Contribution ID: **286** Type: **Poster**

Efficient Fine-Tuning of Visual Language Models for Chemical Image Understanding

Monday 13 October 2025 18:00 (1h 30m)

We investigate how vision-language models (VLMs) can be fine-tuned for chemistry-specific tasks by incorporating both molecular structure images and domain-specific textual descriptions. While general-purpose VLMs lack precision and adaptability in the chemical domain, our study addresses this gap through efficient fine-tuning strategies. In particular, we explore which selective layer tuning methods are most effective. Experimental evaluations using synthetic data and GPT-based assessment, in which accuracy scores were assigned based on the correctness of generated responses, reveal that tuning the Q (query) and V (value) modules of cross-attention layers yields the best performance. Our approach improves multimodal understanding in chemical contexts and presents a step toward lightweight, domain-adapted VLMs that are practical for scientific research and education.

Primary author: CHOI, Hyukjun (Ajou University)

Co-authors: HWANG, Ji Sun; KANG, Juyoung (Ajou university)

Presenter: CHOI, Hyukjun (Ajou University)

Session Classification: Poster Session

Track Classification: SciDataCon Persistent Themes: Data and Research