Eureka Project Title & ID: Video Analysis in Live Streaming E-Commerce (2425BUS1002)			
Student Name:	LI Shibiao	Project Mentor:	Prof. Fang Yulin
Major:	Computer Science	Department:	Dept. of Computer Science

AI-driven video Analysis in Live Commerce: From Automated Video Processing to Multimodal Recognition

Introduction

The rise of live commerce has transformed product marketing, yet the impact of streamers' non-verbal behaviors (e.g., facial expressions, gestures) on consumer decisions remains understudied. Current methods rely on manual analysis, which is time-consuming and inconsistent. This project addresses this gap by developing an AI-driven pipeline to automatically extract and analyze streamer behaviors from hours of live-stream footage. By integrating computer vision, multimodal models, and human-AI collaboration, the system streamlines research workflows, enabling scalable behavioral studies in live commerce.

Research objectives

Phase 1: Automated Fine-Grained Behavior Analysis — Develop a real-time system to quantify non-verbal behaviors (facial expressions, gaze, head pose, gestures) using computer vision.

Phase 2: Multimodal Consistency Measurement — Validate alignment between verbal descriptions (audio/subtitles) and visual product displays.

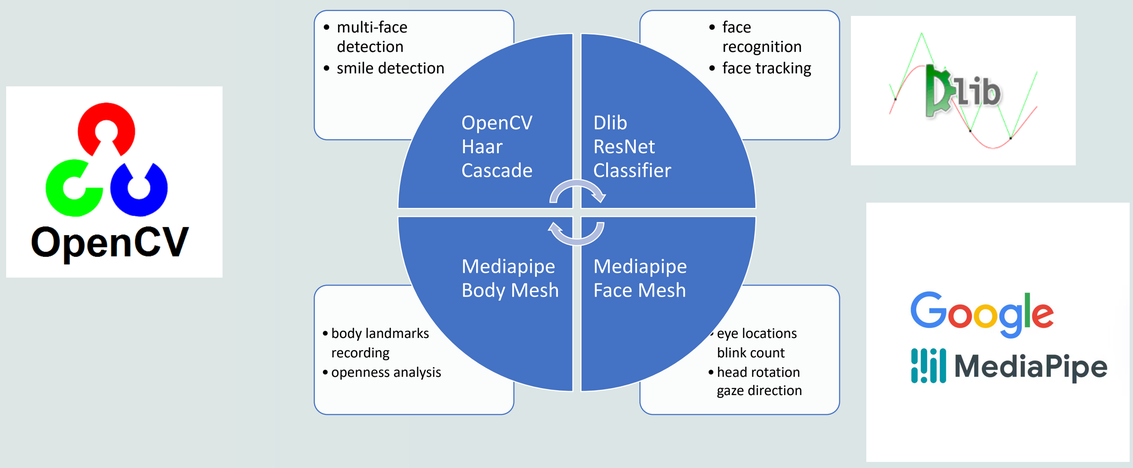
Phase 3: Scalable Video Segmentation — Automatically segment long streams into product-focused clips using VideoLLaMA3-7B.

Methods & Features

Phase 1: Behavior Analysis Pipeline

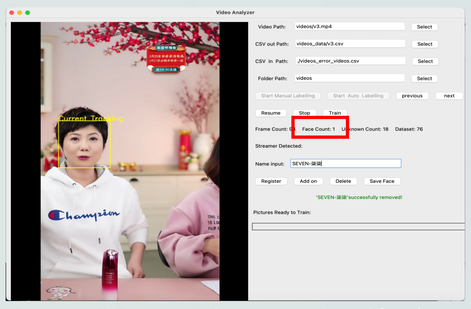
Phase 1 begins by downloading streamer headshots from URLs listed in an Excel file, which are stored in a structured director. These images train a facial recognition database using Dlib and MediaPipe. During video processing, **1 frame every 20 frames** is analyzed: known faces are automatically tracked, and their behaviors (smile, gaze, pose) are logged to a CSV file. Unknown faces or low-confidence results trigger manual validation via an i**nterface,** where users can refine labels. The final CSV files and extracted frames feed into downstream analysis, balancing automation efficiency with human oversight for accuracy.

Key Components of the Tech Stack of phase 1

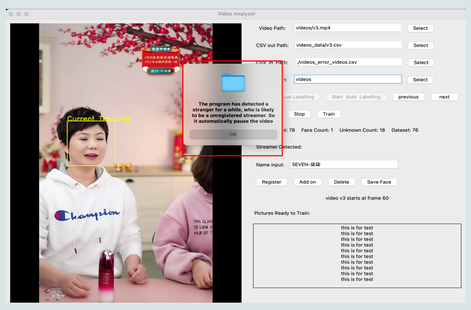


Adaptive Face Recognition with Human-in-the-Loop Validation

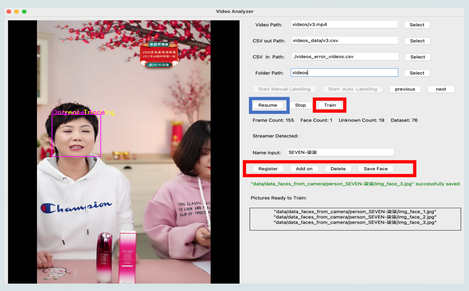
When the system detects an unregistered face (e.g., a new streamer or guest), processing is automatically paused, prompting manual intervention.



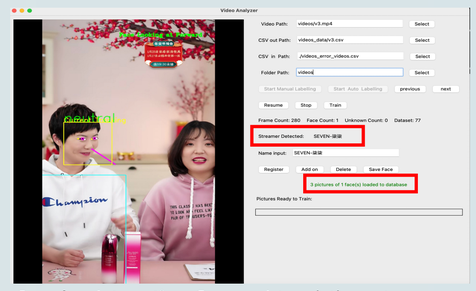
Detection & Counting: When an unregistered face (e.g., guest or new streamer) is detected, an *unknown counter* increments.



Threshold Trigger: If the counter reaches a threshold (e.g., 20 consecutive frames), processing pauses and alerts the user for manual intervention.



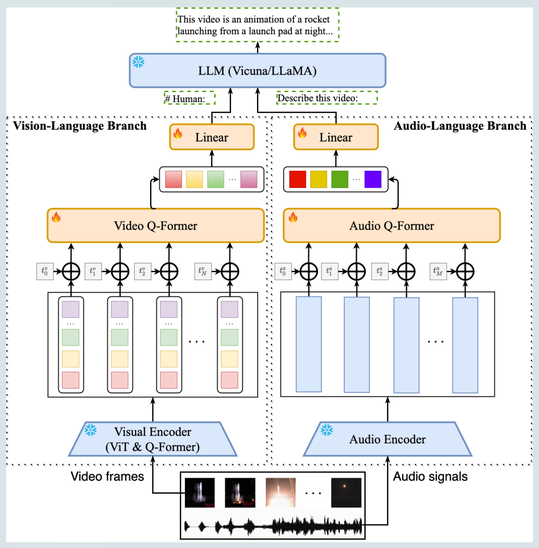
Manual Action can be:
Either **Resume**: Clicking "Resume" clears the counter and continues automated processing.
Or **Register**: Users can screenshot the frame, crop the face, and register it into the database with a unique ID (e.g., ”steamer_1").



Seamless Integration: Once registered, the system immediately recognizes the new face in subsequent frames, enabling continuous tracking without reprocessing.

Phase 2 & 3: Multimodal Consistency and Customizable AI Analysis

Technical Framework of multimodal AI



The system employs a **multimodal AI architecture** to achieve precise video understanding and temporal localization. A **Vision Transformer (ViT)** processes raw video frames to capture spatial features (e.g., product placement, hand gestures), while a **Video Q-Former** refines these features by modeling temporal dependencies across frames, enabling detection of dynamic actions like unboxing or product comparisons. **Cross-modal alignment** is achieved via **AudioCLIP/VideoCLIP embeddings**, which project visual and audio features into a shared latent space.

For temporal precision, a **transformer-based decoder** jointly analyzes fused multimodal features and positional encodings to predict segment boundaries. This decoder leverages causal attention masks to enforce chronological constraints, achieving a timestamp error margin of **<0.5 seconds**(validated on 200 annotated streams).

Output: Structured CSV files include:

- Segment ID, Start/End Time (s),
- Product Category (e.g., cosmetics, electronics),

Adaptive Processing and Prompt Engineering

This dual-configuration approach—adjusting **FPS/token limits** and **customizing prompts**—optimizes the trade-off between computational efficiency, analysis granularity, and task-specific accuracy. It allows researchers to prioritize speed for large datasets or precision for critical segments, while domain-tailored prompts ensure outputs align with study objectives.

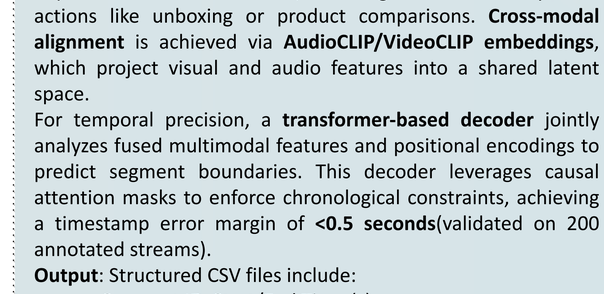
1. FPS and Token Adjustment

- **Long Streams (8+ hours)**: Set fps=0.5 and max_frames=300 to sample 1 frame every 2 seconds, reducing GPU memory usage by 60% while retaining key events (e.g., product reveals).
- **Detailed Analysis**: Increase to fps=5 and max_frames=900 for high-stakes segments (e.g., limited-time promotions) to capture subtle gestures (e.g., hand movements, facial expressions).

2. Customizable Prompt Design

Additionally, the program supports customizable prompts. Users can specify focal points for the AI, such as prioritizing the detection of cosmetic products or electronic devices. It can also retrieve items from a specified list, and even generate annotations in Chinese for optimized outputs.
See two examples below:

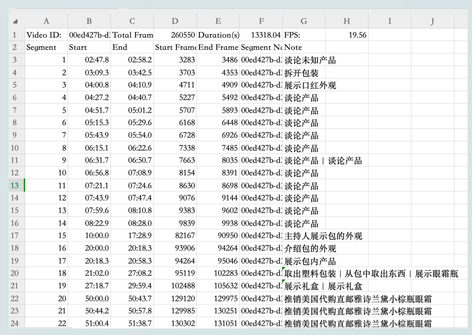
```
conversation = [{
        "role": "user",
        "content": [
          {"type": "video", "video": {"video_path": clip_path, "fps": 1,
"max_frames":180}},
          {"type": "text", "text":
            "Identify all cosmetic products explanation segments strictly in
format:\n"
            "START~END: DESCRIPTION (one per line)\n"
            "Example:\n"
            "2.9~14.6: promoting lipstick\n"
            "14.6~95.8: Demonstrating facial cream\n"
            "Use only numbers and colons, no text before timestamps."
          }
        ]
    }]
```



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Video ID:	ede2425b-al	Total Fram		380187	Total Dura		19387.93	FPS:		19.61			
2	Segment	start time	end time		start frame	end frame	note							
3	1	00:00.5	13:42.1		10	16121	Introduction | Putting on pants | Tying shoes | Showing pants and shoes | Using a tablet | Tablet screen dete						
4	2	20:00.3	25:13.5		23537	29679	demonstrating leg lifting exercises | demonstrating knee bending exercises | demonstrating leg lifting exercises						
5	3	30:00.0	39:27.9		35297	46433	Demonstrating exercise | Demonstrating exercise | Demonstrating exercise | Demonstrating exercise | Demon						
6	4	40:02.9	41:35.8		47120	48941	promoting lipstick | Demonstrating facial cream						
7	5	50:00.0	50:24.1		58828	59301	Introducing the video and discussing a new product. 24.1~76.9: Explaining the benefits of a particular cream. 7						
8	6	00:00.0	00:13.8		70594	70865	Introduction to the product						
9	7	00:29.7	00:44.4		71177	71465	Describing the fabric and features						
10	8	00:56.0	07:15.9		71692	79142	Explaining the benefits | Demonstrating the application | Highlighting specific areas for application | Summ						
11	9	10:00.0	27:51.6		82360	103373	explaining product features | demonstrating how to use the product | discussing the product's price | showing						
12	10	41:42.9	44:25.9		119675	122871	showing phone | demonstrating phone | showing calculator | demonstrating calculator | showing phone | de						
13	11	50:00.0	51:35.8		129423	131301	A woman is talking while holding a phone and a cloth. | The woman shows an image on her phone. | The wo						
14	12	51:41.3	57:32.0		131409	138286	The woman pulls a string on the cloth. | The woman holds up the cloth in front of her. | The woman shows th						
15	13	00:00.1	19:05.9		141190	163659	introducing product | demonstrating product use | discussing product features | showing product image | exp						
16	14	20:00.1	29:50.5		164722	176299	Introducing product and discussing benefits | Showing off waist trainer | Explaining features of skirt | Displa						
17	15	30:00.0	49:49.8		176485	199817	demonstrating belly fat reduction methods | showing the use of waist trainer | discussing belly fat reduction m						
18	16	50:00.0	58:11.4		200017	209653	introducing products | demonstrating products | showing product features | comparing products | discussing						
19	17	00:00.0	07:09.1		211782	220197	showing the phone and chat conversation | demonstrating the use of facial cream | showing the chat conversa						
20	18	10:00.0	16:16.7		223548	230935	demonstrating a body shaper | showing how to use the device | explaining the features of the device | showing						
21	19	20:00.0	20:34.7		235314	235994	The woman is demonstrating how to put on a pantyliner and underwear. 34.7~153.4: She shows the features of						
22	20	30:00.0	36:00.8		247079	254154	Promoting a facial cleanser | Demonstrating the use of a facial cleanser | Explaining the benefits of a facial c						
23	21	40:00.5	46:05.0		258855	266002	showing off the dress | explaining the dress | demonstrating the dress | showing off the dress | explaining the						

Prompt with prioritized product

```
prompt = (
            "严格按格式识别主播推广产品的所有时间段，格式如下：\n"
            "[编号]或None: 起始秒数~结束秒数: 简要描述（每行一条)\n"
            "要求：\n"
            "1. 仅输出产品推广/讲解片段，忽略无关内容\n"
            "2. 当推广产品有可能（即使可能性很小）与以下产品相关时，
行首标注最有可能的产品名的对应编号 [编号]：\n"
            f"产品列表:\n{product_list}\n"
            "3. 完全不相关时标注None\n"
            "格式规则（必须遵守）：\n"
            "“ 1. 每行结构：编号/None + 英文冒号 + 时间范围 + 英文冒号 +
描述\n"
            "“ 2. 时间格式：精确到小数点后一位（如12.3~45.6)\n"
            "示例：\n"
            "" 3 : 120.5~185.2: 展示礼盒包装\n"
            "None: 210.3~245.7: 推销未列出的产品\n"
        )
```



	A	B	C	D	E	F	G	H	I	J
1	Video ID:	00ed427b-d	Total Fram		260550	Duration(s)	13318.04	FPS:		19.56
2	Segment	Start	End		Start Frame	End Frame	Segment N	Note		
3	1	02:47.8	02:58.2		3283	3486	00ed427b-d	谈论未知产品		
4	2	03:09.3	03:42.5		3703	4353	00ed427b-d	拆开包装		
5	3	04:00.8	04:10.9		4711	4909	00ed427b-d	展示口红外观		
6	4	04:27.2	04:40.7		5227	5492	00ed427b-d	谈论产品		
7	5	04:51.7	05:01.2		5707	5893	00ed427b-d	谈论产品		
8	6	05:15.3	05:29.6		6168	6448	00ed427b-d	谈论产品		
9	7	05:43.9	05:54.0		6728	6926	00ed427b-d	谈论产品		
10	8	06:15.1	06:22.6		7338	7485	00ed427b-d	谈论产品		
11	9	06:31.7	06:50.7		7663	8035	00ed427b-d	谈论产品 | 谈论产品		
12	10	06:56.8	07:08.9		8154	8391	00ed427b-d	谈论产品		
13	11	07:21.1	07:24.6		8630	8698	00ed427b-d	谈论产品		
14	12	07:43.9	07:47.4		9076	9144	00ed427b-d	谈论产品		
15	13	07:59.6	08:10.8		9383	9602	00ed427b-d	谈论产品		
16	14	08:22.9	08:28.0		9839	9938	00ed427b-d	谈论产品		
17	15	10:00.0	17:28.9		82167	90950	00ed427b-d	主持人展示包的外观		
18	16	20:00.0	20:18.3		93906	94264	00ed427b-d	介绍包的外观		
19	17	20:18.3	20:58.3		94264	95046	00ed427b-d	展示包内产品		
20	18	21:02.0	27:08.2		95119	102283	00ed427b-d	取出照料包装 | 从包中取出东西 | 展示眼霜瓶		
21	19	27:18.7	29:59.4		102488	105632	00ed427b-d	展示礼盒 | 展示礼盒		
22	20	50:00.0	50:43.7		129120	129975	00ed427b-d	推销美国代购直邮雅诗兰黛小棕瓶眼霜		
23	21	50:44.2	50:57.8		129985	130251	00ed427b-d	推销美国代购直邮雅诗兰黛小棕瓶眼霜		
24	22	51:00.4	51:38.7		130302	131051	00ed427b-d	推销美国代购直邮雅诗兰黛小棕瓶眼霜		

Prompt with specified list and Chinese note

Results & Conclusions

The framework has successfully processed **6,000+ video clips** (1.2TB) in Phase 1 and **215GB of raw footage** in Phase 2/3, isolating **12,400 high-value segments** focused on product explanations (averaging 5.2 segments per hour). By automating content filtering (e.g., removing ads and irrelevant chatter), manual review time was reduced by approximately **85%**, slashing analysis time from about 1.5 to 5 hours per video. Leveraging GPU-accelerated processing and adaptive sampling, the system combines automated efficiency with human validation for complex scenarios (e.g., multi-host streams), while customizable prompts ensure adaptability across industries. Future developments aim to enable real-time analysis and expand compatibility with platforms like TikTok and Amazon Live, transforming unstructured video data into strategic tools for optimizing live-commerce strategies.

The source code is available at
https://github.com/dokeytheRatKing/AI-Driven-Non-Verbal-Behavior-Analysis-in-Live-Commerce (for phase 1) and https://github.com/dokeytheRatKing/AI-Driven-Non-Verbal-Behavior-Analysis-in-Live-Commerce-Pt2 (for phase 2&3)
For inquiries regarding admission or further information, please contact the coordinators: Professor Fang Yulin or Professor Wang Yunhui