DeepVision-103K: A Visually Diverse, Broad-Coverage, and Verifiable Mathematical Dataset for Multimodal Reasoning

Anonymous ACL submission

Abstract

Reinforcement Learning with Verifiable Rewards (RLVR) has been shown effective in enhancing the visual reflection and reasoning capabilities of Large Multimodal Models (LMMs). However, existing datasets are predominantly derived from either small-scale manual construction or recombination of prior resources, which limits data diversity and coverage, thereby constraining further gains in model performance. To this end, we introduce **DeepVision-103K**, a comprehensive dataset for RLVR training that covers diverse K12 mathematical topics, extensive knowledge points, and rich visual elements. Models trained on DeepVision achieve strong performance on multimodal mathematical benchmarks, and generalize effectively to general multimodal reasoning tasks. Further analysis reveals enhanced visual perception, reflection and reasoning capabilities in trained models, validating DeepVision's effectiveness for advancing multimodal reasoning.

1 Introduction

Large language models (LLMs) trained with reinforcement learning from verifiable rewards (RLVR), such as DeepSeek-R1 (DeepSeek-AI et al., 2025) and OpenAI o-series (OpenAI et al., 2024), demonstrate remarkable reasoning capabilities. A key insight is that RLVR incentivizes thinking behaviors—the ability to decompose problems, self-correct in step-by-step reasoning. Recent works (Wang et al., 2025a; Xia et al., 2025; Yang et al., 2025a) extend this paradigm to large multimodal models (LMMs), achieving enhanced visual reflection and reasoning abilities. Central to this progress is high-quality training data, but existing training sets for multimodal RLVR exhibit several key limitations.

- **Synthetically constructed datasets**: Fully synthesized with professional tools like GeoGebra (Lu et al., 2021; Qiao et al., 2025). They provide abundant data for constructible categories (e.g., geometric diagrams, function curves) but lack real-world mathematical scenarios, limiting robust generalization to general tasks.

- **Human-annotated K12 datasets**: Gathered from authentic K12 education scenarios and human-annotated to obtain verifiable answers (Meng et al., 2025; Liu et al., 2024). While offering broader categories, dependence on expert annotation limits its scalability.

- **Recombination of existing datasets**: Filtration (Wang et al., 2025d; Zha et al., 2025) or recombination (Peng et al., 2025; Yang et al., 2025b; Zhang et al., 2025) of prior sources. These approaches create no novel problems, resulting in overlap across datasets and lacking broader data distribution.

To address these limitations, we propose **DeepVision-103K**, a large-scale multimodal mathematical dataset designed for RLVR, featuring:

- **Visual Diversity**: DeepVision-103K covers major visual categories including geometry, analytic plots, charts, and real-world items in mathematical contexts. Within each category, DeepVision offers richer element types than existing open-source datasets (Figure 4).

- **Broad Coverage**: DeepVision-103K incorporates wide-ranging multimodal mathematical problems (Figure 5) and visual logic problems (mazes, chess, tangrams), jointly enhancing mathematical and visual logic reasoning.

- **Automatic Data Curation Pipeline**: We present an automatic curation pipeline (Figure 6) comprising validity filtering, pass-rate stratification and correctness verification,

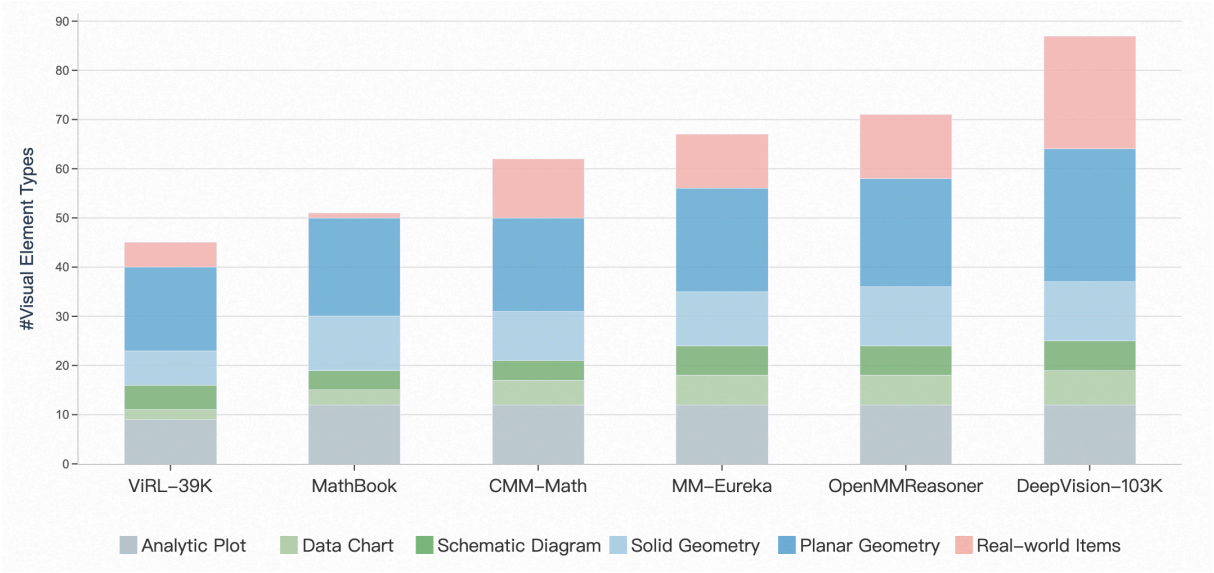


Figure 1: The number of different visual element types of training datasets.

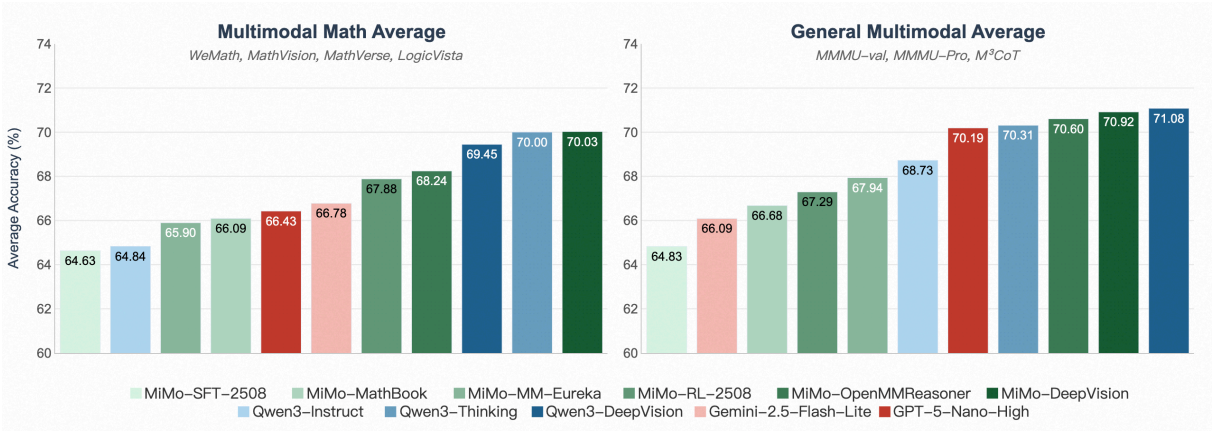


Figure 2: Performance on multimodal math and general multimodal benchmarks, we report averaged Pass@1 accuracy across benchmarks.

which transforms diverse but noisy real-world K12 problems into structured and verifiable QA pairs.

Consequently, models trained on DeepVision-103K achieve top performance (Figure 2) on mathematical and general multimodal reasoning. DeepVison models outperform: (1) models trained on other open-source datasets, (2) the official thinking variant built on the same base model, and (3) strong closed-source baselines. These results underscore the value of DeepVision-103K as a resource for advancing multimodal reasoning. The remainder of this paper is organized as follows:

- Sec. 2 presents an overview of DeepVision-103K, including its format, visual elements distribution, and topics covered.
- Sec. 3 details the data curation pipeline to construct DeepVision-103K, encompassing validity filtering, model-centric difficulty filtering and query correctness verification.
- Sec. 4 describes the training setup and evaluation results of models trained on DeepVision-103K.
- Sec. 5 analyzes the impact of DeepVision-103K on model performance and provides ablation studies on the data curation pipeline.

2 Overview of DeepVision

DeepVision adopts a rich annotation schema to facilitate various downstream tasks in multimodal reasoning. As illustrated in Figure 3, each sample contains the following components:

- **Question & Image:** A multimodal mathe-

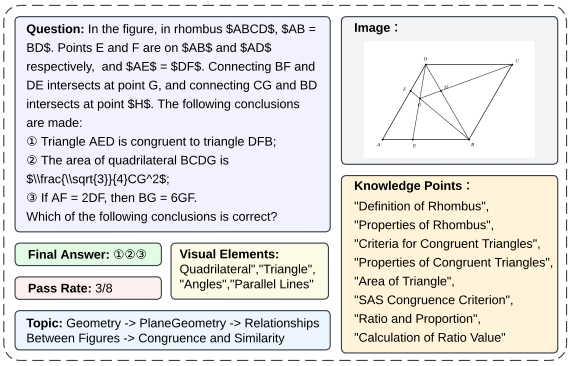


Figure 3: A data sample from DeepVision.

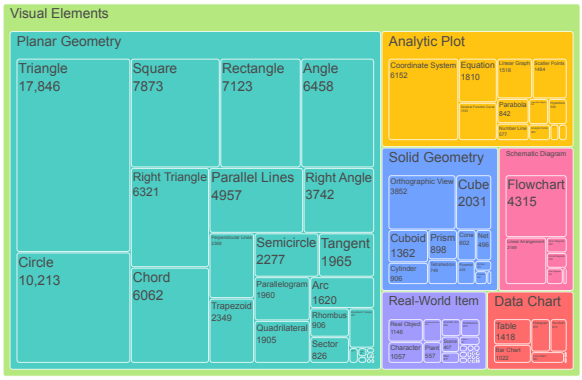


Figure 4: Visual elements distribution in DeepVision.

matical problem consisting of a textual problem statement and the corresponding image.

- **Final Answer:** A unique, verifiable answer that enables rule-based reward computation in RLVR.
- **Pass Rate:** The proportion of correct responses obtained during model rollouts.
- **Topic:** A hierarchical classification indicating which branch of mathematics the problem belongs to.
- **Knowledge Points:** A list of specific mathematical concepts, theorems, or techniques required to solve the problem.
- **Visual Elements:** A list of geometric or graphical objects depicted in the image, describing what visual content should be perceived and interpreted.

2.1 Visual Diversity

To assess the richness of visual content in DeepVision, we annotated each image with both *category* and *fine-grained type* following (Mo et al., 2018; Rosin, 2008). As shown in Figure 4, DeepVision includes diverse visual elements across 6 categories, each presenting unique perceptual challenges.

We summarized the coverage of each category in Table 1. Notably, DeepVision-Math captures **cross-category visual combinations** and real-world items in mathematical contexts, requiring models to reason across multiple visual representations simultaneously. Examples are provides in Appendix A.

2.2 Broad Coverage

DeepVision demonstrates exceptional breadth across mathematical topics and knowledge points.

Category	Key Visual Elements
Planar Geometry	Primitives (Angle, Triangle, Circle, Quadrilateral,Polygon), Relations (Parallelism, Tangency, Chords), Properties (Right Angles, Perpendicularity)
Solid Geometry	3D Primitives (Cube, Prism, Cylinder, Cone), Spatial Representations (Orthographic Views, Nets), Sections (Frustums, Hemispheres)
Analytic Plot	Coordinate Systems, Function Curves (Linear, General), Conic Sections (Parabola, Hyperbola), Scatter Points, Inequality Regions
Data Chart	Statistical Graphs (Bar, Histogram, Pie, Line), Structured Data (Tables, Stem-and-Leaf)
Schematic Diagram	Logical Structures (Flowcharts, Tree Diagrams), Physics/Sets (Force Diagrams, Circuits, Venn Diagrams), Linear Arrangements
Real-World Item	Objects (Characters, Household Items), Contextual Scenes (Architecture, Maps, Scientific Tools)
Cross-category	Combinations of multiple visual categories

Table 1: Visual categories and element coverage in DeepVision-103K.

We categorized each problem using a hierarchical topic structure, following the methodology from Qiao et al. (2025).

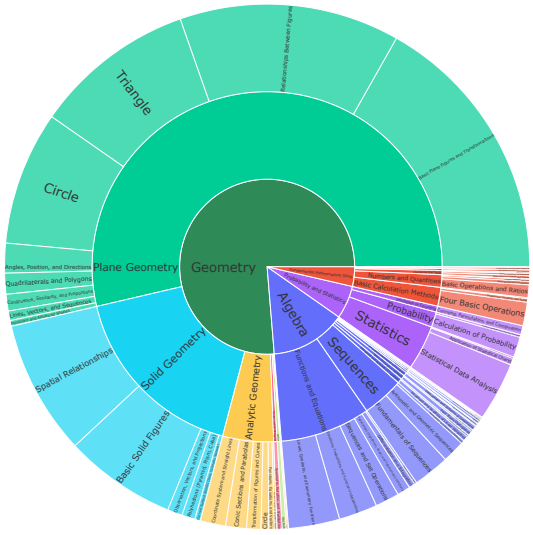


Figure 5: Mathematical topics distribution in DeepVision-103K.

As illustrated in Figure5, our dataset spans four major mathematical disciplines, with **Geometry** dominating the distribution, followed by substantial coverage in **Algebra**, **Probability and Statistics**, and **Fundamental Mathematical Skills**. Within these domains, DeepVision covers an extensive range of over 200 specific topics and nearly 400 distinct knowledge points. This breadth ensures models are exposed to diverse problem-solving paradigms, fostering robust and generalizable reasoning capabilities. Beyond mathematical problems that require applying formulas and theorems, DeepVision incorporates **visual logic problems** from Zebra-CoT (Li et al., 2025)—including mazes, chess problems, tangrams — where solutions emerge primarily from visual perception and logical deduction.

3 Construction of DeepVision

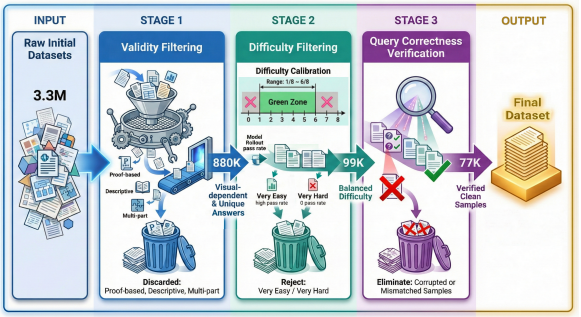


Figure 6: Curation pipeline for mathematical data in DeepVision-103K.

We curated our dataset from open-source multimodal mathematics SFT corpora, including MM-MathInstruct-3M (Wang et al., 2025c) and MultiMath-300K (Peng et al., 2024). Both datasets collect K12 level problems from real educational contexts, forming an initial pool of 3.3M samples. To derive verifiable data from this extensive yet noisy collection, we applied a three-stage curation pipeline in Figure 6:

1. **Validity Filtering:** Remove problems inherently unsuitable for RL training, including proof-based, descriptive and multi-answer questions.

2. **Difficulty Filtering:** Calibrate sample difficulty based on model capability through rollout pass rates.

3. **Query Correctness Verification:** Validate the correctness of image-question pairs and answers to eliminate corrupted samples.

Stage 1: Validity Filtering. Reinforcement learning requires unique and verifiable answers to provide reliable reward signals. In this stage, we first applied rule-based filtering to remove proof or explanation tasks containing keywords such as "prove", "explain", "describe". For the remaining questions, we employed Qwen3-VL-32B-Instruct (Bai et al., 2025) to analyze each sample, counting the number of answers and determining whether visual information is necessary. Only questions with unique answer and genuinely require visual information were retained. After this stage, we obtained 880K questions.

Stage 2: Difficulty Filtering. Data of appropriate difficulty is crucial for efficient RL training (Zeng et al., 2025b). DeepMath (He et al., 2025) employs SOTA models to annotate difficulty based on human-defined standards, which may not align well with model capabilities (Qiao et al., 2025). We adopt an approach similar to Qwen3-VL (Bai et al., 2025). For each question, we performed 8 rollouts using MiMo-VL-7B-SFT (Team et al., 2025) and then calculated accuracy with Math-Verify (Kydlíček, 2025). We only retained data with $[\frac{1}{8}, \frac{6}{8}]$ pass rate. Zero pass rate samples were removed as they are too difficult or unverifiable, while those in $[\frac{7}{8}, 1]$ were dropped because overly easy data can reduce exploration in RL training (Zeng et al., 2025a). For visual logic data, since it's well formed from Zebra-CoT (Li et al., 2025), we perform rollout under the same setting with math data, obtaining 26K clean and verifiable training data. Appendix B provides details of this procedure.

Stage 3: Query Correctness Verification. Correct answers are crucial for providing effective RL rewards, and so are well-formed questions. Although we only retained questions with pass rates greater than zero, models still randomly guessed answers for inherently problematic queries (e.g., garbled text or image-text mismatches). To this end, we prompted Gemini-3-Flash (Google, 2025) to (1) verify that each question is complete and free of garbled text, (2) check potential image-text mismatch, and (3) validate the provided answer. Only samples passing all checks were retained. After this final stage, we obtained 77K correct and verifiable QA pairs for RL training.

Model	Multimodal Math				General Multimodal		
	WeMath	MathVision	$\text{MathVerse}_{\text{vision}}$	LogicVista	MMMU_{val}	MMMU_{Pro}	M^3CoT
Closed-source Models							
GPT-5-Nano-High	78.62	58.75	70.3	58.03	70.78	70.64	69.15
Gemini-2.5-Flash-Lite	83.85	52.47	70.3	60.49	64.77	65.08	68.42
Qwen3-VL-8B Series							
Qwen3-VL-8B-Instruct	79.36	51.44	67.38	61.16	67.66	67.69	70.83
Qwen3-VL-8B-Thinking	84.54	**57.89**	**72.84**	64.73	69.33	**70.29**	71.31
Qwen3-VL-8B-DeepVision	**85.11**	55.49	72.46	**64.73**	**71.33**	**70.29**	**71.61**
MiMo-VL-7B Series							
MiMo-VL-7B-SFT-2508	74.42	50.69	72.71	60.71	63.77	60.69	70.02
MiMo-VL-7B-RL-2508	76.95	53.91	**76.39**	64.28	67.44	63.87	70.57
MiMo-VL-7B-MM-Eureka	79.08	50.00	73.35	61.16	67.67	65.78	70.36
MiMo-VL-7B-MathBook	77.18	51.31	73.60	62.28	66.33	63.47	70.23
MiMo-VL-7B-OpenMMReasoner	**83.45**	52.97	74.87	61.68	66.78	66.82	**78.21** [1]
MiMo-VL-7B-DeepVision	82.98	**55.24**	76.26	**65.62**	**71.00**	**69.19**	72.56

Table 2: Performance comparison across multimodal mathematical reasoning and general multimodal benchmarks. We report Pass@1 accuracy (%). The best results for each model family are shown in **bold**.

4 Experiments

In this section, we presented a comprehensive evaluation of the mathematical and general multimodal reasoning capabilities of models trained on DeepVision.

4.1 Setup

Models We conducted training on LMMs that already possess thinking capabilities, including MiMo-VL-7B-SFT-2508 (Team et al., 2025) and Qwen3-VL-8B-Instruct (Bai et al., 2025). Both models have been exposed to visual reasoning data during the pretrain or midtrain stages, exhibiting native visual thinking abilities.

Algorithm We employed GSPO (Zheng et al., 2025) for RL training, utilizing rule-based rewards based on answer correctness (+1 for correct answers, 0 otherwise). We specified the required response format through prompts, and no additional format reward was applied. Detailed training configurations and prompts are provided in Appendix C.

Baselines We compared against (1) **Closed-source models**: GPT-5-Nano-High, Gemini-2.5-Flash-Lite; (2) **Official thinking variants**: Qwen3-VL-8B-Thinking, MiMo-VL-7B-RL-2508; and (3) **Open-source datasets**: MM-Eureka (Meng et al., 2025), human-annotated real K12 data; MathBook (Qiao et al., 2025), human curated data; OpenMMReasoner (Zhang et al., 2025), filtration and combination of prior sources. We trained MiMo-VL-7B-SFT-2508 on these datasets under the same setting for fair comparison with MiMo-VL-7B-DeepVision.

Evaluation We evaluated our models on the following benchmarks: (1) **Multimodal Math**: WeMath (Qiao et al., 2024), $\text{MathVerse}_{\text{vision}}$ (Zhang et al., 2024), MathVision (Wang et al., 2024), and LogicVista (Xiao et al., 2024). (2) **General Multimodal**: M^3CoT (Chen et al., 2024), MMMU_{VAL} (Yue et al., 2024a) and $\text{MMMU}_{\text{Pro\_full}}$ (Yue et al., 2024b). For inference parameters, we set the maximum token length at 32K for all evaluation. Decoding parameters follow the official recommendations. Complete details are provided in Appendix D.

4.2 Multimodal Mathematics Reasoning Results

As shown in Table 2, training on DeepVision yields strong results in mathematical reasoning.

Consistent gains across benchmarks. Compared to respective Instruct/SFT baselines, Qwen3-VL-8B-DeepVision and MiMo-VL-7B-DeepVision achieve uniform improvements across all evaluated benchmarks, with gains ranging from 2.91% to 8.56%.

Substantial improvements. On WeMath and LogicVista, DeepVision models surpass their official thinking variants and closed-source models. Qwen3-VL-8B-DeepVision reaches sota results on WeMath (85.11%), MiMo-VL-7B-DeepVision reaches sota results on LogicVista (65.62%). On

MathVision and MathVerse, they exceed or substantially narrow the gap with thinking variants.

Superiority over existing open-source datasets. Compared to models trained on other open-source datasets, MiMo-VL-7B-DeepVision demonstrates clear advantages, highlighting the value of DeepVision as a high-quality RL training resource.

4.3 Generalization Beyond Mathematics

Table 2 shows that DeepVision models generalize effectively to general-purpose multimodal tasks, achieving consistent improvements over foundation models and surpassing official thinking variants across all three benchmarks. In contrast, models trained on other open-source datasets show limited improvements in general domains. This disparity suggests that the diverse visual elements and broad domain coverage in DeepVision are crucial for enhancing general multimodal reasoning capabilities, which is further supported by our analysis in Sec. 5.2.

5 Analyses

Our analyses investigate the following key questions:

***Q1**: Enhanced Capabilities.* What capabilities are enhanced after RL on DeepVision-103K?

***Q2**: The Value of Visual Logic Data.* What role do the introduced visual logic tasks (e.g., mazes, tangrams, and games) play in the DeepVision-103K dataset?

***Q3**: Necessity of query correctness verification.* Recent studies (Wu et al., 2025; Shao et al., 2025) suggest that RLVR can work even under random rewards. Is correctness verification step truly necessary in our data curation pipeline?

5.1 Enhanced Capabilities

To investigate how RL on DeepVision improves model capabilities, we systematically compared Qwen3-VL-8B-Instruct and Qwen3-VL-8B-DeepVision across multiple benchmarks. We collected cases where DeepVision succeeds but Instruct fails and asked human annotators to analyze the underlying mechanism following Algorithm 1.

For each sample, annotators cited verbatim evidence from model response (Figure 8). If no evidence supports, the sample was labeled as GUESS. Our analysis reveals three enhancement types, as shown in Figure 7.

Algorithm 1: Human Annotation Protocol

Input: Query $(\text{Image}, \text{Text})$,Ground Truth y, Incorrect Instruct Response R_I, Correct DeepVision Response R_D
Output: Improvement Mechanism C

Analyze visual descriptions in R_I
if *Descriptions contradict Image* **then**
  Root Cause ← **Visual Misperception**
else
  Root Cause ← **Incorrect Reasoning**
end
if *Root Cause is* ***Visual Misperception*** **then**
  if R_D *correct at first observation* **then**
    $C \leftarrow$ VISUAL PERCEPTION
  else
    if R_D *corrected via reflection* **then**
      $C \leftarrow$ VISUAL REFLECTION
    else
      $C \leftarrow$ GUESS
    end
  end
else if *Root Cause is* ***Incorrect Reasoning*** **then**
  if R_D *shows valid reasoning chain* **then**
    $C \leftarrow$ REASONING
  else
    $C \leftarrow$ GUESS
  end
end
return C

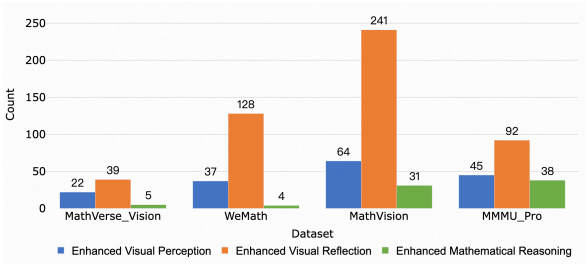


Figure 7: Enhanced Capabilities

Type I: Enhanced Visual Perception. We observed enhanced "one-shot perception"—DeepVision model correctly identifies geometric shapes, numerical values, and spatial relationships in the initial observation, without requiring iterative re-examination (Figure 8).

Type II: Enhanced Visual Reflection. When initial perceptual errors occur, DeepVision demonstrates a stronger capacity for genuine visual

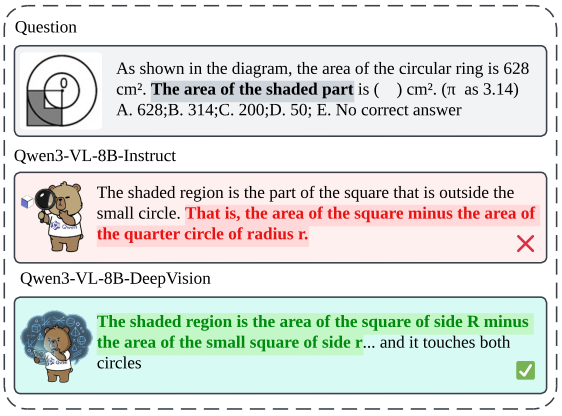


Figure 8: DeepVision model correctly identifies the shaded region on the first attempt.

re-examination—actively recounting elements, remeasuring angles, and re-inspecting spatial relationships—whereas the base model tends to rephrase conclusions without revisiting the visual content (Figure 9).

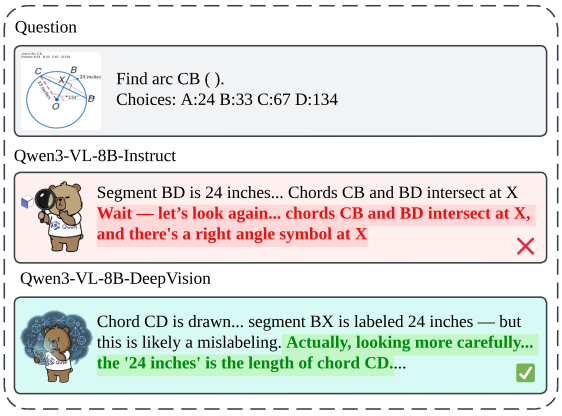


Figure 9: DeepVision actively re-examines visual content to correct errors, while the base model merely rephrases without genuine verification.

Type III: Enhanced Mathematical Reasoning. Beyond visual capabilities, RL fine-tuning also enhances pure mathematical reasoning. In cases where both models correctly extract identical visual information, DeepVision demonstrates more rigorous mathematical reasoning (Figure 10).

5.2 The Value of Visual Logic Data

DeepVision spans two data domains—multimodal math and visual logic, which differ in reasoning paradigms. Multimodal math requires extracting visual evidence and applying mathematical knowledge (e.g., formulas, theorems, computations) to reach an answer. In contrast, visual logic is driven

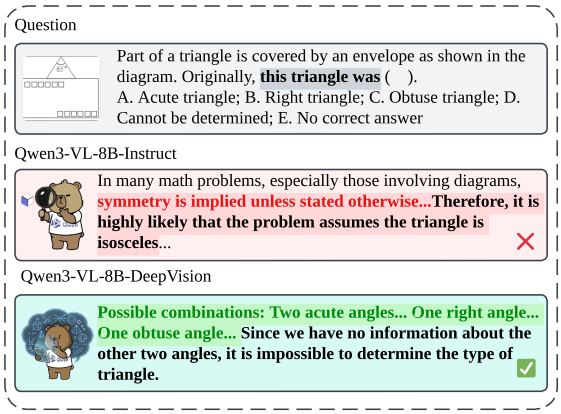


Figure 10: DeepVision systematically enumerates all possible angle combinations and concludes the type cannot be determined, while the Instruct model incorrectly assumes symmetry without justification.

mainly by visual cues (e.g.,object positions, spatial relations, and patterns), with little reliance on explicit mathematical knowledge. Zha et al. (2025) points out that mixing heterogeneous domains may introduce interference and conflicting gradients, potentially harming learning. This motivated us to examine whether introducing visual-logic data is indeed beneficial, and how each domain contributes to the final performance.

We performed controlled ablations by varying the training data composition while keeping the data exposure comparable. In our full setting (DeepVision-103K$_{200}$), our final model, MiMo-VL-7B-DeepVision, was trained for 200 steps on a 3:1 mixture of multimodal math (77K) and visual logic (26K). We evaluated three single-domain counterparts:

- **Math-77K$_{150}$**: math only for 150 steps (same math exposure as DeepVision$_{200}$).
- **Math-77K$_{200}$**: math only for 200 steps (same total exposure as DeepVision$_{200}$).
- **Visual-logic-26K$_{50}$**: visual logic only for 50 steps (same visual logic exposure as DeepVision$_{200}$).

Results in Table 3 show that scaling math training is consistently beneficial: both math-only variants outperform the base model, and extending training from 150 to 200 steps improves every benchmark. However, math alone is not sufficient to reach the best performance. Under the same total exposure, Math-77K$_{200}$ underperforms the mixed

Data Composition	Multimodal Math					General Multimodal			
	WeMath	MathVision	MathVerse	LogicVista	Avg.	MMMU_{val}	MMMU_{pro}	M3CoT	Avg.
MiMo-VL-7B	74.42	50.69	72.71	60.71	64.63	63.77	60.69	70.02	64.83
DeepVision-103K_{200}	**82.98**	55.23	**76.26**	**65.92**	**70.10**	**71.00**	69.19	72.56	**70.92**
w/o visual logic data									
Math-77K_{150}	81.67	54.83	74.23	63.98	68.68	70.00	68.55	72.09	70.21
Math-77K_{200}	82.07	**55.72**	74.74	63.53	69.02	68.50	**69.67**	**72.65**	70.27
w/o multimodal math data									
Visual-logic-26K_{50}	79.54	51.61	73.35	63.98	67.12	68.33	67.34	71.61	69.09
w/o correctness verification									
Unverified-125K_{200}	82.36	53.02	73.47	62.86	67.93	69.33	67.80	71.70	69.61

Table 3: Ablation studies on data composition and quality. We report Pass@1 accuracy (%) across mathematical reasoning and general multimodal benchmarks. All experiments used MiMo-VL-7B-SFT-2508 as the base model.

setting on math average (69.02% vs. 70.10%) with a clear gap on LogicVista (63.53% vs. 65.92%).

These results indicate that introducing visual logic data is valuable, and is further supported by the visual logic-only setting (Visual-logic-26K_{50}), which improves over the foundation model across all benchmarks, demonstrating positive transfer from visual logic to both mathematical and general evaluations. We attribute these gains to two factors: (i) spatial reasoning and pattern recognition are broadly useful primitives shared across mathematical and general multimodal tasks, and (ii) visual logic training directly strengthens these primitives while multimodal math alone does not sufficiently cultivate.

5.3 Necessity of query correctness verification.

After pass-rate filtering, we obtained 99k samples calibrated to the model's capability. To ensure the validity of the reward signals in RLVR, we further applied Gemini-3.0-Flash to remove samples with garbled text or image–text mismatches, and filtered out samples whose answers were inconsistent with Gemini's solutions, discarding an additional 22K samples. However, Wu et al. (2025); Shao et al. (2025) has suggested that LLMs can improve even under spurious rewards, raising doubts about whether strict query correctness is essential for RLVR. To investigate this, we evaluated an unverified variant (Unverified-125K_{200}) which was trained 200 steps on the 99k unverified math data and 26k visual logic data.

Table 3 shows that Unverified_{200} improves over the base model, but remains substantially worse than DeepVision_{200} (67.93% vs. 70.10% on math average; 69.61% vs. 70.92% on general average). This indicates that query correctness verification is necessary because corrupted inputs or incorrect answers hinder the model's progress highlighting that accurate and reliable reward signals are crucial for multimodal RLVR.

6 Conclusion

We present **DeepVision-103K**, a large-scale and verifiable multimodal dataset for RLVR, curated from diverse real-world K12 sources via a three-stage pipeline of validity filtering, pass-rate-based difficulty calibration, and query correctness verification. DeepVision-103K incorporates wide-ranging multimodal mathematical problems and visual logic problems, and covers major visual categories including geometry, analytic plots, charts, and real-world items in mathematical contexts. Training on DeepVision-103K yields top performance on both mathematical and general multimodal tasks. Our further analysis reveals enhanced visual perception, reflection and reasoning capabilities for models trained on DeepVision-103K. We point out multimodal math data and visual logic data contribute to each other in multimodal reasoning, and show the importance of query correctness in multimodal RLVR training.

7 Limitations

While DeepVision-103K substantially increases visual diversity, the distribution is imbalanced (e.g., planar geometry dominates), and some rare element types remain underrepresented. our pipeline relies on strong external models (e.g., Gemini) for query correctness verification, which may introduces potential bias and additional cost, and may filter out a small portion of valid but hard samples. Our dataset focuses on K12-level problems with unique final answers to enable verifiable rewards;

thus it does not fully cover open-ended mathematical tasks (e.g., proof writing, multi-solution problems) that require richer evaluation signals.

References

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. 2025. Qwen3-vl technical report. *Preprint*, arXiv:2511.21631.

Mislav Balunović, Jasper Dekoninck, Ivo Petrov, Nikola Jovanović, and Martin Vechev. 2025. Matharena: Evaluating llms on uncontaminated math competitions.

Qiguang Chen, Libo Qin, Jin Zhang, Zhi Chen, Xiao Xu, and Wanxiang Che. 2024. M^3cot: A novel benchmark for multi-domain multi-step multi-modal chain-of-thought. *Preprint*, arXiv:2405.16473.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, and 181 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

Google. 2025. Gemini3-flash-preiview model card. `https://deepmind.google/models/gemini/flash/`.

Zhiwei He, Tian Liang, Jiahao Xu, Qiuzhi Liu, Xingyu Chen, Yue Wang, Linfeng Song, Dian Yu, Zhenwen Liang, Wenxuan Wang, Zhuosheng Zhang, Rui Wang, Zhaopeng Tu, Haitao Mi, and Dong Yu. 2025. Deepmath-103k: A large-scale, challenging, decontaminated, and verifiable mathematical dataset for advancing reasoning. *Preprint*, arXiv:2504.11456.

Hynek Kydlíček. 2025. Math-Verify: Math Verification Library.

Ang Li, Charles Wang, Kaiyu Yue, Zikui Cai, Ollie Liu, Deqing Fu, Peng Guo, Wang Bill Zhu, Vatsal Sharan, Robin Jia, Willie Neiswanger, Furong Huang, Tom Goldstein, and Micah Goldblum. 2025. Zebra-cot: A dataset for interleaved vision language reasoning. *Preprint*, arXiv:2507.16746.

Wentao Liu, Qianjun Pan, Yi Zhang, Zhuo Liu, Ji Wu, Jie Zhou, Aimin Zhou, Qin Chen, Bo Jiang, and Liang He. 2024. Cmm-math: A chinese multimodal math dataset to evaluate and enhance the mathematics reasoning of large multimodal models. *Preprint*, arXiv:2409.02834.

Pan Lu, Ran Gong, Shibiao Jiang, Liang Qiu, Siyuan Huang, Xiaodan Liang, and Song-Chun Zhu. 2021. Inter-gps: Interpretable geometry problem solving with formal language and symbolic reasoning. *Preprint*, arXiv:2105.04165.

Fanqing Meng, Lingxiao Du, Zongkai Liu, Zhixiang Zhou, Quanfeng Lu, Daocheng Fu, Tiancheng Han, Botian Shi, Wenhai Wang, Junjun He, Kaipeng Zhang, Ping Luo, Yu Qiao, Qiaosheng Zhang, and Wenqi Shao. 2025. Mm-eureka: Exploring the frontiers of multimodal reasoning with rule-based reinforcement learning. *Preprint*, arXiv:2503.07365.

Kaichun Mo, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. 2018. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. *Preprint*, arXiv:1812.02713.

OpenAI, :, Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, Alex Iftimie, Alex Karpenko, Alex Tachard Passos, Alexander Neitz, Alexander Prokofiev, Alexander Wei, Allison Tam, and 244 others. 2024. Openai o1 system card. *Preprint*, arXiv:2412.16720.

Shuai Peng, Di Fu, Liangcai Gao, Xiuqin Zhong, Hongguang Fu, and Zhi Tang. 2024. Multimath: Bridging visual and mathematical reasoning for large language models. *Preprint*, arXiv:2409.00147.

Yingzhe Peng, Gongrui Zhang, Miaosen Zhang, Zhiyuan You, Jie Liu, Qipeng Zhu, Kai Yang, Xingzhong Xu, Xin Geng, and Xu Yang. 2025. Lmm-r1: Empowering 3b lmms with strong reasoning abilities through two-stage rule-based rl. *Preprint*, arXiv:2503.07536.

Runqi Qiao, Qiuna Tan, Guanting Dong, Minhui Wu, Chong Sun, Xiaoshuai Song, Zhuoma GongQue, Shanglin Lei, Zhe Wei, Miaoxuan Zhang, Runfeng Qiao, Yifan Zhang, Xiao Zong, Yida Xu, Muxi Diao, Zhimin Bao, Chen Li, and Honggang Zhang. 2024. We-math: Does your large multimodal model achieve human-like mathematical reasoning? *Preprint*, arXiv:2407.01284.

Runqi Qiao, Qiuna Tan, Peiqing Yang, Yanzi Wang, Xiaowan Wang, Enhui Wan, Sitong Zhou, Guanting Dong, Yuchen Zeng, Yida Xu, Jie Wang, Chong Sun, Chen Li, and Honggang Zhang. 2025. We-math 2.0: A versatile mathbook system for incentivizing visual mathematical reasoning. *Preprint*, arXiv:2508.10433.

PAUL L. Rosin. 2008. *2D Shape Measures for Computer Vision*, pages 347–371.

Rulin Shao, Shuyue Stella Li, Rui Xin, Scott Geng, Yiping Wang, Sewoong Oh, Simon Shaolei Du, Nathan Lambert, Sewon Min, Ranjay Krishna, Yulia Tsvetkov, Hannaneh Hajishirzi, Pang Wei Koh, and Luke Zettlemoyer. 2025. Spurious rewards: Rethinking training signals in rlvr. *Preprint*, arXiv:2506.10947.

Core Team, Zihao Yue, Zhenru Lin, Yifan Song, Weikun Wang, Shuhuai Ren, Shuhao Gu, Shicheng Li, Peidian Li, Liang Zhao, Lei Li, Kainan Bao, Hao Tian, Hailin Zhang, Gang Wang, Dawei Zhu, Cici, Chenhong He, Bowen Ye, and 55 others. 2025. Mimo-vl technical report. *Preprint*, arXiv:2506.03569.

Haozhe Wang, Chao Qu, Zuming Huang, Wei Chu, Fangzhen Lin, and Wenhu Chen. 2025a. Vl-rethinker: Incentivizing self-reflection of vision-language models with reinforcement learning. *Preprint*, arXiv:2504.08837.

Haozhe Wang, Chao Qu, Zuming Huang, Wei Chu, Fangzhen Lin, and Wenhu Chen. 2025b. Vl-rethinker: Incentivizing self-reflection of vision-language models with reinforcement learning. *arXiv preprint arXiv:2504.08837*.

Ke Wang, Junting Pan, Weikang Shi, Zimu Lu, Mingjie Zhan, and Hongsheng Li. 2024. Measuring multimodal mathematical reasoning with math-vision dataset. *Preprint*, arXiv:2402.14804.

Ke Wang, Junting Pan, Linda Wei, Aojun Zhou, Weikang Shi, Zimu Lu, Han Xiao, Yunqiao Yang, Houxing Ren, Mingjie Zhan, and Hongsheng Li. 2025c. Mathcoder-VL: Bridging vision and code for enhanced multimodal mathematical reasoning. In *The 63rd Annual Meeting of the Association for Computational Linguistics*.

Xiyao Wang, Zhengyuan Yang, Chao Feng, Hongjin Lu, Linjie Li, Chung-Ching Lin, Kevin Lin, Furong Huang, and Lijuan Wang. 2025d. Sota with less: Mcts-guided sample selection for data-efficient visual reasoning self-improvement. *Preprint*, arXiv:2504.07934.

Mingqi Wu, Zhihao Zhang, Qiaole Dong, Zhiheng Xi, Jun Zhao, Senjie Jin, Xiaoran Fan, Yuhao Zhou, Huijie Lv, Ming Zhang, Yanwei Fu, Qin Liu, Songyang Zhang, and Qi Zhang. 2025. Reasoning or memorization? unreliable results of reinforcement learning due to data contamination. *Preprint*, arXiv:2507.10532.

Jiaer Xia, Yuhang Zang, Peng Gao, Sharon Li, and Kaiyang Zhou. 2025. Visionary-r1: Mitigating shortcuts in visual reasoning with reinforcement learning. *Preprint*, arXiv:2505.14677.

Yijia Xiao, Edward Sun, Tianyu Liu, and Wei Wang. 2024. Logicvista: Multimodal llm logical reasoning benchmark in visual contexts. *Preprint*, arXiv:2407.04973.

Senqiao Yang, Junyi Li, Xin Lai, Bei Yu, Hengshuang Zhao, and Jiaya Jia. 2025a. Visionthink: Smart and efficient vision language model via reinforcement learning. *Preprint*, arXiv:2507.13348.

Yi Yang, Xiaoxuan He, Hongkun Pan, Xiyan Jiang, Yan Deng, Xingtao Yang, Haoyu Lu, Dacheng Yin, Fengyun Rao, Minfeng Zhu, Bo Zhang, and Wei Chen. 2025b. R1-onevision: Advancing generalized multimodal reasoning through cross-modal formalization. *Preprint*, arXiv:2503.10615.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, and 3 others. 2024a. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. *Preprint*, arXiv:2311.16502.

Xiang Yue, Tianyu Zheng, Yuansheng Ni, Yubo Wang, Kai Zhang, Shengbang Tong, Yuxuan Sun, Botao Yu, Ge Zhang, Huan Sun, Yu Su, Wenhu Chen, and Graham Neubig. 2024b. Mmmu-pro: A more robust multi-discipline multimodal understanding benchmark. *arXiv preprint arXiv:2409.02813*.

Weihao Zeng, Yuzhen Huang, Qian Liu, Wei Liu, Keqing He, Zejun Ma, and Junxian He. 2025a. Simplerl-zoo: Investigating and taming zero reinforcement learning for open base models in the wild. *Preprint*, arXiv:2503.18892.

Yongcheng Zeng, Zexu Sun, Bokai Ji, Erxue Min, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, Haifeng Zhang, Xu Chen, and Jun Wang. 2025b. Cures: From gradient analysis to efficient curriculum learning for reasoning llms. *Preprint*, arXiv:2510.01037.

Yuheng Zha, Kun Zhou, Yujia Wu, Yushu Wang, Jie Feng, Zhi Xu, Shibo Hao, Zhengzhong Liu, Eric P. Xing, and Zhiting Hu. 2025. Vision-g1: Towards general vision language reasoning with multi-domain data curation. *Preprint*, arXiv:2508.12680.

Kaichen Zhang, Keming Wu, Zuhao Yang, Bo Li, Kairui Hu, Bin Wang, Ziwei Liu, Xingxuan Li, and Lidong Bing. 2025. Openmmreasoner: Pushing the frontiers for multimodal reasoning with an open and general recipe. *Preprint*, arXiv:2511.16334.

Renrui Zhang, Dongzhi Jiang, Yichi Zhang, Haokun Lin, Ziyu Guo, Pengshuo Qiu, Aojun Zhou, Pan Lu, Kai-Wei Chang, Peng Gao, and Hongsheng Li. 2024. Mathverse: Does your multi-modal llm truly see the diagrams in visual math problems? *Preprint*, arXiv:2403.14624.

Yifan Zhang and Team Math-AI. 2025. American invitational mathematics examination (aime) 2025.

Chujie Zheng, Shixuan Liu, Mingze Li, Xiong-Hui Chen, Bowen Yu, Chang Gao, Kai Dang, Yuqiong Liu, Rui Men, An Yang, Jingren Zhou, and Junyang Lin. 2025. Group sequence policy optimization. *Preprint*, arXiv:2507.18071.

A Visual Example

In this section, we present cross-category visual combination examples in DeepVision.

Question: Water is poured into a cup (the rate of water pouring remains constant), the relationship between the height of water h in the cup and the pouring time t is shown in the figure. What could be the shape of the cup?
A) ①③ B) ①② C) ①②③ D) ②③

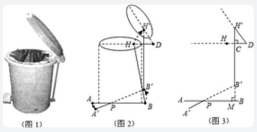
Image :

Visual Elements:
Solid Geometry: Cone, Cylinder, Frustum
Analytic Plots: Coordinate System, Function Curve

Figure 11: Solid Geometry & Analytic Plots.

Question: A household trash can has a horizontal pedal \(AB\) that rotates about a fixed point \(P\)....In the opened position (Figure 3), t find the minimum height of the pedal \(AB\) above the ground so that \(\angle HDH' \ge 60^\circ\). (Answer to two significant figures; \(\sqrt2\approx1.41,\ \sqrt3\approx1.73\).)

Image :



Visual Elements:
Planar Geometry: Triangle, Angles, Parallel Lines, Perpendicular Lines
Solid Geometry: Cylinder
Charts and Diagrams: Mechanical Diagram

Figure 12: Planar Geometry & Solid Geometry & Analytic Plots.

B Data Construction

For math data, wo did use all the data within $[\frac{1}{8}, \frac{4}{8}]$ pass rate. But did not use all the $[\frac{5}{8}, \frac{6}{8}]$ data (as the are too much), we recall them by knowledge points which has a low distribution in $[\frac{1}{8}, \frac{4}{8}]$ pass rate data.

For visual logic data, we only used tangrams, maze, chess data from Zebra-CoT (Li et al., 2025), with pass rate at $[\frac{3}{8}, \frac{4}{8}]$. This choice was made to broaden the training distribution while keeping the dataset size manageable.

We list the data collection protocol of our datasource.

Data Source	License
MM-MathInstruct-3M (Wang et al., 2025c)	Apache 2.0 https://huggingface.co/datasets/MathLLMs/MM-MathInstruct
MultiMath-300K (Peng et al., 2024)	Unset https://huggingface.co/datasets/pengshuai-rin/multimath-300k
Zebra-CoT (Li et al., 2025)	cc-by-nc-4.0 https://huggingface.co/datasets/multimodal-reasoning-lab/Zebra-CoT

Table 4: Licenses and usage permissions for the data sources used in this work.

C Training Details

We used `verl` as the training framework. Configurations for training DeepVision series models are listed in Table 5.

Config	Value
lr	1e-6
kl_coef	1e-3
max_prompt_length	2K
max_response_length	16K
gen_batch_size	512
train_batch_size	256
mini_batch_size	64
micro_batch_size	32
group_filtering	acc
clip_ratio_low	1e-3
clip_ratio_high	1e-4
temperature	1.0
rollout.n	16
total_training_steps	200

Table 5: Configurations for training DeepVision series models.

We used 32 H20 GPU for a single training, a training step cost 0.5h. We used the following prompt template during training and evaluation.

[1]Extremely high because OpenMMReasoner includes ViRL-39K(Wang et al., 2025b), which includes M^{3}CoT.

Training / Evaluation Prompt Template

You are a multimodal reasoning assistant. You receive images and texts, perform step-by-step reasoning (including re-checking the image) before producing the final answer. Please provide a clear, concise answer inside \boxed{} tag. For multiple choice questions, put only the letter like \boxed{A} without any additional text. For fill-in-the-blank and problem-solving questions, put only the final answer.

D Evaluation Details

We provide detailed information about the benchmarks used for evaluation and the inference hyperparameters for each model.

D.1 Benchmarks

We evaluated our models across three categories of benchmarks, as summarized in Table 6.

Table 6: Overview of evaluation benchmarks.

Category	Benchmark	#Samples	Reference
Multimodal Math	WeMath	1,740	(Qiao et al., 2024)
	MathVision	3,040	(Wang et al., 2024)
	$\text{MathVerse}_{\text{vision}}$	788	(Zhang et al., 2024)
	LogicVista	448	(Xiao et al., 2024)
General Multimodal	M^3CoT	2,318	(Chen et al., 2024)
	MMMU_{val}	900	(Yue et al., 2024a)
	$\text{MMMU}_{\text{Pro\_full}}$	1,730	(Yue et al., 2024b)
Text-only Math	AIME 2025	30	(Zhang and Math-AI, 2025)
	HMMT 2025	30	(Balunović et al., 2025)

D.2 Inference Hyperparameters

We used different inference hyperparameters for different model families to ensure optimal performance. The detailed configurations are listed in Table 7.

Table 7: Inference hyperparameters for each model family.

Parameter	Qwen3-VL-Thinking	Qwen3-VL-Instruct	MiMo-VL-(SFT/RL)
top_p	0.95	0.8	0.95
top_k	20	20	–
temperature	1.0	0.7	0.3
repetition_penalty	1.0	1.0	–
presence_penalty	0.0	1.5	–
max_tokens	32,768	32,768	32,768

For Qwen3-VL-DeepVision models, we adopted the same hyperparameters as Qwen3-VL-Instruct. For MiMo-VL-DeepVision, we adopted the same hyperparameters as MiMo-VL.

D.3 Evaluation Method

For each benchmark, we first calculate accuracy with MathVerify (Kydlíček, 2025), then prompt GPT-5-mini to re-judge cases marked as incorrect by MathVerify to reduce false negatives caused by parsing errors, equivalent expressions, or formatting variations. We use the revised judgment as the final label.

E Potential Risks

We do not anticipate significant potential risks from this work. DeepVision-103K is derived from publicly available K12-level educational content and is designed for verifiable-answer multimodal reasoning rather than sensitive decision-making. The dataset contains no personal identifiers, and our curation process filters out corrupted or unsafe samples.