
MiCo: Multi-image Contrast for Reinforcement Visual Reasoning – Supplementary Materials

Anonymous Author(s)

Affiliation

Address

email

1 In the supplementary materials, we first elaborate on more implementation details in Appendix A.
2 Then, then we report more experimental results in Appendix B. Afterwards, we give more qualitative
3 examples in Appendix C to demonstrate and analyze the incentivized reasoning process. Finally, we
4 discuss the potential social impact in Appendix D.

5 A More Implementation Details

6 **Prompt expansion.** In the main paper, we give an example of the user question as follows:

Reasoning Template of MiCo

Reasoning Prompt: First output the thinking process in `<think>` `</think>` and give the final answer in `<answer>` `</answer>` tags.

User Question: Regardless of the augmentation, are image1 and image2 the same? How about image2 and image3, image1 and image3? Only return T(True) or F(False) in `<answer>` `</answer>`, for example `<think>` `</think>` `<answer>`TFT`</answer>`.

7

8 As we verified in Tab.2 (c) that prompt diversity plays a critical role in encouraging the model to
9 generalize its reasoning ability. To this end, we design a systematic prompt expansion strategy along
10 two axes: question phrasing and comparison structure.

11 First, we construct both forward questions (*e.g.*, “Are image1 and image2 the same?”) and reverse
12 questions (*e.g.*, “Are image1 and image2 different?”), allowing the model to reason under varied
13 semantic instructions. Second, we vary the combinations of image pairs being queried (*e.g.*, image1
14 vs. image2, image2 vs. image3, image1 vs. image3), ensuring comprehensive coverage of possible
15 relations. Through controlled sampling, we balance the resulting answer types (*e.g.*, TFT, TTF, FFF,
16 *etc.*), avoiding label bias and promoting robust multi-image understanding across diverse visual
17 scenarios and logical outcomes.

18 **Data filtering.** We apply data filtering When selecting similar but different image pairs from the
19 image editing dataset and video frames. For image editing data, to ensure high-quality training
20 samples for visual reasoning, we implement a pixel-wise comparison strategy to filter out image pairs
21 with overly large differences. Specifically, for each image pair, we compute the absolute difference
22 across corresponding pixels. For RGB images, we first average the per-channel differences to obtain
23 a single grayscale difference map. A pixel is considered “different” if its value exceeds a predefined
24 threshold (30). We then calculate the ratio of differing pixels across the entire image. If this difference
25 ratio exceeds 0.8, the image pair is discarded. This simple yet effective rule ensures that the model
26 focuses on learning from subtle and semantically meaningful variations, rather than from trivially
27 dissimilar or noisy pairs. For video data, we calculate the SSIM between the selected two frames,
28 and remove the samples with an SSIM value greater than 0.95 to filter nearly the same images.

Table 1: **Task analysis on MMStar.** We report performance on different sub-tasks to evaluate the generalization ability of MiCo compared to the Qwen2.5VL baseline.

	Overall	Coarse Perc.	Fine Perc.	Inst. Reason.	Logic Reason.	Math	Sci. & Tech.
Qwen2.5VL-7B	64.07	72.00	60.40	69.60	67.20	65.60	49.60
MiCo-7B	65.33 +	73.20 +	59.60 -	72.00 +	68.80 +	69.20 +	49.20 -

Table 2: **Task analysis on MuirBench.** Performance breakdown across 12 sub-tasks to evaluate the fine-grained generalization ability of MiCo.

	Action	Attr. Sim.	Cartoon	Counting	Diagram	Diff. Spot.
Qwen2.5VL-7B	40.85	58.67	46.15	34.19	77.89	54.41
MiCo-7B	40.85	57.65 -	46.15	34.19	79.90 +	55.29 +
	Geo. Und.	Img-Text	Ordering	Scene Und.	Vis. Grnd.	Vis. Ret.
Qwen2.5VL-7B	49.00	72.63	14.06	61.83	33.33	63.70
MiCo-7B	53.00 +	74.14 +	20.31 +	63.98 +	35.71 +	71.23 +

Table 3: **Task analysis on BLINK.** Performance comparison across 14 sub-tasks to evaluate the generalization ability of MiCo.

	ArtStyle	Counting	Forensic	FuncCorr	IQTest	Jigsaw	MultiView
Qwen2.5VL-7B	69.23	70.83	48.48	27.69	18.00	59.33	54.89
MiCo-7B	72.65 +	70.00 -	47.27 -	30.77 +	26.00 +	69.33 +	42.11 -
	ObjLoc	RelDepth	RelReflect.	SemCorr	SpatialRel	VisCorr	VisSim
Qwen2.5VL-7B	54.10	81.45	40.30	33.09	88.81	52.33	86.67
MiCo-7B	54.92 +	76.61 -	31.34 -	41.73 +	90.21 +	61.05 +	85.19 -

29 **Differences with NoisyRollout [3].** Recently, NoisyRollout [3] also leverages image augmentation to
 30 enhance GRPO [4]. However, our Augmented GRPO is fundamentally different from their strategy.

31 NoisyRollout introduces a hybrid rollout strategy by mixing trajectories from both clean and
 32 moderately distorted images. Specifically, they add Gaussian noise on the images. Then, sample
 33 n trajectories on noisy images and another n trajectories on clean images. In this way, they get $2n$
 34 trajectories in total with more diversity. Afterwards, NoisyRollout calculates the advantages based on
 35 the hybrid trajectories and optimizes the policy model on clean images.

36 Differently, our Augmented GRPO is inspired by semi-supervised learning in computer vision, where
 37 we sample n trajectories with weak image augmentation. We assume that it would be easier for the
 38 policy model to get more high-quality CoTs with weak image augmentation. Then, we calculate the
 39 advantages using these high-quality CoTs and optimize the model with stronger augmented images.
 40 This allows us to obtain correct and informative Chain-of-Thoughts (CoTs) even for samples that
 41 would otherwise be answered incorrectly under strong augmentation, thereby improving the model’s
 42 generalization to more challenging examples.

43 B More Experimental Results

44 B.1 More quantitative results

45 To further evaluate the generalization ability of MiCo, we report its performance on a diverse set
 46 of sub-tasks from three comprehensive benchmarks: MMStar [1], MuirBench [5], and BLINK [2].
 47 As shown in Tables 1, 2, and 3, MiCo consistently improves upon Qwen2.5VL-7B across most
 48 reasoning-related tasks.

49 On **MMStar**, MiCo achieves notable gains in instance reasoning, logical reasoning, and math,
 50 suggesting enhanced multi-step inference and abstraction capabilities. The performance on fine-
 51 grained perception slightly decreases, indicating potential room for improvement in precise visual
 52 attribute understanding.

53 For **MuirBench**, MiCo improves on 8 out of 12 sub-tasks, including Diagram Understanding,
 54 Geographic Understanding, and Visual Retrieval. These tasks involve complex spatial, contextual,
 55 or comparative reasoning, showing the effectiveness of our visual comparison objective. Tasks like
 56 Attribute Similarity and Counting show marginal drops, possibly due to their reliance on absolute
 57 visual matching rather than relational reasoning.

On **BLINK**, MiCo shows strong improvements on Functional Correspondence, IQ Test, Jigsaw, Semantic Correspondence, and Visual Correspondence—all of which require visual logic, spatial matching, or multi-view inference. However, tasks such as Multi-view Reasoning and Relative Reflectance exhibit declines, suggesting future efforts could focus on making the model more robust to challenging viewpoint shifts and subtle appearance variations.

Overall, these quantitative results demonstrate that MiCo is particularly effective at improving tasks involving reasoning, structure, and comparison, while fine-grained low-level perception remains a direction for future enhancement.

B.2 Unsuccessful Attempts

Throughout our exploration, we experimented with several alternative approaches that ultimately did not lead to improved performance. For completeness and to facilitate future research, we summarize these unsuccessful attempts and provide insights into why they may have failed.

Confidence reweighting. Since our task is formulated as answering T/F questions, even when evaluating three comparisons simultaneously, there remains a non-trivial chance (12.5%) of obtaining the correct answer purely by guessing. To reduce the impact of such randomness, we explored adding an additional reward or weight based on the model’s answer confidence. Specifically, we experimented with several approaches to compute confidence scores from the softmax probabilities of the output tokens. However, these confidence-based reweighting strategies did not yield any performance improvements. We analyze that this may be due to the fact that the softmax probability of the predicted token does not reliably reflect the model’s true certainty about the overall answer. In particular, the model may assign high confidence to tokens that are syntactically or semantically unrelated to the actual correctness of the reasoning (e.g., punctuation, or irrelevant words within the output). As a result, the computed “confidence” can be misleading, making it an ineffective signal for reward shaping.

Importance sampling. As in our Augmented GRPO, we sample the trajectories on simple examples with weak augmentations, but we use the trajectory to optimize harder examples with strong augmentations. This might cause misalignment similar to offline reinforcement learning. In this way, we apply importance sampling, which calculates the probability gap between the trajectories for the simple and hard examples as a weight to reweight the reward/advantages. This strategy could not bring improvements. We suspect that although importance sampling is theoretically justified, it may interfere with the core optimization dynamics of GRPO. Specifically, GRPO relies on the relative ranking of trajectories within a group to compute structured advantages. Introducing importance weights—derived from distribution shifts—may distort this internal ranking or inject instability into the reward signals. Additionally, the token-level probability changes caused by visual augmentations can be noisy or poorly calibrated, making the computed importance weights unreliable in practice.

C Qualitative Analysis

We add more visual demonstrations for the reasoning ability of MiCo in Fig. 1 and Fig. 2. These qualitative examples demonstrate the strong reasoning capability of MiCo across various visual tasks. In Figure 1, the model exhibits detailed step-by-step analysis to distinguish visual differences, count distinct objects, and solve jigsaw-like problems. Rather than relying on superficial features, MiCo actively grounds its reasoning in object identity, pose, structure, and scene semantics. For example, in the toy comparison case, it not only detects the number of different objects but also considers subtle variations in assembly, model type, and color configuration. In the jigsaw task, it correctly identifies missing or manipulated segments by referencing spatial consistency and scene-level context.

Figure 2 further highlights MiCo’s ability to tackle abstract reasoning challenges. In the IQ-style pattern recognition question, the model deduces a complex symbol progression rule based on character groupings and positions. For functional correspondence and spatial matching, it accurately aligns image pairs by understanding object affordances and relative part placement. Additionally, in the visual similarity task, it discerns fine-grained geometric and design attributes to match images at a structural level rather than based on superficial texture or color.



Figure 1: **Demonstrations** for detailed comparison and jigsaw solving.

Together, these examples reveal that MiCo does not merely perform image-text matching but is capable of systematic, multi-step reasoning grounded in visual understanding. This reflects its generalization ability across both low-level visual tasks and high-level abstract reasoning challenges.

D Potential Social Impact

MiCo explores a self-supervised and reinforcement learning-based approach to improve multi-image reasoning in vision-language models without relying on human-annotated question-answer pairs. By leveraging intrinsic visual constraints, such as consistency across augmented views and differences



Figure 2: **Demonstrations** for IQ test, functional correspondence, and visual similarity.

115 between similar images, MiCo significantly reduces the need for labor-intensive data curation. This
 116 has the potential to democratize the development of reasoning-capable AI systems, making them more
 117 accessible in low-resource settings or for underrepresented languages and domains where curated
 118 datasets are scarce.

119 However, as with any powerful vision-language technology, there is a risk of misuse, particularly in
 120 applications involving surveillance, misinformation, or unauthorized inference of user intent from
 121 visual data. MiCo’s improved ability to perform fine-grained comparisons across images could be
 122 exploited in privacy-invasive scenarios if deployed irresponsibly. To mitigate such risks, we advocate
 123 for deploying MiCo in alignment with responsible AI guidelines, ensuring transparency, consent,
 124 and clear boundaries in its application domains. In practice, this includes integrating robust sensitive
 125 content filtering, restricting deployment in high-stakes or privacy-sensitive scenarios, and establishing
 126 human-in-the-loop mechanisms for critical decision-making processes.

References

- [1] Lin Chen, Jinsong Li, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Zehui Chen, Haodong Duan, Jiaqi Wang, Yu Qiao, Dahua Lin, et al. Are we on the right way for evaluating large vision-language models? *NeurIPS*, 2024. 2
- [2] Xingyu Fu, Yushi Hu, Bangzheng Li, Yu Feng, Haoyu Wang, Xudong Lin, Dan Roth, Noah A Smith, Wei-Chiu Ma, and Ranjay Krishna. Blink: Multimodal large language models can see but not perceive. In *ECCV*, 2024. 2
- [3] Xiangyan Liu, Jinjie Ni, Zijian Wu, Chao Du, Longxu Dou, Haonan Wang, Tianyu Pang, and Michael Qizhe Shieh. Noisyrollout: Reinforcing visual reasoning with data augmentation. *arXiv:2504.13055*, 2025. 2
- [4] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv:2402.03300*, 2024. 2
- [5] Fei Wang, Xingyu Fu, James Y Huang, Zekun Li, Qin Liu, Xiaogeng Liu, Mingyu Derek Ma, Nan Xu, Wenxuan Zhou, Kai Zhang, et al. Muirbench: A comprehensive benchmark for robust multi-image understanding. In *ICLR*, 2025. 2