

OpenVAU: Open-World Video Anomaly Understanding

Aditya Mittal, Reno Kriz, Dave Etters, Cameron Carpenter
Contact: Email: mittalaa@uci.edu | Phone: +1 (669)-304-8483



JOHNS HOPKINS
UNIVERSITY

human language technology
center of excellence

Introduction

THE OPEN-WORLD ANOMALY CHALLENGE

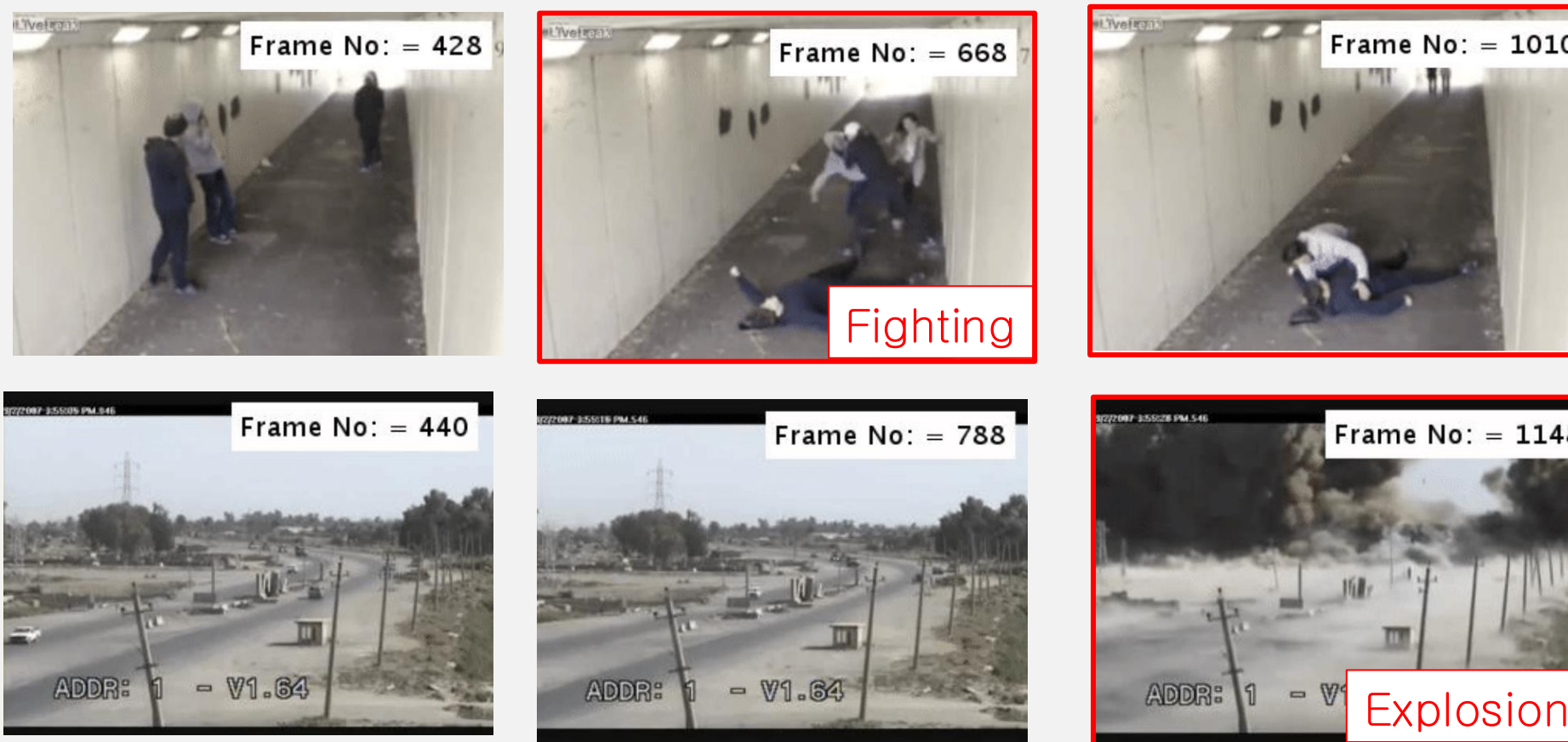
- **Goal:** Leverage pre-trained MLLMs for open-world anomaly **understanding** and **detection**.

Current problems:

- **Anomaly understanding:** what is an anomaly?
- **Anomaly detection:** how to find anomalous videos for retrieval?
- Can we do it **efficiently**?

This is especially challenging for **raw, unedited videos**.

Existing Anomaly Benchmarks: UCF CRIME → Predefined Anomaly Classes



- ✓ Closed-set anomaly classes
- ✓ Fixed train/test split
- ✓ In-distribution evaluation

SCALE 2026 Benchmark: Multivent-Raw → Undefined Anomaly Classes

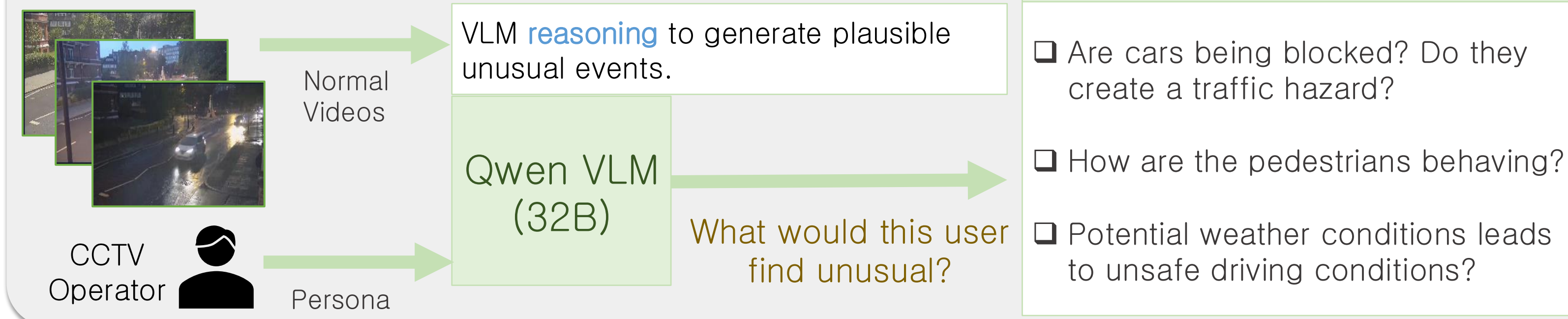


- ❑ Open-set anomaly classes
- ❑ No train/test split
- ❑ Out-of-distribution evaluation

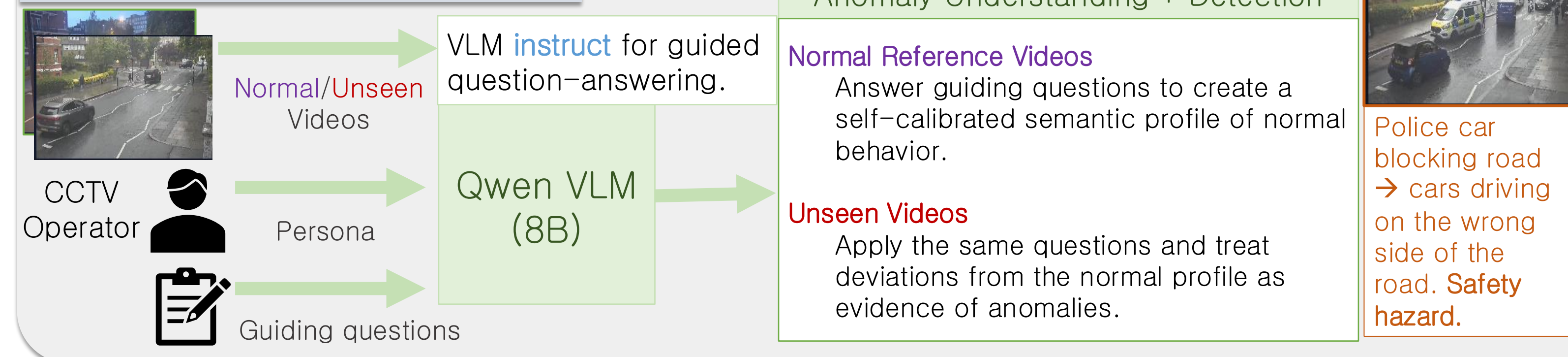
Method & Results

Methodology: Semi-Supervised Learning

Step 1: Semantic Calibration



Step 2: Anomaly Understanding



Comparison with existing work

- Needs **train/test split**.
- **Distribution-level** anomaly detection.
- SCALE baseline **two-pass approach**: distribution anomaly profile → comparing new videos against base profile.
- OpenVAU uses a subset of reference videos to generate **semantic anomaly profiles** based on guiding questions.
- VLM answers these questions directly for new unseen videos

Anomaly Understanding

OpenVAU defines anomalies relative to a persona and context. Guiding questions turn vague definitions into concrete criteria for the VLM.

Anomaly Detection

MLLM reasoning identifies anomalous videos and timestamps, supporting both video-level retrieval and within-video temporal localization.

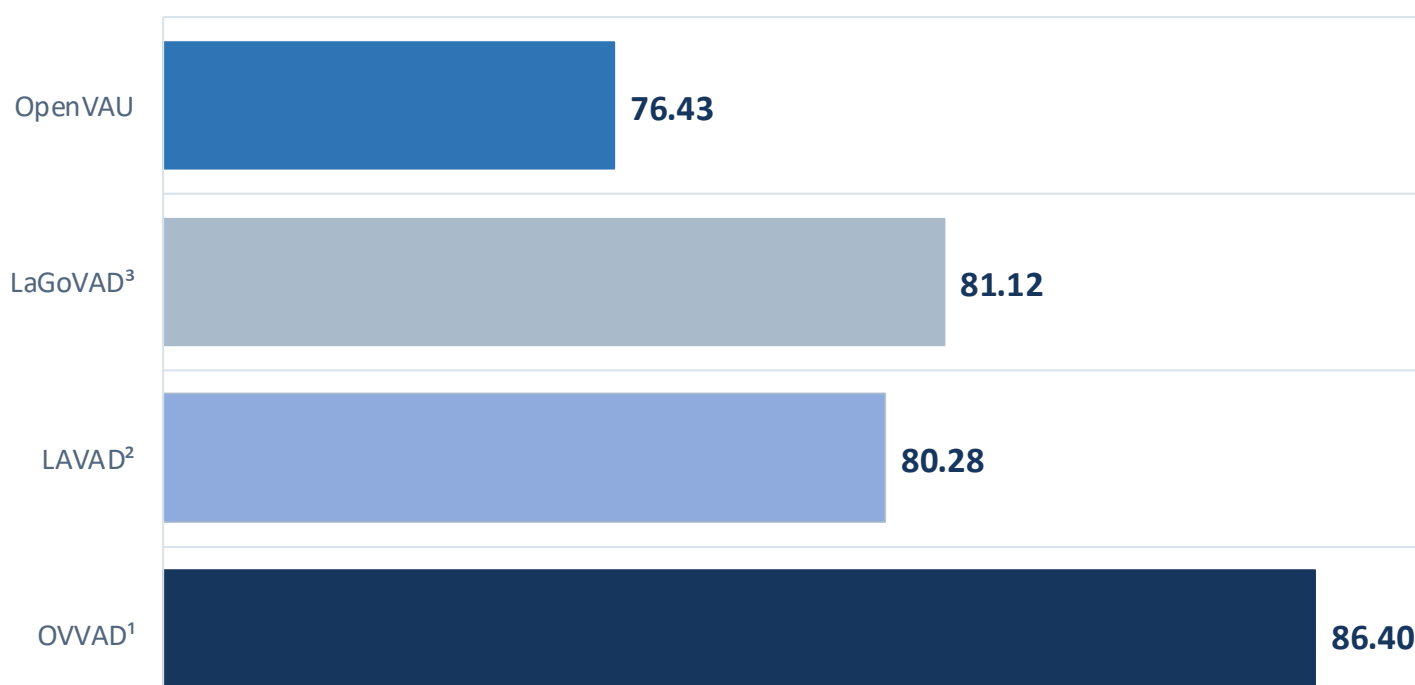
Efficient Inference

A large reasoning model generates guiding questions once. A smaller instruction-tuned VLM answers them across unseen videos to reduce repeated computation.

Results

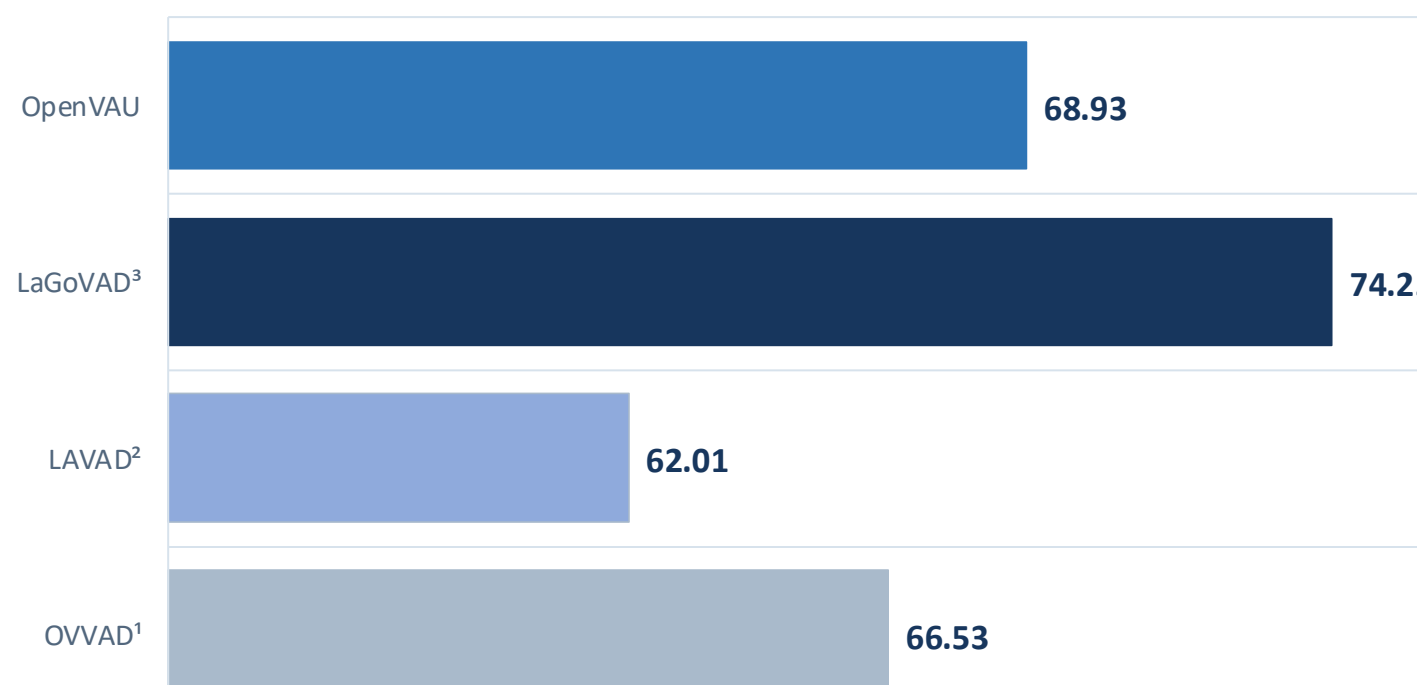
On MultiVENT, OpenVAU improves over the SCALE baseline on both nDCG and InfoF1.

UCF-Crime · AUC (%) ↑



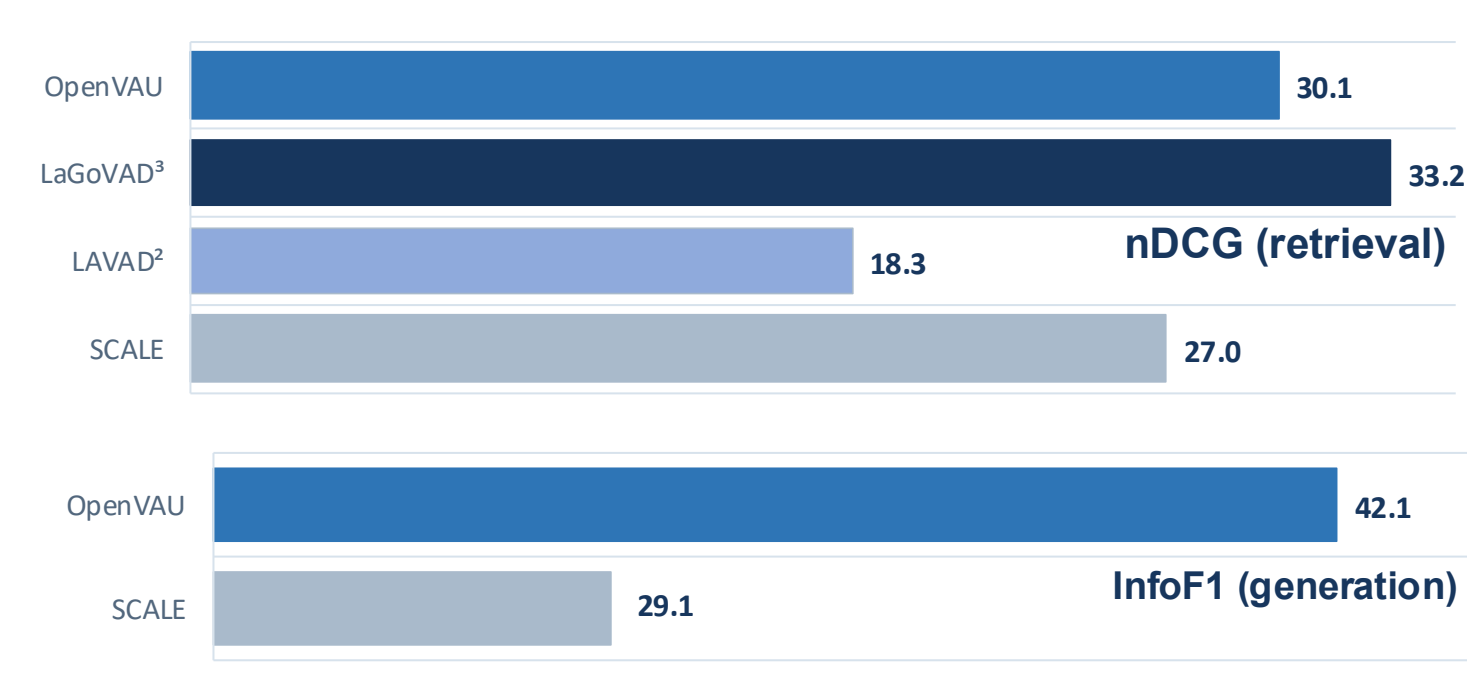
AUC = area under the frame-level ROC curve

XD-Violence · AP (%) ↑



AP = frame-level average precision

MultiVENT-Raw-Anomaly



nDCG = retrieval ranking quality | InfoF1 = description quality

Methods: 1) OVVAD (CVPR, 2024), 2) LAVAD (CVPR, 2024), 3) LaGoVAD (ICLR, 2026), SCALE two-pass baseline, OpenVAU (Ours)