.5 (-29.8)	55.8 (-20.7)

indicating substantial rater consistency.

Beyond absolute scores, Figure 4(c) presents a pairwise preference study: for each survey data, we randomly sample 50 cases and ask astronomers to compare explanation quality between Spec-o3 and o3, labeling each comparison as win/tie/loss. Across datasets, Spec-o3 is preferred at least as often as o3, with a low loss rate. Together, these results support using LLM Judges as a scalable quality filter (e.g., for ranking or flagging low-confidence trajectories) to reduce expert workload in large surveys.

4.6 Ablation Study

Impact of Two-Stage Training. We conduct ablation studies to analyze the specific contributions of cold-start SFT and agentic reinforcement learning (RL), with results summarized in Table 4. We observe that applying pure RL (#1) yields only limited improvements, as the model lacks the foundational spectral interpretation priors necessary for effective tool utilization. While cold-start SFT alone (#2) achieves performance comparable to pure RL using only sparse expert trajectories, its performance remains bounded by the scarcity of high-quality demonstrations and its weak generalization (Chu et al., 2025). Notably, integrating both stages (#0) triggers a substantial performance leap, nearly doubling the F_1 scores compared to the single-stage baselines (#1, #2). This confirms that injecting domain priors via cold start is a prerequisite for RL to effectively optimize tool usage strategies and unlock robust performance.

Impact of Tool Access. To enable fine-grained inspection, we provide a visualization tool for on-demand re-rendering of local wavelength regions. To validate the necessity of this component, we evaluated a variant (#4) trained without tool access, forcing the model to rely solely on textual Chain-of-Thought (CoT). As shown in Table 4, removing tool access results in a significant performance degradation for both model sizes, despite the application of full SFT and RL procedures. This decline shows that a static global view is insufficient for reliable verification, making interactive re-rendering of local evidence essential for detecting subtle diagnostic features.

5 Conclusion

In this work, we introduce Spec-o3 to mitigate the manual inspection bottleneck in modern spectroscopic survey workflows. Spec-o3 is a tool-augmented vision-language agent that performs interleaved multimodal chain-of-thought reasoning in an astronomer-like inspection loop. Trained with a two-stage post-training strategy, Spec-o3 achieves state-of-the-art performance and remains robust under distribution shifts. Human evaluations confirm that its inspection trajectories are physically consistent and aligned with expert reasoning, positioning Spec-o3 as a scalable solution for the data deluge in future large-scale spectroscopic surveys.

6 Limitations

A limitation of this work is that our evaluation focuses on a limited set of rare-object types, and does not yet cover broader spectral subclasses or the most challenging observational conditions. In addition, we abstract expert vetting as an interactive zoom-and-reason loop with a spectral visualization tool, while real catalog construction often requires cross-matching external databases and incorporating other modalities (e.g., photometry, imaging, or time-domain evidence) for confirmation. Although outcome-based RL can scale with label-only data, our approach still relies on expert demonstration trajectories for cold start, which introduces a nontrivial barrier when extending to new tasks or surveys. Finally, we do not yet provide production-oriented risk controls such as calibration, abstention, or triage mechanisms for deferring uncertain cases to human experts.

References

Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, and 1 others. 2016. The desi experiment part i: science, targeting, and survey design. *arXiv preprint arXiv:1611.00036*.

David M Alexander, Tamara M Davis, E Chaussidon, VA Fawcett, Alma X Gonzalez-Morales, Ting-Wen Lan, Christophe Yeche, S Ahlen, JN Aguilar, E Armengaud, and 1 others. 2023. The desi survey validation: Results from visual inspection of the quasar survey spectra. *The Astronomical Journal*, 165(3):124.

Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. 2023. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv preprint arXiv:2308.12966*.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, and 45 others. 2025a. Qwen3-VL Technical Report. *arXiv preprint arXiv:2511.21631*.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, and 1 others. 2025b. Qwen2.5-vl technical report. *arXiv preprint arXiv:2502.13923*.

EF Borra. 2015. Validation of observations obtained with a liquid mirror telescope by comparison with sloan digital sky survey observations. *The Astronomical Journal*, 149(6):185.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, and 1 others. 2024. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24185–24198.

Tianzhe Chu, Yuexiang Zhai, Jihan Yang, Shengbang Tong, Saining Xie, Dale Schuurmans, Quoc V. Le, Sergey Levine, and Yi Ma. 2025. Sft memorizes, rl generalizes: A comparative study of foundation model post-training. *arXiv preprint*.

Xiang-Qun Cui, Yong-Heng Zhao, Yao-Quan Chu, Guo-Ping Li, Qi Li, Li-Ping Zhang, Hong-Jun Su, Zheng-Qiu Yao, Ya-Nan Wang, Xiao-Zheng Xing, and 1 others. 2012. The large sky area multi-object fiber spectroscopic telescope (lamost). *Research in Astronomy and Astrophysics*, 12(9):1197.

Yue Fan, Xuehai He, Diji Yang, Kaizhi Zheng, Ching-Chen Kuo, Yuting Zheng, Sravana Jyothi Narayanaraju, Xinze Guan, and Xin Eric Wang. 2025. Grit: Teaching mllms to think with images. *arXiv preprint arXiv:2505.15879*.

Ziyu Fang, Xiangru Li, and Haining Li. 2025. A catalog of 12,766 carbon-enhanced metal-poor stars from lamost data release 8. *The Astrophysical Journal Supplement Series*, 277(1):30.

Christopher J Fluke, Sarah E Hegarty, and CO-M MacMahon. 2020. Understanding the human in the design of cyber-human discovery systems for data-driven astronomy. *Astronomy and Computing*, 33:100423.

CJ Fluke, D Vohl, VA Kilborn, and C Murugeshan. 2023. Survey-scale discovery-based research processes: Evaluating a bespoke visualisation environment for astronomical survey data. *Publications of the Astronomical Society of Australia*, 40:e035.

Chaoyou Fu, Haojia Lin, Zuwei Long, Yunhang Shen, Meng Zhao, Yifan Zhang, Shaoqi Dong, Xiong Wang, Di Yin, Long Ma, Xiawu Zheng, Ran He, Rongrong Ji, Yunsheng Wu, Caifeng Shan, and Xing Sun. 2024. Vita: Towards open-source interactive omni multimodal llm. *arXiv preprint arXiv:2408.05211*.

Leah M Fulmer, Stephanie Juneau, Catherine Merrill, Adam S Bolton, David L Nidever, Robert Nikutta, Stephen T Ridgway, Knut AG Olsen, and Benjamin A Weaver. 2023. Astro data lab spectral viewer requirements for wide-area spectroscopic surveys. *arXiv preprint arXiv:2302.08906*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, and 1 others. 2025. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638.

YiMing He, Zhong Cao, Hui Deng, Feng Wang, Ying Mei, and Lei Tan. 2024. Identification of carbon stars

in lamost dr9 based on deep learning. *The Astrophysical Journal Supplement Series*, 274(1):6.

Samuel R Hinton, Tamara M Davis, Chris Lidman, Karl Glazebrook, and Geraint F Lewis. 2016. Marz: Manual and automatic redshifting software. *Astronomy and Computing*, 15:61–71.

Jack Hong, Shilin Yan, Jiayin Cai, Xiaolong Jiang, Yao Hu, and Weidi Xie. 2025a. Worldsense: Evaluating real-world omnimodal understanding for multimodal llms. *arXiv preprint arXiv:2502.04326*.

Jack Hong, Chenxiao Zhao, ChengLin Zhu, Weiheng Lu, Guohai Xu, and Xing Yu. 2025b. Deepeyesv2: Toward agentic multimodal model. *arXiv preprint arXiv:2511.05271*.

Keith Inight, Boris T Gänsicke, Axel Schwope, Scott F Anderson, Elmé Breedt, Joel R Brownstein, Sebastian Demasi, Susanne Friedrich, JJ Hermes, Knox S Long, and 1 others. 2025. Cataclysmic variables from sloan digital sky survey–v (2020–2023) identified using machine learning. *Monthly Notices of the Royal Astronomical Society*, 536(2):1057–1076.

Chaoya Jiang, Yongrui Heng, Wei Ye, Han Yang, Haiyang Xu, Ming Yan, Ji Zhang, Fei Huang, and Shikun Zhang. 2025. Vlm-r ^{3}: Region recognition, reasoning, and refinement for enhanced multimodal chain-of-thought. *arXiv preprint arXiv:2505.16192*.

S Juneau, R Canning, DM Alexander, R Pucha, VA Fawcett, AD Myers, J Moustakas, O Ruiz-Macias, S Cole, Z Pan, and 1 others. 2025. Identifying missing quasars from the desi bright galaxy survey. *The Astronomical Journal*, 169(3):157.

Stéphanie Juneau, Alice Jacques, Steve Pothier, Adam S Bolton, Benjamin A Weaver, Ragadeepika Pucha, Sean McManus, Robert Nikutta, and Knut Olsen. 2024. Sparcl: Spectra analysis and retrievable catalog lab. *arXiv preprint arXiv:2401.05576*.

Young-Lo Kim, Isobel Hook, Andrew Milligan, Lluís Galbany, Jesper Sollerman, Umut Burgaz, Georgios Dimitriadis, Christoffer Fremling, Joel Johansson, Tomás E Müller-Bravo, and 1 others. 2024. How accurate are transient spectral classification tools?—a study using 4646 sedmachine spectra. *Publications of the Astronomical Society of the Pacific*, 136(11):114501.

Xiao Kong and A-Li Luo. 2021. Identification of white dwarfs from gaia edr3 via spectra from lamost dr7. *Research Notes of the AAS*, 5(10):249.

Xin Lai, Junyi Li, Wei Li, Tao Liu, Tianjian Li, and Hengshuang Zhao. 2025. Mini-o3: Scaling up reasoning patterns and interaction turns for visual search. *arXiv preprint arXiv:2509.07969*.

Ting-Wen Lan, R Tojeiro, E Armengaud, J Xavier Prochaska, TM Davis, David M Alexander, Anand Raichoor, Rongpu Zhou, Christophe Yeche, C Balland, and 1 others. 2023. The desi survey validation: Results from visual inspection of bright galaxies, luminous red galaxies, and emission-line galaxies. *The Astrophysical Journal*, 943(1):68.

Martin Landriau, Erin Mentuch Cooper, Dustin Davis, Karl Gebhardt, Robin Ciardullo, Éric Armengaud, Arjun Dey, Anand Raichoor, David J Schlegel, Michael Wilson, and 1 others. 2025. Desi spectroscopy of hetdex emission-line candidates i: Line discrimination validation. *arXiv preprint arXiv:2503.02229*.

François Lanusse and 1 others. 2023. The dawes review 10: The impact of deep learning for the analysis of galaxy surveys. *Publications of the Astronomical Society of Australia*, 40:e001.

Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, and 1 others. 2024a. Llava-onevision: Easy visual task transfer. *arXiv preprint arXiv:2408.03326*.

Jincheng Li, Chufan Lai, Youfen Wang, Ali Luo, and Xiaoru Yuan. 2024b. Spectrumva: Visual analysis of astronomical spectra for facilitating classification inspection. *IEEE Transactions on Visualization and Computer Graphics*, 30(8):5386–5403.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven C. H. Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International Conference on Machine Learning*.

Yadong Li, Jun Liu, Tao Zhang, Song Chen, Tianpeng Li, Zehuan Li, Lijun Liu, Lingfeng Ming, Guosheng Dong, Da Pan, and 1 others. 2025. Baichuan-omni-1.5 technical report. *arXiv preprint arXiv:2501.15368*.

Maggie Lieu. 2025. A comprehensive guide to interpretable ai-powered discoveries in astronomy. *Universe*, 11(6):187.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26296–26306.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36, pages 34892–34916. Curran Associates, Inc.

Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

OpenAI. 2025a. Introducing GPT-4.1 in the API.

OpenAI. 2025b. Introducing gpt-5. https://openai.com/index/introducing-gpt-5/. Accessed: 2026-01-04.

OpenAI. 2025c. Think with images.

Liam Parker, Francois Lanusse, Siavash Golkar, Leopoldo Sarra, Miles Cranmer, Alberto Bietti, Michael Eickenberg, Geraud Krawezik, Michael McCabe, Rudy Morel, and 1 others. 2024. Astroclip: a cross-modal foundation model for galaxies. *Monthly Notices of the Royal Astronomical Society*, 531(4):4990–5011.

J Ratajczak, KS Dawson, N Weaverdyck, J Aguilar, S Ahlen, E Armengaud, S Bailey, D Bianchi, D Blanco, A Brodzeller, and 1 others. 2025. The compilation and validation of the spectroscopic redshift catalogs for the desi-cosmos and desi-xmmlss fields. *arXiv preprint arXiv:2508.09286*.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Karina Rojas, Thomas E Collett, Daniel Ballard, Mark R Magee, Simon Birrer, Elizabeth Buckley-Geer, James HH Chan, Benjamin Clément, José M Diego, Fabrizio Gentile, and 1 others. 2023. The impact of human expert visual inspection on the discovery of strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 523(3):4413–4430.

Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Fiorenzo Stoppa, Turan Bulmus, Steven Bloemen, Stephen J Smartt, Paul J Groot, Paul Vreeswijk, and Ken W Smith. 2025. Textual interpretation of transient image classifications from large language models. *Nature Astronomy*, pages 1–10.

Yongkang Sun, Zhenghao Cheng, Shuo Ye, Ruobin Ding, Yijiang Peng, Jiawen Zhang, Zhenyan Huo, Wenyuan Cui, Xiaofeng Wang, Jianrong Shi, and 1 others. 2021. A catalog of 323 cataclysmic variables from lamost dr6. *The Astrophysical Journal Supplement Series*, 257(2):65.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR.

Lei Tan, Hui Deng, Ying Mei, Yixing Chen, Tianhang Liu, Feng Wang, and 1 others. 2025. A robust method for identifying be stars in the lamost data release 11 based on deep-learning approach. *arXiv preprint arXiv:2511.02221*.

Lei Tan, Ying Mei, Zhicun Liu, Yangping Luo, Hui Deng, Feng Wang, Linhua Deng, and Chao Liu. 2022. A robust hot subdwarfs identification method based on deep learning. *arXiv preprint arXiv:2201.08967*.

Songjun Tu, Qichao Zhang, Jingbo Sun, Yuqian Fu, Linjing Li, Xiangyuan Lan, Dongmei Jiang, Yaowei Wang, and Dongbin Zhao. 2025. Perception-consistency multimodal large language models reasoning via caption-regularized policy optimization. *arXiv preprint arXiv:2509.21854*.

Olivier Vincent, Pierre Bergeron, and Patrick Dufour. 2023. Data-driven selection and spectral classification of white dwarf stars. *Monthly Notices of the Royal Astronomical Society*, 521(1):760–771.

Haozhe Wang, Alex Su, Weiming Ren, Fangzhen Lin, and Wenhu Chen. 2025. Pixel reasoner: Incentivizing pixel-space reasoning with curiosity-driven reinforcement learning. *arXiv preprint arXiv:2505.15966*.

Sebastian Johann Wetzel, Seungwoong Ha, Raban Iten, Miriam Klopotek, and Ziming Liu. 2025. Interpretable machine learning in physics: A review. *arXiv preprint arXiv:2503.23616*.

Shuo Ye, Wen-Yuan Cui, Yin-Bi Li, A-Li Luo, and Hugh RA Jones. 2025. Deep learning interpretability analysis for carbon star identification in gaia dr3. *Astronomy & Astrophysics*, 697:A107.

Donald G York, Jennifer Adelman, John E Anderson Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, and 1 others. 2000. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579.

Hailong Yuan, Yanxia Zhang, Yue Wu, Yajuan Lei, Yiqiao Dong, Zongrui Bai, Guangwei Li, Haotong Zhang, and Yongheng Zhao. 2016. A team spectral inspection platform based on asera. *Proceedings of the International Astronomical Union*, 12(S325):320–323.

Shengwen Zhang, Yanxia Zhang, and Chao Liu. 2025a. A white dwarf catalog from lamost dr11 using deep learning. *The Astrophysical Journal Supplement Series*, 279(2):36.

Xintong Zhang, Zhi Gao, Bofei Zhang, Pengxiang Li, Xiaowen Zhang, Yang Liu, Tao Yuan, Yuwei Wu, Yunde Jia, Song-Chun Zhu, and 1 others. 2025b. Adaptive chain-of-focus reasoning via dynamic visual search and zooming for efficient vlms. *arXiv preprint arXiv:2505.15436*.

Yi-Fan Zhang, Xingyu Lu, Shukang Yin, Chaoyou Fu, Wei Chen, Xiao Hu, Bin Wen, Kaiyu Jiang, Changyi Liu, Tianke Zhang, and 1 others. 2025c. Thyme: Think beyond images. *arXiv preprint arXiv:2508.11630*.

Jiaxing Zhao, Xihan Wei, and Liefeng Bo. 2025a. R1-omni: Explainable omni-multimodal emotion recognition with reinforcement learning. *arXiv preprint arXiv:2503.05379*.

Shitian Zhao, Haoquan Zhang, Shaoheng Lin, Ming Li, Qilong Wu, Kaipeng Zhang, and Chen Wei. 2025b. Pyvision: Agentic vision with dynamic tooling. *arXiv preprint arXiv:2507.07998*.

Ziwei Zheng, Michael Yang, Jack Hong, Chenxiao Zhao, Guohai Xu, Le Yang, Chao Shen, and Xing Yu. 2025. Deepeyes: Incentivizing" thinking with images" via reinforcement learning. *arXiv preprint arXiv:2505.14362*.

Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, and 1 others. 2025. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*.

A Spectral Visualization Software

Professional spectroscopic surveys typically provide dedicated visualization software to support expert-driven spectral inspection. Figure 5 shows a representative example from the official LAMOST spectrum viewer, which is widely used by astronomers during catalog construction and verification.

The interface is organized into three functional panels. The left panel displays object-level metadata, including observational identifiers, coordinates, and acquisition parameters. The central panel renders the spectrum as a flux-wavelength plot, enabling experts to assess the global spectral morphology and identify prominent features. The right panel provides interactive controls for spectral analysis, including line annotations, smoothing options, data export, and wavelength-range selection.

A key capability of this software is the ability to re-render localized wavelength regions on demand. By specifying a wavelength interval through the range selection tool (highlighted in red), astronomers can zoom into diagnostically relevant regions and examine fine-grained spectral structures that may not be discernible from the global view alone. This process is typically performed iteratively: experts alternate between global inspection and multiple localized zooms to verify candidate-specific features before reaching a final decision.

The right panel of Figure 5 illustrates an example of such localized re-rendering, where the spectrum is restricted to the 6400-6700 Å range to facilitate detailed inspection of emission or absorption features in this region. This interactive, evidence-driven workflow closely reflects real-world expert practice and motivates the tool-augmented inspection paradigm adopted in Spec-o3.

B Spectral Visualization Tool

To let the VLM follow an astronomer-like, interactive inspection routine with spectral visualization software (i.e., repeatedly zooming into informative wavelength ranges), we implement a lightweight spectral visualization tool that can be invoked during inference via function calling. At the beginning of each inspection session, the tool caches the target's original one-dimensional wavelength-flux array and renders an initial spectrum plot spanning the full wavelength coverage as the starting view. In subsequent steps, the model may request a new visualization by specifying a wavelength range (e.g., $[\lambda_{\min}, \lambda_{\max}]$). The tool then slices the corresponding segment from the cached array, renders the localized spectrum view, and returns the resulting plot image to the model for the next reasoning step.

C SpecVI-Bench Statistics

This appendix summarizes the detailed statistics and data specification of SpecVI-Bench used throughout our experiments. SpecVI-Bench is designed to mirror the cataloging workflow in practice, where astronomers verify candidates one type at a time by inspecting spectra and deciding whether each spectrum should be accepted as the target class. Accordingly, we formulate SpecVI-Bench as five independent binary verification tasks, one for each rare type.

Despite its binary form, SpecVI-Bench remains challenging. Real survey spectra exhibit diverse noise patterns and instrumental artifacts, and many non-target objects can closely resemble the target types. As a result, reliable verification typically requires substantial domain experience, even for professional astronomers. Table 5 reports the number of positive and negative samples for each of the five rare-type spectral verification tasks, covering both training and test splits. For each task, positive samples are drawn from the official LAMOST catalogs, while negative samples are constructed via rejection sampling to approximate the high-confusion setting encountered during expert visual inspection. The training sets are approximately balanced to facilitate stable optimization, whereas the test sets are deliberately imbalanced, reflecting the extreme rarity of true targets in real survey scenarios.

Table 5: Statistics of SpecVI-Bench for the five rare-type verification tasks.

Type	**Train** (Pos / Neg)	**Test** (Pos / Neg)
CV	343 / 368	229 / 1229
CS	355 / 316	237 / 1084
SS	341 / 365	228 / 1214
MG	327 / 331	218 / 1103
WD	315 / 303	210 / 1002
Total	1681 / 1683	1122 / 5632

In addition to the binary label, each example in SpecVI-Bench provides the original 1D wavelength-flux arrays of the spectrum for on-demand visualization, together with a task prompt that specifies the verification question and the diag-

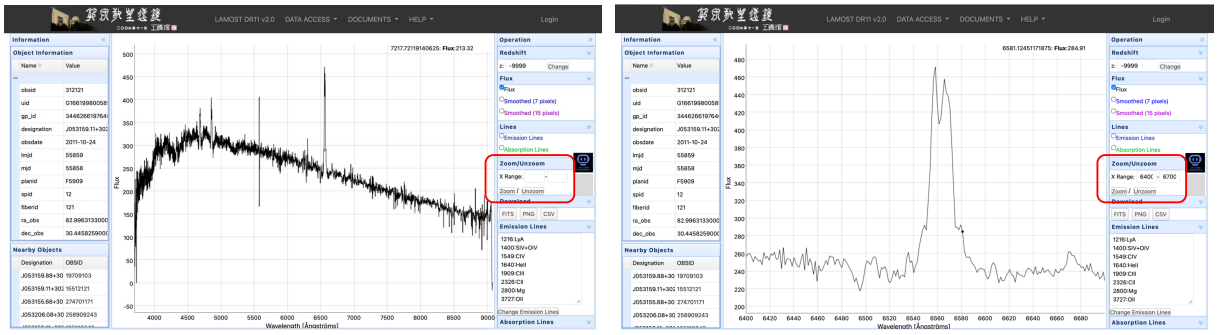

Figure 5: Screenshot of the official LAMOST spectrum viewer. The left panel lists object metadata, the center panel visualizes the spectrum, and the right panel provides interactive controls. The red box highlights the wavelength-range selection tool for zooming and re-rendering spectral details. Left: global spectrum view. Right: zoomed view re-rendered over 6400-6700 Å.

nostic criteria. An example prompt is provided in Table 6.

D Background on LAMOST, SDSS, and DESI

LAMOST The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; also known as the Guoshoujing Telescope) is a quasi-meridian reflecting Schmidt telescope located at the Xinglong Station of the National Astronomical Observatories of China.[2] It provides a 5° field of view and a highly multiplexed spectroscopic system with 4000 fibers feeding 16 spectrographs (250 fibers per spectrograph). In its low-resolution survey configuration, LAMOST covers 370–900 nm with typical resolving power $R = 500$–1500.[3]

SDSS. The Sloan Digital Sky Survey (SDSS) is a long-running program that has delivered large-scale spectroscopic datasets through multiple generations of public data releases.[4] In the original SDSS spectroscopic system, observations are taken using fiber plug plates that enable 640 spectra per exposure, with a wavelength coverage of ∼3800–9200 Å and a resolving power of about $R \approx 1800$. Later phases employ upgraded spectrographs (e.g., the BOSS spectrographs) and continue SDSS's emphasis on broad community access to calibrated spectra and derived products.[5]

DESI. The Dark Energy Spectroscopic Instrument (DESI) is a 5000-fiber multi-object spectrograph conducting the DESI Survey on the Mayall 4-meter telescope at Kitt Peak National Observatory.[6] According to the DESI instrument documentation, a wide-field corrector provides an ∼8 square-degree field of view, and the fibers feed 10 triple-arm spectrographs that simultaneously cover 360–980 nm. DESI is designed to obtain optical spectra for tens of millions of galaxies and quasars to build a 3D map of the Universe and constrain the physics of cosmic acceleration.[7]

E Generalization Evaluation

To evaluate the generalization ability of Spec-o3, we construct two types of datasets: cross-survey datasets and cross-task datasets.

Cross-Survey Datasets The cross-survey evaluation is conducted on spectra from SDSS and DESI, which differ from LAMOST in instrumentation, data reduction pipelines, and observational conditions. Starting from the test split of SpecVI-Bench samples constructed on LAMOST, we perform cross-matching based on sky coordinates to identify corresponding observations in the SDSS and DESI archives. A matching radius of 3 arcseconds is adopted, which is a commonly used tolerance for cross-survey astrometric matching. For each verification task, we retain approximately 50 positive samples and 250 negative samples after matching. Because we could not obtain enough matched samples for SS and WD, we only evaluate cross-survey generalization on CV, CS, and MG.

Cross-Task Datasets. To assess zero-shot generalization to unseen inspection tasks, we additionally construct cross-task datasets targeting O-, B-,

[2] https://www.lamost.org/public/instrument?locale=en
[3] https://www.lamost.org/public/instrument?locale=en
[4] https://www.sdss.org/
[5] https://www.sdss4.org/dr17/spectro/spectro_basics/
[6] https://www.desi.lbl.gov/
[7] https://www.desi.lbl.gov/

Table 6: Example prompt for a SpecVI-Bench sample. Here, <image> is a placeholder indicating that the model receives an input spectral image.

Prompt Example.

<image>

Analyze the provided spectrum to determine if it is a **"Cataclysmic Variables (CV)"**.

- **Key Indicators for CV (Check for either state):**
  - **State 1: Quiescent / Low-State (Emission-Dominated)**
    - **(Primary):** Strong and *very broad* Hydrogen Balmer emission lines (e.g., $H\alpha$ at 6563 Å, $H\beta$ at 4861 Å). The "broadness" (high velocity dispersion, often $\mathrm{FWHM} > 1000\ \mathrm{km\,s^{-1}}$) is the key diagnostic, indicating a high-velocity accretion disk.
    - **(Secondary):** Broad neutral Helium (He I) emission lines (e.g., 5876 Å, 6678 Å).
    - **(Key Confirmation):** Presence of high-excitation *ionized* Helium (He II) emission at 4686 Å. This is a strong indicator of accretion onto a white dwarf.
    - **(Line Profile):** Emission lines may exhibit a "double-peaked" profile, which is a classic signature of a rotating accretion disk viewed at high inclination.
  - **State 2: Outburst / High-State (Absorption-Dominated)**
    - **(Primary):** Spectrum is dominated by a bright, blue continuum (looks like a hot star).
    - **(Secondary):** The emission lines from State 1 are weak, absent, or "filled in", and are replaced by broad, shallow *absorption* lines (primarily Balmer and He I).
    - **(Morphology):** The overall spectrum mimics a hot B/A-type star. The crucial difference is that the absorption lines are significantly broader, shallower, and more "washed-out" (or "smeared") than the sharp lines of a normal stellar photosphere, due to high rotational speeds and pressure broadening in the disk.

Think first, call **spectral_visualization_tool** if needed to examine features in detail, then provide your final answer. Your response must adhere to the following strict rules:

1. **Overall Structure:** Your response must follow the format <think>...</think> <tool_call>...</tool_call> (if a tool is used) <answer>...</answer>.
2. **Tool Usage Constraint:** You may only make **one** tool call per turn.
3. **Final Answer Format:** In the <answer> tag, your conclusion must be inside a \boxed{} command (e.g., \boxed{YES}), followed by your justification.

and A-type spectra. Positive samples for each stellar type are drawn from the corresponding official catalogs released by LAMOST. Negative samples are collected from the same survey to preserve realistic background contamination. For each task, we similarly retain around 50 positive samples and 250 negative samples.

Across all settings, these datasets are used exclusively for evaluation. No samples from SDSS, DESI, or the O/B/A-type tasks are included during training, ensuring a strict zero-shot evaluation regime.

F Training Dynamics

Figure 6 illustrates the evolution of agent behavior during the RL stage. See Algorithm 1 for the rollout procedure used during agentic RL. We observe a distinct convergence pattern where task performance (Reward and F1 score) steadily improves, while the average number of tool calls gradually decreases throughout the training process. This trend suggests that as the model masters the task, it progressively learns to pinpoint high-value diagnostic features with fewer redundant observations. Additionally, Spec-o3-7B demonstrates faster convergence and superior stability compared to the 3B model, confirming that a stronger backbone enables more effective policy optimization.

G Human Evaluation Details

Expert Annotators Background We recruited six expert astronomers for the evaluation. To ensure rigorous verification, eligibility was strictly limited to individuals holding a Ph.D. in astronomy and possessing a track record of peer-reviewed publications on spectroscopic analysis. This prerequisite guarantees that all annotators command the deep domain expertise required for accurate physical interpretation and reliable verification.

Scoring Rubric Reasoning quality was evaluated on a 0–5 scale targeting coherence and physical consistency. The criteria, summarized in Table 7, span from invalid hallucinations to scientifically sound interpretations. This metric rigorously captures both the factual correctness of the answer and the internal logic of the reasoning process.

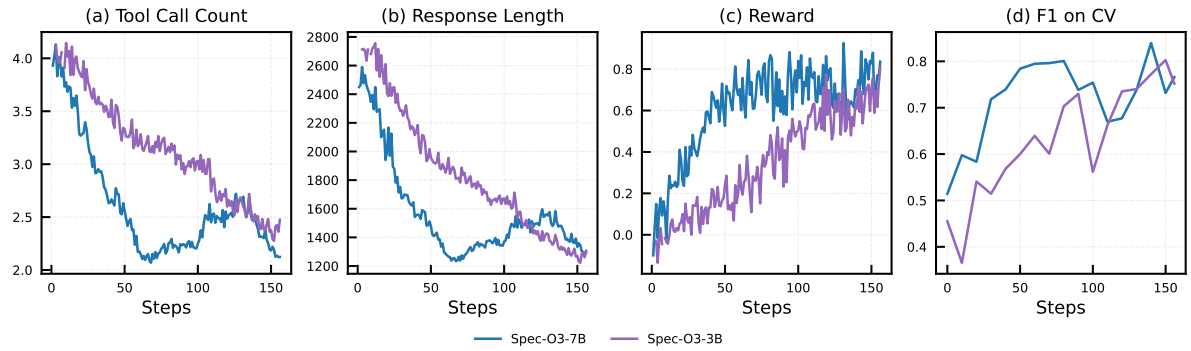

Figure 6: Training dynamics of Spec-o3-3B and Spec-o3-7B during agentic rl. Panels show (a) average number of tool calls per trajectory, (b) average response length, (c) average reward, and (d) F1 score on the CV verification set as training proceeds.

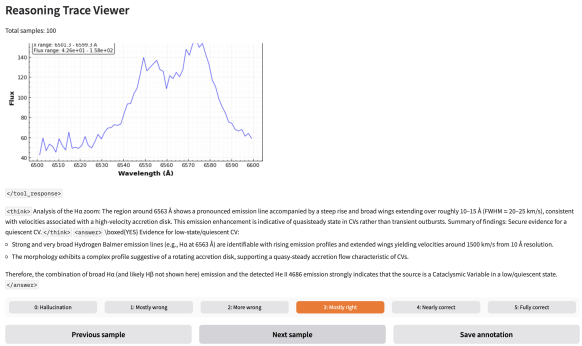

Figure 7: Custom annotation interface for expert evaluation. The interface includes a scrollable central panel that displays the full, interleaved reasoning trajectory and tool outputs. Experts review the complete chain of thought before assigning a coherence score (0-5) via the bottom control bar.

Annotation Interface To streamline the evaluation process, we developed a custom annotation interface shown in Figure 7. The interface features a scrollable reasoning viewer designed to accommodate the model's full multi-turn trajectory. This allows experts to seamlessly inspect the complete history of interleaved textual analysis and tool-generated spectral plots within a unified view. The bottom panel provides standardized controls for navigation and scoring, ensuring a consistent workflow across all annotators.

Table 7: Human evaluation rubric for reasoning coherence and physical consistency.

Score	Definition
0	Complete hallucination. The explanation is entirely incorrect or physically implausible.
1	Mostly wrong. The majority of the explanation is incorrect or misleading.
2	More wrong than correct. Some valid elements exist, but errors dominate the reasoning.
3	Mostly correct. The main reasoning is sound but contains noticeable errors.
4	Nearly correct. The explanation is largely accurate with only minor issues.
5	Fully correct. A coherent, accurate, and physically consistent description.

Algorithm 1 Rollout for agentic RL

Require: Spectrum array S (wavelength–flux), task prompt T_0, policy model π_θ, visualization tool $\mathcal{V}$, max steps T

Ensure: Trajectory τ for RL update

$\mathcal{C} \leftarrow \textsc{CacheSpectrum}(S)$ ▷ store the raw numeric array for this session

$I_0 \leftarrow \mathcal{V}(\mathcal{C}, \textsc{FullRange}, \textsc{label} = \varnothing)$

$\tau \leftarrow [(T_0, I_0)]$

for $t \leftarrow 0$ to $T - 1$ **do**

  $y_t \sim \pi_\theta(\tau)$ ▷ generate next model output conditioned on history

  if $\textsc{IsToolCall}(y_t)$ **then**

    $(\lambda_{\min}, \lambda_{\max}, \ell) \leftarrow \textsc{ParseArgs}(y_t)$

    $I_{t+1} \leftarrow \mathcal{V}(\mathcal{C}, [\lambda_{\min}, \lambda_{\max}], \textsc{label} = \ell)$

    $\tau \leftarrow \tau \cup [(y_t, I_{t+1})]$ ▷ interleave tool call and returned visualization

  else if $\textsc{IsFinalAnswer}(y_t)$ **then**

    $\tau \leftarrow \tau \cup [(y_t)]$

    break

  else

    $\tau \leftarrow \tau \cup [(y_t)]$ ▷ pure text reasoning step (optional)

  end if

end for

$r \leftarrow \textsc{ComputeOutcomeReward}(\tau)$ ▷ e.g., correctness and format constraints

return (τ, r)