Final Presentation

07/30/2025

Multi-modal Large Language Models for Video Anomaly Detection

Lead : Dr. Tian

Eric Li

Video Anomaly Detection

- Distinguish
  - Normal
  - Abnormal
- **Challenge**
  - The definition of anor
    ambiguous.
  - Abnormal behaviors
    context
  - Difficult to recognize from a single video frame.
- Advantage
  - **Real-time warning**
    → Faster response of security incidents
  - **Replacement of manual surveillance**
    → Improve resource utilization efficiency.

Area:

- **Surveillance**
- **Security**

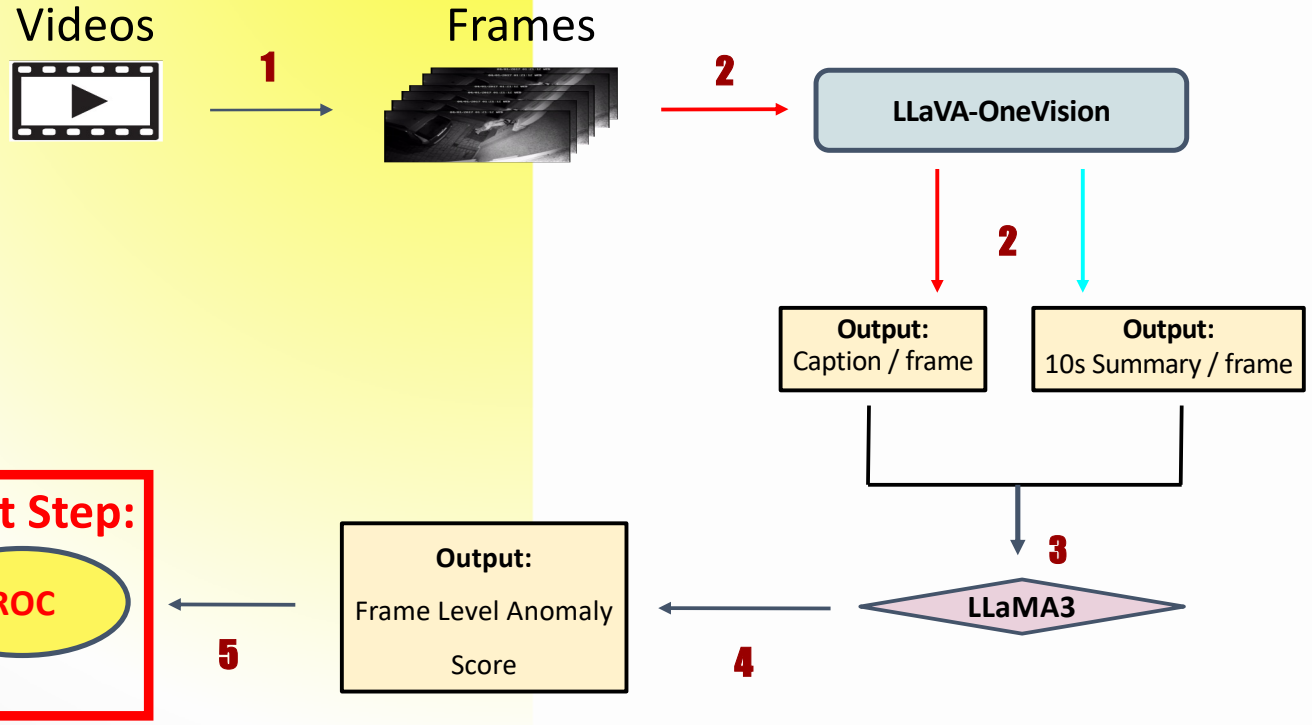


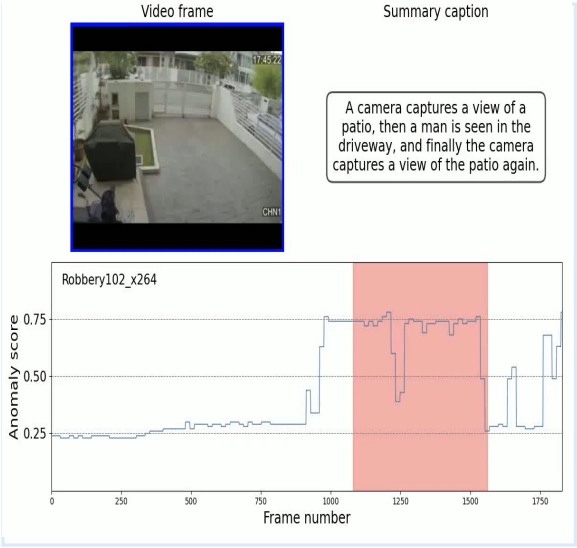
The goal of my project

- Create a new **training-free LVLM** with enhanced Video Anomaly Detection capabilities.

(Currently, the highest **ROC AUC** achieved on the UCF-Crime dataset is nearly 0.8)

Our Model:


Videos
1
Frames
2
LLaVA-OneVision
2
Output:
Caption / frame
Output:
10s Summary / frame
3
LLaMA3
4
Output:
Frame Level Anomaly
Score
5
Last Step:
ROC


Video frame
Summary caption
A camera captures a view of a patio, then a man is seen in the driveway, and finally the camera captures a view of the patio again.
Robbery102_x264
Anomaly score
0.75
0.50
0.25
0
250
500
750
1000
1250
1500
1750
Frame number

Our current outcome

Method	AUC (%)
ZS CLIP	53.16
ZS IMAGEBIND (IMAGE)	53.65
ZS IMAGEBIND (VIDEO)	55.78
LAVAD	78.32
Our	**81.13**

Our Future Plan

1. Help our model **distinguish normal & abnormal behaviors** by using strategies like memory bank

2. **Extend our framework** to process audio signals in addition to visual inputs.

Thank you!