~~Senior Projects~~ Knowledge Sharing

JubJones

Intelligent Multi-Camera Person Tracking and Analytics System

2025 Presentation

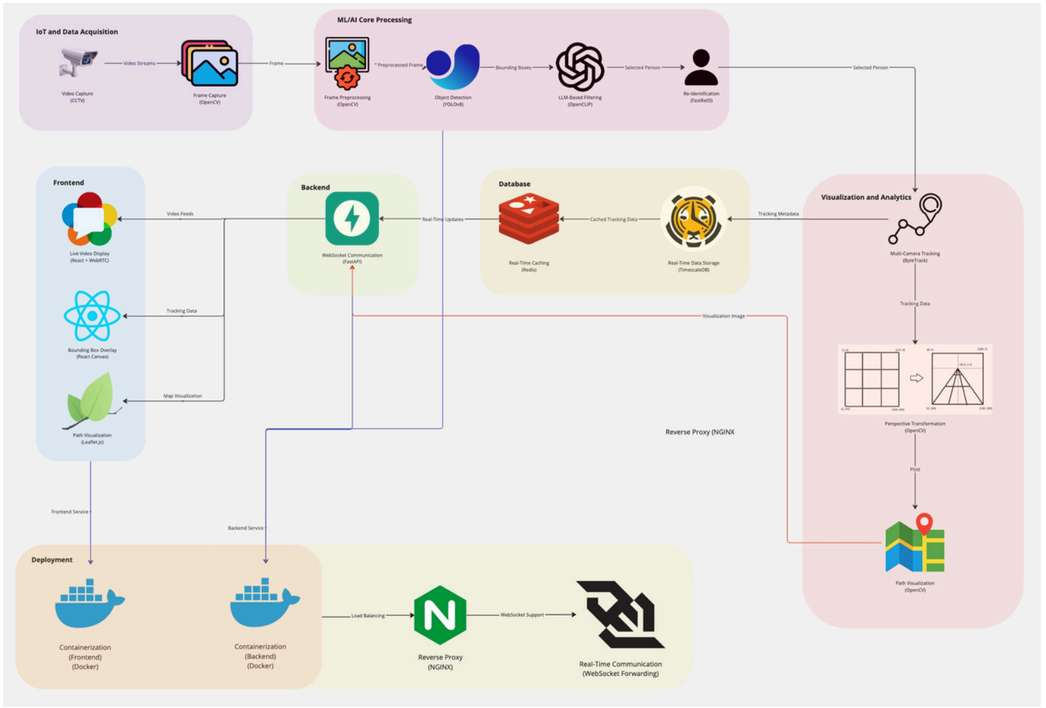
Next Slide

Recap

2025 Presentation

Core	Optional
LLM-Powered Person Selection	Multi-Object Tracking
Real-Time Person Detection and Re-Identification	Voice Command Integration
Path Tracking and Visualization	Dynamic Analytics and Alerts
	Automatically generate map

Features



Architecture

Today’s topic

- Hmm GTA V?
- Alternative Data Source
- Datasets
- Paper
- What changes
- ~~SRS Progress~~

We’re here

Week 4	2025-01-28	Experience in AI-enabled systems development	invited speaker (Top-Kampol)	
Week 5	2025-02-04	Pitching slides prep Kaset Fair Week	NO CLASS	1 Tentative title/advisor
Week 6	2025-02-11	Pitching slides prep CPE Project Day	NO CLASS	
Week 7	2025-02-18	Pitching (5 minutes)	Kitsana/Tana	2 Pitch slides, topic 3 Peer feedback
Week 8		Midterm Examination Week		
Week 9	2025-03-04	How to write Software Requirement Specs (SRS)	Jitti/Tana	
Week 10	2025-03-11	SRS workshop 1	Kitsana/Tana	SRS draft
Week 11	2025-03-18	SRS workshop 2	Kitsana/Tana	SRS draft

What is GTA V?

2025 Presentation

Altenative Data Source?

Why

Privacy Concerns (PDPA Compliance)

Technical Challenges with CCTV Integration

Lack of Visual Diversity

How

Create a Controlled Environment in GTA V (Optional)

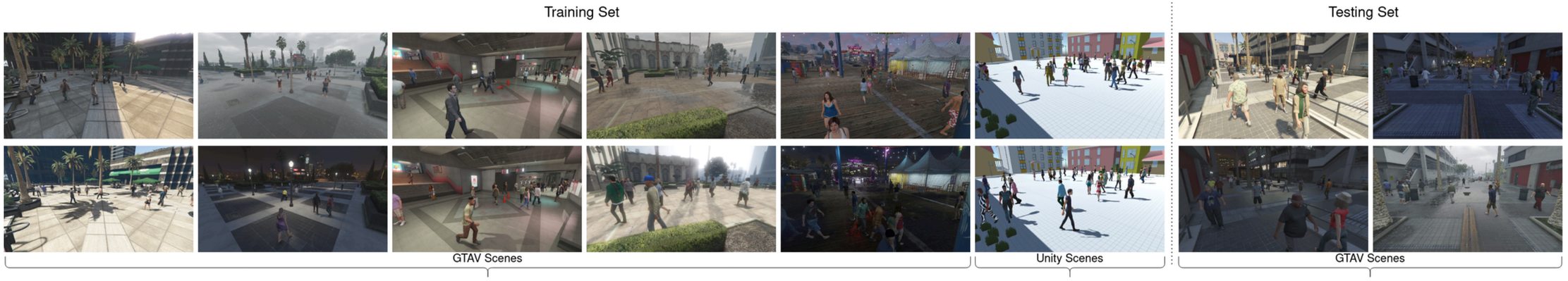
Capture Multiple Camera Views

Stream Video to the Backend

Datasets

GMVD: The Generalized Multi-View Detection

- Designed for multi-view pedestrian detection and tracking in **synthetic** environments.
- How many of them?
  - Scenes: 7 distinct scenes (1 indoor, 6 outdoor).
  - Frames:
    - Training Frames: 4,983
    - Testing Frames: 1,012
  - Cameras: Configurations vary between 3 to 8 cameras per scene.
  - Resolution: High-resolution images (1920x1080).
- **Bounding boxes** for pedestrians and vehicles.



Datasets

PoseTReID Dataset

Datasets

PoseTReID Dataset

- A **synthetic** dataset designed for people tracking and **re-identification (Re-ID)** in distributed contexts.
- It is built using GTA V and focuses on tracking individuals across multiple cameras or video scenes in both **short-term** and **long-term** scenarios.
- How many of them?
  - The PoseTReID dataset contains approximately 2,957 images in total, ~~with around 600 face images per character for the 5 characters~~ included in the dataset.
  - Annotations:
    - Includes **bounding boxes** for NPCs.
    - Provides **unique IDs** for each NPC across frames for re-identification.
    - Frame index.
    - ~~NPC name/ID (e.g., "Michael," "Trevor").~~
  -

2025 Presentation

Paper

- Generalization Challenges in MVD
- ~~Proposed GMVD Dataset~~
- Proposed Method
- ~~Comprehensive Experiments~~
- ~~Benchmarking on GMVD~~

GMVD: The Generalized Multi-View Detection

Bringing Generalization to Deep Multi-View Pedestrian Detection

Jeet Vora[1], Swetanjal Dutta[1], Kanishk Jain[1], Shyamgopal Karthik[2], Vineet Gandhi[1]

***Abstract*— Multi-view Detection (MVD) is highly effective for occlusion reasoning in a crowded environment. While recent works using deep learning have made significant advances in the field, they have overlooked the generalization aspect, which makes them *impractical for real-world deployment*. The key novelty of our work is to *formalize* three critical forms of generalization and *propose experiments to evaluate them*: generalization with i) a varying number of cameras, ii) varying camera positions, and finally, iii) to new scenes. We find that existing state-of-the-art models show poor generalization by overfitting to a single scene and camera configuration. To address the concerns: (a) we propose a novel Generalized MVD (GMVD) dataset, assimilating diverse scenes with changing day-time, camera configurations, varying number of cameras, and (b) we discuss the properties essential to bring generalization to MVD and propose a barebones model to incorporate them. We perform a comprehensive set of experiments on the WildTrack, MultiViewX and the GMVD datasets to motivate the necessity to evaluate generalization abilities of MVD methods and to demonstrate the efficacy of the proposed approach. The code and the proposed dataset can be found at https://github.com/jeetv/GMVD**

I. Introduction

"Essentially all models are wrong, but some are useful."

— George E. P. Box

In this work, we pursue the problem of Multi-View Detection (MVD), a mainstream solution for dealing with occlusions, especially when detecting humans/pedestrians in crowded settings. The input to MVD methods is images from multiple calibrated cameras observing the same area from different viewpoints with an overlapping field of view. The predicted output is an occupancy map [1] in the ground plane (bird's eye view). The solutions of MVD has evolved from classical methods [1], [2], [3], to hybrid approaches [4] to end-to-end trainable deep learning architectures [5]. Expectedly, the current landscape of MVD is dominated by end-to-end trainable deep learning methods [5], [6], [7]. We argue that by *training and testing on homogeneous data*, current deep MVD methods have overlooked critical fundamental

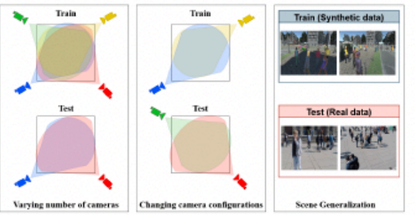


Fig. 1. Three forms of generalization required in MVD: (a) varying number of cameras, (b) different camera configurations, and (c) generalizing to new scenes.

2) *Varying configuration:* The model should not overfit to the specific camera configuration. The performance should be similar even with altered camera positions, as long as they span the dedicated area.
3) *Varying scenes:* Models trained on one scene should work on another (model trained on a traffic signal should work on a setup inside a university).

Surprisingly, the existing deep learning-based MVD methods are primarily trained and tested with the same camera configuration, on the same scene, using the same number of cameras. Even the environmental conditions (time, weather, etc.) are similar across train and test splits. For instance, the most commonly used Wildtrack dataset [8] includes a 200 second recording from all cameras, where the first 3 minutes are used for training and the rest of the 20 seconds are used for testing. We argue that the current State Of The Art (SOTA) methods are seriously hindered from the deployment perspective. The current models [5], [6], [7] will break if a camera malfunctions. They will need retraining if a camera needs to be added to the setup. Furthermore, our experiments

2025 Presentation

Paper

Generalization Challenges in MVD

- SOTA methods fail to generalize well
  - Trained on homogeneous datasets
  - Fixed camera setups and scenes
- Critical form of generalizations
  - Varying number of cameras
  - Varying camera configurations
  - New scenes

GMVD: The Generalized Multi-View Detection

Bringing Generalization to Deep Multi

Jeet Vora[1], Swetanjal Dutta[1], Kanishk Jain[1], Shyar

***Abstract*— Multi-view Detection (MVD) is highly effective for occlusion reasoning in a crowded environment. While recent works using deep learning have made significant advances in the field, they have overlooked the generalization aspect, which makes them *impractical for real-world deployment*. The key novelty of our work is to *formalize* three critical forms of generalization and *propose experiments to evaluate them*: generalization with i) a varying number of cameras, ii) varying camera positions, and finally, iii) to new scenes. We find that existing state-of-the-art models show poor generalization by overfitting to a single scene and camera configuration. To address the concerns: (a) we propose a novel Generalized MVD (GMVD) dataset, assimilating diverse scenes with changing day-time, camera configurations, varying number of cameras, and (b) we discuss the properties essential to bring generalization to MVD and propose a barebones model to incorporate them. We perform a comprehensive set of experiments on the WildTrack, MultiViewX and the GMVD datasets to motivate the necessity to evaluate generalization abilities of MVD methods and to demonstrate the efficacy of the proposed approach. The code and the proposed dataset can be found at https://github.com/jeetv/GMVD**

Fig. 1. T
of camera
scenes.

2) *Va*
to
sh
as

I. Introduction

2025 Presentation

Paper

Proposed Method

GMVD: The Generalized Multi-View Detection

- Average pooling
- DropView regularization
- KL-Divergence + Cross-Correlation loss

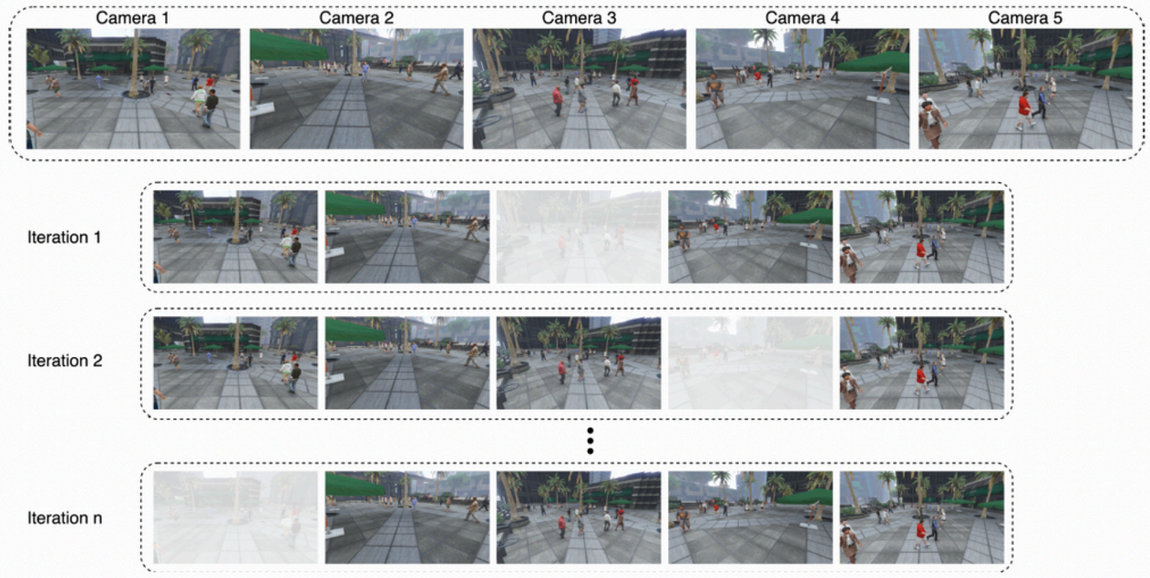
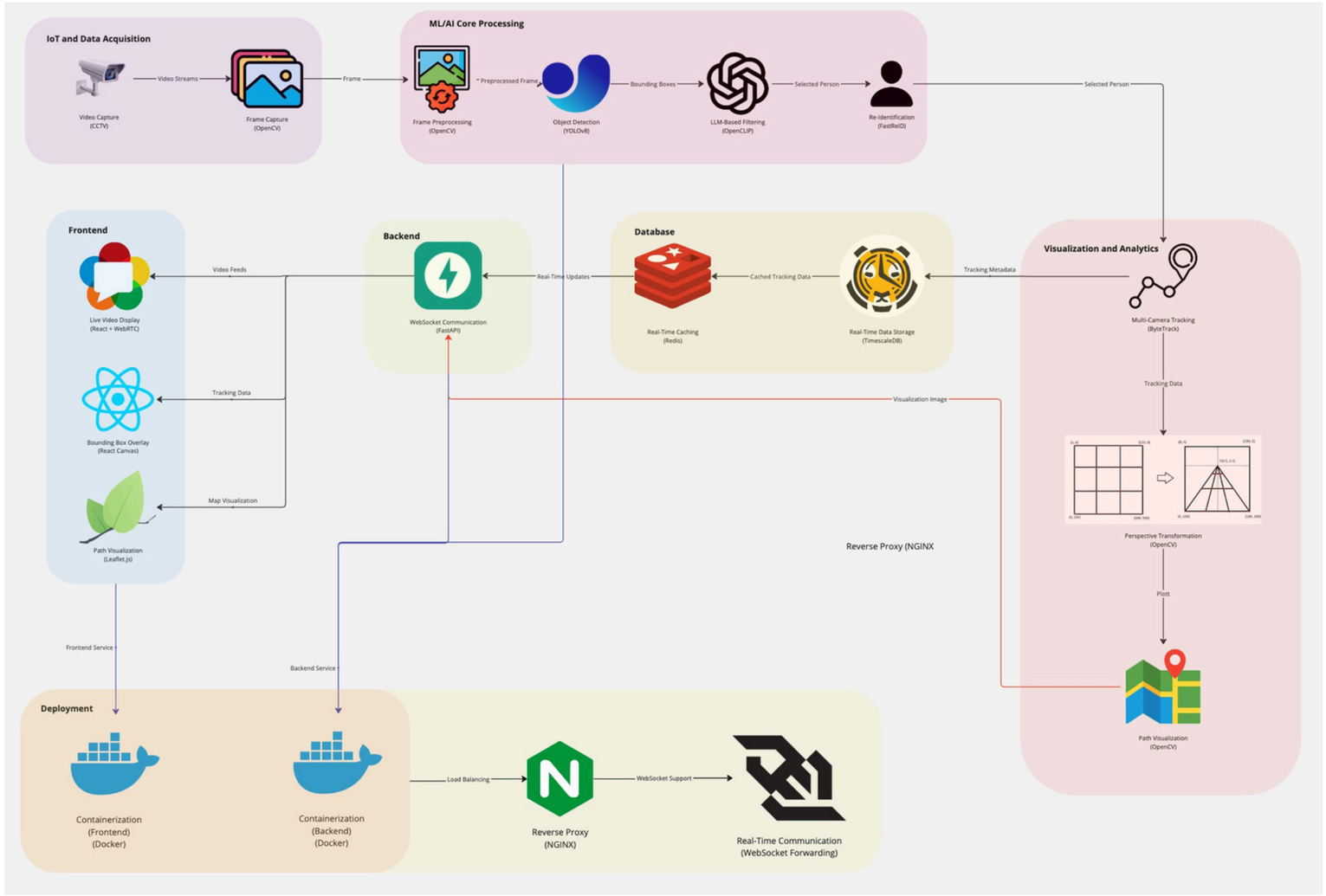


Fig. 5. An illustration of our proposed DropView regularization

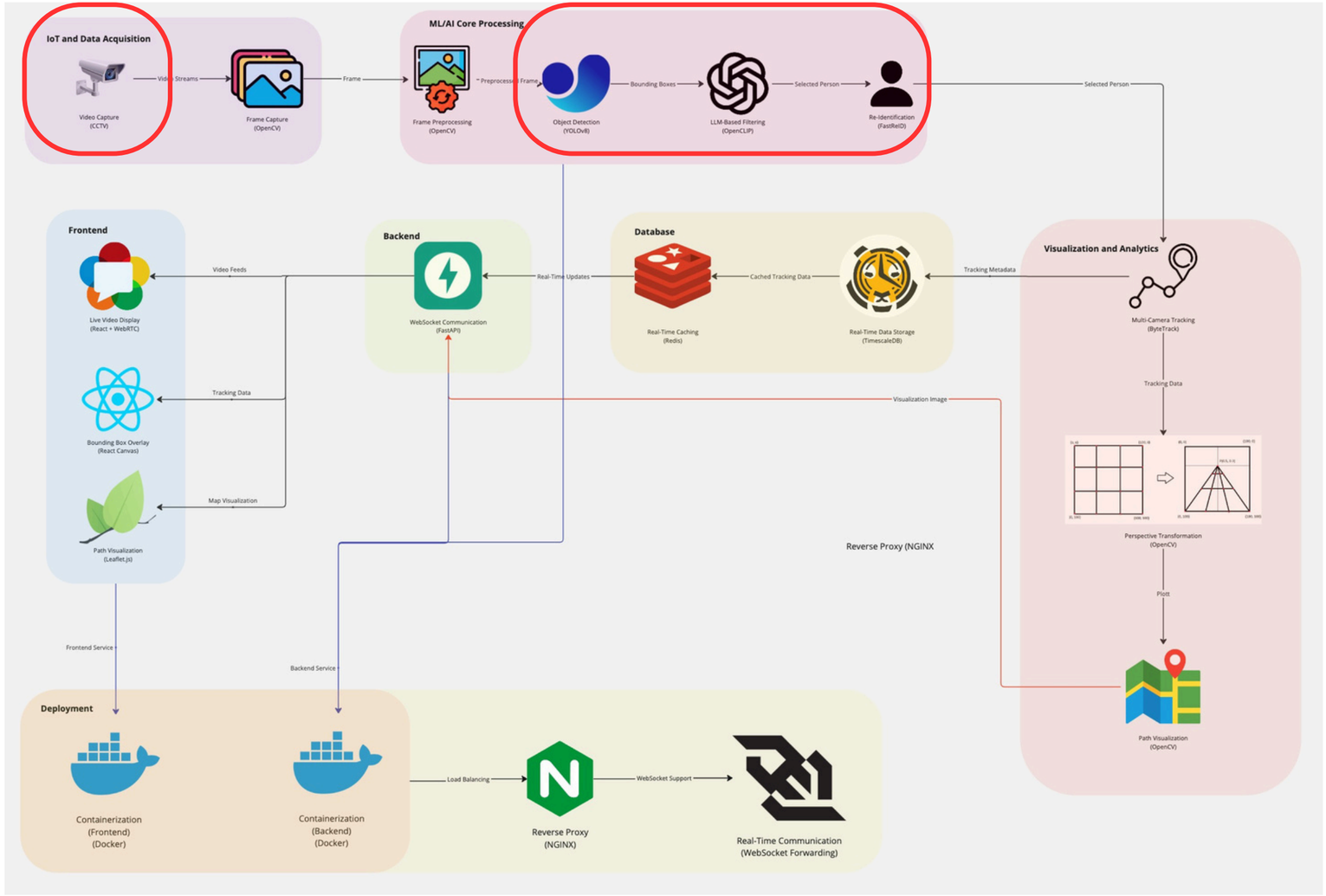
2025 Presentation

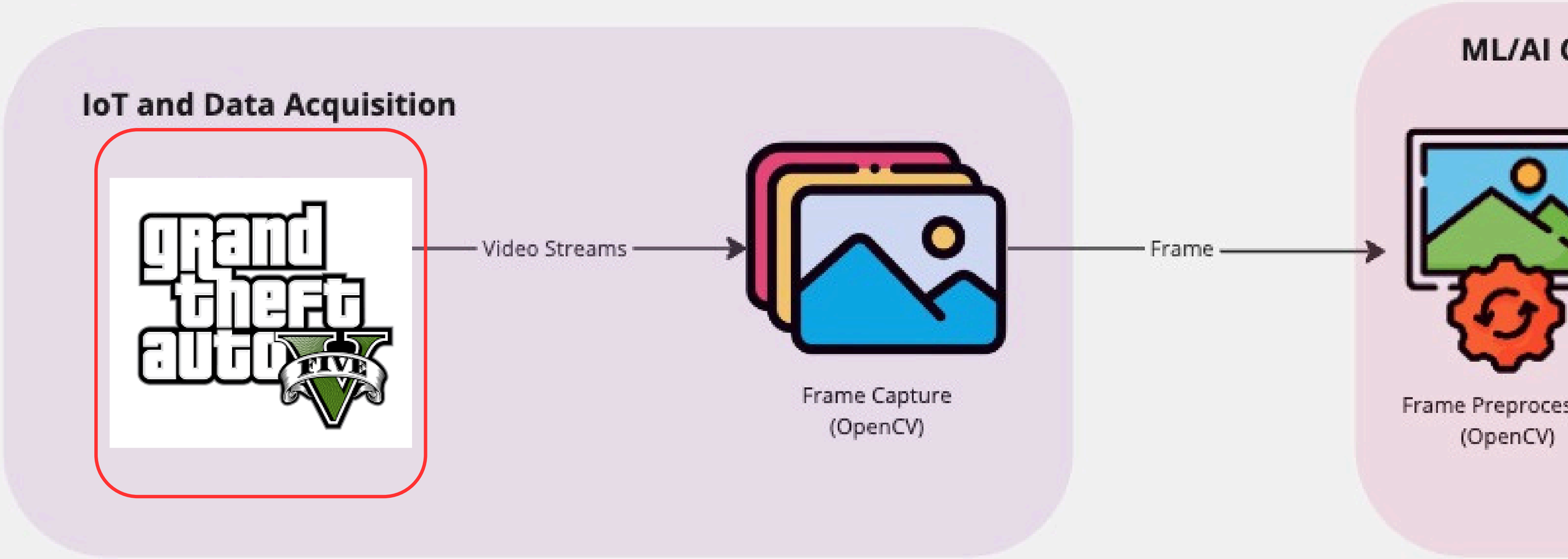
What changes?



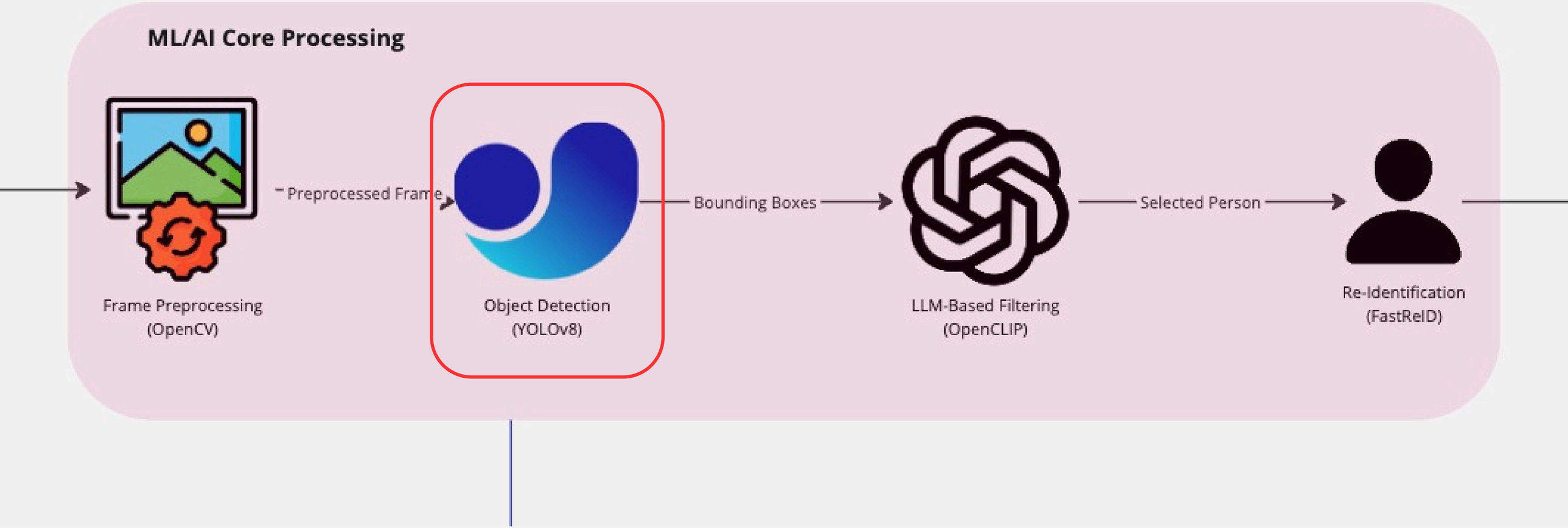
2025 Presentation

What changes?

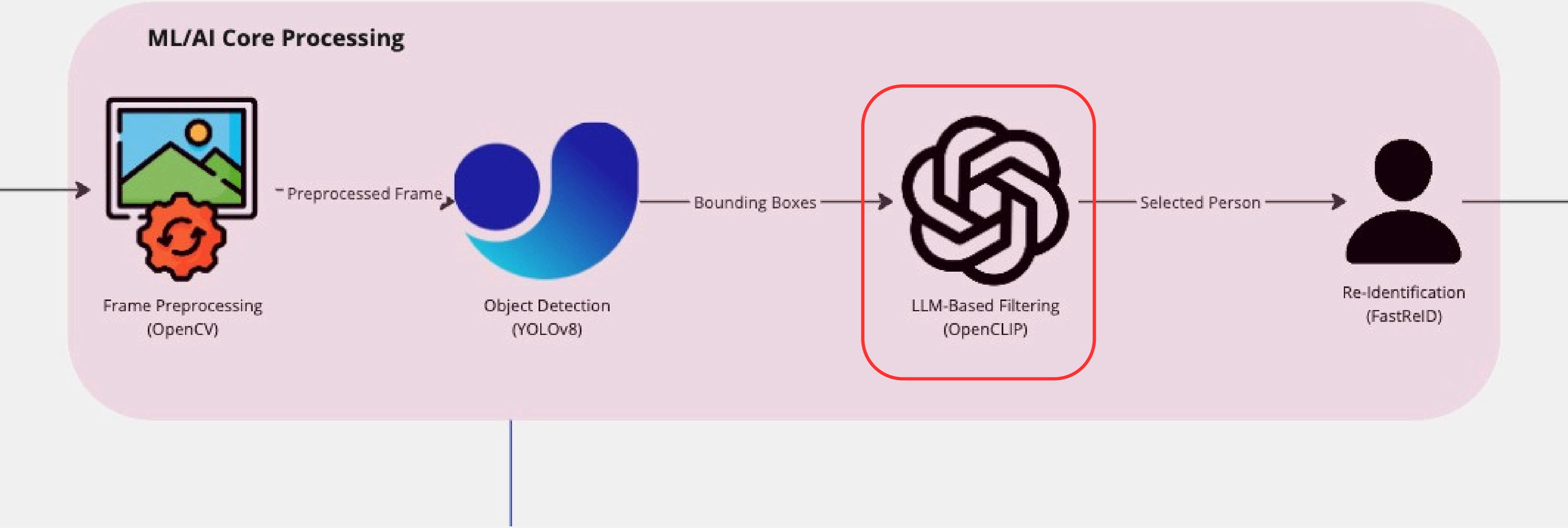




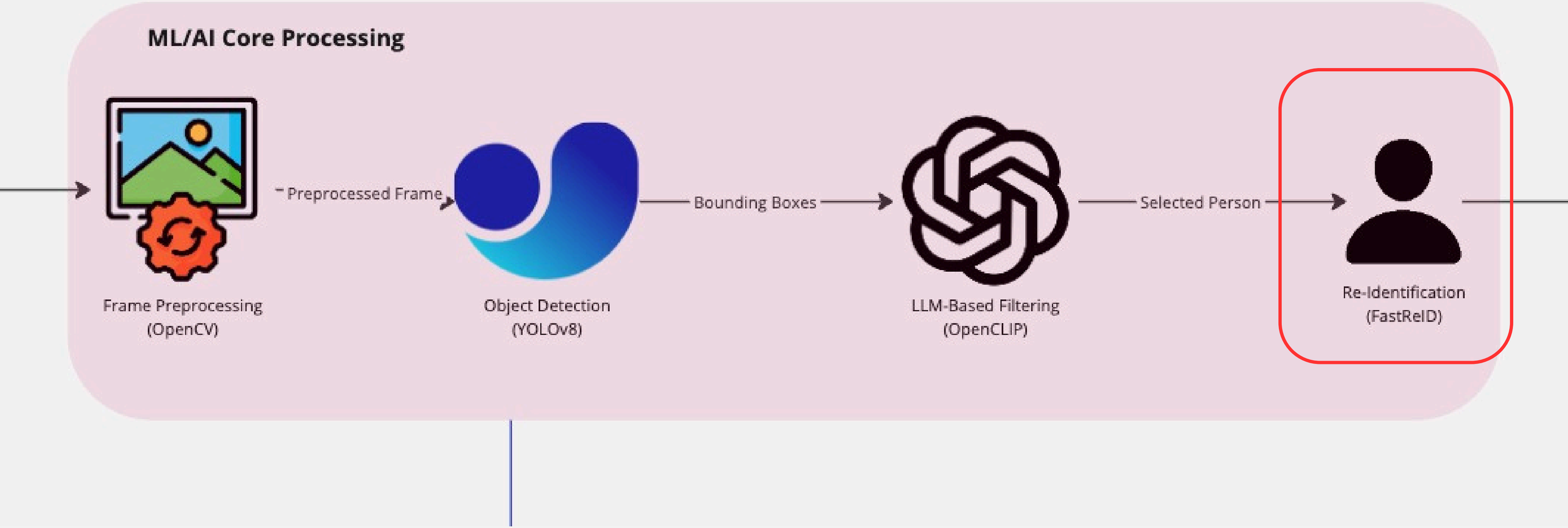
- Set up multiple in-game cameras using ScriptHookV or Native GTA V functions to simulate a multi-camera surveillance system.
- Position cameras at different angles and locations to cover the entire area, ensuring overlapping fields of view for re-identification testing.
- Record or stream the footage from these cameras in real time, saving it as video files or directly passing frames to the backend.



- Fine-tune YOLOv8 on a dataset like **GMVD** to improve detection accuracy for GTA V NPCs.
- These datasets include bounding box annotations for NPCs in various scenes, lighting conditions, and weather variations, making them ideal for synthetic environments.



- Fine-tune OpenCLIP on cropped images of NPCs from the **PoseTReID** datasets.
- Use prompts tailored to GTA V NPCs, such as "a person wearing a red jacket and jeans" or "a person with a backpack."
- Ensure the model learns to associate synthetic NPC features with textual descriptions.



- Train FastReID on the **PoseTReID** dataset, which includes multi-view images of NPCs with unique IDs.
- Use the dataset's synchronized multi-camera views to improve the model's ability to recognize NPCs across different angles and lighting conditions.

Thank You

2025 Presentation