Research Statement

Sule Bai

My research interests lie in Multimodal Large Language Models (MLLMs), which I believe hold immense potential for a wide range of real-world applications. With the increasing popularity of reinforcement learning (RL), recent studies have shown its effectiveness in significantly enhancing the reasoning capabilities of language models. My recent submission, UniVG-R1, proposes a reasoning-guided MLLM for universal visual grounding, optimized via reinforcement learning. This work further strengthens my belief in the promise of combining MLLMs with RL to build more capable, interpretable, and generalizable AI systems.

Looking ahead, I plan to focus my PhD research on the following key directions:

1. Enhancing Reasoning in MLLMs: I aim to explore how to continuously incentivize reasoning through RL-based optimization, enabling models to perform well on increasingly complex and abstract multimodal tasks.
2. Unified Multimodal Understanding and Generation: I am particularly interested in building models that natively handle both generation and understanding in a tightly coupled, mutually beneficial framework.
3. MLLMs as Agents: I envision MLLMs evolving into general-purpose agents capable of assisting users with daily tasks through multimodal perception, memory, planning, and action.

I see strong synergy between my prior work and your group's pioneering research on MLLMs. I would be excited to continue this line of inquiry under your guidance, contributing to the development of the next generation of controllable, unified, and generalizable multimodal agents.