LASER: A Corrective Lens for LVLMs via Visual Attention Preservation and Sink Suppression

Bowen Yuan, Zijian Wang, Yadan Luo, Shijie Wang, and Zi Huang

The University of Queensland
{bowen.yuan, zijian.wang, y.luo, shijie.wang, helen.huang}@uq.edu.au

Abstract. Large vision–language models (LVLMs) exhibit strong reasoning ability but suffer from visual forgetting during long-horizon decoding, where attention progressively drifts away from visual evidence. Existing methods largely treat this issue as a late-stage attention decay problem or attempt to mitigate it through heuristic reminders or post-hoc attention lifting. Through systematic empirical analysis, we find that performance degradation under visual forgetting is largely driven by two overlooked factors: early-stage attention decay disrupts evidence acquisition, and attention concentration on a subset of task-irrelevant visual sink tokens. Motivated by these insights, we propose LASER, a post-training framework that regulates both the visual attention trajectory and intra-visual token attention distribution during reasoning. Technically, LASER introduces two complementary rewards: a Visual Grounding Reward, which encourages the model to maintain attention on semantically salient visual tokens throughout decoding, and a Sink Suppression Reward, which penalizes excessive attention concentration on visual sink tokens. Together, these rewards preserve early-stage grounding while preventing attention collapse onto uninformative regions. Extensive experiments on eight benchmark datasets demonstrate that LASER consistently outperforms strong baselines, validating attention-aware training as an effective remedy for visual forgetting. The code is available at `https://github.com/KeViNYuAn0314/LASER`.

Keywords: Large Vision Language Models · Visual Forgetting · Chain-of-Thought Reasoning

1 Introduction

Reasoning has emerged as a transformative capability in large language models (LLMs), enabling LLMs to solve complex tasks by decomposing problems into explicit steps before producing a final answer [22, 27, 28, 48, 53, 67, 68]. This paradigm has been driven by reinforcement learning (RL) [9, 22]. Rather than relying on supervised CoT data, RL incentivizes LLMs to develop sophisticated cognitive behaviors, including self-reflection, verification, and "aha" moments. Building on this success, recent efforts have extended RL-based reasoning to Large Vision Language Models (LVLMs), aiming to unlock analogous capabilities in the vision domain [16, 21, 41, 51, 60]. LVLMs are trained to generate detailed

chain-of-thought (CoT) rationales that jointly attend to both visual and textual information [45, 58, 65], yielding stronger problem-solving capabilities and improved generalization across challenging benchmarks [8, 11, 35].

A notable phenomenon that emerges alongside extended visual reasoning is *visual forgetting*: as generation progresses, LVLMs progressively reduce attention to visual tokens and increasingly rely on self-generated textual context. Recent studies [19, 33, 52, 61] have linked this attention decay to performance degradation in long-horizon visual reasoning tasks. To mitigate this issue, prior work has explored two main strategies: (i) post-training methods that directly encourage sustained visual attention in the later decoding stage [23], and (ii) introducing hand-crafted [42] or learned visual reminder [46, 59] stages that either re-inject visual inputs (e.g., image replay or token reactivation) or prompt the model to re-focus attention on visual components during extended reasoning.

While these strategies yield empirical improvements, they primarily intervene at isolated stages without regulating attention dynamics across the reasoning trajectory. Through comprehensive empirical analysis, we argue that most of the existing works overlook two critical dimensions of preventing visual forgetting. First, the temporal sensitivity of grounding *(When)*. Our analysis reveals that the degradation often begins in early decoding stages, where visual evidence establishes the semantic scaffold for subsequent reasoning. Errors introduced at this stage would propagate autoregressively and amplify throughout the trajectory. Second, the distributional allocation of attention *(Where)*. We find that even when total visual attention is preserved, attention can collapse onto a small subset of persistent sink tokens: semantically uninformative tokens that absorb disproportionate attention mass across decoding steps. Such concentration restricts effective visual grounding. These observations suggest that mitigating visual forgetting requires both temporal preservation starting from early-stage grounding and explicit regulation of intra-visual attention allocation.

To address the limitations, we propose LASER, an **L**VLM **A**ttention-aware **S**ight-**E**nhanced **R**easoning framework built on Group Relative Policy Optimization (GRPO) [9]. LASER explicitly regulates both the temporal and distributional dynamics of visual attention during training. It introduces a two-component visual attention reward. The first component, a grounding preservation term, sustains early-stage non-sink visual attention and prevents premature disengagement from visual evidence. The second component, a sink suppression term, discourages disproportionate concentration on persistent sink tokens, ensuring that preserved visual attention is redistributed toward informative content. Together, these components provide a principled solution to visual forgetting in extended multimodal reasoning. The contributions of this work are threefold:

- We provide a causal analysis of visual forgetting in LVLMs, showing that performance degradation originates in early-stage grounding and that attention collapses onto persistent sink tokens, with magnitude-preserving strategies often amplifying this imbalance.

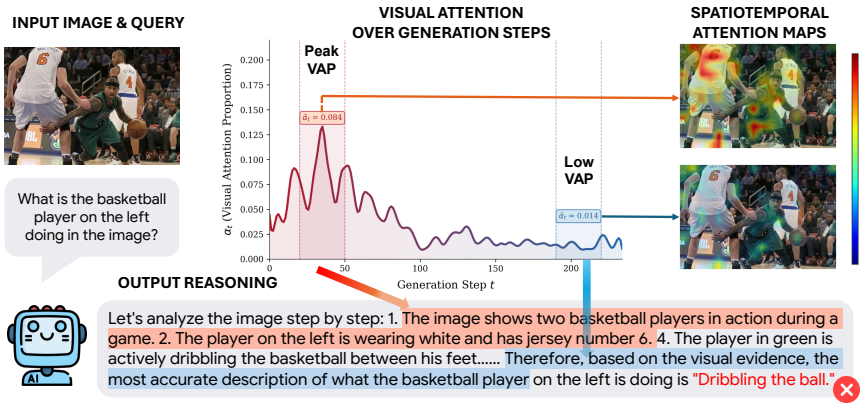


Fig. 1: Illustration of visual forgetting during long-horizon reasoning in LVLMs. *Left:* an input image and the corresponding question. *Middle:* the Visual Attention Proportion (VAP) over generation steps shows that attention to visual tokens peaks early and progressively decays during reasoning. *Right:* difference in attention maps reveals that later decoding stages could attend less to task relevant visual regions.

– We introduce LASER, a trajectory-level attention shaping framework that jointly regulates the temporal dynamics and intra-visual allocation of attention during long-horizon reasoning.
– LASER consistently outperforms strong baselines across benchmarks, achieving 1.9% and 5.7% improvements on visual reasoning and visual perception tasks, respectively. Post-hoc attention analysis shows that the improvements are accompanied by sustained visual grounding throughout generation, validating the effectiveness of our attention-centric training framework.

2 Related Work

2.1 Reasoning in Large Vision-Language Models

Large Vision-Language Models (LVLMs) [1–3, 31, 32] extend large language models (LLMs) with visual perception by integrating image representations into the language generation process. Given the increasing requirement for handling complex, multi-step inference tasks, eliciting LVLM reasoning capabilities has become paramount. Early works address this through chain-of-thought (CoT) prompting [20, 36, 66] or supervised fine-tuning (SFT) on curated multimodal reasoning traces [12, 45, 58]. While effective, SFT-based methods rely on fixed reasoning demonstrations and promote imitation, which can limit generalization to unseen reasoning patterns and complex multimodal scenarios.

To address these limitations, recent works draw inspiration from reinforcement learning (RL)–based reasoning in LLMs, notably DeepSeek-R1 [9], which

demonstrates that Group Relative Policy Optimization (GRPO) [40] can elicit strong reasoning behaviors through outcome-driven optimization. Building on this paradigm, several studies [10, 41, 49, 55, 60, 62] extend RL-based training to LVLMs, enabling models to explore diverse reasoning trajectories and learn from outcome-based feedback. Specifically, Vision-SR1 [29] introduces a self-rewarding framework that decouples visual perception from language reasoning, while VL-Rethinker [49] addresses optimization challenges in large-scale RL via selective sample replay and forced rethinking mechanisms to promote explicit self-reflection.

2.2 Visual Forgetting

During extended response generation, LVLMs exhibit a notable degradation in visual grounding. Attention to visual tokens progressively diminishes, causing later responses to rely increasingly on language priors [5, 39]. This phenomenon, referred to as *visual forgetting*, has been shown to correlate with the emergence of hallucinations. Training-free techniques operate at test time by explicitly re-emphasizing visual information, such as contrastive or vision-aware decoding strategies [15, 26], or by reallocating attention mass toward visual tokens through attention manipulation [17, 47]. Instruction fine-tuning approaches [42, 54] focus on preserving visual representations during text-heavy instruction tuning to reduce modality imbalance introduced by language-dominant supervision. RL-based methods [23, 46] encourage visually grounded trajectories via reward design, to strengthen visual reflection during response generation.

3 Diagnosing Visual Forgetting in LVLM Reasoning

Visual forgetting is a structural vulnerability for LVLMs during the long-horizon reasoning: as generation length increases, attention to visual evidence progressively decays [23, 42, 46]. However, existing characterizations primarily document the overall decay of visual attention, without examining its temporal sensitivity or internal allocation patterns. To uncover these finer-grained dynamics, we re-examine attention trajectories throughout the reasoning process. Through case studies and attention trajectory analysis (Fig. 1), we observe that this decay is not uniform throughout the reasoning process. Early decoding steps exhibit a summarization-like role, where visual information is consolidated into intermediate textual representations. When visual

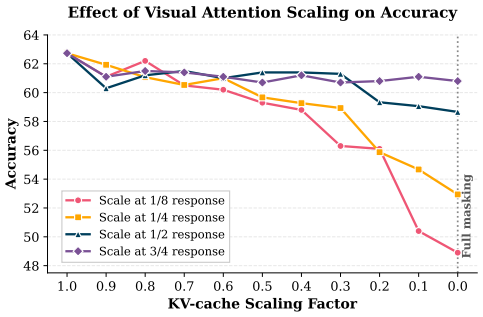


Fig. 2: Accuracy on MMStar under visual KV-cache scaling factor at various reasoning stages.

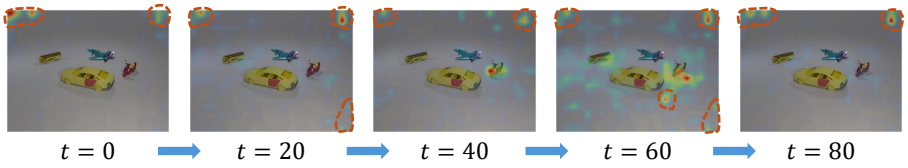


Fig. 3: Visualizations of attention maps over reasoning steps. We identify image regions that receive high attention score despite lacking semantic relevance to the task.

grounding errors happen at this stage, errors in evidence acquisition propagate autoregressively throughout the remaining reasoning process. At the same time, we observe that the visual attention that persists is not always directed toward semantically relevant regions; a portion concentrates on visually irrelevant areas, indicating suboptimal allocation. These observations motivate a deeper investigation into two fundamental questions: *when* this decay most impacts reasoning performance, and *where* attention should be distributed across visual tokens to enable effective grounding.

When is visual forgetting more consequential? To test the causal impact of visual attention decay, we perform controlled decoding-time interventions by applying a one-time scaling of the value states of visual tokens in the KV cache by a factor κ at specific decoding steps. Importantly, this intervention is applied only once at the designated stage and does not affect subsequent KV updates. Because value states govern the contribution of visual tokens to downstream decoding, reducing $\kappa < 1$ transiently suppresses visual influence at that moment. We evaluate this intervention on MMStar [6], a benchmark requiring sustained visual reasoning. As shown in Fig. 2, even a single early-stage suppression produces disproportionately large performance degradation. When full masking is applied at the $\frac{1}{8}$ stage, accuracy drops from 62.8% to 48.9%, whereas an identical one-time intervention at the $\frac{3}{4}$ stage results in only a 2.0% decrease. The degradation at $\frac{1}{8}$ is over three times larger than at $\frac{1}{2}$ and nearly seven times larger than at $\frac{3}{4}$. Even under moderate suppression ($\kappa = 0.5$), early-stage intervention induces a 3.7% absolute drop, compared to only 1.3% at later stages. This pronounced asymmetry indicates that early-stage visual grounding plays a decisive role in forming intermediate reasoning states. Notably, despite being a one-time perturbation, early suppression induces persistent performance degradation, demonstrating that mis-grounded representations formed early continue to condition the entire trajectory.

> *Finding 1*: Visual forgetting exhibits strong early-stage sensitivity: early-stage attention decay causally impairs downstream reasoning through autoregressive propagation.

Where should visual attention be allocated? To investigate this question, we analyze the distribution of visual attention across individual tokens. As shown

in Fig. 3, we observe the emergence of a persistent subset of visual tokens that consistently receive disproportionately high per-token attention across generation steps. We refer to them as visual sink tokens $\mathcal{S} \subset \mathcal{V}$: a phenomenon [24, 30, 56] where certain tokens absorb disproportionate attention mass independent of semantic contribution. Quantitatively, visual sink tokens receive 2–3× more per-token attention than the visual-token average throughout generation (Fig. 4, right), and this concentration emerges as early as the initial decoding stages.

To assess the causal impact of distributional allocation, we compare two inference-time interventions at each scaling factor. Uniform scaling proportionally reduces the value states of all visual tokens, preserving their relative attention distribution. In contrast, redistributed scaling preserves the same total visual attention mass but reallocates attention from sink tokens to non-sink tokens. In Fig. 4 left, redistributed scaling achieves better results than uniform scaling across suppression levels. This result demonstrates that reasoning performance depends not only on the magnitude of visual attention but critically on its allocation across visual tokens. Simply preserving total attention mass is insufficient; effective grounding requires directing attention toward informative visual content.

Finding 2: Visual attention is distributionally imbalanced: concentration on persistent sink tokens degrades reasoning performance even when total visual attention is preserved.

Together, our findings characterize visual forgetting as a dual pathology: magnitude decay, in which global visual attention diminishes over the generation horizon, and distributional collapse, in which the remaining attention disproportionately concentrates on semantically uninformative sink tokens. Our controlled inference-time interventions demonstrate that both pathologies causally impair reasoning performance. However, such interventions operate only at inference and do not alter the model's learned attention dynamics. As a result, they cannot provide a persistent solution that enforces reliable grounding behavior. These observations motivate a training-time approach that explicitly regulates both the temporal evolution and distributional allocation of visual attention.

4 Methodology

Motivated by our findings of visual forgetting that stem from (i) insufficient visual attention and (ii) attention collapse to visual sink tokens, we propose LASER, a post-training framework that explicitly reshapes visual attention dynamics during GRPO training. Our approach introduces two complementary rewards: *visual grounding reward* that sustains attention to informative visual tokens, and *sink suppression reward* that penalizes excessive attention to uninformative sink tokens. The overall framework is illustrated in Fig. 5.

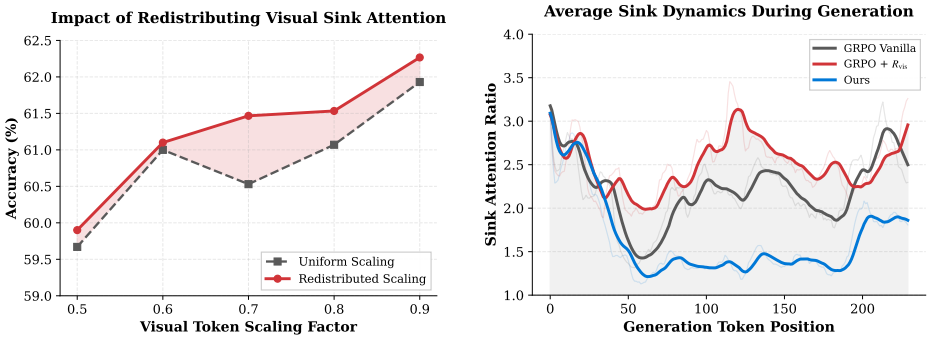


Fig. 4: Analysis of visual sink token attention. **Left.** Redistributing attention away from sink tokens yields consistent accuracy gains over uniform scaling. **Right.** Visual sink attention ratio across generation steps for GRPO vanilla vs. GRPO + R_{vis}.

4.1 Preliminaries

We build on Group Relative Policy Optimization (GRPO) [40], which eliminates the value function and estimates advantages by comparing rewards within groups of sampled outputs. In the multimodal setting, each training instance consists of a triplet (v, q, a) comprising a visual input v, a textual question q, and a ground-truth answer a. The policy π_θ samples a group of G independent responses $\{o_i\}_{i=1}^{G}$. Each response o_i receives a scalar reward R_i, and the advantage is computed via group-level normalization:

$$\hat{A}_i = \frac{R_i - \text{mean}\big(\{R_j\}_{j=1}^{G}\big)}{\text{std}\big(\{R_j\}_{j=1}^{G}\big)}. \tag{1}$$

GRPO updates the policy by maximizing the clipped surrogate objective with a KL penalty:

$$\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{(v,q,a)\sim\mathcal{D},\, \{o_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(\cdot|v,q)} \left[\frac{1}{G}\sum_{i=1}^{G} \frac{1}{|o_i|} \sum_{t=1}^{|o_i|} \left(\min\Big(\rho_{i,t}(\theta)\,\hat{A}_i,\ \text{clip}\big(\rho_{i,t}(\theta), 1-\epsilon, 1+\epsilon\big)\hat{A}_i\Big) - \beta\, D_{\text{KL}}(\pi_\theta \| \pi_{\text{ref}}) \right) \right], \tag{2}$$

where $\rho_{i,t}(\theta) = \pi_\theta(o_{i,t} \mid v, q, o_{i,<t}) \,/\, \pi_{\theta_{\text{old}}}(o_{i,t} \mid v, q, o_{i,<t})$ is the importance sampling ratio, and π_{ref} is the reference policy.

4.2 Identifying Visual Sink Tokens

We establish a principled criterion to identify visual sink tokens based on the massive activation patterns observed in hidden states [24, 43]. Our identification

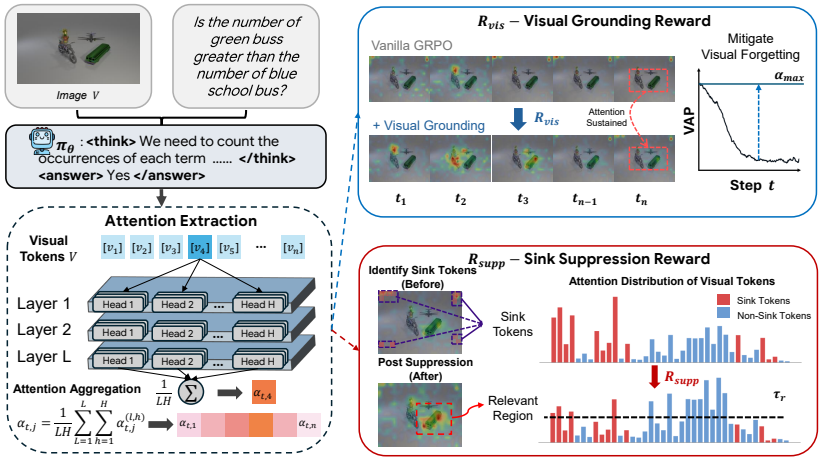


Fig. 5: Overview of LASER. Visual token attention is extracted during generation to compute two complementary rewards: R_{vis} sustains overall VAP across reasoning steps, while R_{supp} identifies and suppresses sink tokens to rectify visual attention from visual sink tokens.

process consists of two stages. First, we isolate a subset of massive activation dimensions $\mathcal{D}^* \subseteq [D]$, where $[D]$ denotes the full set of hidden state dimensions. Specifically, a dimension is included in $\mathcal{D}^*$ if its mean absolute activation magnitude, averaged across all visual tokens and layers, anomalously exceeds the average magnitude over all dimensions. Subsequently, a visual token j is classified as a sink token if its mean activation magnitude within $\mathcal{D}^*$ surpasses a predefined threshold η. Formally, the set of sink tokens $\mathcal{S}$ is defined as:

$$\mathcal{S} = \left\{ j \in \mathcal{V} \mid \frac{1}{|\mathcal{D}^*|} \sum_{d \in \mathcal{D}^*} \left| h_{j,d}^{(l^*)} \right| > \eta \right\}, \tag{3}$$

where $h_{j,d}^{(l^*)}$ denotes the d-th component of the hidden state for token j at the final layer l^*, and η is determined by the top-k percentile over all visual tokens $\mathcal{V}$. The remaining non-sink tokens are denoted as $\mathcal{V}' = \mathcal{V} \setminus \mathcal{S}$.

4.3 Visual Grounding Reward

To counteract the attention decay identified in **Finding 1**, we introduce a Visual Grounding Reward R_{vis} that incentivizes the model to maintain focus on semantically-salient visual tokens. Specifically, R_{vis} is formulated over the non-sink token set $\mathcal{V}'$ rather than the full set $\mathcal{V}$. This design prevents the reinforcement of uninformative sink activations, thereby ensuring the grounding signal remains concentrated on task-critical regions and mitigating error propagation.

Mathematically, we first quantify the visual attention allocated to semantically-salient regions. At each generation step t, the attention proportion restricted to the non-sink token set $\mathcal{V}'$ is defined as:

$$\alpha'_t = \frac{1}{LH} \sum_{l=1}^{L} \sum_{h=1}^{H} \sum_{j \in \mathcal{V}'} a_{t,j}^{(l,h)}, \tag{4}$$

where L and H denote the number of transformer layers and attention heads, $a_{t,j}^{(l,h)}$ is the attention weight from the generated token at step t to visual token j at layer l and head h.

To mitigate the premature disengagement from visual evidence, we formulate the visual grounding reward R_{vis} to penalize deviations from the peak attention level observed during the sequence. The reward is assigned as:

$$R_{\text{vis}} = \sum_{t=1}^{T} \exp\left(-\lambda \left(1 - \frac{\alpha'_t}{\alpha'_{\max}}\right)\right) + \beta, \tag{5}$$

where $\alpha'_{\max} = \max_t \alpha'_t$ represents the peak non-sink attention proportion, and $\lambda > 0$ governs the sensitivity to attention decay. Furthermore, we introduce β as a penalty term to discourage reward hacking via excessively redundant reasoning sequences, thereby ensuring both grounding quality and conciseness.

4.4 Visual Sink Suppression Reward

While R_{vis} successfully enhances the attention weights of non-sink tokens, it remains insufficient to override the inherent attentional dominance of sink tokens. As identified in **Finding 2**, a specific subset of sink tokens exhibits excessively high activations; despite the relative gains in non-sink attention, these sink tokens continue to attract an excessive amount of the model's focus, thereby interfering with the reasoning process. Hence, we introduce the Sink Suppression Reward R_{supp} as a necessary constraint. Unlike R_{vis}, which merely promotes non-sink visual regions, R_{supp} explicitly penalizes the absolute concentration on sink tokens.

We measure this concentration as the ratio of the average per-token attention within the sink set $\mathcal{S}$ relative to the entire visual set $\mathcal{V}$:

$$\bar{a}_{\mathcal{S}} = \frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} a_j, \qquad \bar{a}_{\mathcal{V}} = \frac{1}{|\mathcal{V}|} \sum_{j \in \mathcal{V}} a_j, \tag{6}$$

where a_j denotes the attention weight of token j averaged across all generation steps, layers, and heads. Specifically, R_{supp} is formulated to suppress responses where this ratio exceeds a tolerance threshold τ:

$$R_{\text{supp}} = \exp\left(-\gamma \cdot \max\left(0,\ \frac{\bar{a}_{\mathcal{S}}}{\bar{a}_{\mathcal{V}}} - 1\right)\right), \tag{7}$$

where $\gamma > 0$ governs the penalty intensity. We set a constraint that sink tokens should not, on average, receive more attention than standard visual tokens.

4.5 Combined Reward Objective

R_{vis} and R_{supp} work in harmony: the former encourages sustained attention to semantically-salient regions, while the latter suppresses attention to uninformative sink tokens. Visual rewards are only granted if the model's answer is correct. This gating mechanism ensures that the model is rewarded for meaningful grounding rather than just increasing attention patterns on incorrect reasoning paths. Incorporating the format check into the accuracy reward R_{acc} as a necessary condition for correctness, the total reward is:

$$R = R_{\text{acc}} + \omega \cdot \mathbf{1}_{[R_{\text{acc}}>0]} \cdot \big(R_{\text{vis}} + R_{\text{supp}}\big), \tag{8}$$

where $\omega > 0$ is a single scalar that weights the attention-shaping signals.

5 Experiment

5.1 Experimental Setup

Implementation Details. We use open-source LVLM Qwen-2.5-VL-7B-Instruct to evaluate our method, which is consistent with baseline methods. For cold-start, we train our model for 2 epochs following [7]. The model is subsequently trained using GRPO with visual grounding reward and suppression reward for 2 epochs, based on verl training framework [14]. We source and filter our training data from the established [25] and [7], resulting in 45K RL training samples. All experiments are conducted on 4 NVIDIA H100 GPUs. All experiments are evaluated using VLMEvalKit [13] under identical inference settings with temperature 0.6 and top-p 0.95. The detailed composition of training data and training hyperparameters is provided in supplementary materials.

Benchmark Datasets. We evaluate our model on 8 benchmarks to comprehensively assess visual reasoning capabilities across diverse tasks. These benchmarks span diverse domains including mathematical reasoning, general multimodal understanding, and visual perception:

- Math Reasoning: MathVista [34], MathVision [50], MathVerse [64], We-Math [38], which evaluate visual math problem-solving abilities.
- General Reasoning: MMMU [63], MMStar [6], LogicVista [57], which assess multimodal understanding and logical inference across multiple disciplines.
- Visual Perception: HallusionBench [18], which diagnoses language hallucination and visual illusion by testing model robustness to image manipulations through edited image pairs.

Baseline Methods. We compare our model to a diverse set of existing reasoning models. For proprietary models, we include GPT5 [37] and Gemini-2.5-Pro [44] to represent the current state of the art in multimodal reasoning. We also include open-source LVLMs including InternVL2.5 [4] and Qwen2.5-VL [2]. Additionally, we benchmark against recent reasoning-enhanced models: VisionR1 [21], R1-Onevision [60], OpenVLThinker [10]. Among these, Reflection-V [23] and VAPO-Thinker [46] are specifically designed to address visual forgetting in LVLMs.

Table 1: Performance comparison across various visual reasoning benchmarks. Best results are highlighted in **red bold**, and the second-best results are highlighted in blue underlining.

Model	Math Reasoning				General Reasoning			Visual
	MathVista	MathVerse	MathVision	WeMath	MMMU	MMStar	LogicVista	Hallu-Bench
Proprietary Model								
GPT5 [37]	81.9	81.2	72.0	71.1	85.4	75.7	70.0	65.2
Gemini-2.5-Pro [44]	80.9	76.9	69.1	78.0	84.0	73.6	73.8	64.1
Open-source Vision-Language Model								
InternVL2.5-8B [4]	68.2	34.3	25.6	38.6	49.0	60.8	38.3	49.0
Qwen2.5-VL-7B [2]	66.7	40.7	26.4	33.1	52.7	54.9	42.6	53.6
VisionR1-7B [21]	73.5	52.4	28.2	41.6	57.6	61.4	49.7	49.5
R1-Onevision-7B [60]	64.1	46.4	29.9	**44.6**	54.3	54.1	45.6	52.5
OpenVLThinker-7B [10]	72.3	50.3	25.9	-	-	61.9	-	52.2
Reflection-V-7B [23]	73.3	-	33.9	-	61.3	-	-	53.9
VAPO-Thinker-7B [46]	**75.6**	53.3	31.9	43.6	60.2	63.0	50.9	57.4
LASER (Ours)	72.9	**55.2**	**44.6**	44.4	**62.1**	**64.1**	**51.2**	**60.7**

We designate Qwen2.5-VL-7B as our primary baseline to directly quantify the improvements introduced by our approach. Unless otherwise noted, baseline results are taken from the original publications; we reproduce results only when official numbers are unavailable.

5.2 Main Result

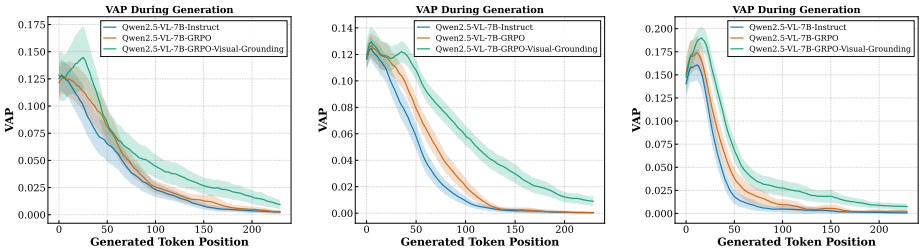
As demonstrated in Tab. 1, LASER achieves consistent performance improvements across vision-centric, hallucination, and visual reasoning benchmarks, validating the effectiveness of jointly regulating the temporal and distributional dynamics of visual attention. Specifically, LASER shows a substantial 3.3% absolute gain on HallusionBench and reaches a new best of 64.1 on MMStar, outperforming specialized methods targeting visual forgetting [21, 46]. LASER shows its strongest advantage in complex scenarios like MathVision, outperforming the closest competitor by a large margin of 12.7%. The results show that these improvements are not merely incremental. The synergy between R_{vis} and R_{supp} ensures that task-relevant visual evidence remains active as a semantic scaffold during reasoning, while preventing attention from being dominated by uninformative sink tokens. By strengthening visual attention and maintaining sustained focus on visual inputs, our method is particularly effective on perceptually demanding tasks, demonstrating more grounded and reliable visual reasoning.

5.3 Ablation Study

We investigate the individual contributions of each reward within the GRPO framework by constructing incremental variants on the Qwen2.5-VL-7B base.

Table 2: Ablation study with various VQA tasks on math reasoning, general visual reasoning, and visual perceptual tasks. Best results are highlighted in **red bold**, and the second-best results are highlighted in blue underlining.

Model	GRPO	R_{vis}	R_{supp}	Math	General	Visual	Avg.
Qwen2.5-VL-7B [2]				41.7	50.1	53.5	48.4
	✓			44.0	54.7	54.6	51.1
	✓	✓		45.7	55.1	55.8	52.2
	✓	✓	✓	**46.0**	**58.6**	**56.6**	**53.7**
Revisual-R1 [7]				48.7	53.4	54.5	52.2
	✓			52.1	57.7	59.7	56.5
	✓	✓	✓	**54.3**	**59.1**	**60.7**	**58.0**



(a) MathVista **(b)** HallusionBench **(c)** MMStar

Fig. 6: Analysis of VAP during generation on various datasets.

As reported in Tab. 2, both R_{vis} and R_{supp} yield consistent performance gains across all benchmarks. Specifically, integrating R_{vis} improves the average score from 51.1 to 52.2 by reducing visual forgetting and maintaining attention on task-relevant regions. The subsequent addition of R_{supp} further improves performance, particularly in general reasoning, increasing the score from 55.1 to 58.6 by reducing attention to uninformative sink tokens. These results demonstrate that while R_{vis} ensures visual persistence throughout the reasoning trajectory, R_{supp} optimizes the distributional quality of attention, collectively ensuring that visual evidence is redirected toward task-critical evidence.

To further demonstrate the effectiveness of our attention-shaping rewards, we evaluate their impact when integrated with the Supervised Fine-Tuning (SFT) base model, Revisual-R1. As observed in Tab. 2, LASER continues to provide substantial performance enhancements, raising the overall average from 52.2 to 58.0. These results show that controlling when and where the model pays visual attention improves performance on the cold-start model, demonstrating the robustness and general applicability of our approach in complex reasoning tasks.

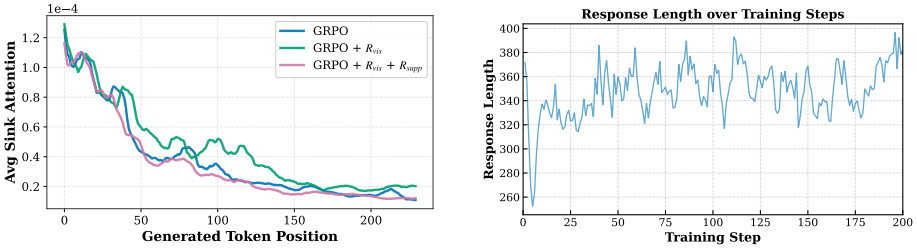


Fig. 7: Left. Visual sink token attention across generation steps for model trained by LASER. **Right.** Response length during training process.

5.4 Further Analysis

LASER preserves visual attention throughout generation. Fig. 6 presents the VAP α'_t across generation steps on MathVista, HallusionBench, and MMStar for three models: the Qwen2.5-VL-Instruct model, Qwen2.5-VL with GRPO and Qwen.2.5-VL trained by our method. Across all benchmarks, the base model and vanilla GRPO exhibit lower VAP and sharp decay throughout generation, suggesting that standard RL training with outcome-based rewards provides no corrective signal for visual forgetting. In contrast, our method consistently sustains higher VAP throughout the generation process, maintaining visual attention beyond the early generation stages where the base model and vanilla GRPO have already disengaged from visual input.

LASER reduces disproportionate attention to visual sink tokens. To verify our method effectively suppresses attention allocated to visual sink tokens, we compare the average sink token attention and sink attention ratio across GRPO, GRPO with R_{vis}, and GRPO with both R_{vis} and R_{supp} (Full LASER). As shown in Fig. 7 left, while R_{vis} encourages the model to maintain higher overall attention on visual tokens, it simultaneously inflates attention on visual sink tokens relative to baseline. Incorporating R_{supp} effectively mitigates this side effect, yielding visual sink token attention even lower than baseline throughout generation. Fig. 4 right further corroborates this through measuring the proportion of visual attention allocated to sink tokens relative to all visual tokens. The model trained with R_{supp} maintains consistently lower sink attention ratios, indicating that it redistributes attention away from semantically uninformative tokens toward more meaningful visual regions.

Reasoning quality is maintained without length inflation or visual re-injection. Fig. 7 right shows the average response length over training on Qwen2.5-VL-7B-Instruct model. Our method maintains response lengths comparable to the base model throughout training, without being prompted to generate extended chain-of-thought reasoning. This reflects that model attends more effectively to visual evidence throughout generation without incurring additional inference cost.

Sustained visual grounding improves perception accuracy. Fig. 8 presents a representative example illustrating the improvements of our method. In this

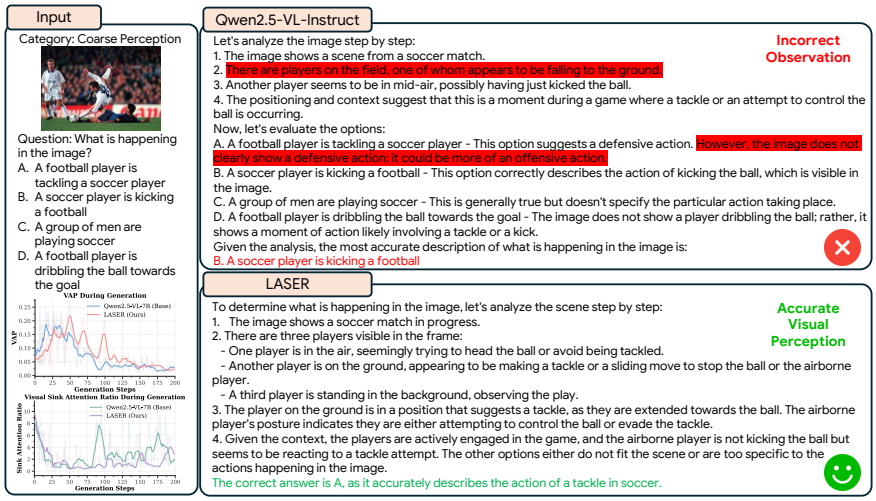


Fig. 8: A qualitative comparison between Qwen2.5-VL-Instruct and LASER on a visual perception task. Premature visual disengagement in the base model leads to an erroneous inference, while LASER maintains visual attention and answers correctly.

example, the base Qwen2.5-VL-Instruct model makes an erroneous summarization in its early-stage reasoning chain, resulting in an incorrect answer after long-horizon reasoning. In contrast, the model trained with LASER maintains fine-grained visual engagement throughout generation, and reasons coherently to the correct answer. This contrast highlights that the base model's error is not attributed to the lack of reasoning power but rather to focusing on visual evidence. By sustaining attention on semantically meaningful visual regions across generations, our method reduces the risk of hallucinated or misattributed observations.

6 Conclusion

In this work, we investigate the problem of visual forgetting in LVLMs, where models steadily disengage from visual evidence during extended reasoning. Through empirical analysis, we demonstrate that the onset of visual attention decay is most harmful, posing a critical impediment to reasoning performance. Beyond attention decay, we reveal a structural flaw in visual forgetting: a subset of visual sink tokens persistently dominates attention in task-irrelevant areas. Motivated by these findings, we propose LASER, a GRPO-based training framework that mitigates visual forgetting through two complementary rewards: a visual grounding reward that sustains attention on informative visual tokens, and a sink suppression reward that discourages spurious attention on sink tokens, steering focus toward non-sink visual evidence. Experiments show that LASER consis-

tently outperforms strong baselines on visual reasoning benchmarks, effectively mitigating visual forgetting with improved visual grounding in attention.

Acknowledgements

This research is partially supported by the Australian Research Council (IH230100013, DP230101196, DE240100105, DP240101814, IE250100108).

References

1. Alayrac, J., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., Ring, R., Rutherford, E., Cabi, S., Han, T., Gong, Z., Samangooei, S., Monteiro, M., Menick, J.L., Borgeaud, S., Brock, A., Nematzadeh, A., Sharifzadeh, S., Binkowski, M., Barreira, R., Vinyals, O., Zisserman, A., Simonyan, K.: Flamingo: a visual language model for few-shot learning. In: NeurIPS (2022)
2. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., Zhong, H., Zhu, Y., Yang, M., Li, Z., Wan, J., Wang, P., Ding, W., Fu, Z., Xu, Y., Ye, J., Zhang, X., Xie, T., Cheng, Z., Zhang, H., Yang, Z., Xu, H., Lin, J.: Qwen2.5-vl technical report. CoRR (2025). https://doi.org/10.48550/ARXIV.2502.13923, https://doi.org/10.48550/arXiv.2502.13923
3. Beyer, L., Steiner, A., Pinto, A.S., Kolesnikov, A., Wang, X., Salz, D., Neumann, M., Alabdulmohsin, I., Tschannen, M., Bugliarello, E., Unterthiner, T., Keysers, D., Koppula, S., Liu, F., Grycner, A., Gritsenko, A.A., Houlsby, N., Kumar, M., Rong, K., Eisenschlos, J., Kabra, R., Bauer, M., Bosnjak, M., Chen, X., Minderer, M., Voigtlaender, P., Bica, I., Balazevic, I., Puigcerver, J., Papalampidi, P., Hénaff, O.J., Xiong, X., Soricut, R., Harmsen, J., Zhai, X.: Paligemma: A versatile 3b VLM for transfer. CoRR (2024), https://doi.org/10.48550/arXiv.2407.07726
4. Cai, Z., Cao, M., Chen, H., Chen, K., Chen, K., Chen, X., et al.: Internlm2 technical report. CoRR (2024), https://doi.org/10.48550/arXiv.2403.17297
5. Chen, L., Zhao, H., Liu, T., Bai, S., Lin, J., Zhou, C., Chang, B.: An image is worth 1/2 tokens after layer 2: Plug-and-play inference acceleration for large vision-language models. In: ECCV. pp. 19–35. Springer (2024)
6. Chen, L., Li, J., Dong, X., Zhang, P., Zang, Y., Chen, Z., Duan, H., Wang, J., Qiao, Y., Lin, D., Zhao, F.: Are we on the right way for evaluating large vision-language models? In: NeurIPS (2024)
7. Chen, S., Guo, Y., Su, Z., Li, Y., Wu, Y., Chen, J., Chen, J., Wang, W., Qu, X., Cheng, Y.: Advancing multimodal reasoning: From optimized cold start to staged reinforcement learning. arXiv preprint arXiv:2506.04207 (2025)
8. Chu, T., Zhai, Y., Yang, J., Tong, S., Xie, S., Schuurmans, D., Le, Q.V., Levine, S., Ma, Y.: Sft memorizes, rl generalizes: A comparative study of foundation model post-training. arXiv preprint arXiv:2501.17161 (2025)
9. DeepSeek-AI: Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. CoRR **abs/2501.12948** (2025), https://doi.org/10.48550/arXiv.2501.12948
10. Deng, Y., Bansal, H., Yin, F., Peng, N., Wang, W., Chang, K.: Openvlthinker: An early exploration to complex vision-language reasoning via iterative self-improvement. CoRR (2025), https://doi.org/10.48550/arXiv.2503.17352

11. Du, J., Li, Y., Li, Y., Liao, L., Zhao, Z., Ye, G.: Medfuse: a multi-source data fusion framework for diabetic retinopathy lesion segmentation. Frontiers of Computer Science (2026)
12. Du, Y., Liu, Z., Li, Y., Zhao, W.X., Huo, Y., Wang, B., Chen, W., Liu, Z., Wang, Z., Wen, J.: Virgo: A preliminary exploration on reproducing o1-like MLLM. CoRR (2025), https://doi.org/10.48550/arXiv.2501.01904
13. Duan, H., Yang, J., Qiao, Y., Fang, X., Chen, L., Liu, Y., Dong, X., Zang, Y., Zhang, P., Wang, J., et al.: Vlmevalkit: An open-source toolkit for evaluating large multi-modality models. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 11198–11201 (2024)
14. Engine, V.: verl: Volcano engine reinforcement learning for llms. https://github.com/verl-project/verl (2025)
15. Favero, A., Zancato, L., Trager, M., Choudhary, S., Perera, P., Achille, A., Swaminathan, A., Soatto, S.: Multi-modal hallucination control by visual information grounding. In: CVPR. pp. 14303–14312. IEEE (2024)
16. Fu, Y., Zhang, Z., Zhang, Y., Wang, Z., Huang, Z., Luo, Y.: Mergevla: Cross-skill model merging toward a generalist vision-language-action agent. In: CVPR (2026)
17. Gong, X., Ming, T., Wang, X., Wei, Z.: DAMRO: dive into the attention mechanism of LVLM to reduce object hallucination. In: EMNLP. pp. 7696–7712. Association for Computational Linguistics (2024)
18. Guan, T., Liu, F., Wu, X., Xian, R., Li, Z., Liu, X., Wang, X., Chen, L., Huang, F., Yacoob, Y., Manocha, D., Zhou, T.: Hallusionbench: An advanced diagnostic suite for entangled language hallucination and visual illusion in large vision-language models. In: CVPR. IEEE (2024)
19. Hao, Z., Li, Z., Li, W., Liu, F., Zhang, M., Li, J.: Echoes as anchors: Probabilistic costs and attention refocusing in llm reasoning. In: The Fourteenth International Conference on Learning Representations (ICLR) (2026)
20. Hu, Y., Shi, W., Fu, X., Roth, D., Ostendorf, M., Zettlemoyer, L., Smith, N.A., Krishna, R.: Visual sketchpad: Sketching as a visual chain of thought for multimodal language models. In: NeurIPS (2024)
21. Huang, W., Jia, B., Zhai, Z., Cao, S., Ye, Z., Zhao, F., Xu, Z., Hu, Y., Lin, S.: Vision-r1: Incentivizing reasoning capability in multimodal large language models. CoRR (2025), https://doi.org/10.48550/arXiv.2503.06749
22. Jaech, A., Kalai, A., Lerer, A., Richardson, A., El-Kishky, A., Low, A., Helyar, A., Madry, A., Beutel, A., Carney, A., et al.: Openai o1 system card. arXiv preprint arXiv:2412.16720 (2024)
23. Jian, P., Wu, J., Sun, W., Wang, C., Ren, S., Zhang, J.: Look again, think slowly: Enhancing visual reflection in vision-language models. CoRR (2025), https://doi.org/10.48550/arXiv.2509.12132
24. Kang, S., Kim, J., Kim, J., Hwang, S.J.: See what you are told: Visual attention sink in large multimodal models. In: ICLR (2025)
25. Leng, S., Wang, J., Li, J., Zhang, H., Hu, Z., Zhang, B., Jiang, Y., Zhang, H., Li, X., Bing, L., et al.: Mmr1: Enhancing multimodal reasoning with variance-aware sampling and open resources. arXiv preprint arXiv:2509.21268 (2025)
26. Leng, S., Zhang, H., Chen, G., Li, X., Lu, S., Miao, C., Bing, L.: Mitigating object hallucinations in large vision-language models through visual contrastive decoding. In: CVPR. pp. 13872–13882. IEEE (2024)
27. Li, Q., Zhang, Y., Mai, Z., Chen, Y., Huang, H., Zhang, J., Zhang, Z., Wen, Y., Li, W., Fu, H., et al.: Can large multimodal models understand agricultural scenes? benchmarking with agromind. Advances in Neural Information Processing Systems (2026)

28. Li, X., et al.: Llm inductive reasoning through multi-agent enhanced monte carlo tree search. In: ACL (2026)
29. Li, Z., Yu, W., Huang, C., Liu, R., Liang, Z., Liu, F., Che, J., Yu, D., Boyd-Graber, J.L., Mi, H., Yu, D.: Self-rewarding vision-language model via reasoning decomposition. CoRR (2025), https://doi.org/10.48550/arXiv.2508.19652
30. Lin, Z., Lin, M., Lin, L., Ji, R.: Boosting multimodal large language models with visual tokens withdrawal for rapid inference. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 5334–5342 (2025)
31. Liu, H., Li, C., Li, Y., Lee, Y.J.: Improved baselines with visual instruction tuning. In: CVPR. IEEE (2024)
32. Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) NeurIPS (2023)
33. Liu, Z., Liu, Y., Zhu, G., Xie, C., Li, Z., Yuan, J., Wang, X., Li, Q., Cheung, S.C., Zhang, S., et al.: Infi-mmr: Curriculum-based unlocking multimodal reasoning via phased reinforcement learning in multimodal small language models. arXiv preprint arXiv:2505.23091 (2025)
34. Lu, P., Bansal, H., Xia, T., Liu, J., Li, C., Hajishirzi, H., Cheng, H., Chang, K., Galley, M., Gao, J.: Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts. In: ICLR. OpenReview.net (2024)
35. Meng, F., Du, L., Liu, Z., Zhou, Z., Lu, Q., Fu, D., Han, T., Shi, B., Wang, W., He, J., et al.: Mm-eureka: Exploring the frontiers of multimodal reasoning with rule-based reinforcement learning. arXiv preprint arXiv:2503.07365 (2025)
36. Mondal, D., Modi, S., Panda, S., Singh, R., Rao, G.S.: Kam-cot: Knowledge augmented multimodal chain-of-thoughts reasoning. In: AAAI. pp. 18798–18806. AAAI Press (2024)
37. OpenAI: Openai gpt-5 system card (2025), https://arxiv.org/abs/2601.03267
38. Qiao, R., Tan, Q., Dong, G., Wu, M., Sun, C., Song, X., Wang, J., Gongque, Z., Lei, S., Zhang, Y., Wei, Z., Zhang, M., Qiao, R., Zong, X., Xu, Y., Yang, P., Bao, Z., Diao, M., Li, C., Zhang, H.: We-math: Does your large multimodal model achieve human-like mathematical reasoning? In: ACL. pp. 20023–20070. Association for Computational Linguistics (2025)
39. Rohrbach, A., Hendricks, L.A., Burns, K., Darrell, T., Saenko, K.: Object hallucination in image captioning. In: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.) EMNLP. pp. 4035–4045. Association for Computational Linguistics (2018)
40. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Zhang, M., Li, Y.K., Wu, Y., Guo, D.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. CoRR (2024), https://doi.org/10.48550/arXiv.2402.03300
41. Shen, H., Liu, P., Li, J., Fang, C., Ma, Y., Liao, J., Shen, Q., Zhang, Z., Zhao, K., Zhang, Q., Xu, R., Zhao, T.: VLM-R1: A stable and generalizable r1-style large vision-language model. CoRR (2025), https://doi.org/10.48550/arXiv.2504.07615
42. Sun, H., Sun, Z., Peng, H., Ye, H.: Mitigating visual forgetting via take-along visual conditioning for multi-modal long cot reasoning. In: ACL. pp. 5158–5171. Association for Computational Linguistics (2025)
43. Sun, M., et al.: Massive activations in large language models. arXiv (2024)
44. Team, G.: Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. CoRR **abs/2507.06261** (2025), https://doi.org/10.48550/arXiv.2507.06261
45. Thawakar, O., Dissanayake, D., More, K.P., Thawkar, R., Heakl, A., Ahsan, N., Li, Y., Zumri, M., Lahoud, J., Anwer, R.M., Cholakkal, H., Laptev, I., Shah, M.,

Khan, F.S., Khan, S.H.: Llamav-o1: Rethinking step-by-step visual reasoning in llms. In: ACL. pp. 24290–24315. Association for Computational Linguistics (2025)
46. Tian, X., Zou, S., Yang, Z., He, M., Waschkowski, F., Wesemann, L., Tu, P.H., Zhang, J.: More thought, less accuracy? on the dual nature of reasoning in vision-language models. CoRR (2025), https://doi.org/10.48550/arXiv.2509.25848
47. Tu, C., Ye, P., Zhou, D., Bai, L., Yu, G., Chen, T., Ouyang, W.: Attention reallocation: Towards zero-cost and controllable hallucination mitigation of mllms. CoRR (2025), https://doi.org/10.48550/arXiv.2503.08342
48. Wang, D., Qiu, R., Huang, Z.: When to commit? towards variable-size self-contained blocks for discrete diffusion language models. arXiv preprint arXiv:2604.23994 (2026)
49. Wang, H., Qu, C., Huang, Z., Chu, W., Lin, F., Chen, W.: Vl-rethinker: Incentivizing self-reflection of vision-language models with reinforcement learning. CoRR (2025), https://doi.org/10.48550/arXiv.2504.08837
50. Wang, K., Pan, J., Shi, W., Lu, Z., Ren, H., Zhou, A., Zhan, M., Li, H.: Measuring multimodal mathematical reasoning with math-vision dataset. In: NeurIPS (2024)
51. Wang, S., Yu, X., Luo, Y., Wang, Z., Zhang, P., Huang, Z.: Language-driven fine-grained retrieval. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2682–2692 (2026)
52. Wang, Y., Wang, S., Cheng, Q., Fei, Z., Ding, L., Guo, Q., Tao, D., Qiu, X.: Visuothink: Empowering lvlm reasoning with multimodal tree search. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 21707–21719 (2025)
53. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al.: Chain-of-thought prompting elicits reasoning in large language models. NeurIPS **35**, 24824–24837 (2022)
54. Wu, J., Xiong, Y., Li, X., Xia, Y., Wang, R., Wang, Y., Yu, T., Kim, S., Rossi, R.A., Yao, L., Shang, J., McAuley, J.J.: Mitigating visual knowledge forgetting in MLLM instruction-tuning via modality-decoupled gradient descent. CoRR (2025), https://doi.org/10.48550/arXiv.2502.11740
55. Xia, J., Zang, Y., Gao, P., Li, Y., Zhou, K.: Visionary-r1: Mitigating shortcuts in visual reasoning with reinforcement learning. CoRR (2025), https://doi.org/10.48550/arXiv.2505.14677
56. Xiao, G., Tian, Y., Chen, B., Han, S., Lewis, M.: Efficient streaming language models with attention sinks. arXiv preprint arXiv:2309.17453 (2023)
57. Xiao, Y., Sun, E., Liu, T., Wang, W.: Logicvista: Multimodal LLM logical reasoning benchmark in visual contexts. CoRR **abs/2407.04973** (2024), https://doi.org/10.48550/arXiv.2407.04973
58. Xu, G., Jin, P., Li, H., Song, Y., Sun, L., Yuan, L.: Llava-cot: Let vision language models reason step-by-step. CoRR (2024), https://doi.org/10.48550/arXiv.2411.10440
59. Yang, S., Niu, Y., Liu, Y., Ye, Y., Lin, B., Yuan, L.: Look-back: Implicit visual re-focusing in mllm reasoning. arXiv preprint arXiv:2507.03019 (2025)
60. Yang, Y., He, X., Pan, H., Jiang, X., Deng, Y., Yang, X., Lu, H., Yin, D., Rao, F., Zhu, M., Zhang, B., Chen, W.: R1-onevision: Advancing generalized multimodal reasoning through cross-modal formalization. CoRR (2025), https://doi.org/10.48550/arXiv.2503.10615
61. Yao, H., Yin, Q., Zhang, J., Yang, M., Wang, Y., Wu, W., Su, F., Shen, L., Qiu, M., Tao, D., et al.: R1-sharevl: Incentivizing reasoning capability of multimodal large language models via share-grpo. arXiv preprint arXiv:2505.16673 (2025)

62. Yuan, B., Song, S., Fernandez, J., Luo, Y., Baktashmotlagh, M., Wang, Z.: Wiswheat: A three-tiered vision-language dataset for wheat management. In: Proceedings of the 33rd ACM International Conference on Multimedia (2025)
63. Yue, X., Ni, Y., Zheng, T., Zhang, K., Liu, R., Zhang, G., Stevens, S., Jiang, D., Ren, W., Sun, Y., Wei, C., Yu, B., Yuan, R., Sun, R., Yin, M., Zheng, B., Yang, Z., Liu, Y., Huang, W., Sun, H., Su, Y., Chen, W.: MMMU: A massive multi-discipline multimodal understanding and reasoning benchmark for expert AGI. In: CVPR. pp. 9556–9567. IEEE (2024)
64. Zhang, R., Jiang, D., Zhang, Y., Lin, H., Guo, Z., Qiu, P., Zhou, A., Lu, P., Chang, K., Qiao, Y., Gao, P., Li, H.: MATHVERSE: does your multi-modal LLM truly see the diagrams in visual math problems? In: ECCV. pp. 169–186. Springer (2024)
65. Zhao, Z., Chen, Z., Huang, Z., Sadiq, S., Chen, T.: Continual text-to-video retrieval with frame fusion and task-aware routing. In: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (2025)
66. Zheng, G., Yang, B., Tang, J., Zhou, H., Yang, S.: Ddcot: Duty-distinct chain-of-thought prompting for multimodal reasoning in language models. In: NeurIPS (2023)
67. Zhou, D., Schärli, N., Hou, L., Wei, J., Scales, N., Wang, X., Schuurmans, D., Cui, C., Bousquet, O., Le, Q., et al.: Least-to-most prompting enables complex reasoning in large language models. arXiv preprint arXiv:2205.10625 (2022)
68. Zhou, Y., Li, W., Lu, Y., Li, J., Liu, F., Zhang, M., Wang, Y., He, D., LIU, H., Zhang, M.: Reflection on knowledge graph for large language models reasoning. In: Findings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL) (2025)