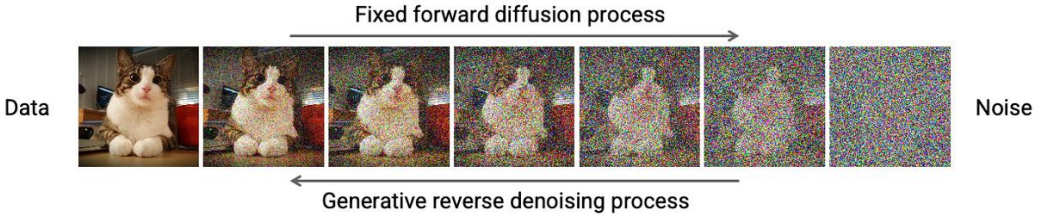
MILR: Improving Multimodal Image Generation via Test-Time Latent Reasoning

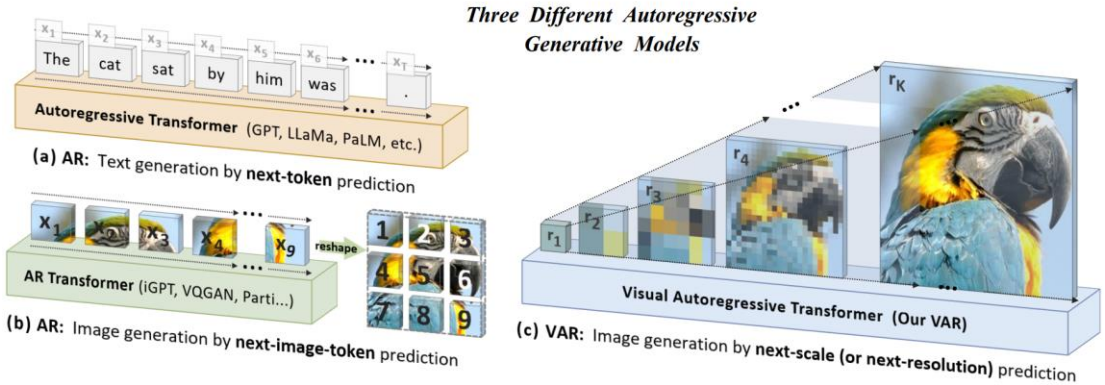
Yapeng Mi[1,2] **Yanpeng Zhao**[2 †✉] **Hengli Li**[2,3] **Chenxi Li**[2,4]
Huimin Wu[2] **Xiaojian Ma**[2] **Song-Chun Zhu**[2,3] **Ying Nian Wu**[5] **Qing Li**[2 ✉]
[1]University of Science and Technology of China
[2]State Key Laboratory of General Artificial Intelligence, BIGAI
[3]Peking University [4]Tsinghua University [5]University of California, Los Angeles

Text-guided image generation



Diffusion

Eg. Stable Diffusion3, FLUX.1-dev…



Autoregressive

Eg. LlamaGen, Emu3, Show-o, Janus-Pro

making it feasible to build unified multimodal understanding and generation (MUG)

Image generation needs reasoning ability

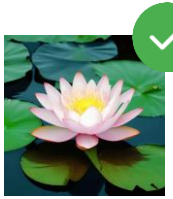
Prompt: A flower that symbolizes purity in China.

Answer: [Referring to lotus]

only prompt

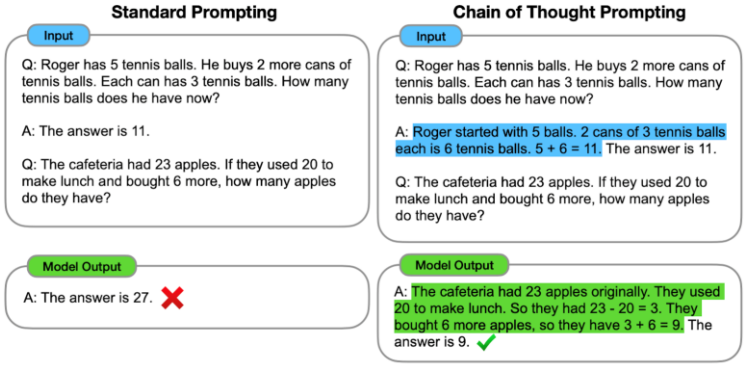
SD3-medium

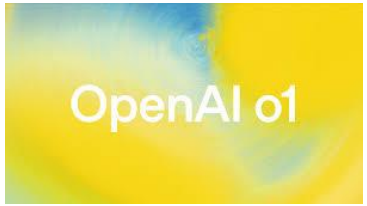
with answer in prompt



SD3-medium

Traditional models are limited in generating images in a single-shot fashion and thus are unable to resolve potential defects



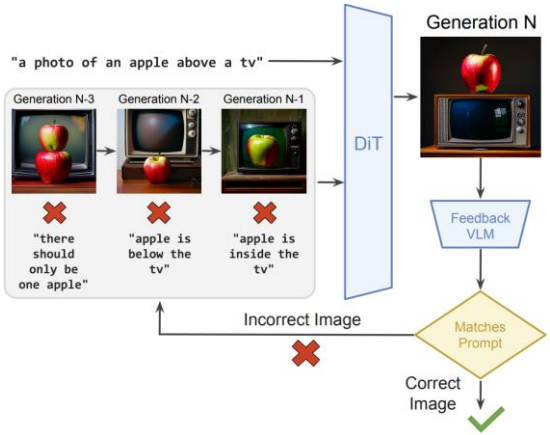


deepseek R1

Inspired by the success of reasoning-augmented LLMs.

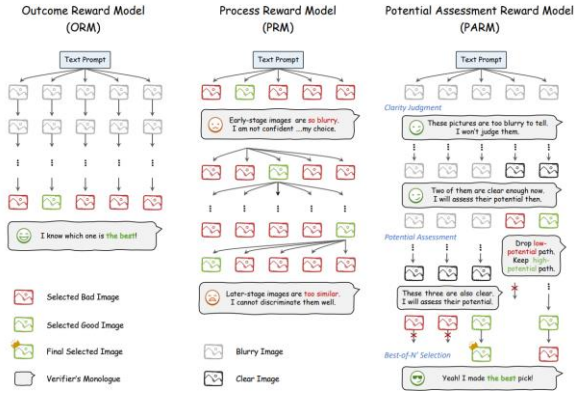
Reasoning-augmented image generation

Language Space



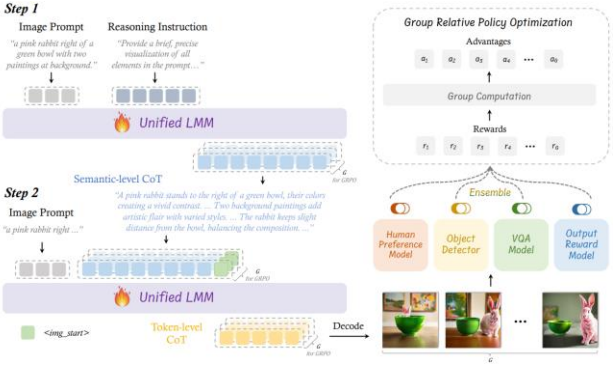
Reflect-DiT

Image Space



PARM

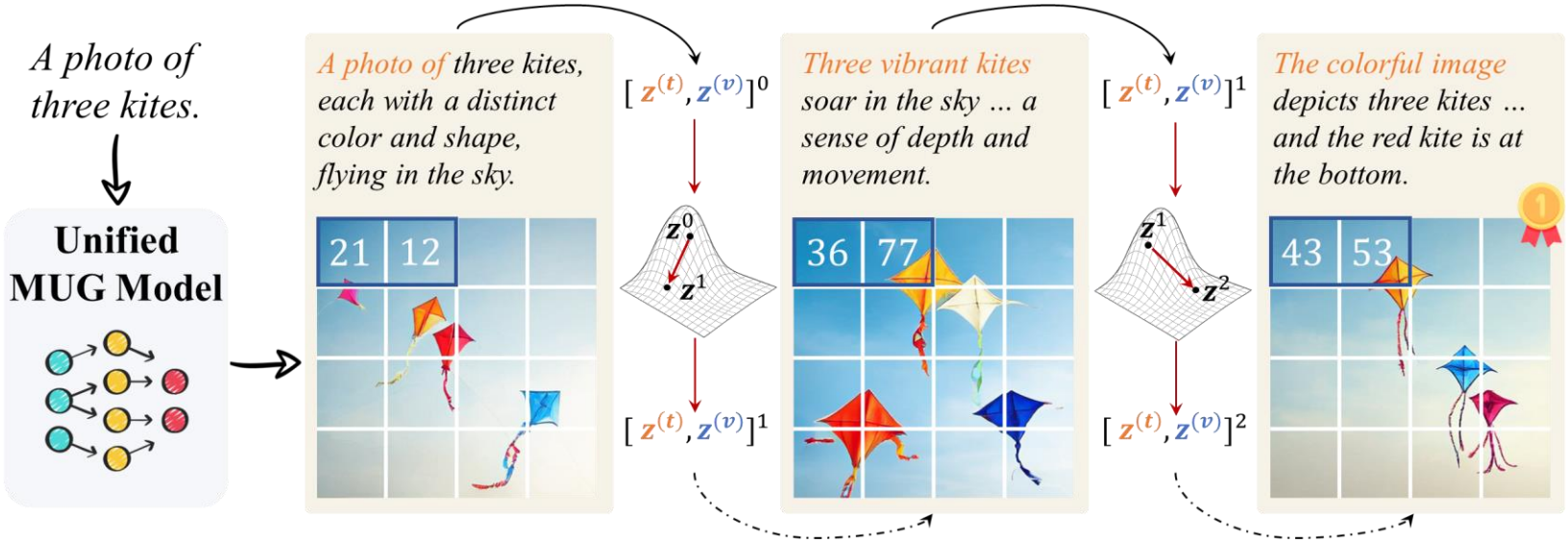
Language & Image Space



T2I-R1

1. Early works lack a mechanism for synergistic reasoning across the two spaces.
2. require carefully curated reasoning data and depend on model fine-tuning, complex and costly.

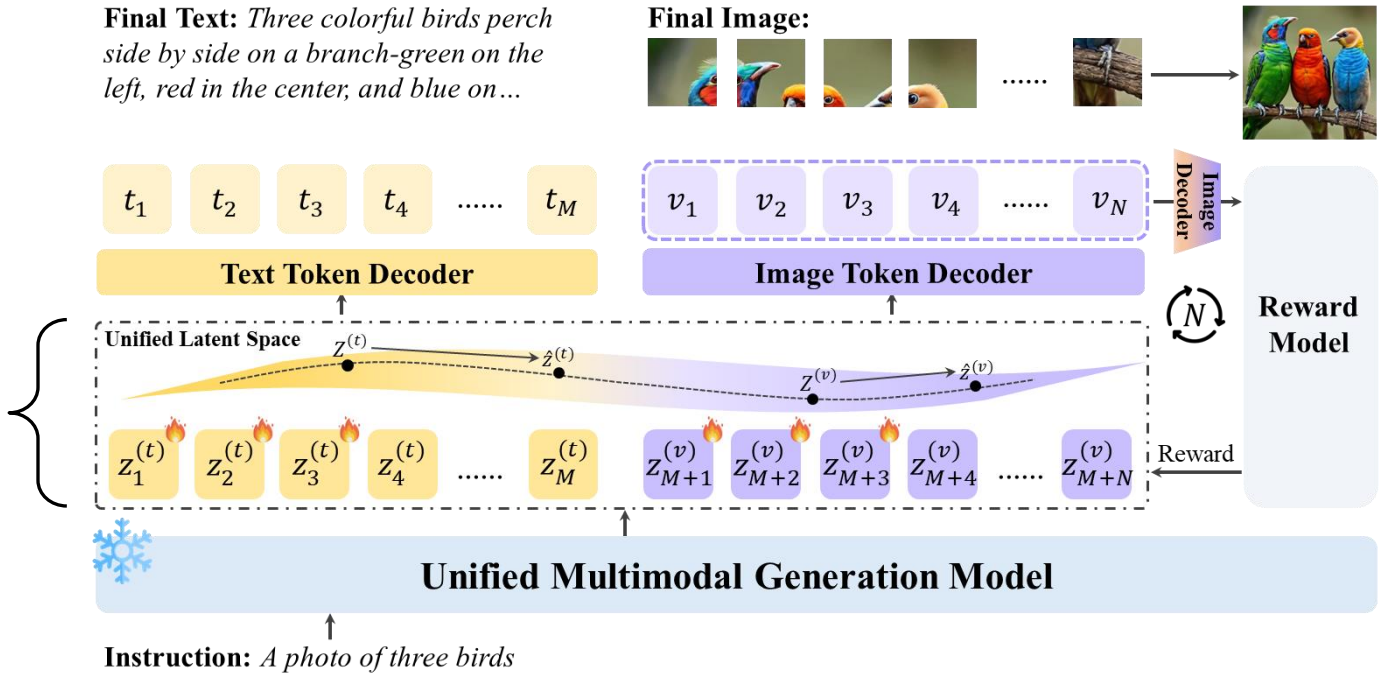
MILR: Multimodal Image generation via test-time Latent Reasoning



1. reason in a unified latent vector space that encodes both text and images

2. use the policy gradient method to perform test-time image generation without finetuning.

Framework



the outputs of the last Transformer layer

Core policy gradient optimization

REINFORCE (Williams, 1992)

$$\mathbf{z}^{k+1} \leftarrow \mathbf{z}^k + \eta \cdot \mathcal{J}(\mathbf{z}^k),$$

$$\mathcal{J}(\mathbf{z}^k) = \mathbb{E}_{V_f \sim p(\cdot|\mathbf{z}^k, c)}[R(V_f, c)\nabla_{\mathbf{z}} \log(p(\mathbf{t}, \mathbf{v}|\mathbf{z}^k))],$$

Experiment Implementation

Algorithm 1 MILR

Require: Instruction c, Learning rate η, MUG model p, reward threshold τ, text and image fraction $\lambda_t, \lambda_v \in (0, 1]$, optimization steps K, Reasoning strategy $\mathcal{S} \in \{\text{JOINT}, \text{ALT}, \text{T2V}\}$

$\mathbf{z}, \mathbf{t}, \mathbf{v} \leftarrow p(\mathbf{t}, \mathbf{v}|c)$ ▷ Initial latent vectors: Equation (1)

$V_f \sim p(\cdot|\mathbf{v})$

$r \leftarrow R(V_f, c)$ ▷ Reward Calculation

$\mathbf{z}^0 \leftarrow (z^{(t)}_{1:\lambda_t|\mathbf{t}|}; z^{(v)}_{1:\lambda_v|\mathbf{v}|})$ ▷ Set λ_t and λ_v fraction

$k \leftarrow 1$ ▷ Starting step index

while $k \leq K$ and $r \leq \tau$ **do**

  $V_f, \mathbf{z}^k \leftarrow \mathit{LatentReasoning}_{\mathcal{S}}(\mathbf{z}^{k-1}, \eta, k, K, c, p)$ ▷ Select a strategy: Algorithm 2

  $r \leftarrow R(V_f, c)$

  $k \leftarrow k + 1$

end while

return V_f

Algorithm 2 LatentReasoning (default: JOINT)

Note. $\mathcal{J}(\mathbf{z})$ as defined in Equation (6); V_f sampled as in Equation (4); T_f is the full reasoning text. $\mathbf{z}^{(t)}_k$ and $\mathbf{z}^{(v)}_k$ denote the text and visual latents at iteration k.

(a) JOINT

$\mathbf{z}^k \leftarrow \mathbf{z}^{k-1} + \eta \cdot \mathcal{J}(\mathbf{z}^{k-1})$

$V_f \sim p(V_f \mid \mathbf{z}^k, c)$

return $V_f, \mathbf{z}^k$

(b) ALT

if $k \bmod 2 = 1$ **then**

  $\mathbf{z}^{(t)}_k \leftarrow \mathbf{z}^{(t)}_{k-1} + \eta \cdot \mathcal{J}(\mathbf{z}^{(t)}_{k-1})$

  $T_f \sim p(T_f|\mathbf{z}^{(t)}_k, c)$

  $V_f \sim p(V_f|\mathbf{z}^{(v)}_{k-1}, T_f, c)$

else

  $\mathbf{z}^{(v)}_k \leftarrow \mathbf{z}^{(v)}_{k-1} + \eta \cdot \mathcal{J}(\mathbf{z}^{(v)}_{k-1})$

  $V_f \sim p(V_f|\mathbf{z}^{(v)}_k, T_f, c)$

end if

return $V_f, \mathbf{z}^k$

(c) T2V

if $k \leq \lfloor K/2 \rfloor$ **then**

  $\mathbf{z}^{(t)}_k \leftarrow \mathbf{z}^{(t)}_{k-1} + \eta \cdot \mathcal{J}(\mathbf{z}^{(t)}_{k-1})$

  $T_f \sim p(T_f|\mathbf{z}^{(t)}_k, c)$

  $V_f \sim p(V_f|\mathbf{z}^{(v)}_{k-1}, T_f, c)$

else

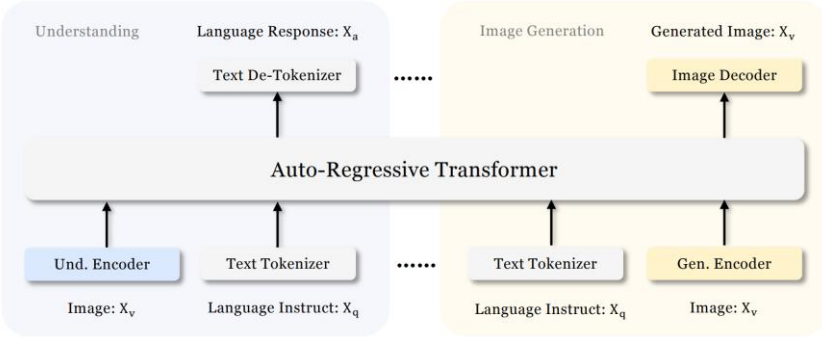
  $\mathbf{z}^{(v)}_k \leftarrow \mathbf{z}^{(v)}_{k-1} + \eta \cdot \mathcal{J}(\mathbf{z}^{(v)}_{k-1})$

  $V_f \sim p(V_f|\mathbf{z}^{(v)}_k, T_f, c)$

end if

return $V_f, \mathbf{z}^k$

Base model: Janus-Pro



fraction set: $\lambda_t = 0.2$ $\lambda_v = 0.02$

Adam optimizer, learning rate = 0.03

a single NVIDIA A100 80GB GPU

Main results

Table 1: Results on GenEval. The best score is in bold, and the second best is underlined.

Method	Single Obj. ↑	Two Obj. ↑	Counting ↑	Colors ↑	Position ↑	Attr. Binding ↑	Overall ↑
Non-reasoning Models							
LlamaGen (Sun et al., 2024)	0.71	0.34	0.21	0.58	0.07	0.04	0.32
Emu3 (Wang et al., 2024)	0.98	0.71	0.34	0.81	0.17	0.21	0.54
FLUX.1-dev (Black Forest Labs, 2024)	0.98	0.79	0.73	0.77	0.22	0.45	0.66
DALL-E 3 (Betker et al., 2023)	0.96	0.87	0.47	0.83	0.43	0.45	0.67
SD3-Medium (Esser et al., 2024)	0.99	0.94	0.72	0.89	0.33	0.60	0.74
BAGEL (Deng et al., 2025)	0.99	0.94	0.81	0.88	0.64	0.63	0.82
GPT-4o (OpenAI, 2025)	0.99	0.92	0.85	0.91	0.75	0.66	0.85
Training-based Reasoning Models							
GoT-R1 (Duan et al., 2025)	0.99	0.94	0.50	0.90	0.46	0.68	0.75
T2I-R1 (Jiang et al., 2025a)	0.99	0.91	0.53	0.91	0.76	0.65	0.79
Flow-GRPO (Liu et al., 2025a)	**1.00**	**0.99**	**0.95**	0.92	**0.99**	0.86	**0.95**
ReasonGen-R1 (Zhang et al., 2025b)	0.99	0.94	0.62	0.90	0.84	0.84	0.86
Janus-Pro-7B(+GRPO) (Tong et al., 2025)	0.99	0.87	0.61	0.87	0.82	0.68	0.81
Janus-Pro-7B(+DPO) (Guo et al., 2025)	0.99	0.89	0.65	0.92	0.82	0.72	0.83
Test-time Reasoning Models							
Reflect-DiT (Li et al., 2025b)	0.98	0.96	0.80	0.88	0.66	0.60	0.81
ReflectionFlow (Zhuo et al., 2025)	**1.00**	0.98	0.90	0.96	0.93	0.72	0.91
Janus-Pro-7B(+Text Enhanced Reasoning)	0.98	0.91	0.55	0.89	0.74	0.67	0.79
Janus-Pro-7B(+Best-of-N)	0.99	0.96	0.89	0.93	0.92	0.80	0.91
Janus-Pro-7B(+PARM) (Guo et al., 2025)	**1.00**	0.95	0.80	0.93	0.91	0.85	0.91
Janus-Pro-1B (Chen et al., 2025)	0.98	0.82	0.51	0.89	0.65	0.56	0.73
Janus-Pro-1B+MILR	**1.00**	0.91	0.78	0.92	0.86	0.86	0.89
Janus-Pro 7B (Chen et al., 2025)	0.98	0.85	0.56	0.89	0.77	0.64	0.78
Janus-Pro-7B+MILR	**1.00**	0.96	0.90	**0.98**	0.98	**0.91**	**0.95**

Table 2: Results on T2I-CompBench and WISE. The best is in bold, and the second is in underlined.

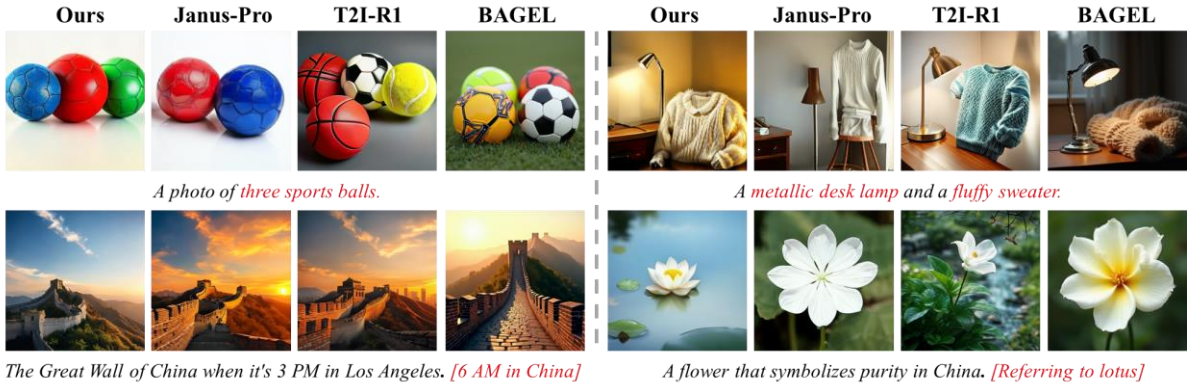
Method	T2I-CompBench							WISE
	Color ↑	Shape ↑	Texture ↑	Spatial ↑	Non-Spatial ↑	Complex. ↑	Overall ↑	Avg ↑
Non-reasoning Models								
PixArt-α Chen et al. (2023)	0.6690	0.4927	0.6477	0.2064	0.3197	0.3433	0.4465	0.47
FLUX.1-dev Black Forest Labs (2024)	0.7407	0.5718	0.6922	0.2863	0.3127	0.3703	0.4957	0.50
DALL-E 3 Betker et al. (2023)	0.7785	**0.6209**	0.7036	0.2865	0.3003	0.3773	0.5112	-
SD3-Medium Esser et al. (2024)	0.8132	0.5885	0.7334	0.3200	0.3140	0.3771	0.5244	0.42
Show-o Xie et al. (2024)	0.5600	0.4100	0.4600	0.2000	0.3000	0.2900	0.3700	0.30
BAGEL Deng et al. (2025)	0.8027	0.5685	0.7021	0.3488	0.3101	0.3824	0.5191	0.52
Training-based Reasoning Models								
T2I-R1 Jiang et al. (2025a)	0.8130	0.5852	0.7243	0.3378	0.3090	0.3993	0.5281	0.54
GoT-R1 Duan et al. (2025)	0.8139	0.5549	**0.7339**	0.3306	0.3169	**0.3944**	0.5241	-
Janus-Pro-7B(+GRPO) Tong et al. (2025)	0.7721	0.5366	0.7317	0.2869	0.3087	0.3697	0.5010	-
Test-time Reasoning Models								
Show-o + PARM Guo et al. (2025)	0.7500	0.5600	0.6600	0.2900	0.3100	0.3700	0.4900	-
Janus-Pro-7B(+Text Enhanced Reasoning)	0.7087	0.4419	0.5821	0.2597	0.3072	0.3761	0.4459	0.46
Janus-Pro-7B(+Best-of-N)	0.7089	0.4925	0.7089	0.3542	**0.3262**	0.3721	0.4938	0.52
Janus-Pro-1B Chen et al. (2025)	0.3411	0.2261	0.2696	0.0968	0.2808	0.2721	0.2478	0.26
Janus-Pro-1B+MILR	0.6066	0.2796	0.4177	0.2796	0.2622	0.2613	0.3512	0.40
Janus-Pro-7B Chen et al. (2025)	0.6359	0.3528	0.4936	0.2061	0.3085	0.3559	0.3921	0.35
Janus-Pro-7B+MILR	**0.8508**	0.5117	0.6949	**0.4613**	0.3078	0.3684	**0.5325**	**0.63**

1. Our method achieve good performance across GenEval, T2I-CompBench, WISE benchmark

2. MILR surpasses frontier non-reasoning models (SD3-Medium, BAGEL), training-based reasoning models (e.g., GoT-R1 and T2I-R1), test-time reasoning models (ReflectionFlow and PARM)

Qualitative results

MILR is capable of geometric, temporal, and cultural reasoning.

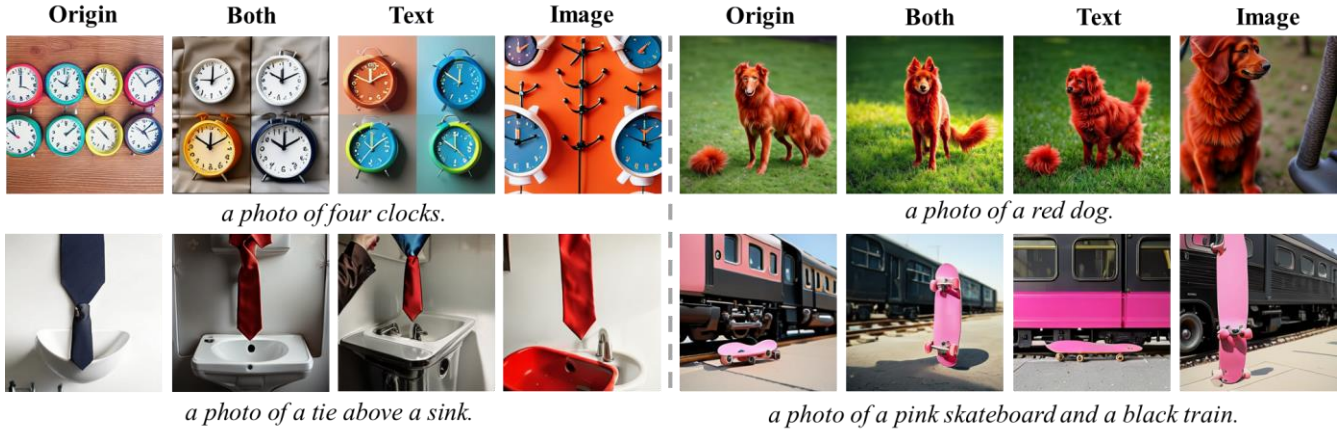


Ablations

Table 3: Ablations of MILR on GenEval, T2I-CompBench, and WISE

Method	GenEval							T2I-CompBench	WISE
	Single Obj.	Two Obj.	Counting	Color	Pos.	Attr. Binding	Overall	Overall	Avg
Janus-Pro-7B+MILR (ours)	1.00	0.96	0.90	0.98	0.98	0.91	**0.95**	**0.5325**	**0.63**
w/o MILR	0.98	0.85	0.56	0.89	0.77	0.64	0.78	0.3921	0.35
w/o Image	1.00	1.00	0.91	0.95	0.95	0.88	0.94	0.5210	0.61
w/o Text	1.00	0.95	0.88	0.91	0.97	0.89	0.93	0.5043	0.56

Joint image-text reasoning in the unified latent space leads to the best performance.



Analysis

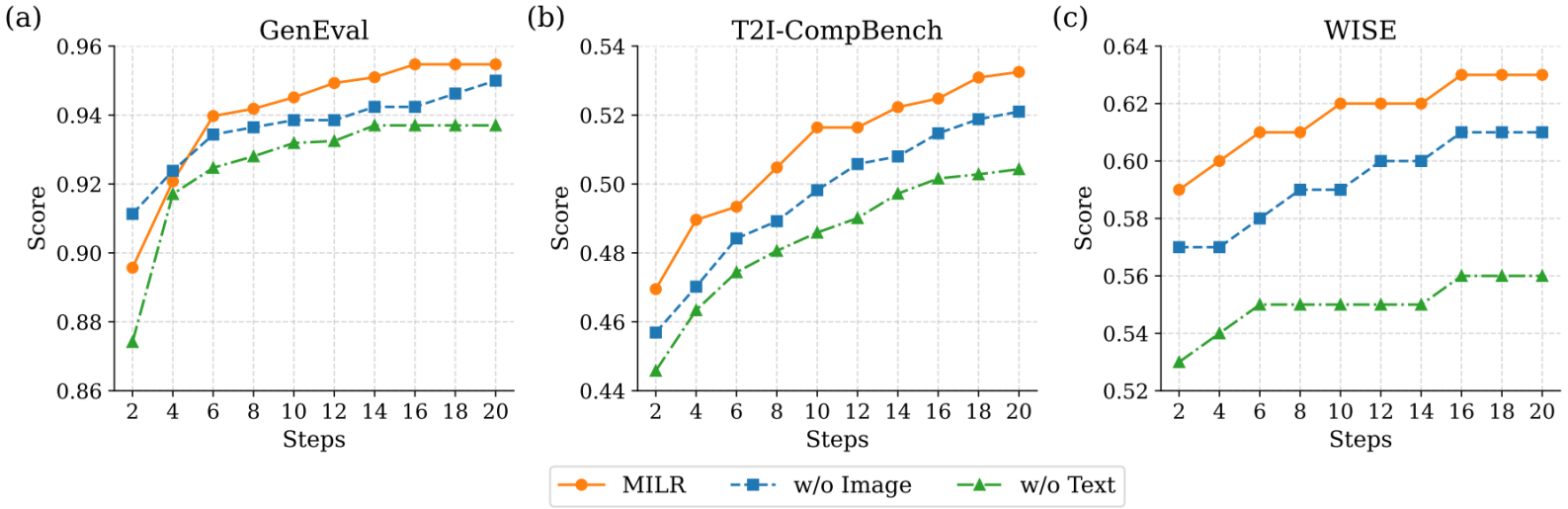


Figure 4: Performance across three benchmarks for varying optimization steps.

Scaling up the number of optimization steps improves performance on all benchmarks.

Analysis

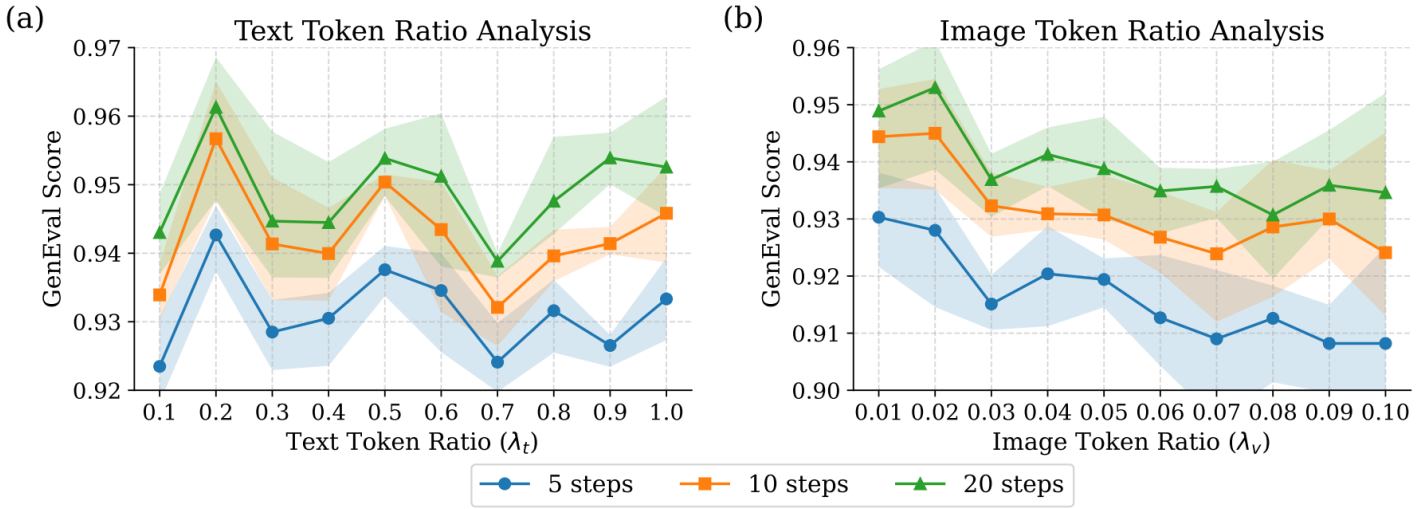


Figure 5: GenEval scores with varying optimization ratios of text and image tokens.

Optimizing a moderate amount (e.g., 20%) of text tokens leads to the best performance.

Optimizing a tiny amount (e.g., 2%) of image tokens gives rise to the best result.

Reward Model Analysis

In real-world scenarios, oracle rewards are usually unknown. To show that MILR is effective without the reliance on oracle , we test it with a set of off-the-shelf reward models

Table 6: GenEval results with different rewards. The best performance is in bold.

Reward	Single Obj.	Two Obj.	Counting	Colors	Position	Attr. Binding	Overall
Baseline	0.98	0.85	0.56	0.89	0.77	0.64	0.78
SelfReward	**1.00**	0.90	0.50	0.88	0.77	0.65	0.79
GPT-4o	0.98	**0.94**	0.80	0.88	0.77	0.68	0.84
UnifiedReward	**1.00**	0.92	0.76	0.89	0.81	0.66	0.84
MixedReward	**1.00**	0.90	**0.83**	**0.89**	**0.88**	**0.75**	**0.87**
OracleReward	1.00	0.96	0.90	0.98	0.98	0.91	0.95

Table 7: T2I-CompBench results with different reward models. The best performance is in bold.

Reward	Color	Shape	Texture	Spatial	Non-Spatial	Complex	Overall
Baseline	0.6359	0.3528	0.4936	0.2061	0.3085	0.3559	0.3921
Self Reward	0.7196	0.4620	0.5887	0.2379	0.3055	0.3868	0.4501
GPT-4o	0.7624	0.4701	0.6561	0.2778	**0.3102**	**0.3881**	0.4775
UnifiedReward	0.8208	0.4609	0.5835	0.3210	0.3044	0.3700	0.4651
MixedReward	0.8009	0.4765	0.6608	0.4077	0.3055	0.3745	0.5043
OracleReward	**0.8508**	**0.5117**	**0.6949**	**0.4613**	0.3078	0.3684	**0.5325**

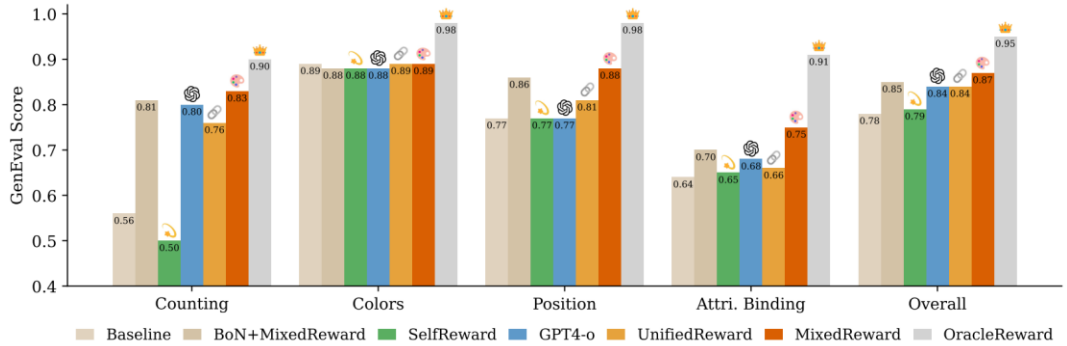


Figure 6: Performance with different reward models on GenEval.

It shows that MILR is robust to different reward models.

Error cases

Error1: Textual Reasoning Collapse

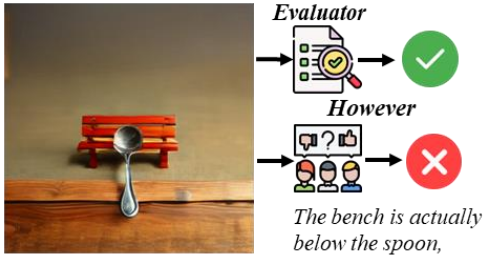
\n white glove positioned beige distinctly beige glove beige glove beige glove beige ……

a photo of four baseball gloves.

Error2: Visual Reasoning Collapse

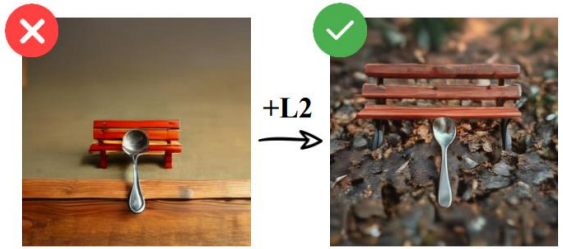
a photo of four handbags

Error3: Reward Hacking



a photo of a bench above a spoon

Adding L2 regularization helps mitigate reward hacking



a photo of a bench above a spoon

a photo of a dining table above a suitcase.

Future Research

- **Generalizability**: It is interesting to see if the effectiveness of MILR transfers to diffusion-based MUG.

- **Better reward models**: our experimental results have revealed that the strongest non-oracle reward model still lags behind the oracle reward. Thus, future work is well-suited for designing reward models that can generalize.

- **Another paradigm** : Maybe it is interesting to see another paradigm of latent reasoning in image generation.