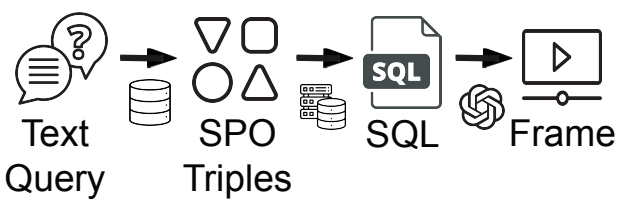
LazyVLM: Neuro-Symbolic Approach to Video Analytics

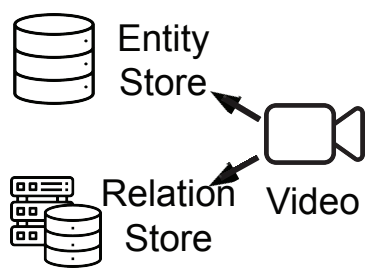
Xiangru Jian, Wei Pang, Zhengyuan Dong, Chao Zhang, M. Tamer Özsu
Cheriton School of Computer Science, University of Waterloo

WATERLOO DSG Data Systems Group #14

Motivation

Video analytics requires both expressive querying and efficient execution





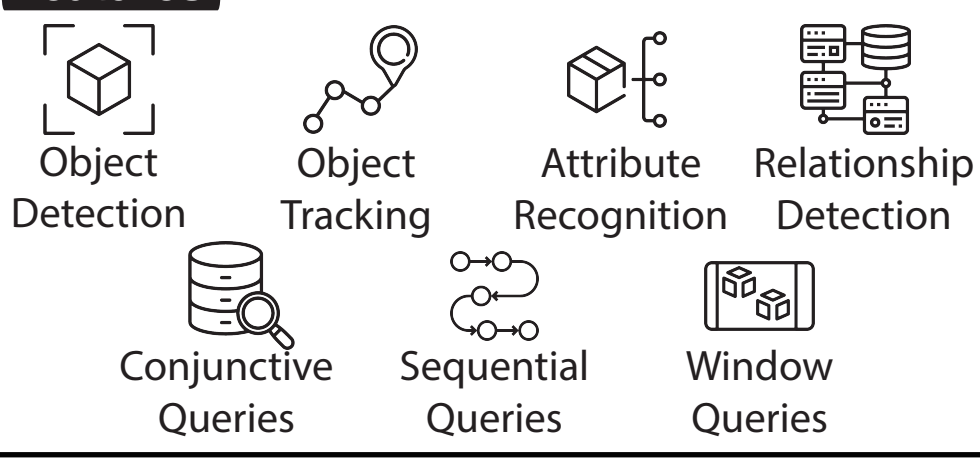
Takeaway

- LazyVLM bridges the gap between end-to-end VLM-based video querying and traditional symbolic systems by combining intuitive query specification with structured execution.
- LazyVLM shows that scalable video analytics might not sacrifice intuitive query specification.

Contribution

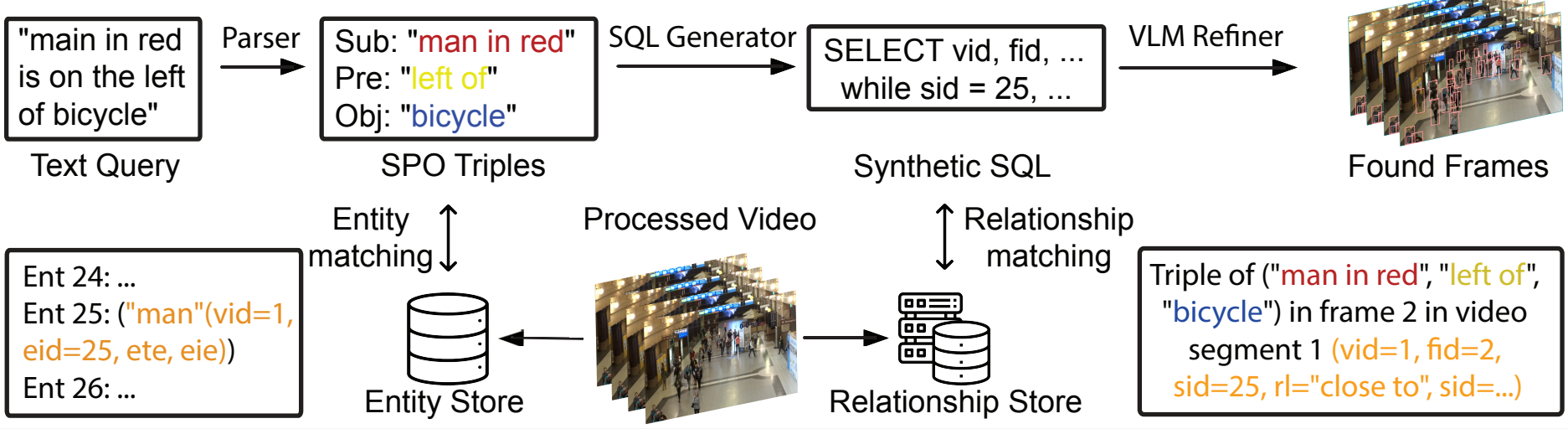
- LazyVLM provides a semi-structured interface for complex multi-frame video queries.
- LazyVLM decomposes video queries into semantic & symbolic search, and lightweight VLM refinement.
- LazyVLM enables scalable, user-friendly open-domain video analytics through relational execution and vector similarity search.

Features



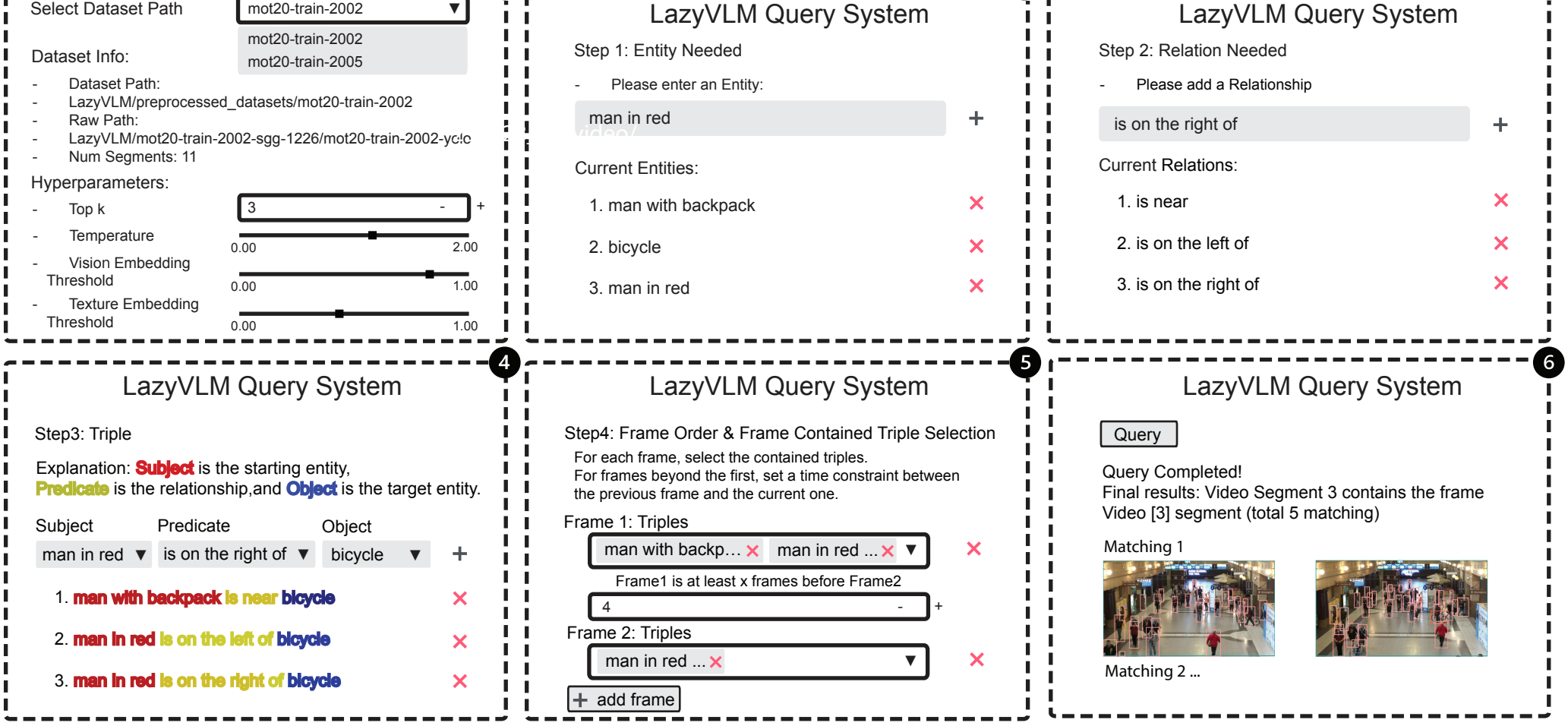
Pipeline

LazyVLM decomposes complex video queries into structured execution for scalable analytics.



UI

LazyVLM enables intuitive multi-frame video querying through structured event specification.



Reference

[1] Bai et al. (2025). Qwen2.5-VL technical report. arXiv.
[2] Daum et al. (2023). VOCALExplore: Pay-as-you-go video data exploration and model building. PVLDB, 16(13), 4188–4201.