Multi-Scale Video Chunking for VLM-Based Event Detection

Nelida Salgado, University of California Merced, Amar Saini & Carmen Carrano, Lawrence Livermore National Laboratory

ABOUT ME

Major: Computer Science & Engineering

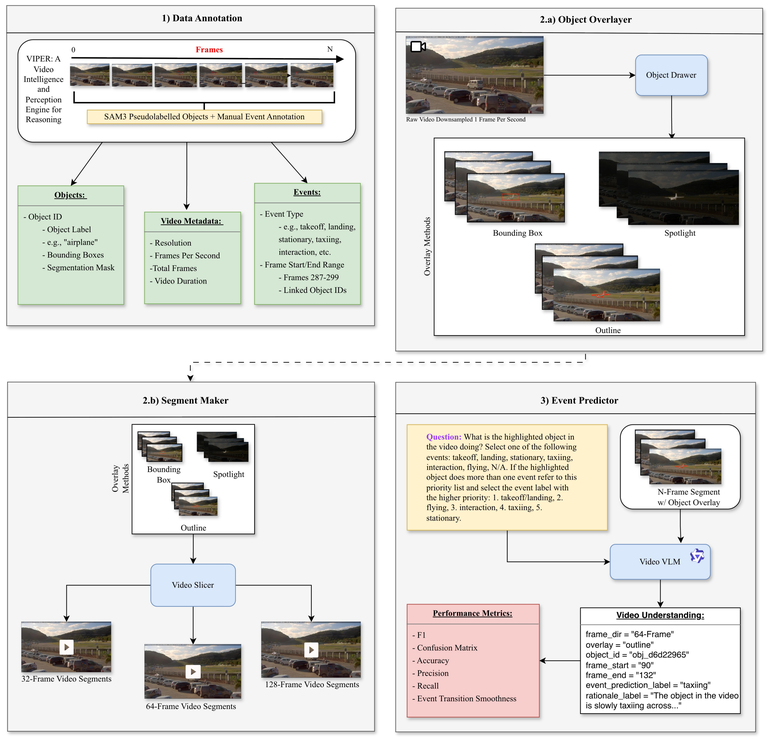
Career Goal: Data Scientist

"I chose LLNL because allows me to grow and strengthen my technical skills, while working alongside passionate researchers to tackle real-world problems."

Interests: Nintendo!

SUMMER RESEARCH

Motivation & Methodology: Vision Language Models (VLMs) excel at short videos but fail on long-duration footage due to memory constraints, token limits, and multi-object ambiguity. We address this by segmenting videos into fixed chunks (32/64/128 frames) and applying object-focused overlays (bounding box/outline/spotlight) for event classification across seven categories, enabling scalable analysis while maintaining computational efficiency.



Future work: Planned ablation studies to evaluate impact on event classification accuracy:

- *Chunk size:* Performance across temporal scales (32/64/128 frames)
- *Overlay Methods:* Comparison of bounding boxes, outlines, spotlight, and color variations
- *Model Scaling:* Evaluation across VLM architectures (Qwen3/3.5/3.6: 0.8B-35B; Gemma-4 series)
- *Prompt Engineering:* Iterative refinement of multi-step reasoning prompts

LLF AWARD

The funds that LLF has provided me with have played a vital role in supporting my summer in Livermore by helping me cover housing and my commute.

This fellowship has allowed me to strengthen my data science skills and most importantly it has provided me with valuable hands-on research experience and the opportunity to build meaningful connections that will support my future goals.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

LLNL-POST-2022359

Livermore Lab FOUNDATION

UNIVERSITY OF CALIFORNIA MERCED

Lawrence Livermore National Laboratory