Master in Computer Vision Barcelona

Module: C6 - Final

Project: Ai City Challenge

Coordinator: J. Ruiz-Hidalgo / R. Morros

Team: G. Grigoryan, S. van der Linde, V. Heuer, P. Zetterberg

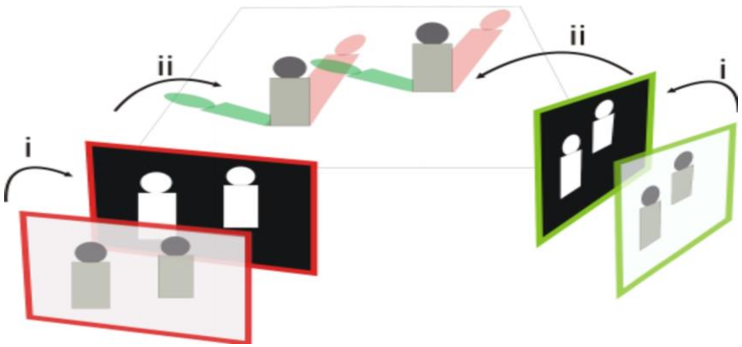
Grigor Grigoryan
1732910

Table of Contents

1. Introduction and Motivation

Grigor Grigoryan
1732910

Introduction & Motivation



Multi-camera view tracking

Introduction & Motivation

Introduction

- This week's assignment focuses on multi-camera vehicle tracking and speed estimation using video estimation.
- The goal is to find an algorithm that is capable of identifying and tracking vehicles across multiple camera views.
- The project aligns with the CPVR 2022 AI City Challenge (Track 1), which aims to improve traffic monitoring and urban mobility analysis.

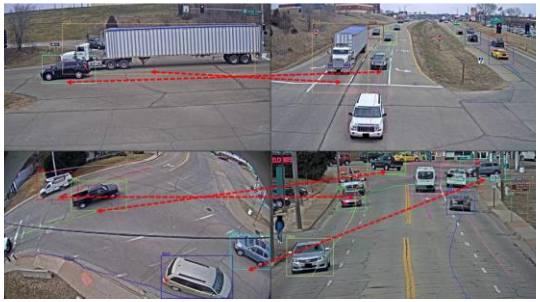
Motivation

- Real-world applications: Multi-camera tracking enhances traffic managment, traffic safety and congestion control.
- Challenges: Tackling challenges such as the re-identifcation of vehicles, occlusions and varying camera angles tests the boundaries of current AI models.
- Computer Vision applications: This project shows how deep learning and image processing can be leveraged for real-world traffic monitoring.

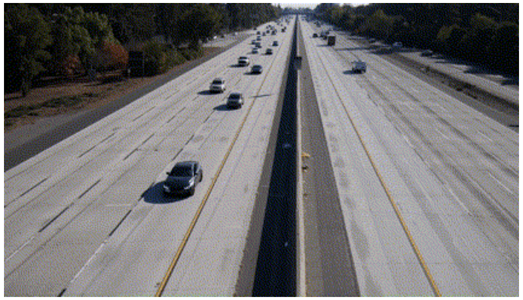
Grigor Grigoryan
1732910

2. Dataset Overview

Dataset Overview



Multiple views



Frame from AI City Challenge

Dataset Overview

Dataset: AI City Challenge (Track 1)

- Source: Naphade, M., Wang, S., Anastasiu, D. C., Tang, Z., Chang, M., Yao, Y., ... & Chellappa, R. The 6th ai city challenge. In 2022 IEEE. In *CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 3346-3355).
- Purpose: multi-target, multi-camera vehicle tracking across urban intersections
- Sequences used:
  - Test sequence: S03
  - Train sequences: S01, S04

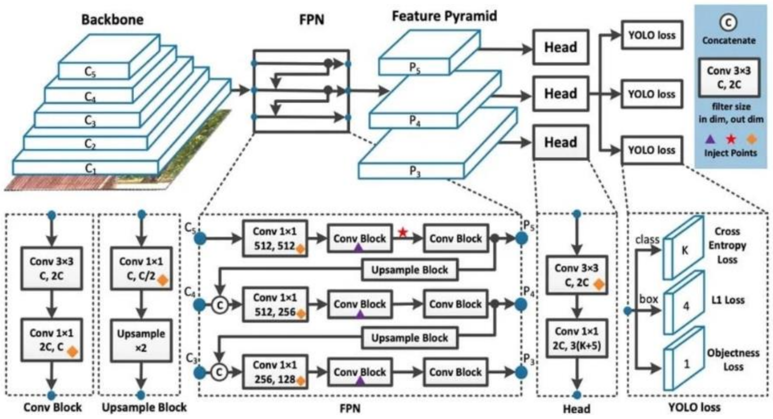
Key Characteristics

- Multiple camera views covering different intersections.
- Real-world traffic data with varying weather/lighting conditions.
- Challenges:
  - High vehicle density
  - Different camera angles and perspectives
  - Vehicle occlussions
  - Identity switching

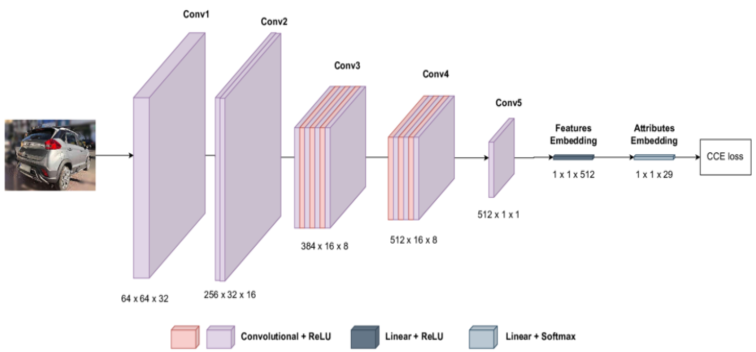
3.1 Algorithm Overview

Saranne van der Linde
1733548

YOLOv8n and OSNet



YOLOv8n Architecture



OSNet_x1_0 Architecture

Algorithm Overview

YOLOv8n

- Lightweight deep learning network used for real-time object tracking.
- Used to detect vehicles in video frames.
- Vehicles are assigned bounding boxes and class information, which is saved to .txt files so it can later be accessed for multi-camera tracking.

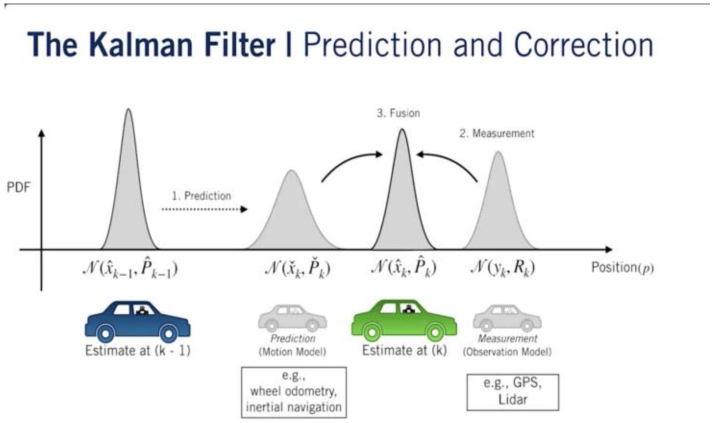
OSNet_x1_0

- Pre-trained re-identification model used for distinguishing objects across multiple video streams.
- After vehicle detection, each detection is compared with objects from other video sequences to check for similarity using OSNet.
- This enables us to match vehicles across multiple cameras.
- Once the matching is finished, objects that do not appear in the matched_objects.txt file are filtered out.
- We keep only the correctly identified vehicles for the final analysis.

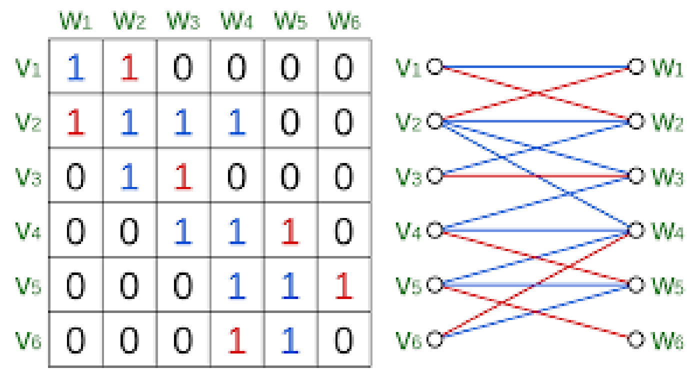
3.2 Algorithm Overview

Saranne van der Linde
1733548

Kalman Filter & Hungarian Matching



Kalman Filtering process



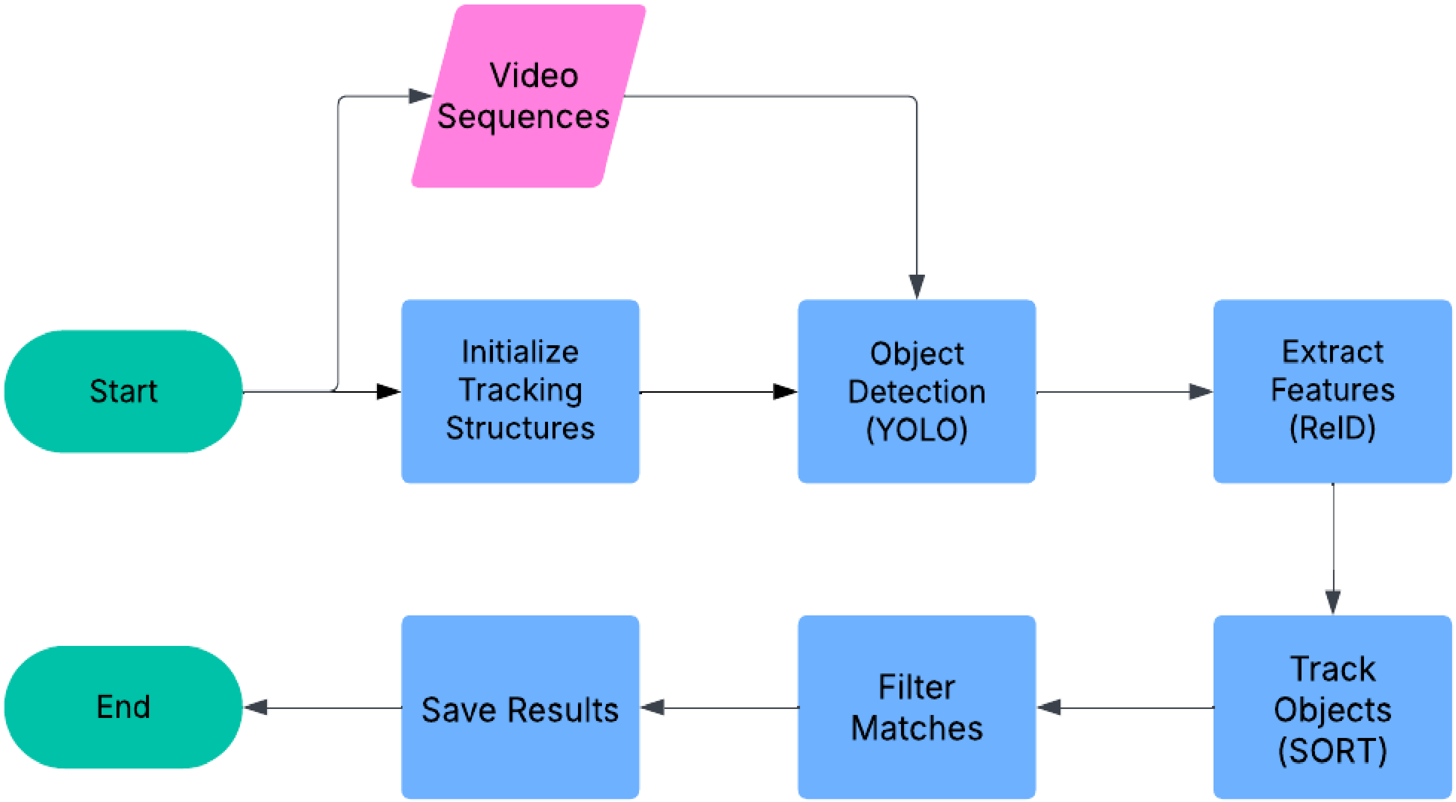
Hungarian Matching

Algorithm Overview

- The **Kalman filter** is used to predict and update object positions in tracking:
- It uses the past state of an object to estimate the next position, and it does so by accounting for velocity and acceleration.
- It then compares the prediction with the actual detection and corrects errors using a measurement update.
- The Kalman gain adjusts the trust in the prediction compared to the new prediction.
- The Kalman Filter is used to predict the object IDs.
- We used the **SORT** library to achieve this.
- Then, **Hungarian matching** is used to measure the similarity between the detected objects and existing trackers.
- Hungarian matching assigns a cost matrix, which is then used to minimize the total tracking error.

4. Our Workflow

Saranne van der Linde
1733548



5.1 Results for Sequence S01

Name	**Vincent Heuer**
NIU	**1732005**

Table of best results for sequence S01		
Camera	Single Camera (Hota / IDF1)	Multiple Camera (Hota / IDF1)
001	32.7 / 38	34.1 / 39.2
002	12.3 / 10.5	13.4 / 11.4
003	16.2 / 17.2	16.9 / 17.8
004	23.3 / 25.6	24.5 / 26.5
005	12.3 / 15.8	14.6 / 19.2
Average	**19.4 / 21.4**	**20.7 / 22.8**

Discussion

- **Multiple-camera tracking outperforms single-camera tracking** in both HOTA and IDF1 metrics, demonstrating improved identity association across views.
- The improvement is **consistent across all cameras**, with the highest gain seen in C005 (+2.3 HOTA, +3.4 IDF1).
- The overall tracking performance remains **moderate,** there is still room for improvement in detection robustness and re-identification accuracy.

5.2 Results for Sequence S03

Name	Vincent Heuer
NIU	1732005

Table of best results for sequence S03		
Camera	Single Camera (Hota / IDF1)	Multiple Camera (Hota / IDF1)
010	33 / 45.5	34.3 / 46.9
011	30.8 / 43.4	33 / 47
012	70.5 / 84.3	70.3 / 84
013	59.6 / 73.6	70.3 / 82.1
014	20.1 / 28.8	22.9 / 31.1
015	94.5 / 100	94.5 / 100
Average	51.4 / 62.6	54.2 / 65.2

Discussion

- **Multiple-camera tracking outperforms single-camera tracking** in both HOTA and IDF1 metrics, demonstrating improved identity association across views.
- We achieved the best results on camera 15.
- The overall tracking performance remains **moderate,** there is still room for improvement in detection robustness and re-identification accuracy.

5.3 Results for Sequence S04

Name	**Vincent Heuer**
NIU	**1732005**

Table of best results for sequence S04		
Camera	Single Camera (Hota / IDF1)	Multiple Camera (Hota / IDF1)
…		
019	73.6 / 84.9	82.1 / 90.2
024	65.8 / 84.2	68.3 / 86.1
033	5 / 2.3	5.1 / 2.8
034	8.3 / 5.2	8.3 / 5.3
…		
Average	**26.5 / 31.5**	**28.3 / 33.1**

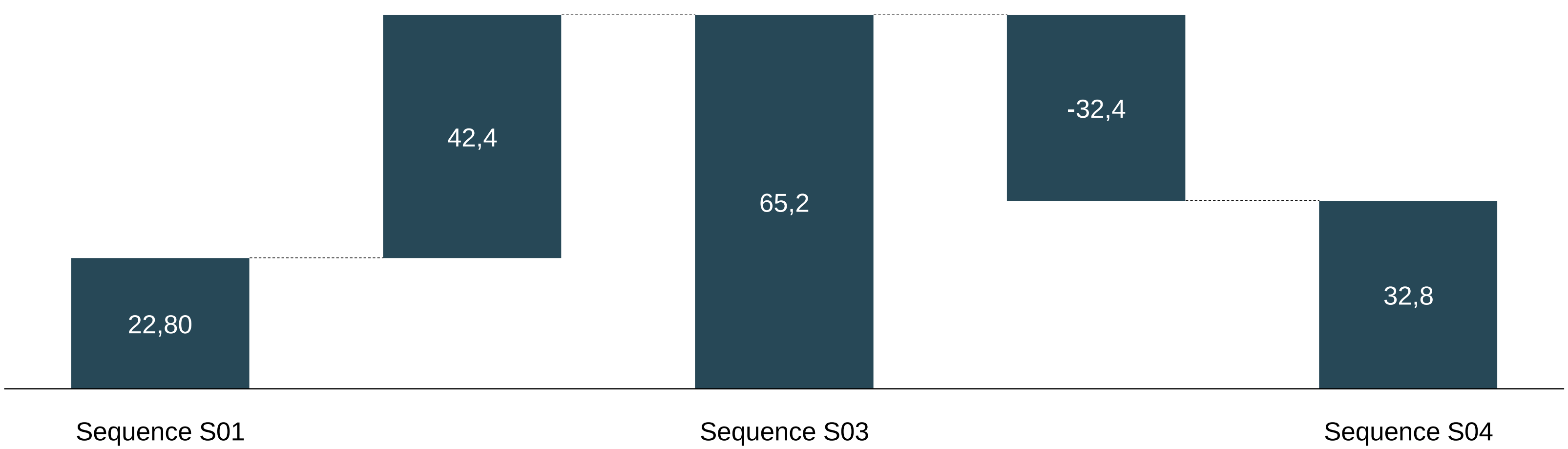
Discussion

- **Multiple-camera tracking improves overall performance**, particularly for **C019** (+8.5 HOTA, +5.3 IDF1) and **C024** (+2.5 HOTA, +1.9 IDF1), showing strong identity matching across views.
- **C033 and C034 remain low performers**, with minimal improvement, indicating challenges in detection or identity consistency.
- **Average tracking performance improves slightly** with multiple-camera tracking (+1.85 HOTA, +1.62 IDF1), but the increase is modest compared to other sequences.
- The large gap between **high-performing (C019, C024) and low-performing (C033, C034) cameras** suggests varying levels of detection reliability, possibly due to scene complexity or occlusions.
- The disparity between high-performing and low-performing cameras suggests that certain scenes or object characteristics are harder to track reliably.

5.4 Comparison of Results and Discussion

Name	Vincent Heuer
NIU	1732005

Performance Comaparison



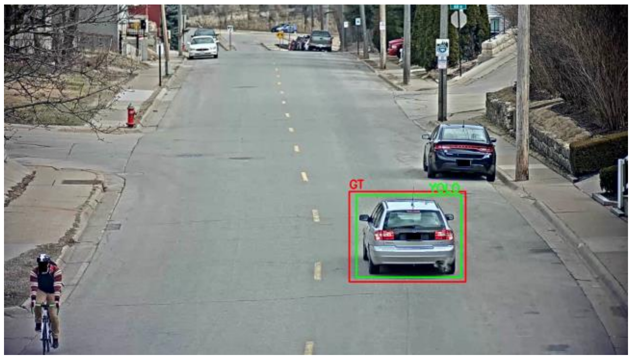
Discussion

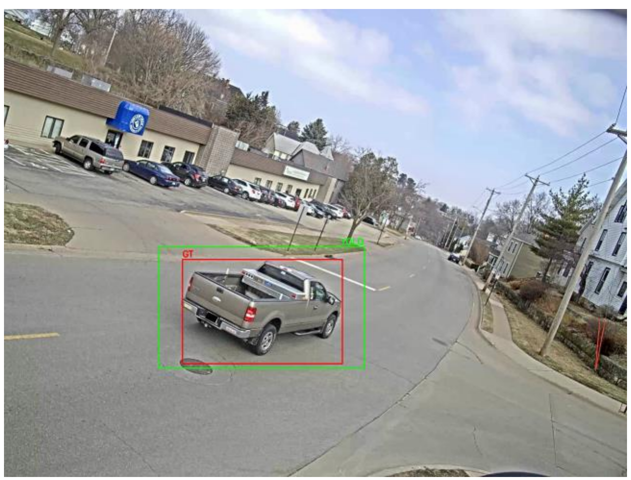
Performance Variability: S03 (65.2%) performed best, likely due to fewer occlusions and stable viewpoints, while S01 (22.8%) and S04 (32.8%) suffered from identity switches, occlusions, and challenging conditions.
Algorithm Limitations: YOLO's detection threshold affects tracking, ReID struggles with viewpoint changes, and SORT's reliance on motion consistency causes ID switches in dense or occluded scenes.
Potential Improvements: Fine-tune ReID with dataset-specific data, upgrade to stronger detectors (e.g., YOLOv8), and enhance tracking with DeepSORT or ByteTrack for better identity preservation.

Name	Philip Zetterberg
NIU	1733746

6.1 Qualitative Analysis – Filtering YOLO Detections
False Negatives

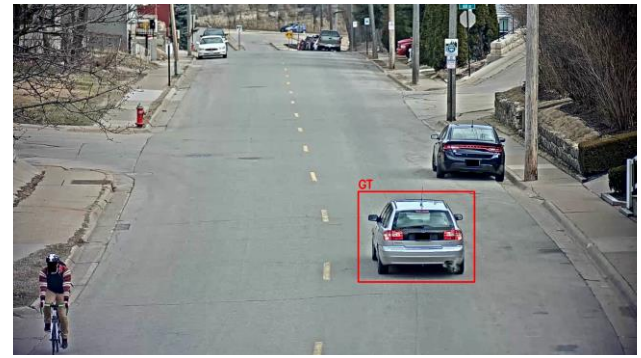
YOLO Detections





SORT - Kalman

After Filtering



6.2 Qualitative Analysis – Filtering YOLO Detections
False Positives

Name	Philip Zetterberg
NIU	1733746

YOLO Detections

SORT - Kalman

After Filtering

6.3 Qualitative Analysis – Kalman Filter – ID Switch

Name	Philip Zetterberg
NIU	1733746

Frame: 1283 - Det ID: **773**

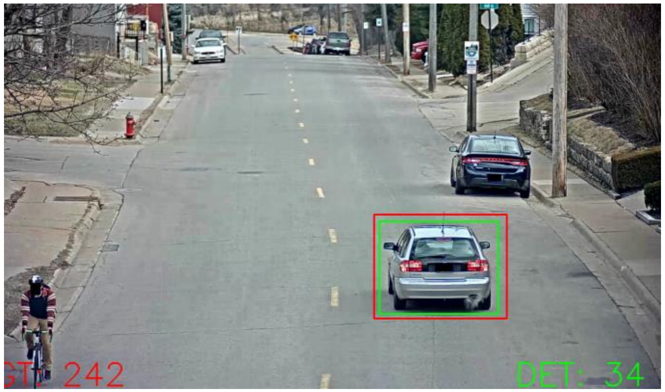
Frame: 1284 - Det ID: **772**

Frame: 1286 - Det ID: **771**

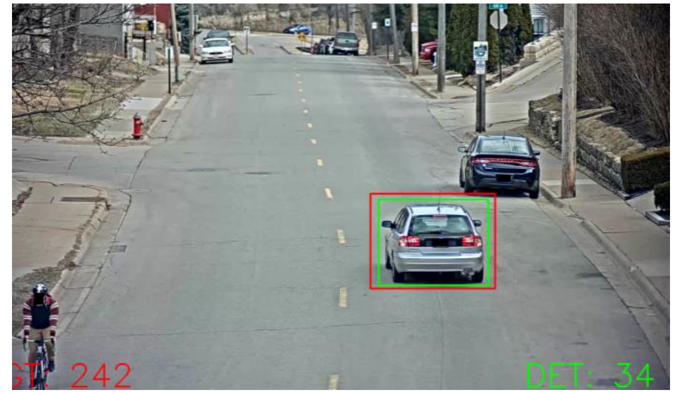
Frame: 1287 - Det ID: **771**

6.4 Qualitative Analysis – Correct ID Tracking

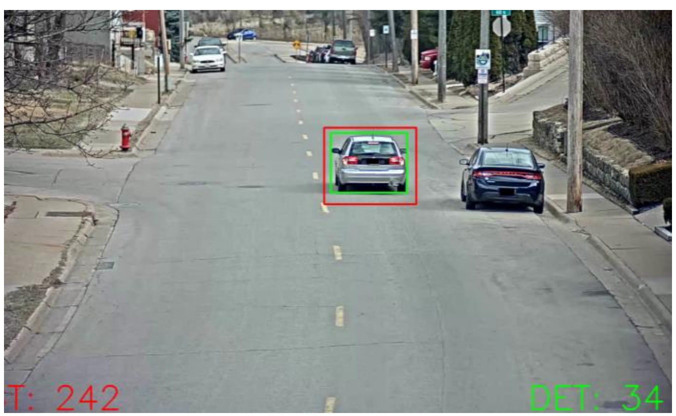
Name	Philip Zetterberg
NIU	1733746



Frame: 224 - Det ID: **34**



Frame: 227 - Det ID: **34**



Frame: 244- Det ID: **34**

Frame: 274 - Det ID: **34**

Name Philip Zetterberg
NIU 1733746

6.5 Qualitative Analysis – Matching

Matches

Cam: **10**

Cam: **12**

Cam: **12**

Cam: **13**

Mismatches

Cam: **10**

Cam: **12**

Unclear

Cam: **12**

Cam: **13**

Conclusion

Name	**Philip Zetterberg**
NIU	**1733746**

Key Findings

1. Many YOLO detections are discarded by the Kalman filter.
2. Object IDs frequently switch, causing inconsistencies.
3. The matching process is not always reliable, often resulting in significant mismatches.

Improvements

1. Train YOLOv8 on the dataset to reduce the number of filtered-out boxes.
2. Enhance the filtering method to minimize mismatches ID switches.
3. Train OSNet to achieve even more accurate matching results.