UNIVERSITY OF MILANO - BICOCCA
Department of Informatics, Systems and Communication
Master's Degree in Data Science

Design and Evaluation of Multi-Camera Multi-Object Tracking Pipelines in Calibrated Environments

Supervisor: Dr. Marco Buzzelli
Co-supervisor: Prof. Simone Bianco

Master's degree thesis by:
Julius Maliwat
student ID number: 864520

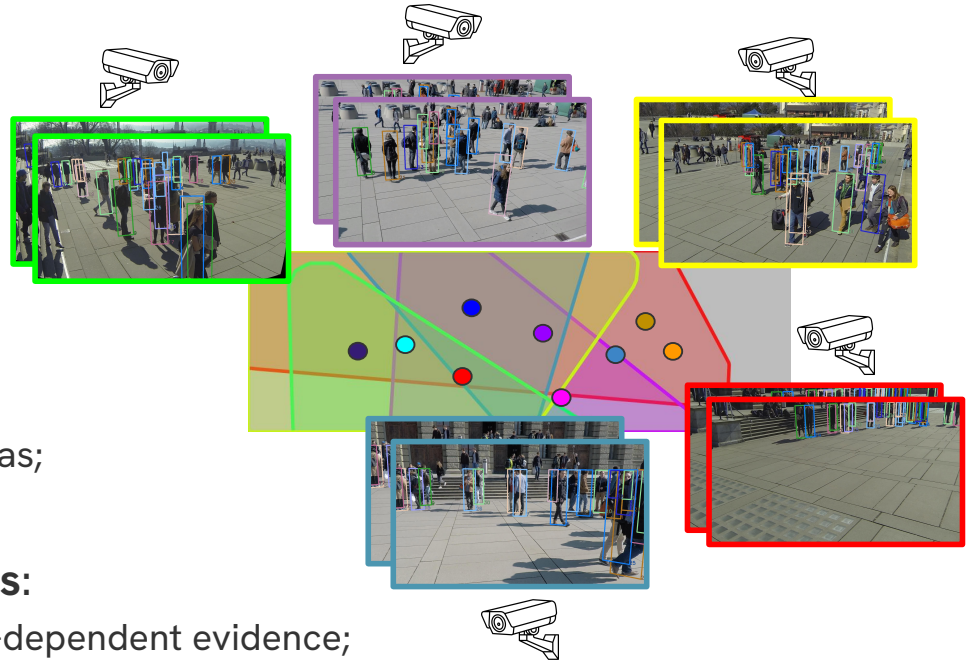
Academic Year 2024/2025

Multi-Camera Multi-Object Tracking: Context & Setting

Task: estimate identity-consistent trajectories over time for multiple objects across multiple cameras.

Domains:

- crowd & mobility analytics;
- public-space safety;
- sports/venue operations;
- autonomous platforms



Our focus:

- pedestrians;
- fixed, calibrated, time-synchronized cameras;
- ground-plane BEV.

Challenges:

1. fuse view-dependent evidence;
2. preserve identities across views/time;
3. resolve occlusion/overlap ambiguities.

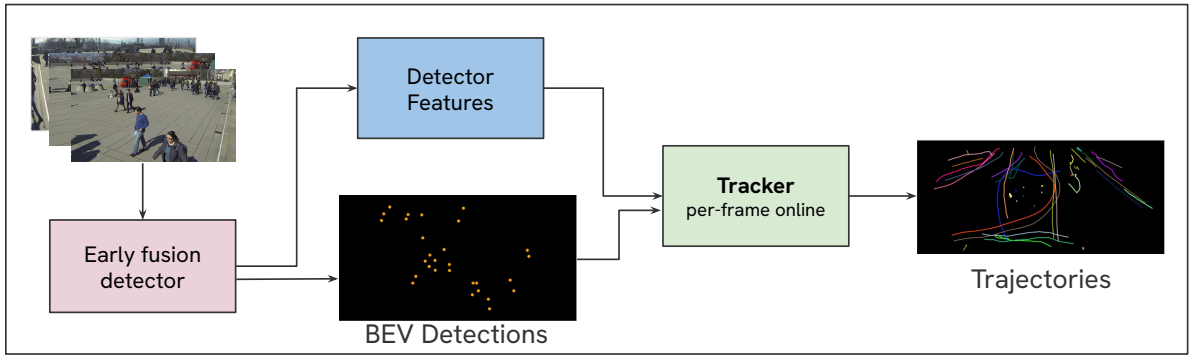
Related work

- Tracking by detection
- **State of art:** fuse multi-camera evidence *before detection* in a shared BEV.

Approach 1

- online: motion + geometric gating
- Appearance from detector features

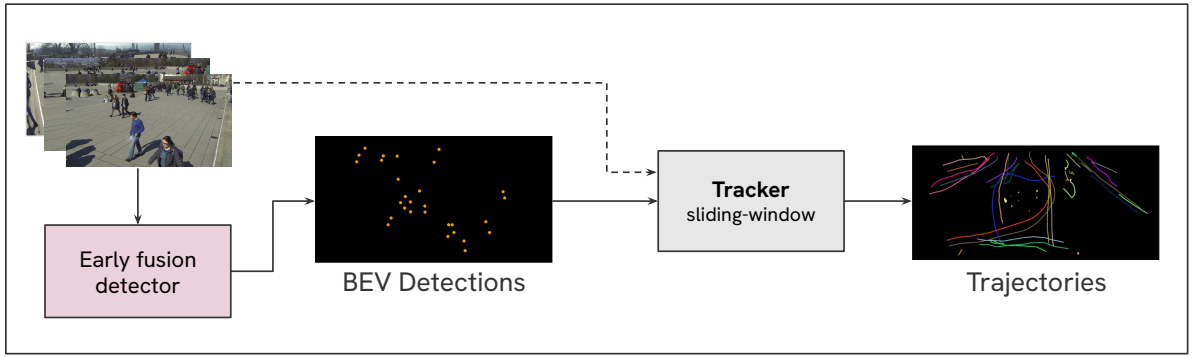
Limitation: tied to the detector



Approach 2

- Appearance from multi-view crops
- Temporal-window association with look-ahead

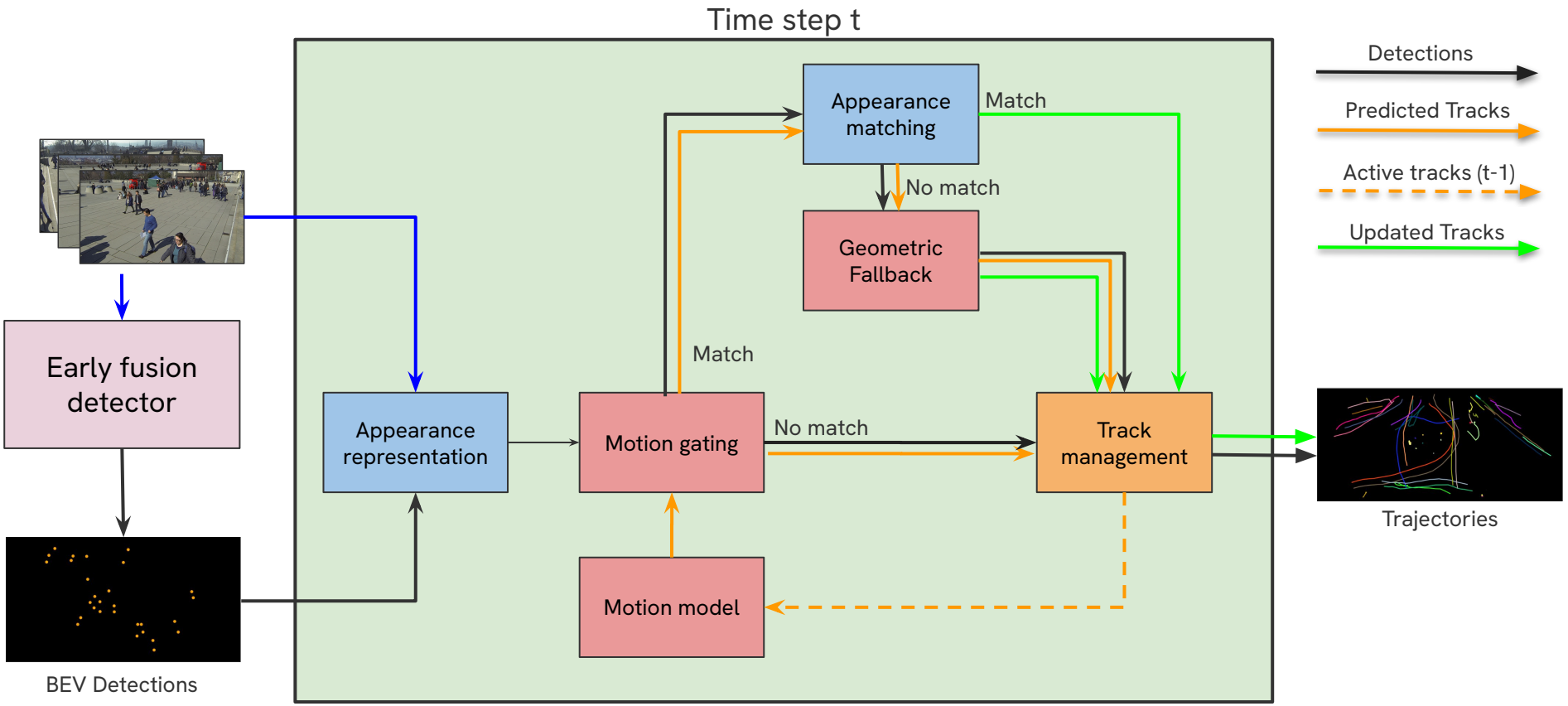
Limitation: not strictly online



Gap: online, detector-agnostic BEV tracker

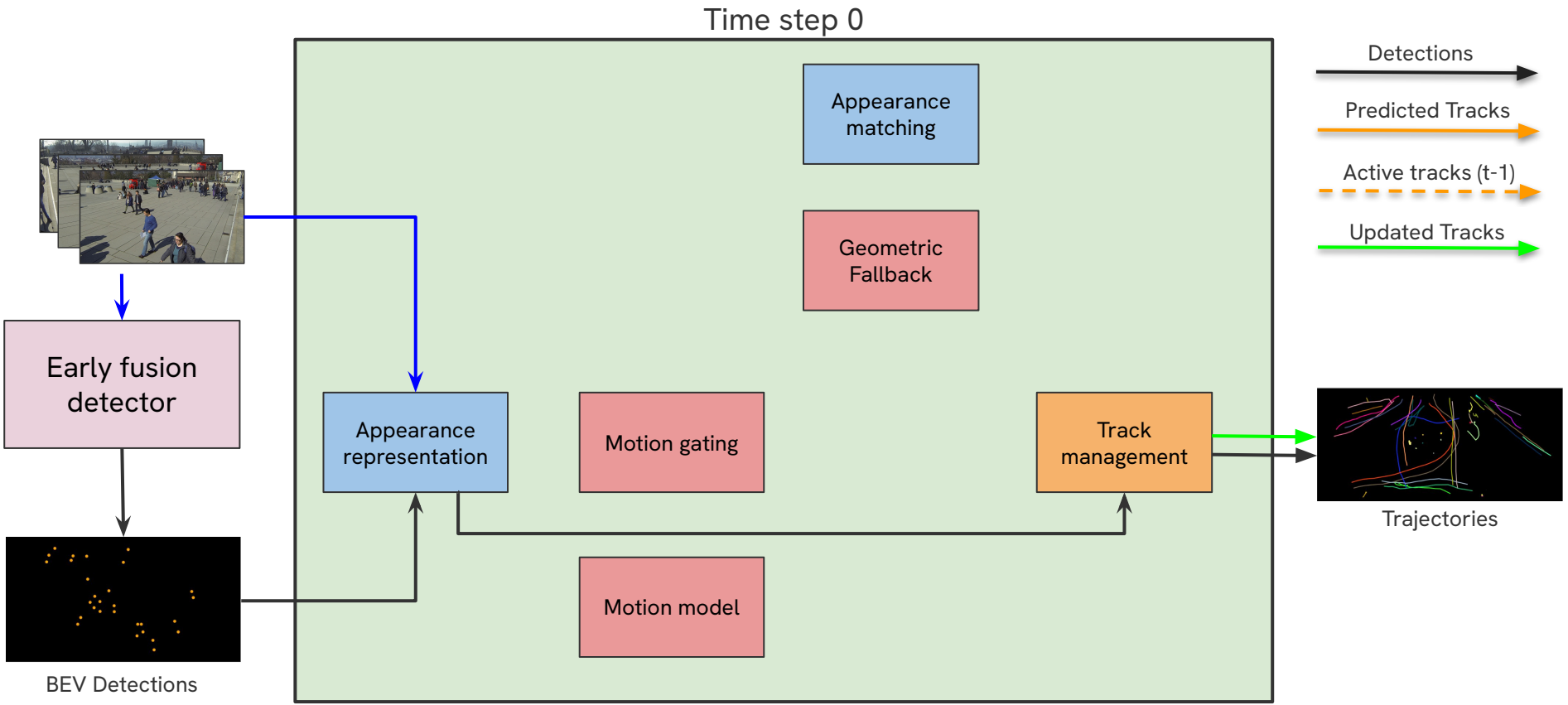
Proposed Tracker

online, detector-agnostic



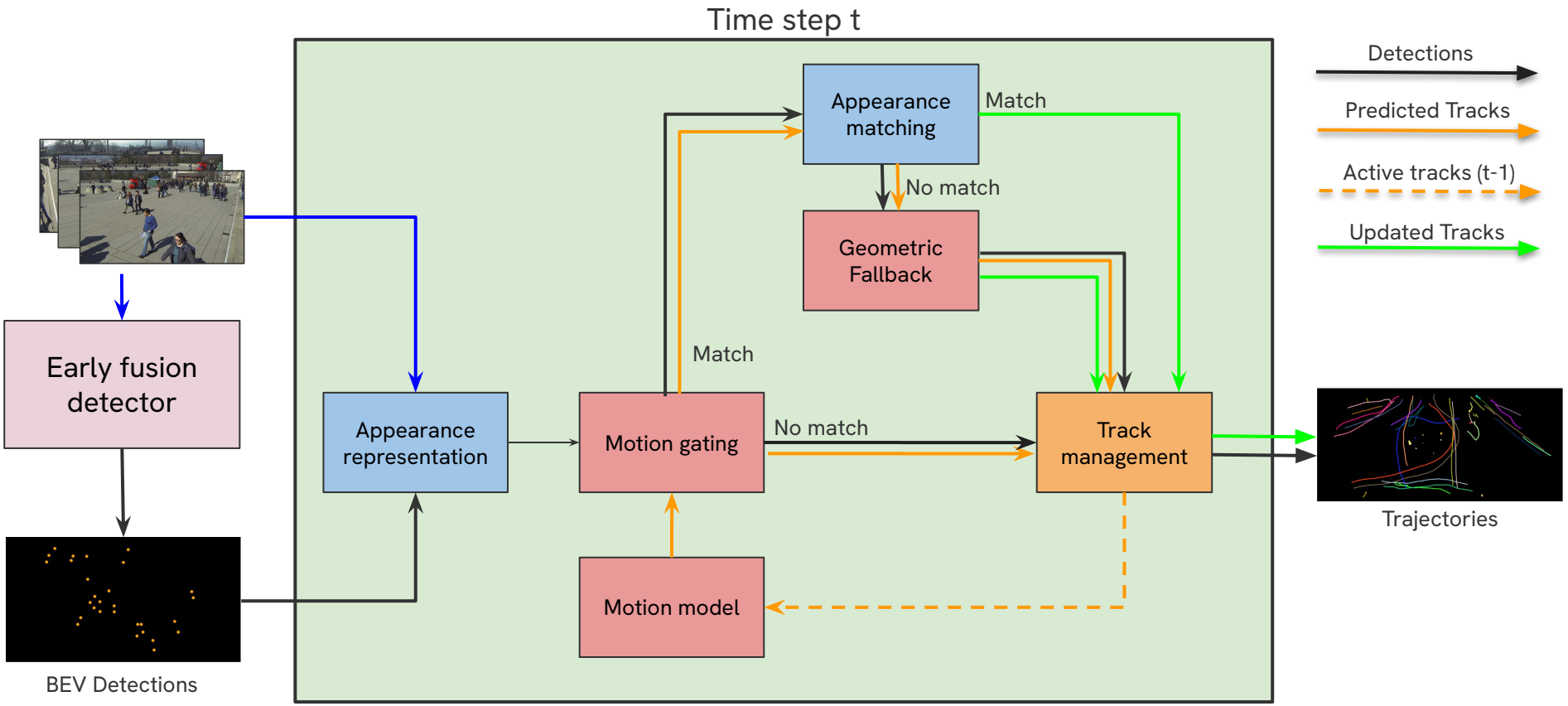
Proposed Tracker - initialization

online, detector-agnostic



Proposed Tracker

online, detector-agnostic



Datasets & Evaluation

Wildtrack: real outdoor plaza at ETH

Statistics

7 cameras;
400 frames/cam
12m×36m ground plane

~20 people per frame
~3.7 cameras per location

Method	IDF1↑	MOTA↑
EarlyBird	92,3	84,9
BEV-SUSHI	95,6	92,6
Attention-Aware	**96,1**	**92,7**
Ours	95,7	91,7

MultiviewX: synthetic, controlled scene

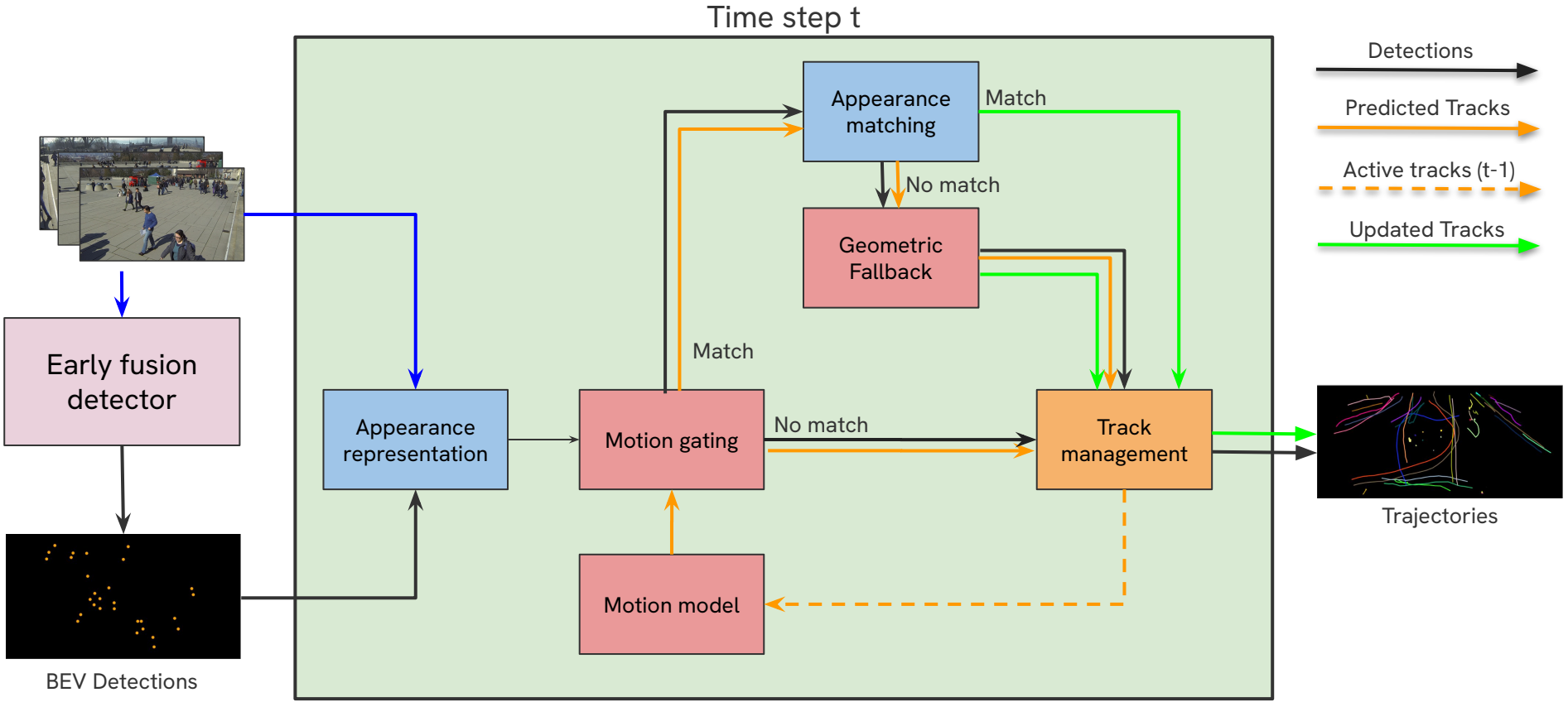
Statistics

6 cameras;
400 frames/cam
16m×25m ground plane

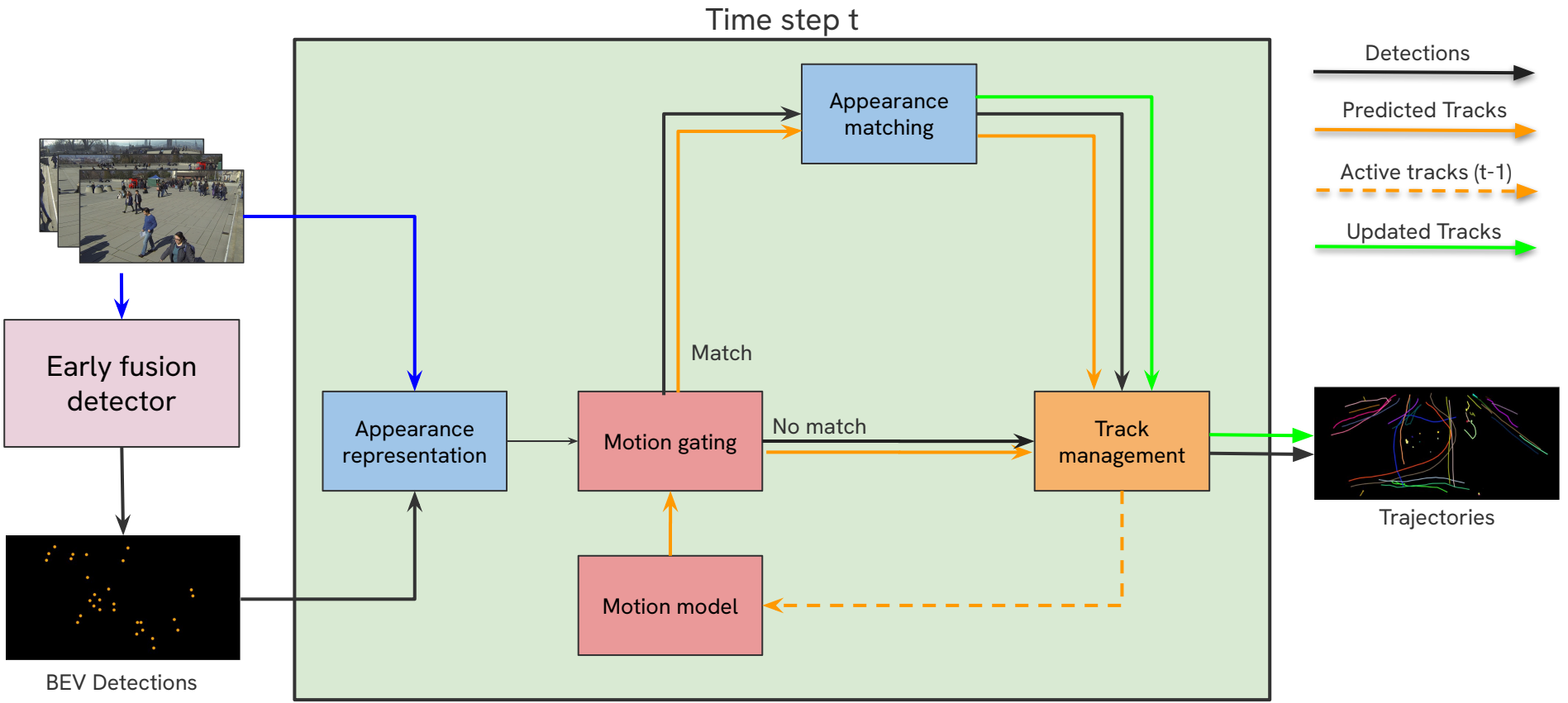
~40 people per frame
~4.4 cameras per location

Method	IDF1↑	MOTA↑
EarlyBird	82,4	88,4
Attention-Aware	85,7	91,3
Ours	**92,0**	**92,2**

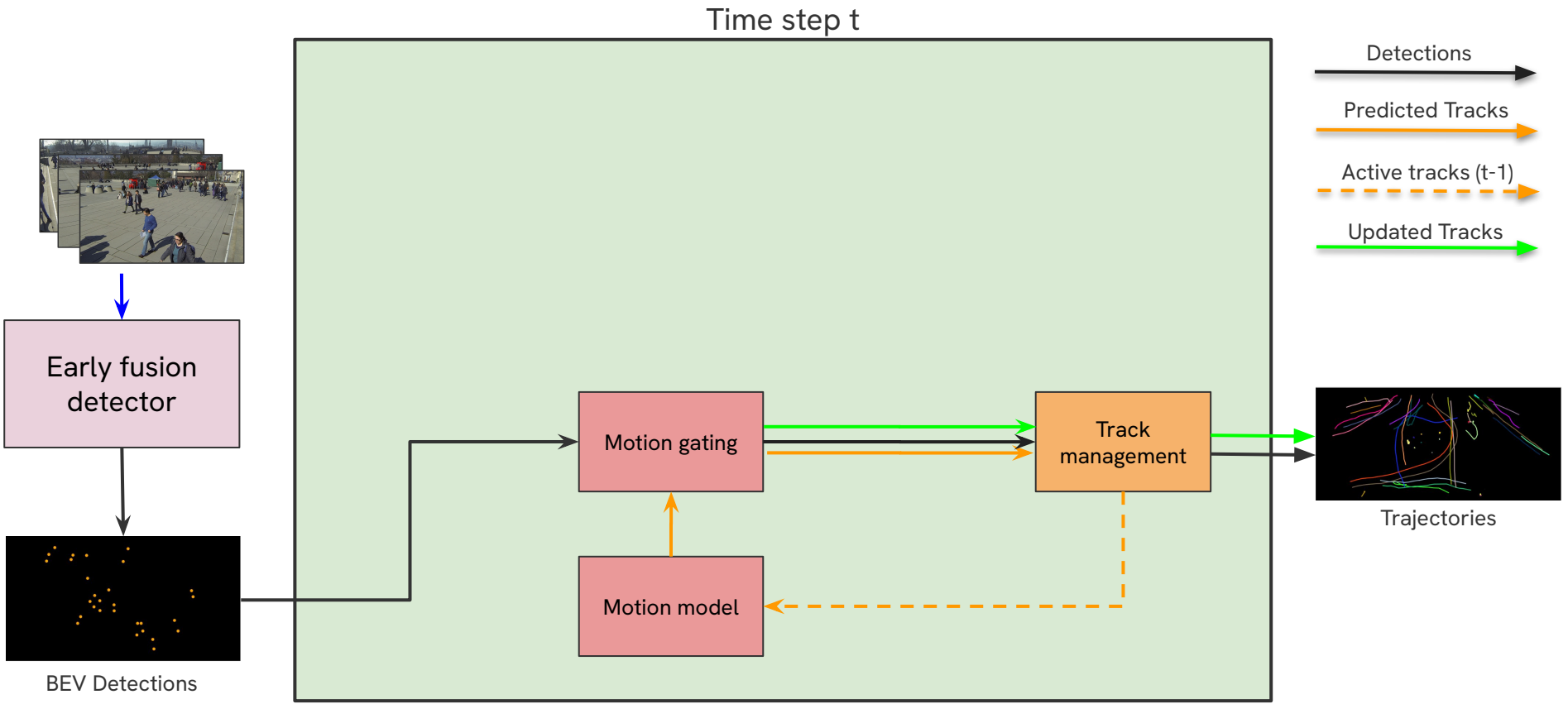
Ablation Study



Ablation Study: appearance-only

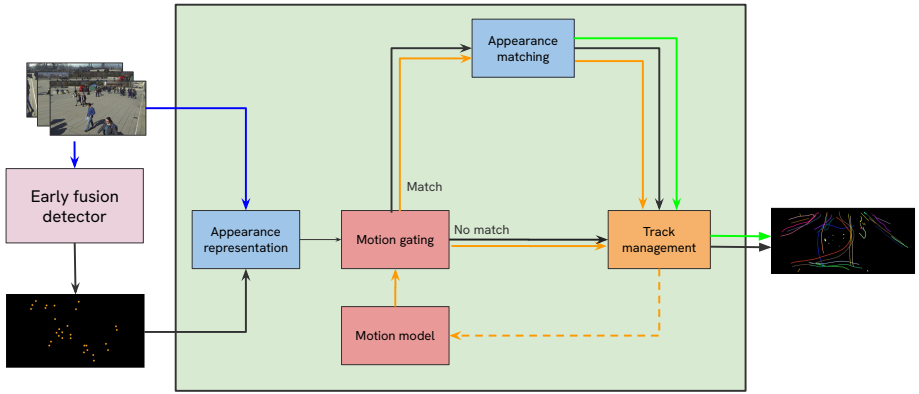


Ablation Study: motion-only

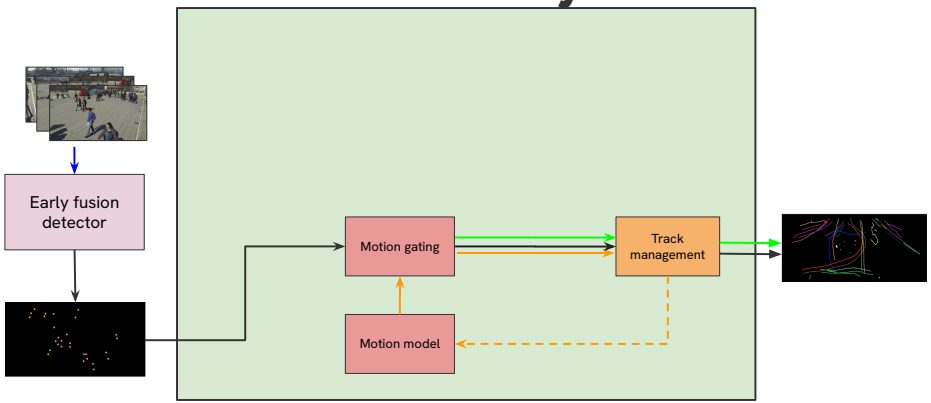


Ablation Study: results

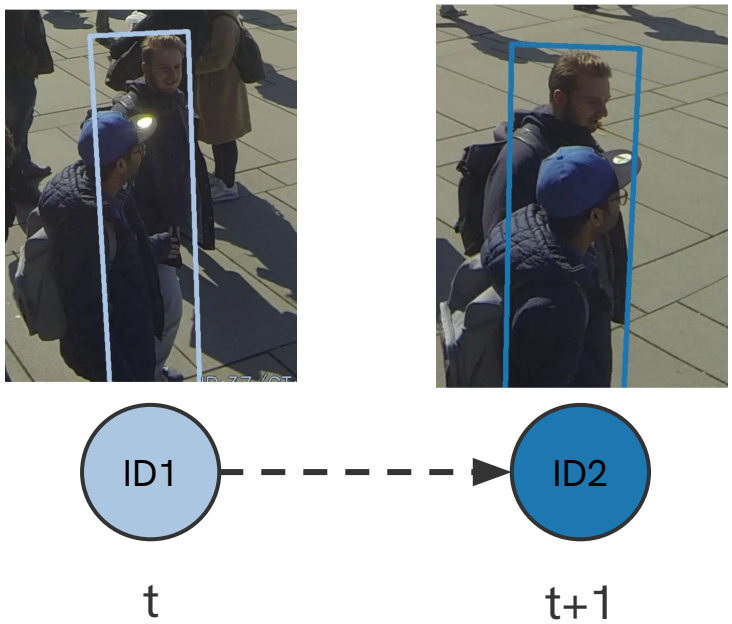
Appearance-only



Motion-only

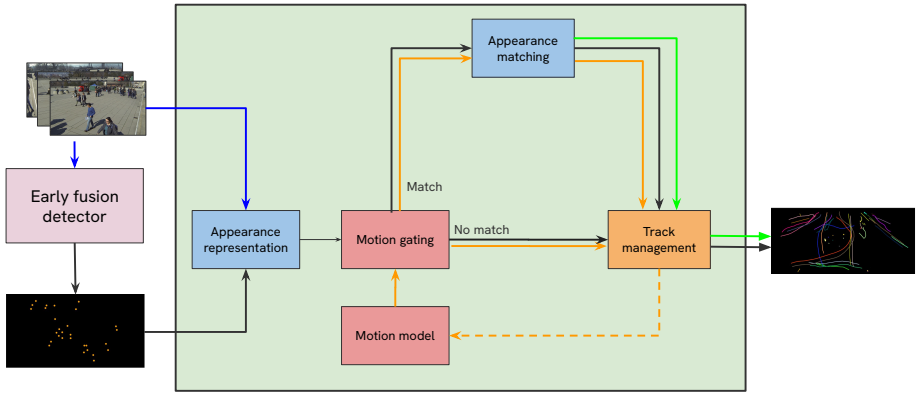


ID Switch

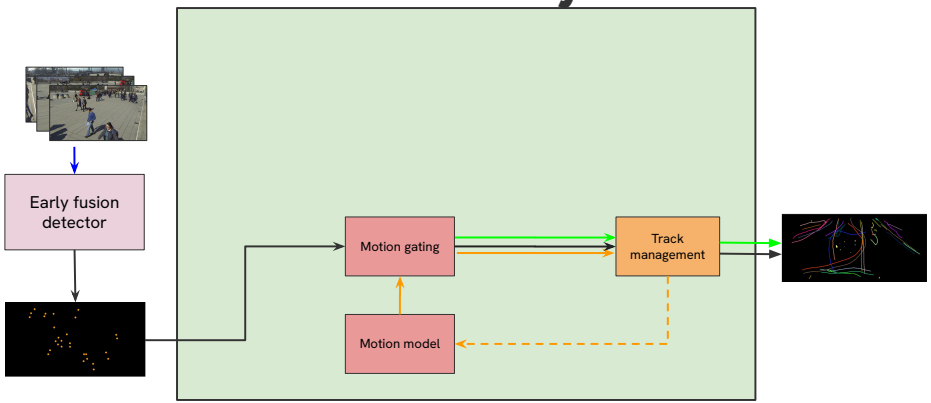


Ablation Study: results

Appearance-only



Motion-only



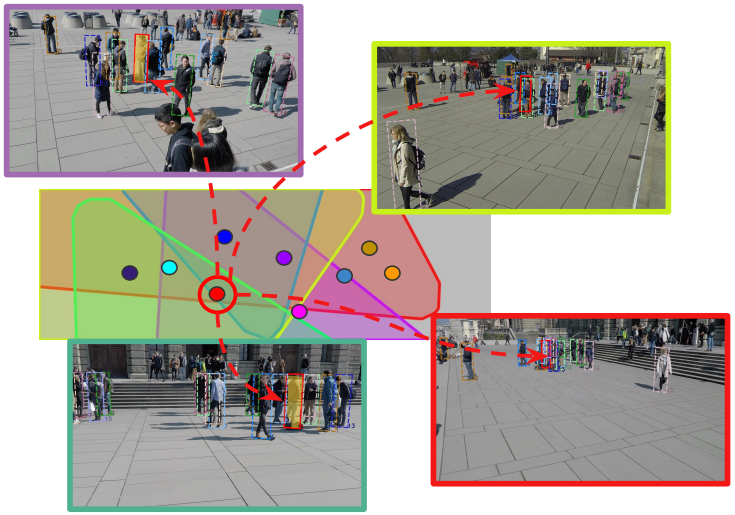
Wildtrack

Method	IDF1↑	MOTA ↑	IDSW↓
Motion-only	90,0	90,6	18
Appearance-only	95,6	91,6	8
Full	**95,7**	**91,7**	**7**

MultiviewX

Method	IDF1↑	MOTA ↑	IDSW↓
Motion-only	89,8	91,0	40
Appearance-only	91,8	92,0	24
Full	**92,0**	**92,2**	**22**

Occlusion-aware visibility score



visible area of the target inside the bounding box after occlusion masking

Diagnose failures

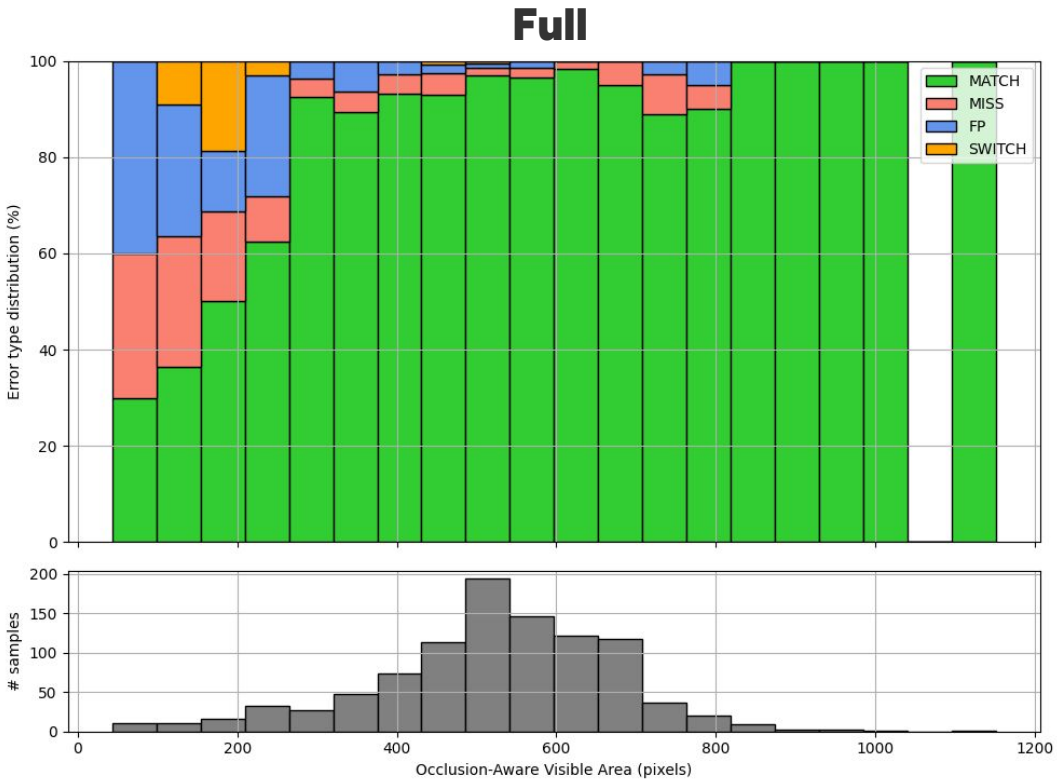
Visibility-aware policy

Reproject each BEV target into all cameras and take the resulting per-view boxes.

Order people by depth, mask pixels covered by nearer subjects

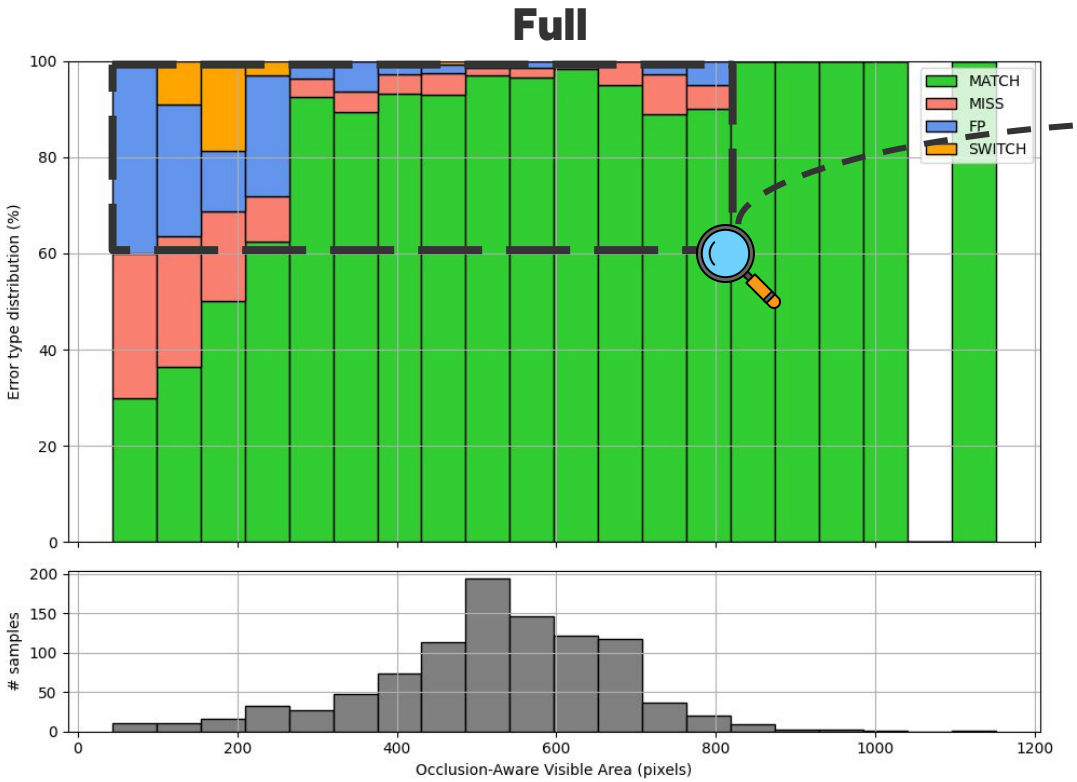
Diagnostics (Wildtrack)

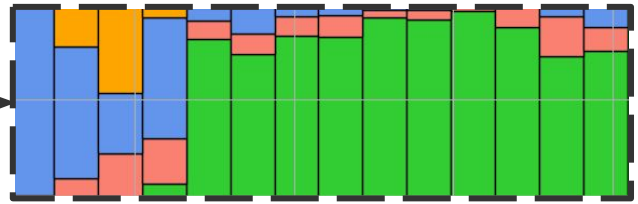
Error breakdown by occlusion-aware visible score

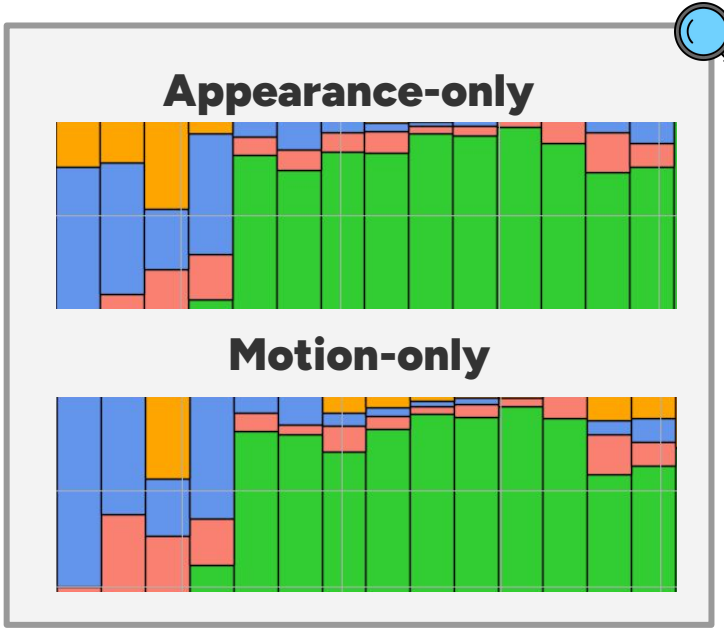


Diagnostics (Wildtrack)

Error breakdown by occlusion-aware visible score

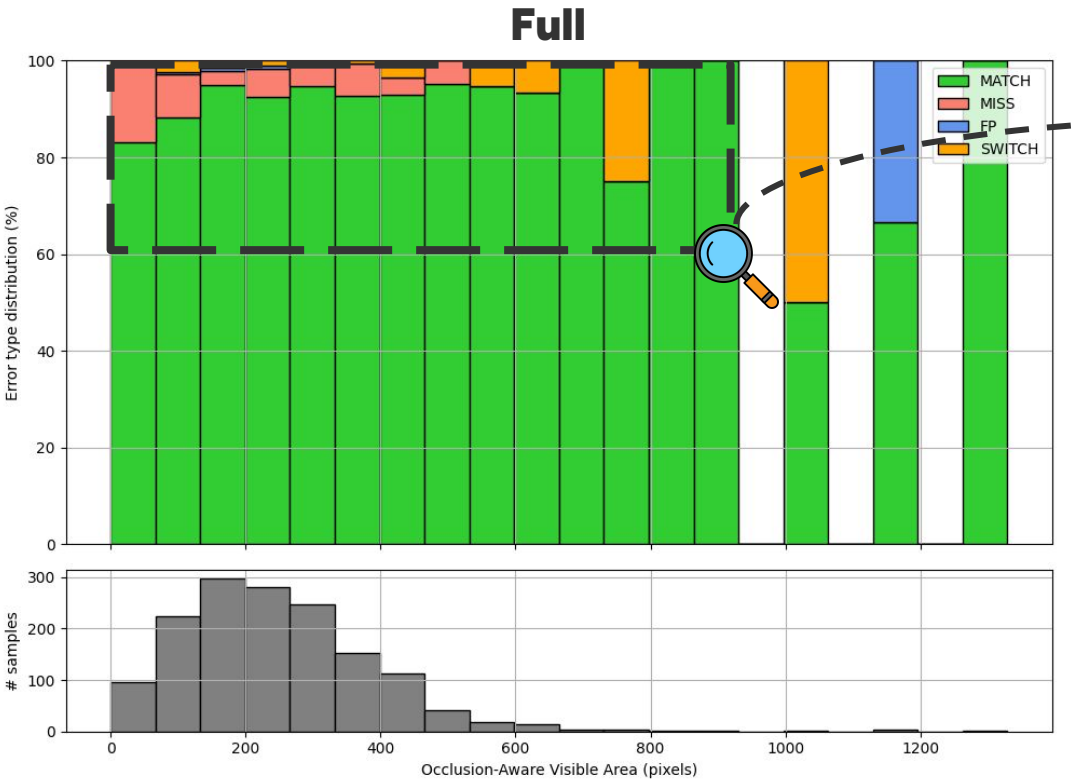


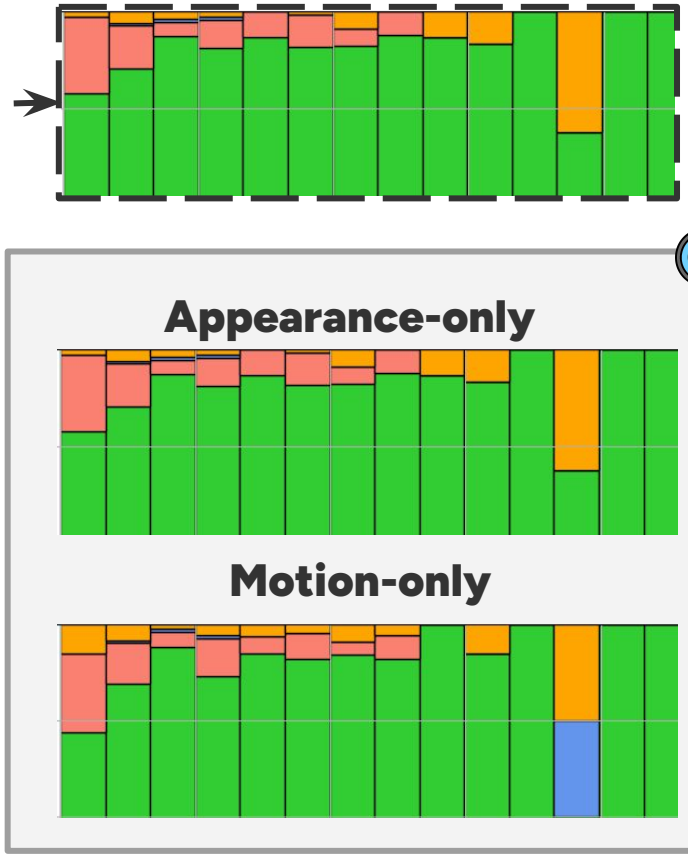




Diagnostics (MultiviewX)

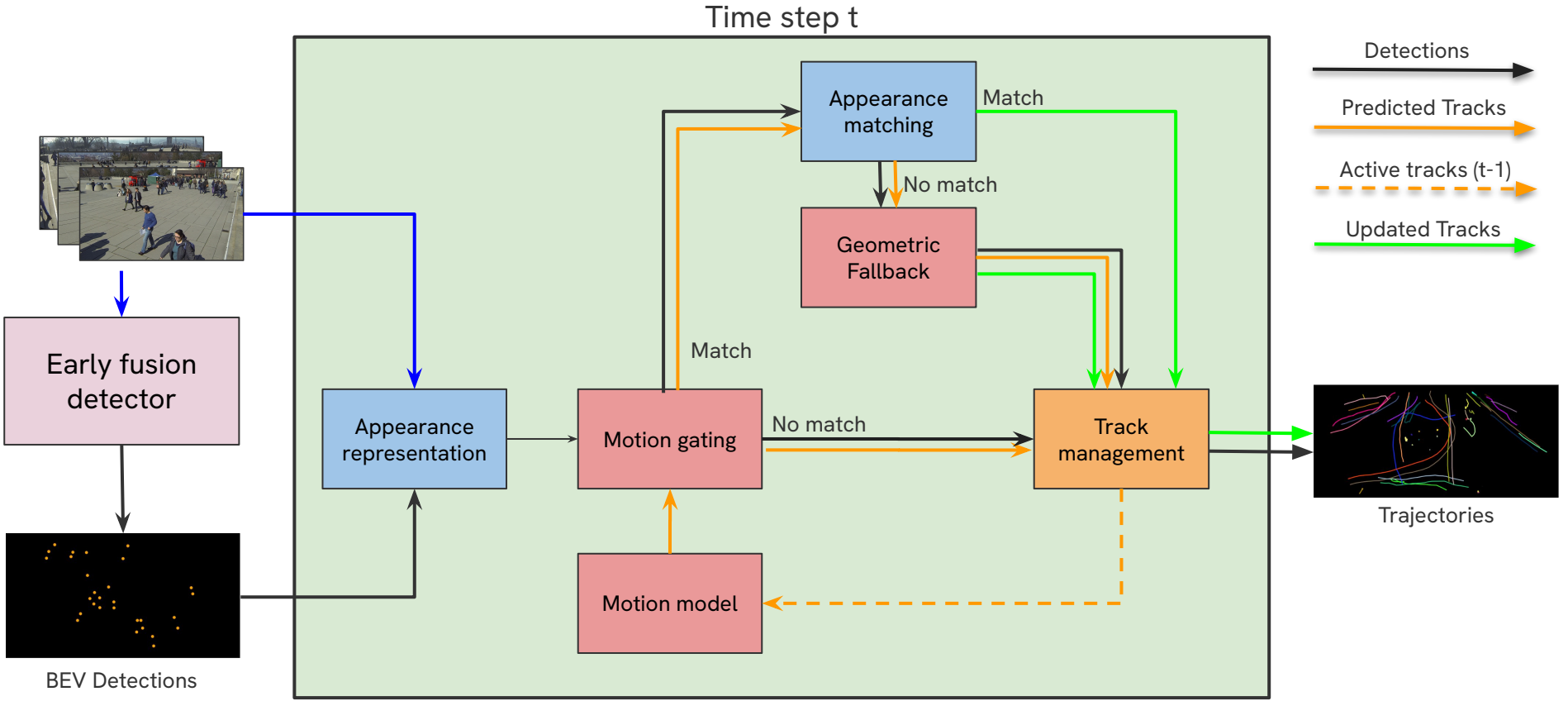
Error breakdown by occlusion-aware visible score





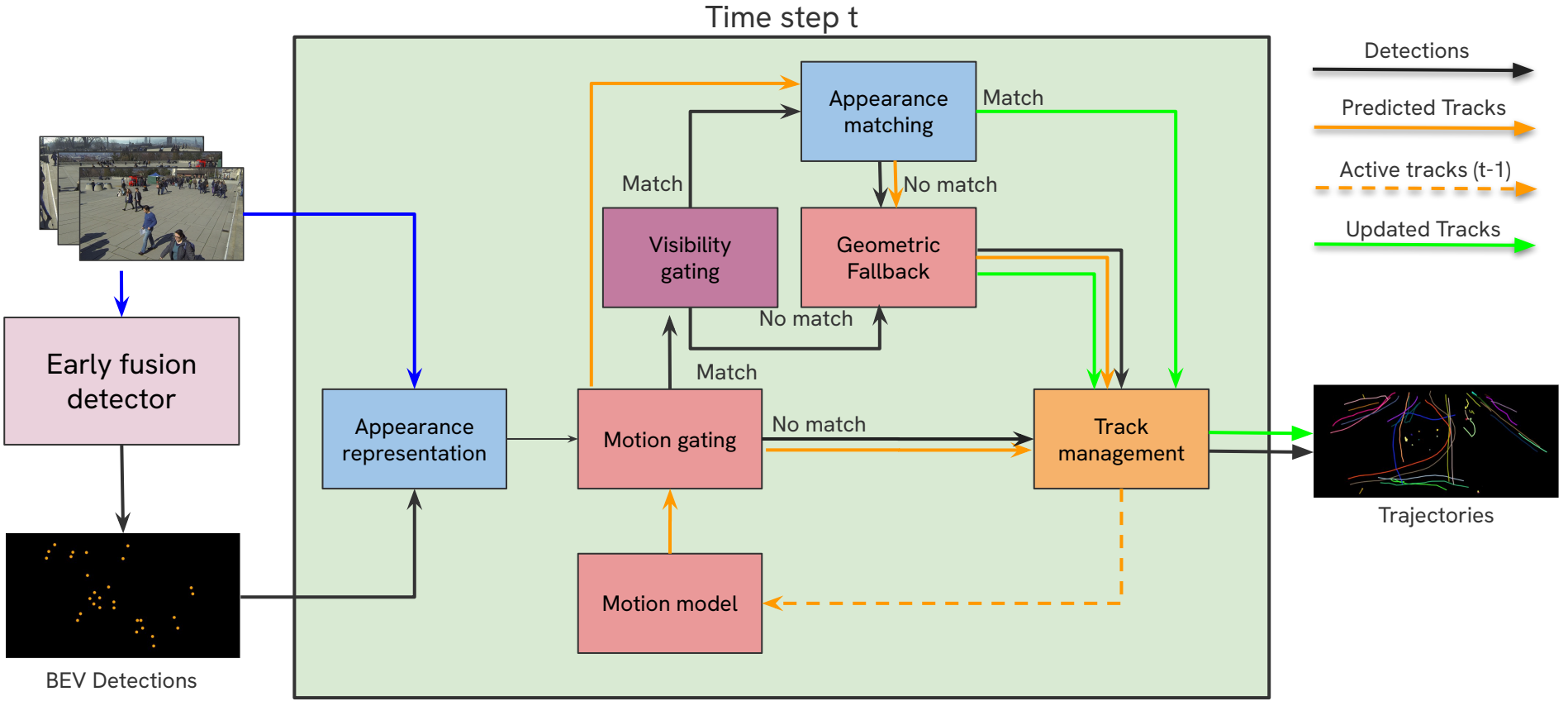
Visibility-aware variant

Policy: appearance when visibility is sufficient; otherwise geometry



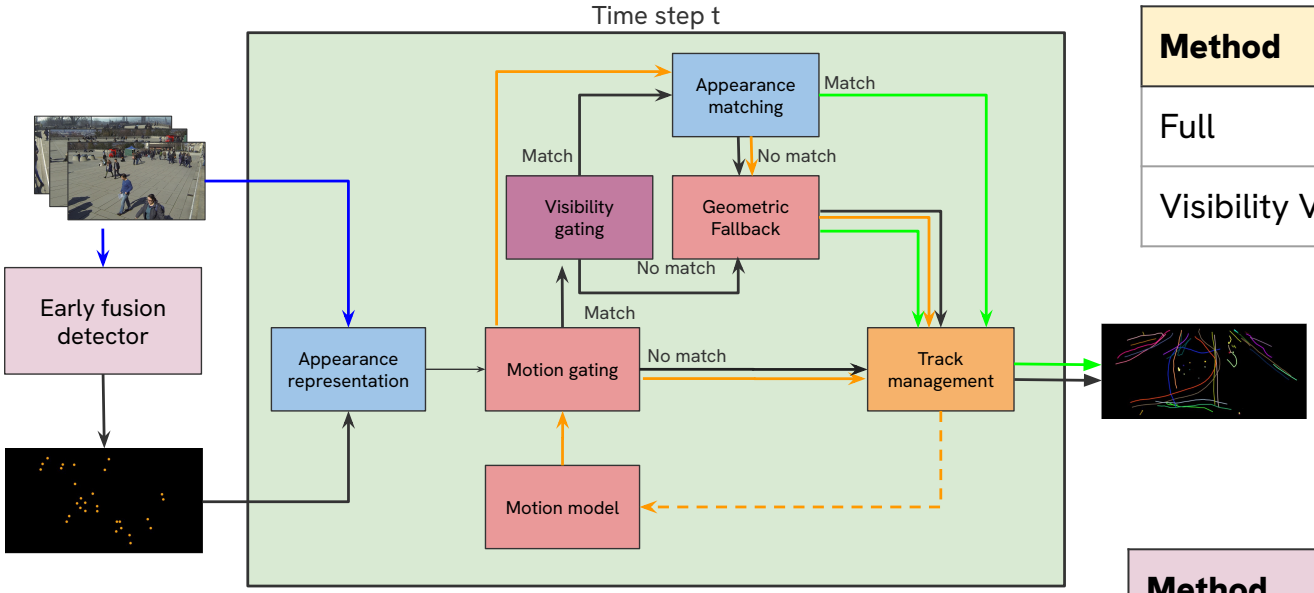
Visibility-aware variant

Policy: appearance when visibility is sufficient; otherwise geometry



Visibility-aware variant

Policy: appearance when visibility is sufficient; otherwise geometry



Wildtrack

Method	IDF1↑	MOTA ↑	IDSW↓
Full	95,7	91,7	7
Visibility Variant	**95,8**	**92,0**	**4**

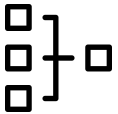
MultiviewX

Method	IDF1↑	MOTA ↑	IDSW↓
Full	**92,0**	**92.2**	**22**
Visibility Variant	91,8	91,8	28

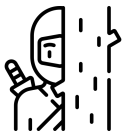
Provides only modest gains and is highly sensitive to the visibility threshold

Limitations

Evaluated on only **two short outdoor benchmarks.**



Detector-agnostic by design, **validated with only one upstream detector.**



Visibility score is geometric and detection-driven; **ignores static occluders, inherits FN/FP.**

Future Directions

Evaluate on **indoor scenes and longer sequences.**

Assess sensitivity to upstream detectors by **testing multiple early-fusion detectors.**

Longer term:
extend to **multi-camera vehicle tracking**.

Thanks!

Do you have any questions?

Detection Quality (MVDeTr)

	MODA↑	MODP↑	Prec.↑	Rec.↑	FN↓	FP↓
Wildtrack	92,0	82,3	96,5	95,5	43	33
MultiviewX	93,6	91,4	99,6	94,0	89	6

Tracking Results (MVDeTr upstream + Ours)

	IDF1↑	MOTA↑	MOTP↑	MT↑	ML↓	IDSW↓
Wildtrack	95,7	91,7	89,8	87,8	4,9	7
MultiviewX	92,0	92,2	95,0	93,4	0	22

$$\mathrm{IDF1} = \frac{2\,\mathrm{IDP} \cdot \mathrm{IDR}}{\mathrm{IDP} + \mathrm{IDR}}$$

$$\mathrm{IDF1} = \frac{2\,\mathrm{IDTP}}{2\,\mathrm{IDTP} + \mathrm{IDFP} + \mathrm{IDFN}}$$

$$\mathrm{IDP} = \frac{\mathrm{IDTP}}{\mathrm{IDTP} + \mathrm{IDFP}}$$

$$\mathrm{IDR} = \frac{\mathrm{IDTP}}{\mathrm{IDTP} + \mathrm{IDFN}}$$

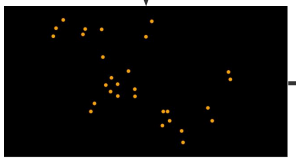
$$\mathrm{MODA} = 1 - \frac{\mathrm{FN} + \mathrm{FP}}{\mathrm{GT}}.$$

$$\mathrm{MOTA} = 1 - \frac{\mathrm{FN} + \mathrm{FP} + \mathrm{IDSW}}{\mathrm{GT}}$$

Proposed Tracker

online, detector-agnostic

Early fusion detector



BEV Detections

ReID Encoder

BEV-to-Image Back Projection

Crop Extraction

ReID Feature Extraction

Multi-View Aggregation

Match

Appearance representation

Motion gating

No match

Track management

Motion model

Trajectories