ZILYU YE

✉ zilyuye@foxmail.com · ☎ (+86) 18138209188 · % Homepage

🎓 EDUCATION

South China University of Technology (SCUT), Guangzhou, China   2022 – present
Undergrad in Artificial Intelligence (AI), expected June 2026   GPA: 86/100

👥 EXPERIENCE

Adaptive Scheduler for Diffusion Models   https://arxiv.org/abs/2412.01243   First Author   **CVPR 2025**

This research proposes a lightweight diffusion time prediction module that forecasts optimal denoising timesteps. Without modifying original model parameters, we efficiently optimize the entire denoising process using the PPO reinforcement learning algorithm, accelerating generation speed and enhancing output quality for text-to-image models like SD3 and FLUX. The approach demonstrates improvements in human preference alignment and text-image matching metrics. Additionally, the module can be integrated with fine-tuning methods like GFlowNet and DDPO to further enhance generation performance through model adaptation.

MLLM Storytelling Enhancement   https://arxiv.org/abs/2408.03695   First Author   **CVPRW 2024 oral**

Proposed *Openstory++*, the first large-scale open-domain visual storytelling dataset (5,000+ downloads on HuggingFace), featuring automated extraction of frames combined with vision-language models and LLMs to generate instance-level annotations and coherent narratives. Significantly improves cross-frame entity consistency in MLLMs. Designed the *Cohere-Bench* evaluation framework to systematically assess model performance in long-context background, style, and entity consistency. Experimental results demonstrate the dataset's superiority in training high-quality open-domain narrative models.

MoE-based Multi-modal Model with Modal Experts   2024

Developed a novel Mixture-of-Experts (MoE) fine-tuning approach for multi-modal models, introducing modality-specific gating networks to enable optimal expert selection. Through simple fine-tuning procedures, we successfully upgraded existing unified generation-comprehension models (e.g., Chameleon, Show-O) into fine-grained MoE architectures. This method enhances multi-modal processing capabilities through specialized expert utilization while maintaining lower inference latency compared to conventional approaches.

RL-Finetuned LLMs for Complex Reasoning   2025

Investigated DeepSeek-R1-style training methodologies using GRPO and rule-based rewards to push the capability boundaries of 0.5B-7B parameter language models in solving Sudoku puzzles. Explored chain-of-thought distillation techniques where larger models' reasoning processes assist in training smaller models, achieving enhanced performance through knowledge transfer.

⚙ TECHNICAL SKILLS

- Linux, Python, C++, C (Embedded Development)
- Proficient in PyTorch, transformers, diffusers, trl and other development frameworks
- Experienced with distributed training frameworks including Accelerate, DeepSpeed, and FSDP
- Familiar with LLM, Diffusion Models, MLLM, and reinforcement learning fine-tuning for these models

♡ HONORS AND AWARDS

Third Prize, "Alibaba Cloud" Cup Algorithm Design Competition	2023
Meritorious Winner, Mathematical Contest in Modeling (MCM)	2024
Third Prize, National ROBOCON Robotics Competition	2024
Academic Scholarship, SCUT	2023, 2024

ℹ ADDITIONAL INFORMATION

- Languages: English - Proficient (CET-6)
- Hobbies: Badminton, Table Tennis, Gourmet Food