

Unifying Short and Long-Term Tracking with Graph Hierarchies



Orcun Cetintas*



Guillem Brasó*



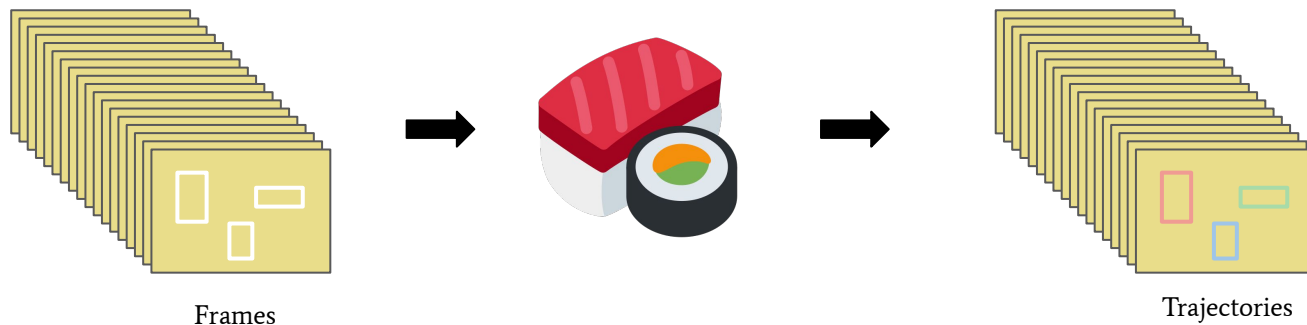
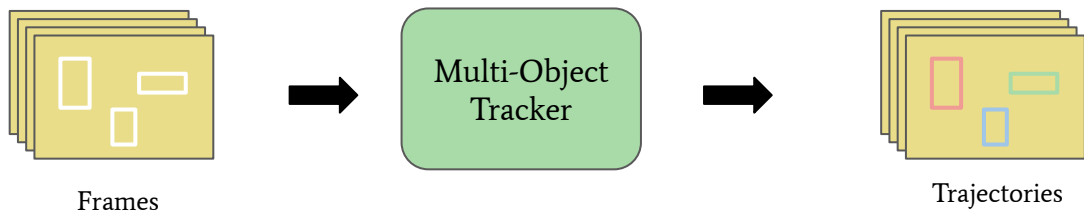
Laura Leal-Taixé'

Poster: THU-PM-217

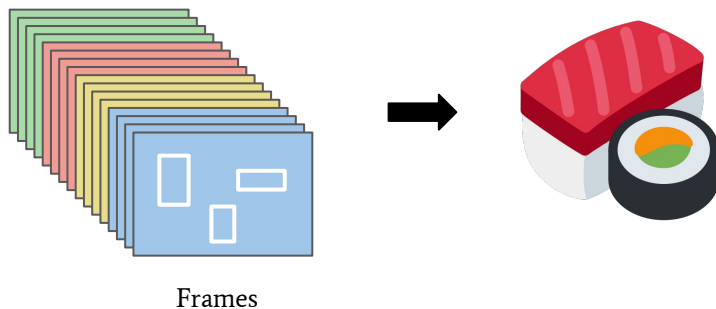


Overview

SUSHI  is a **Strong and Unified Scalable Hierarchical** multi-object tracker



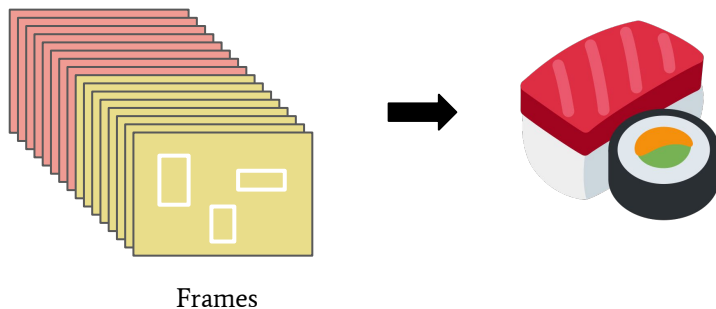
Overview



Key Features:

- Hierarchical processing of long clips (512 frames)
- Shared GNN architecture and weights
- SOTA in 4 diverse datasets

Overview

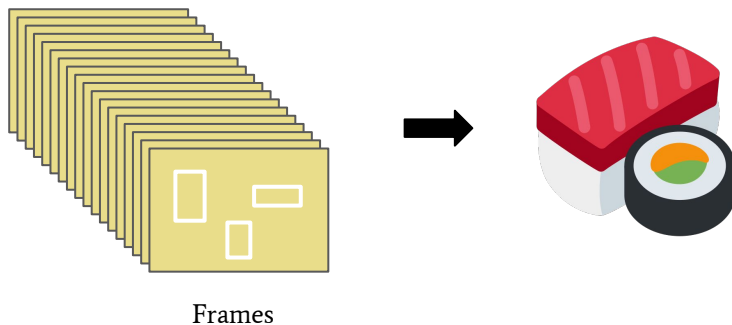


Frames

Key Features:

- Hierarchical processing of long clips (512 frames)
- Shared GNN architecture and weights
- SOTA in 4 diverse datasets

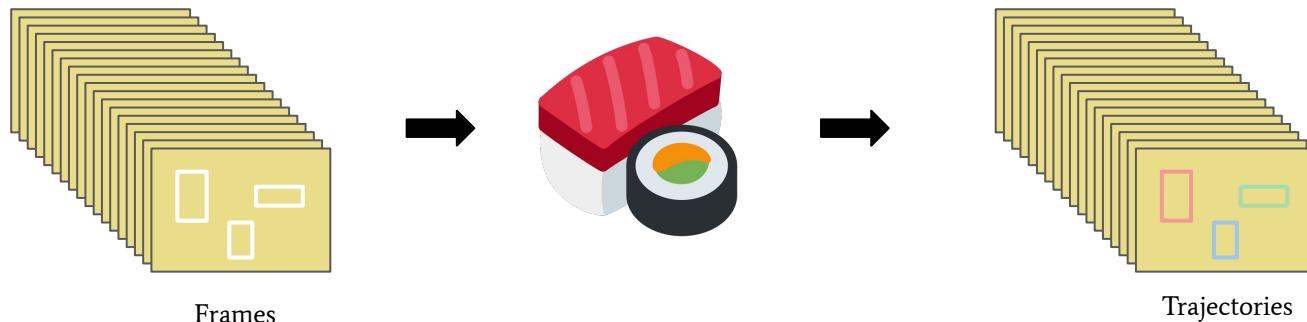
Overview



Key Features:

- Hierarchical processing of long clips (512 frames)
- Shared GNN architecture and weights
- SOTA in 4 diverse datasets

Overview



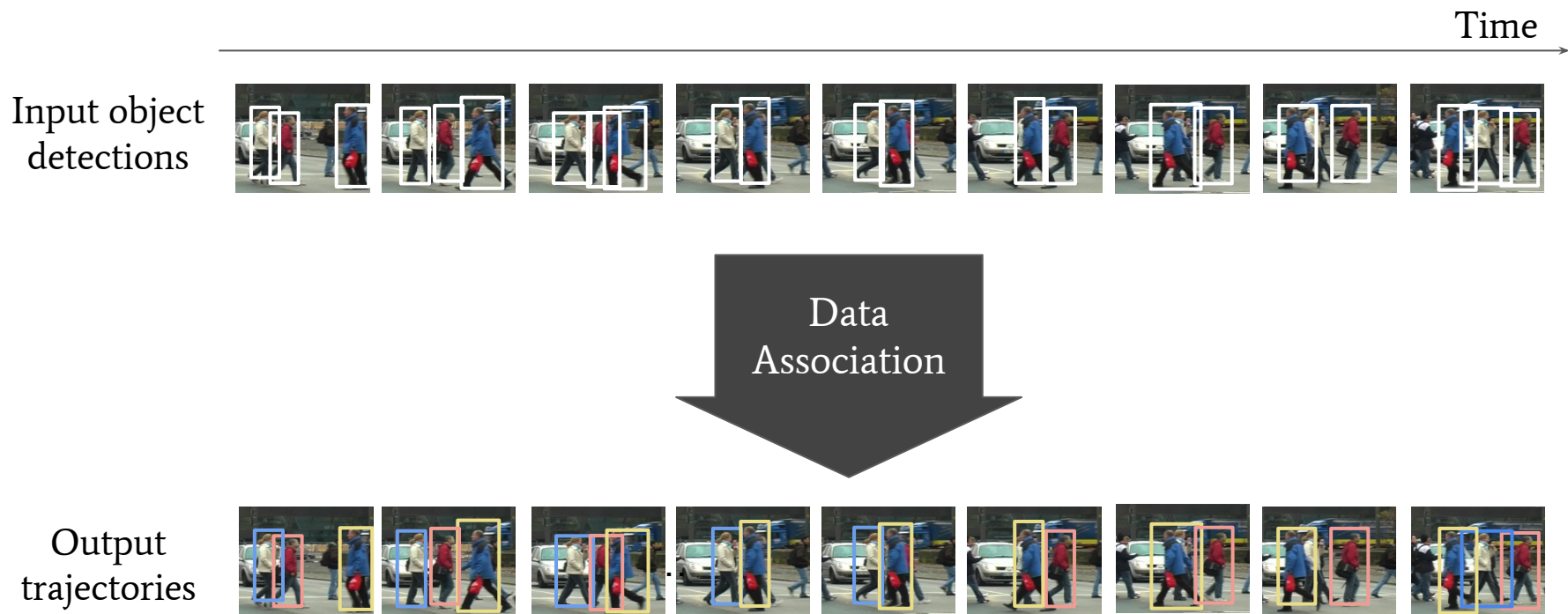
Key Features:

- Hierarchical processing of long clips (512 frames)
- Shared GNN architecture and weights
- SOTA in 4 diverse datasets

Multi-Object Tracking



Tracking-by-detection



Tracking-by-detection: short-term association

[Bergmann et al., 2019, Zhou et al., 2020]



Tracking-by-detection: short-term association [Bergmann et al., 2019, Zhou et al., 2020]



Tracking-by-detection: short-term association

[Bergmann et al., 2019, Zhou et al., 2020]



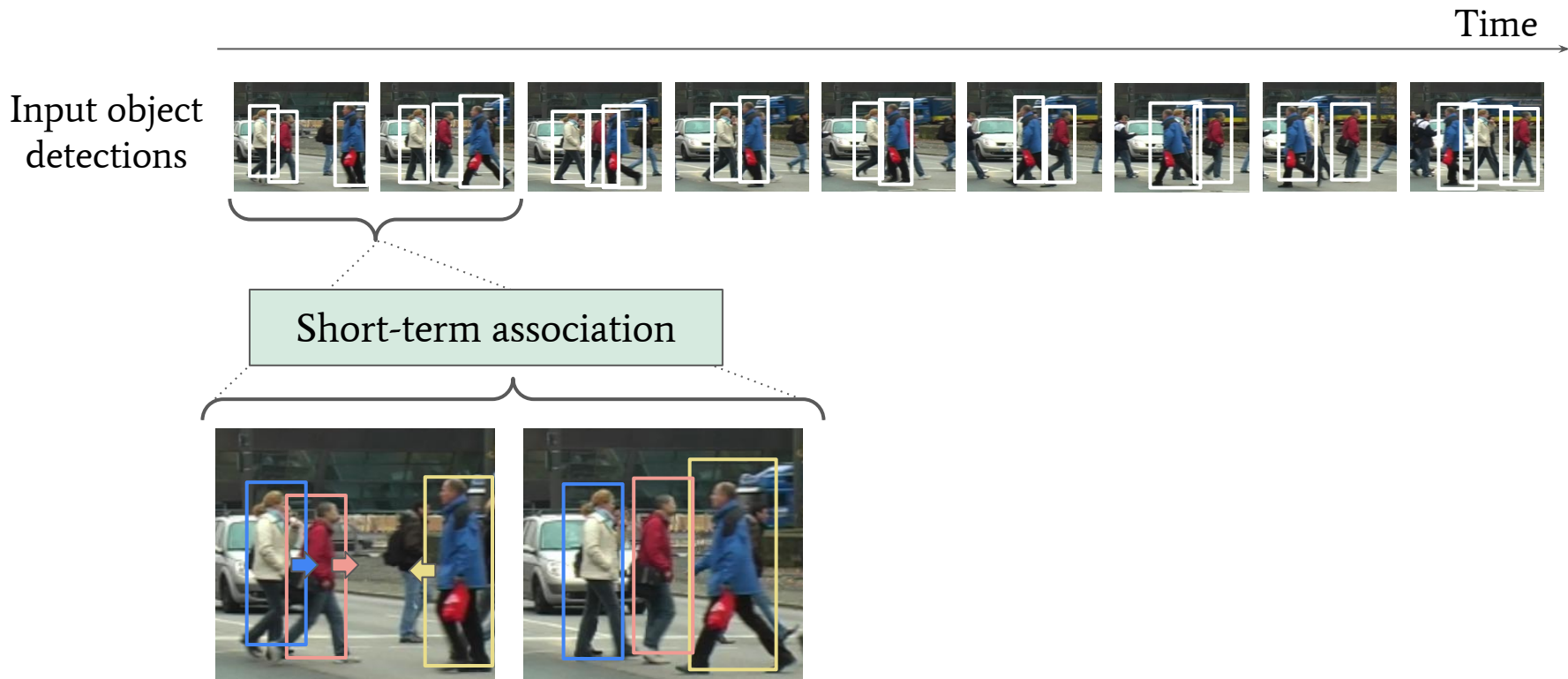
Tracking-by-detection: short-term association

[Bergmann et al., 2019, Zhou et al., 2020]



Tracking-by-detection: short-term association

[Bergmann et al., 2019, Zhou et al., 2020]



Tracking-by-detection: long-term association

[Brasó et al., 2020, Hornakova et al., 2020]

Input object
detections



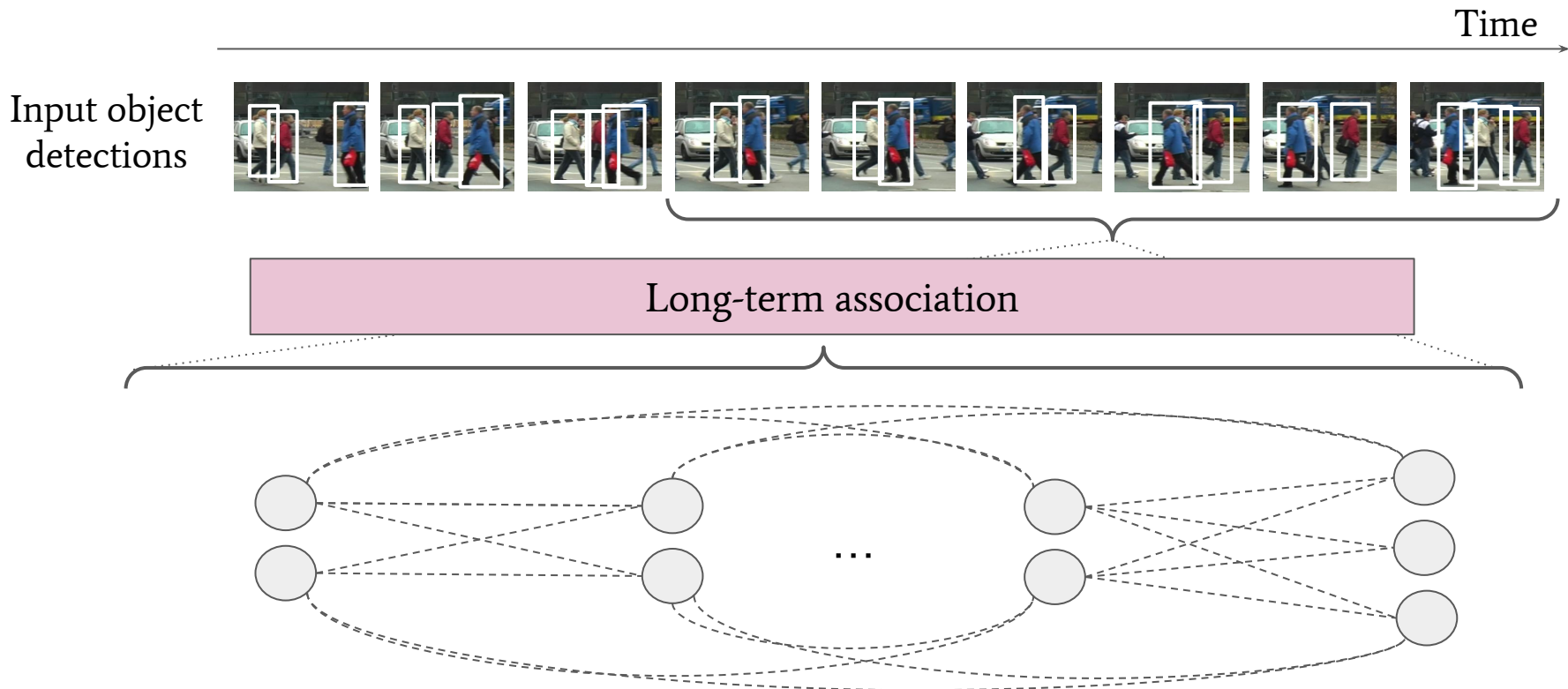
Tracking-by-detection: long-term association

[Brasó et al., 2020, Hornakova et al., 2020]



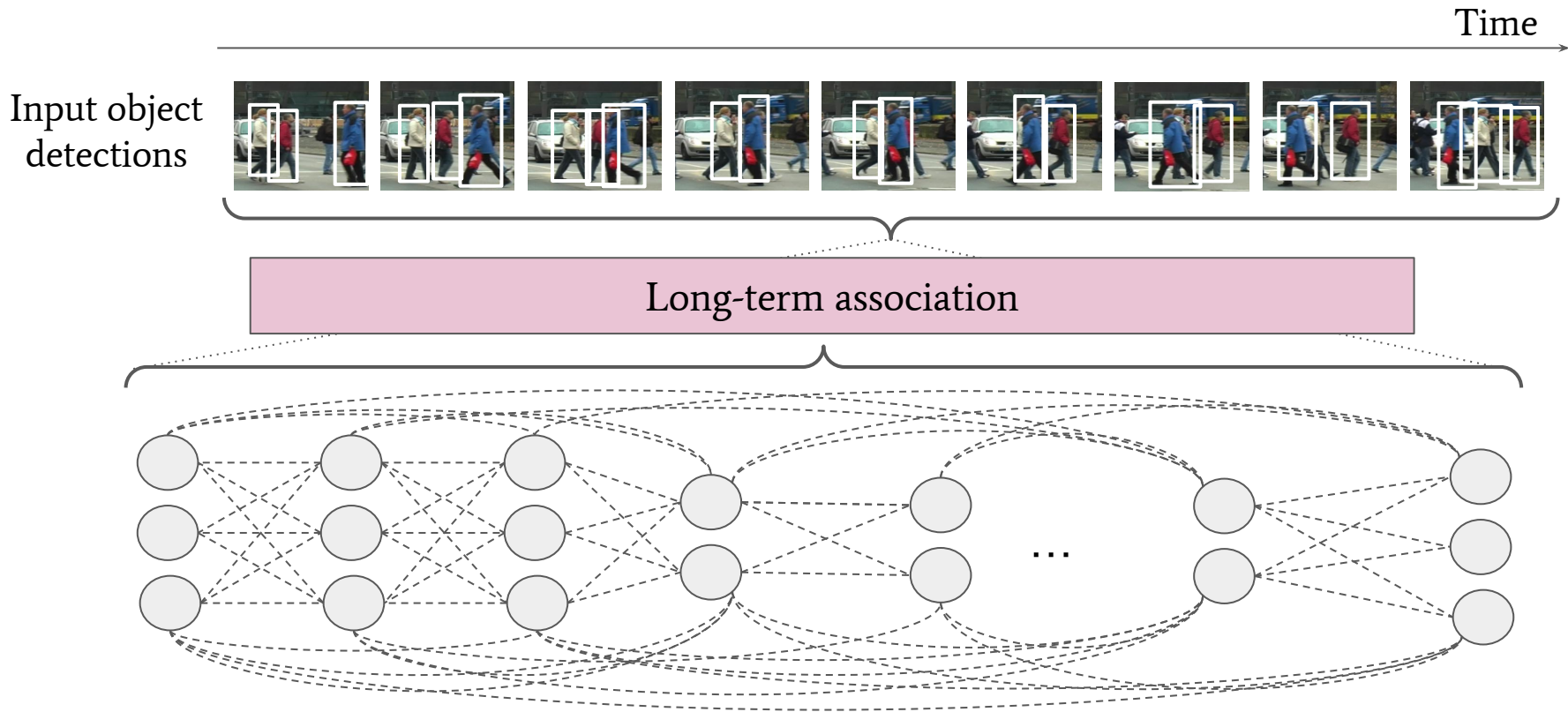
Tracking-by-detection: long-term association

[Brasó et al., 2020, Hornakova et al., 2020]



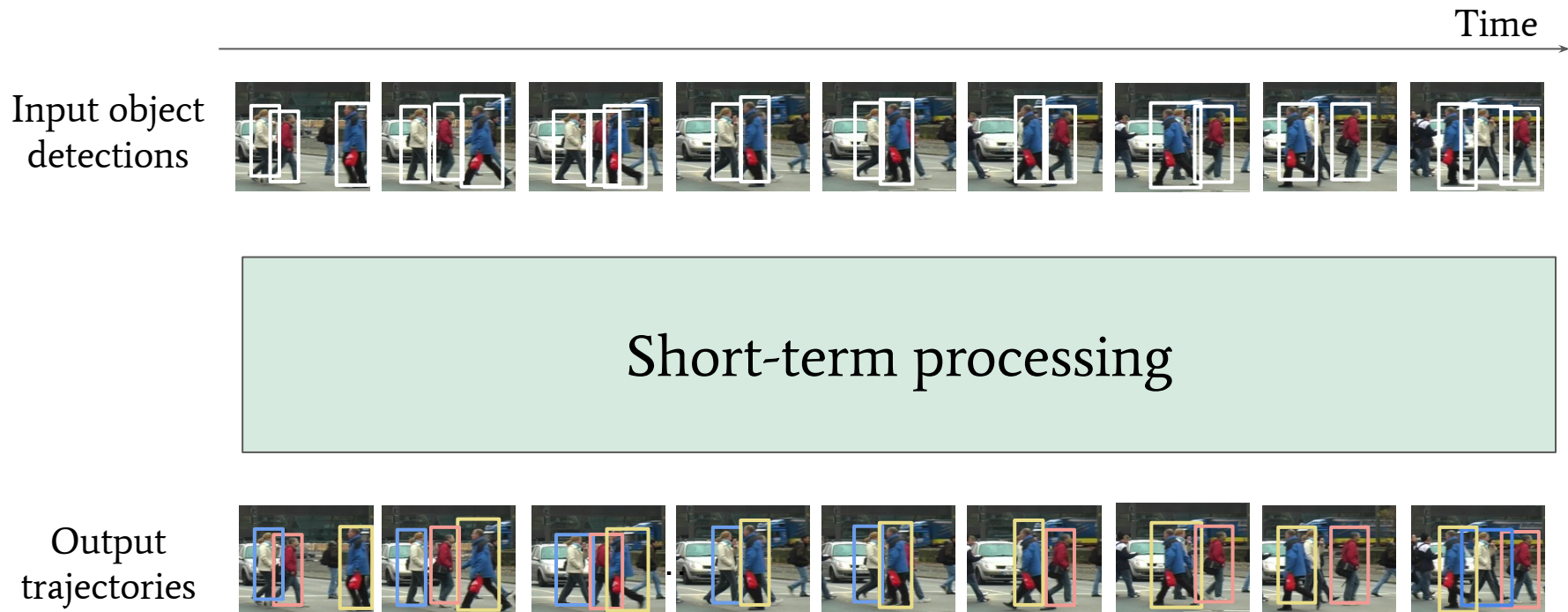
Tracking-by-detection: long-term association

[Brasó et al., 2020, Hornakova et al., 2020]



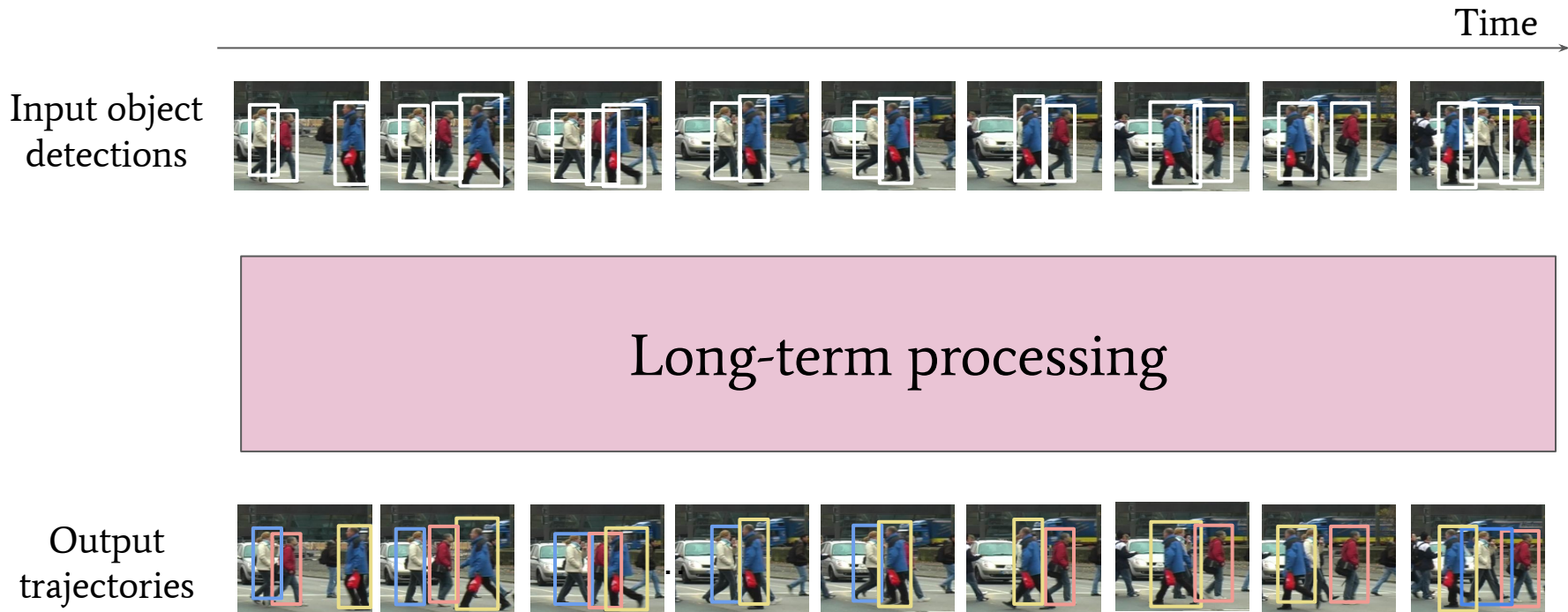
Short-term trackers

✗ Lack robustness under occlusion



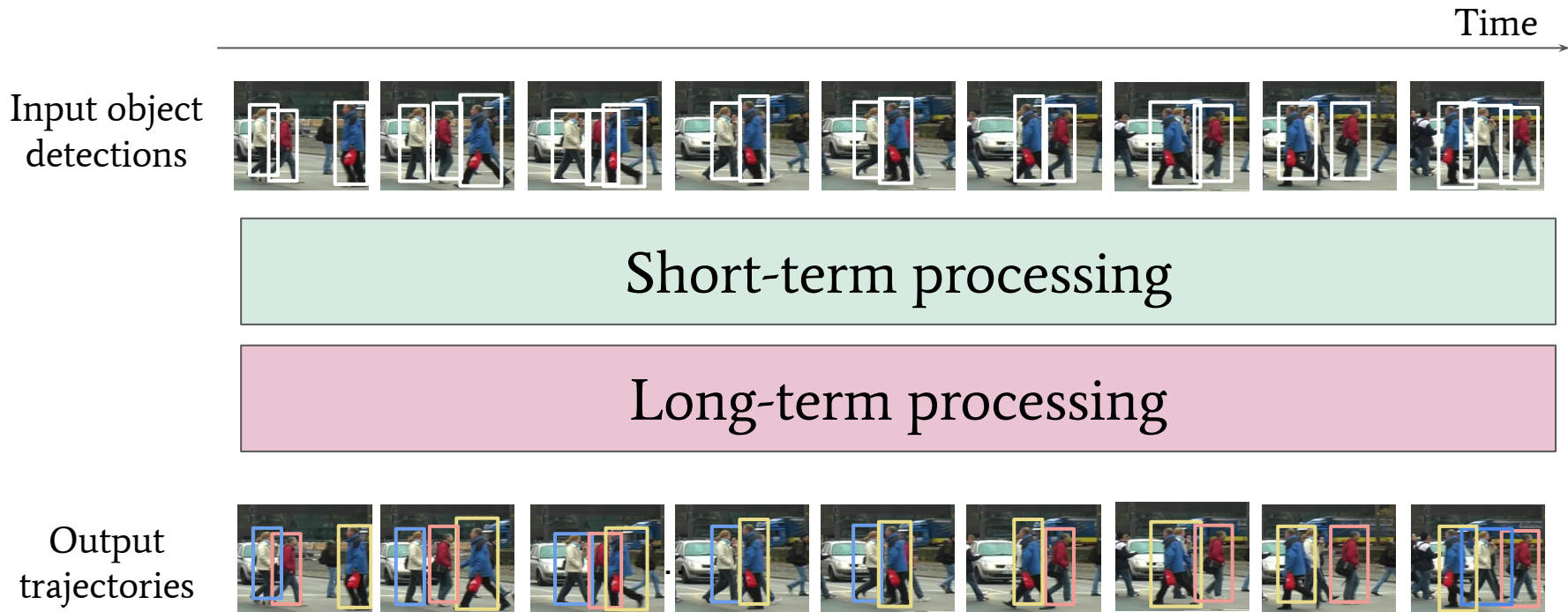
Graph-based trackers

✗ Limited Scalability



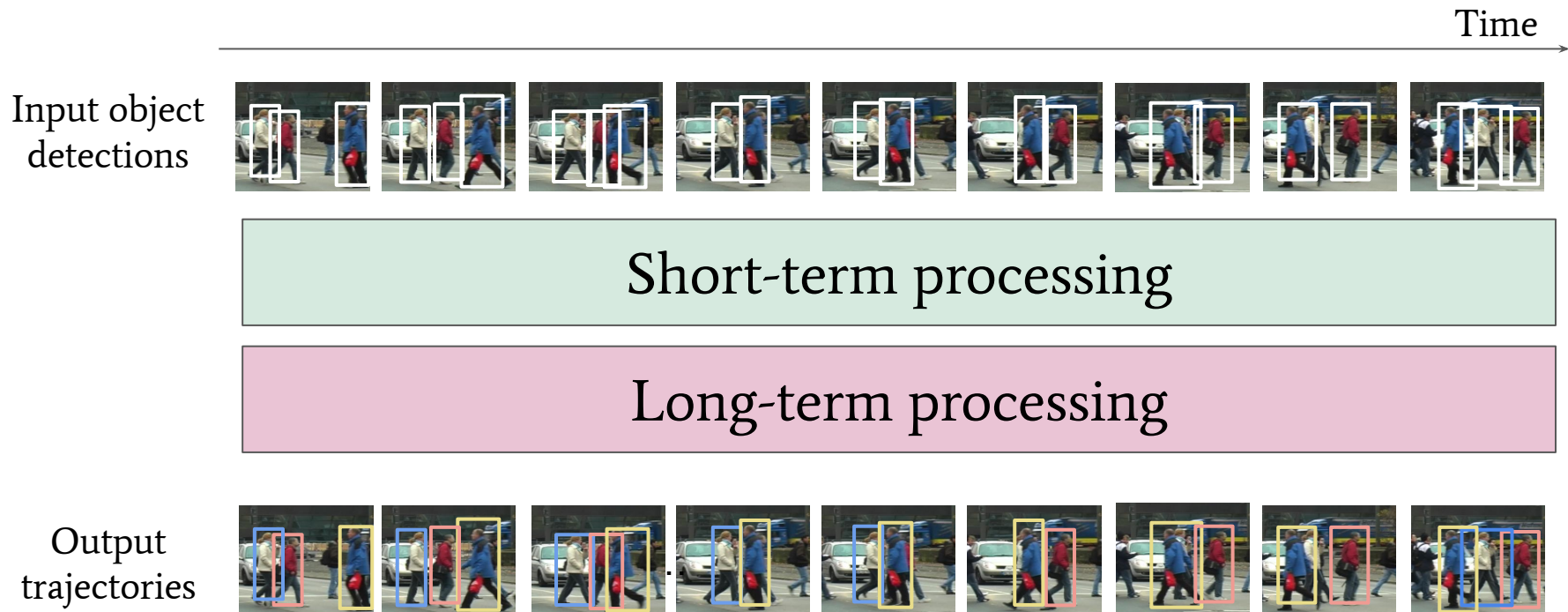
Hybrid multi-level trackers

✗ Limited Scalability



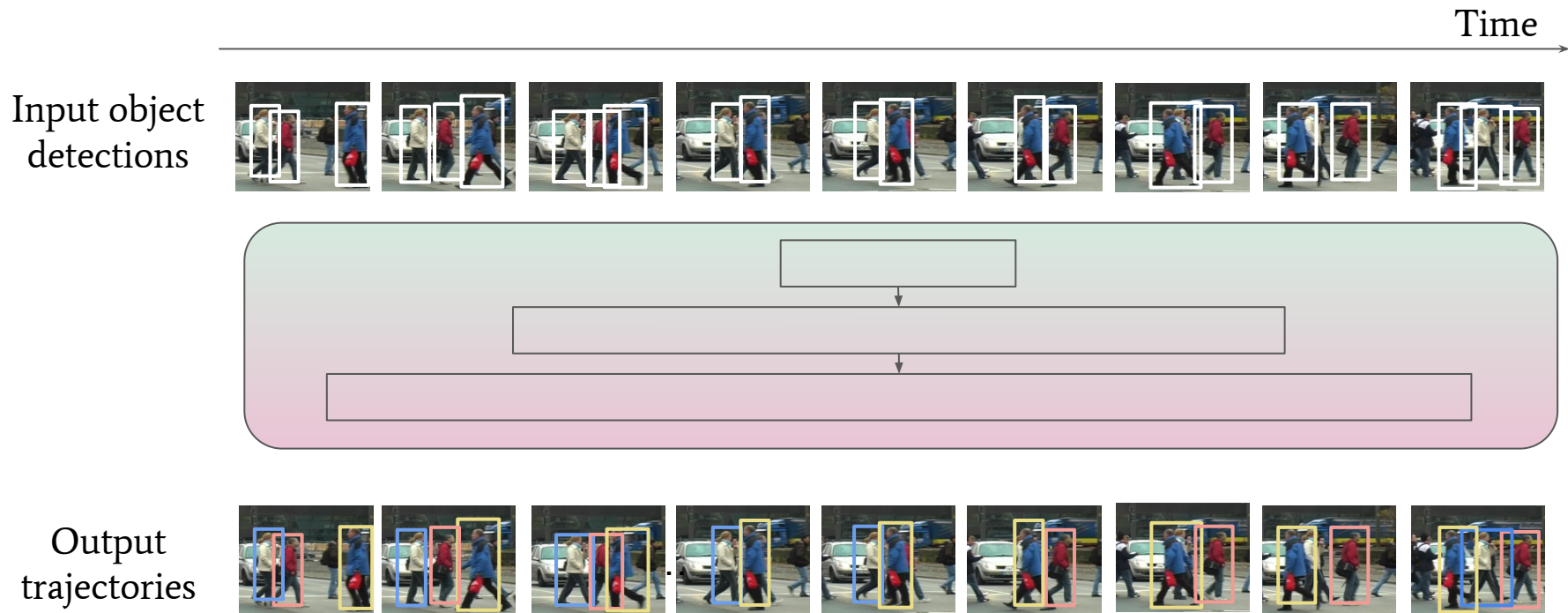
Hybrid multi-level trackers

- ✗ Limited Scalability
- ✗ Require per-level engineering

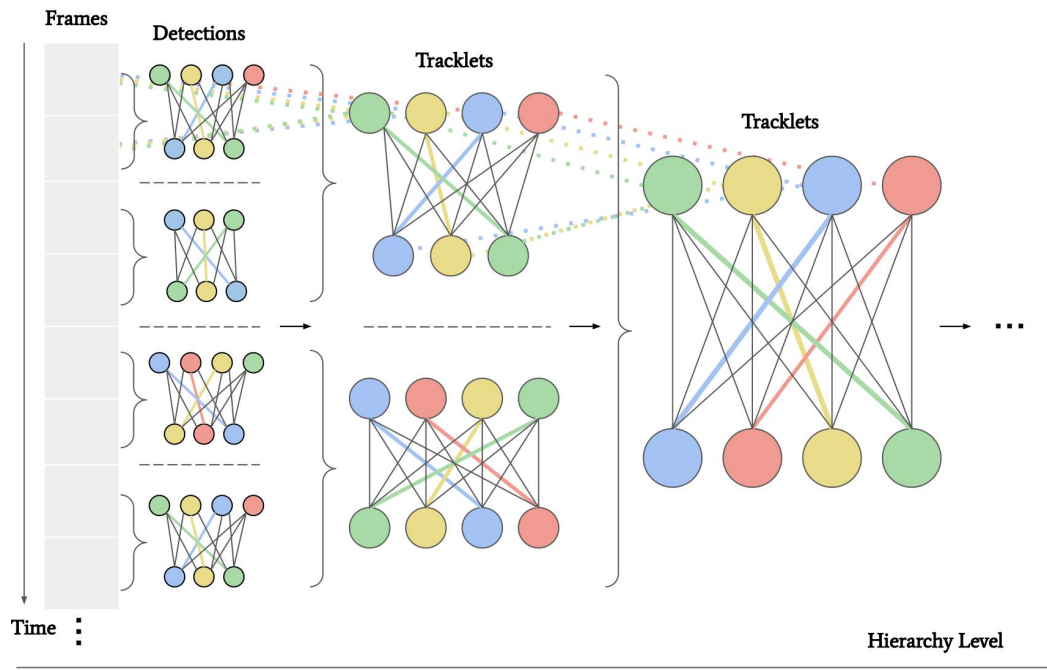


Our unified hierarchical tracker

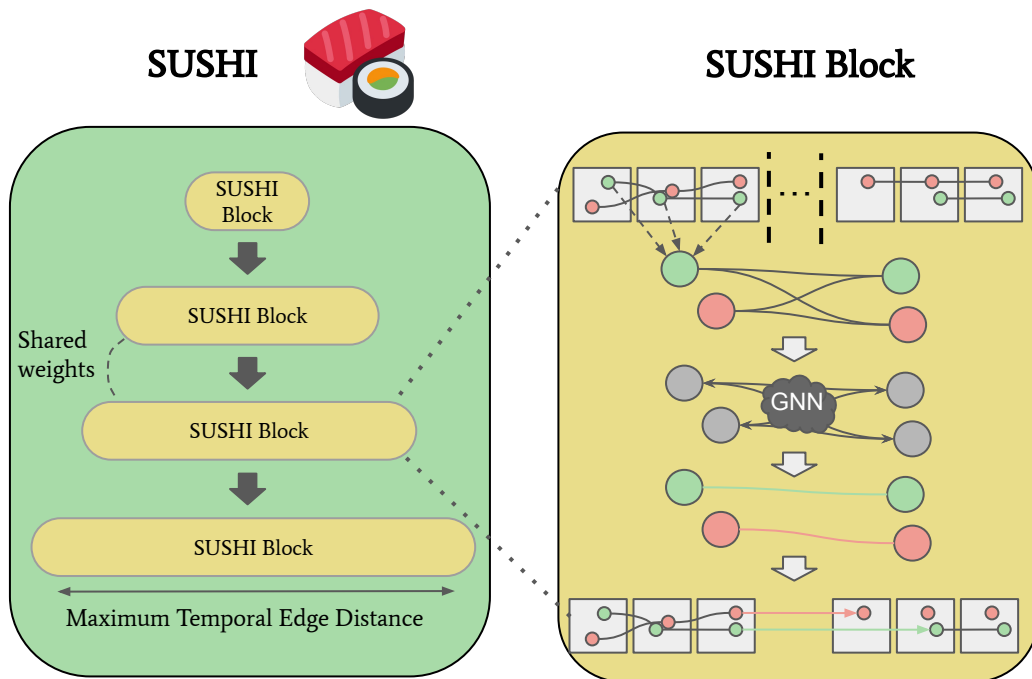
- ✓ Unified across levels
- ✓ Highly scalable



Constructing a hierarchy of tracking graphs



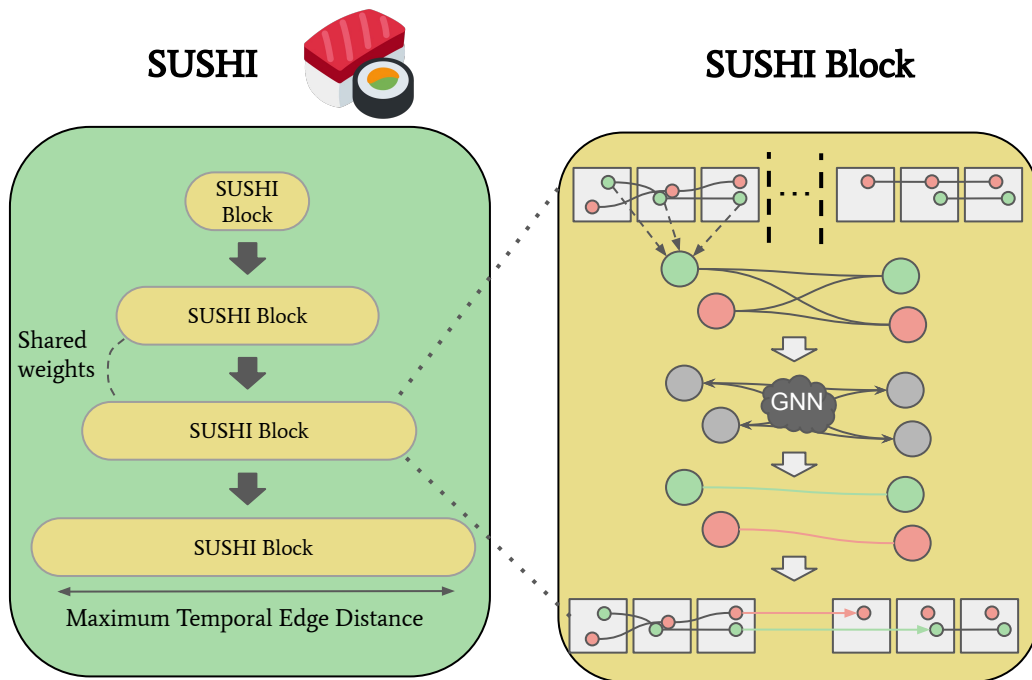
Learning a unified hierarchical tracker



SUSHI Block:

- Graph with tracklets as nodes
- Propagate node and edge embeddings
- Classify edges and merge nodes into longer tracklets

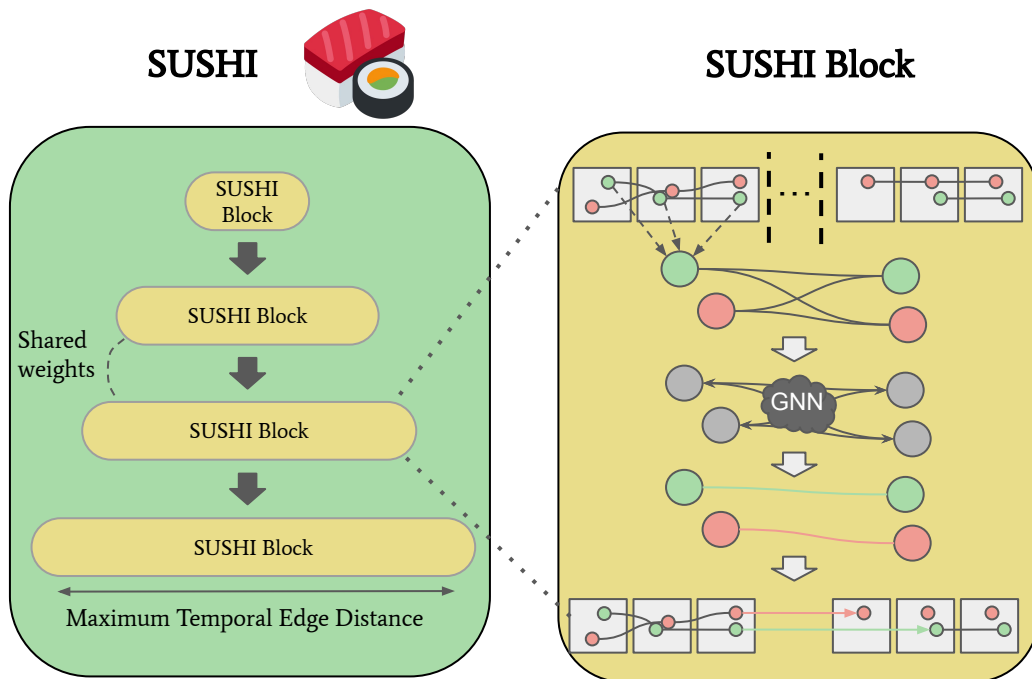
Learning a unified hierarchical tracker



Edge features:

- Based on spatial, motion, time and appearance similarity

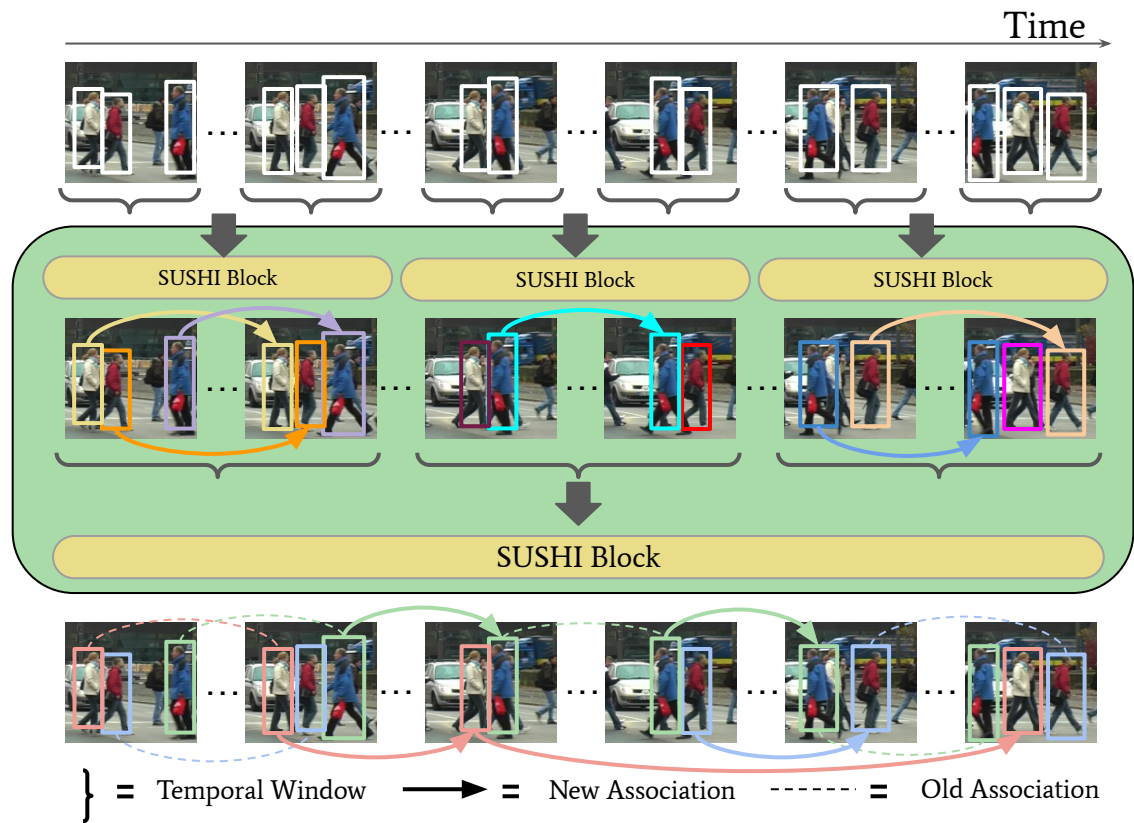
Learning a unified hierarchical tracker



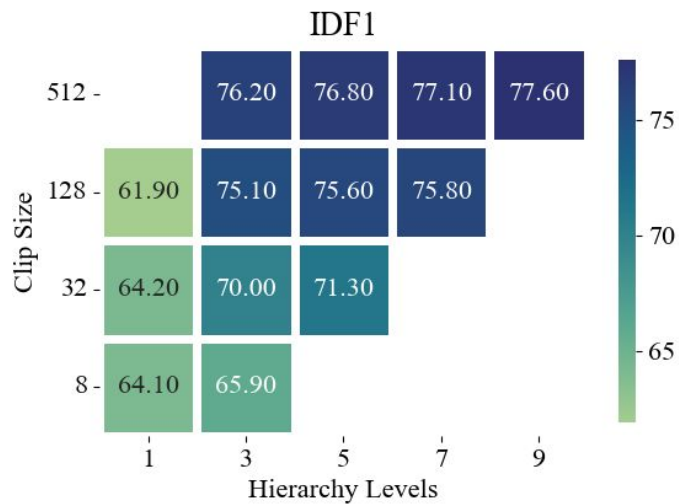
Weight sharing:

- Same architecture and weights for SUSHI blocks
- Learnable level embedding

SUSHI - Pipeline

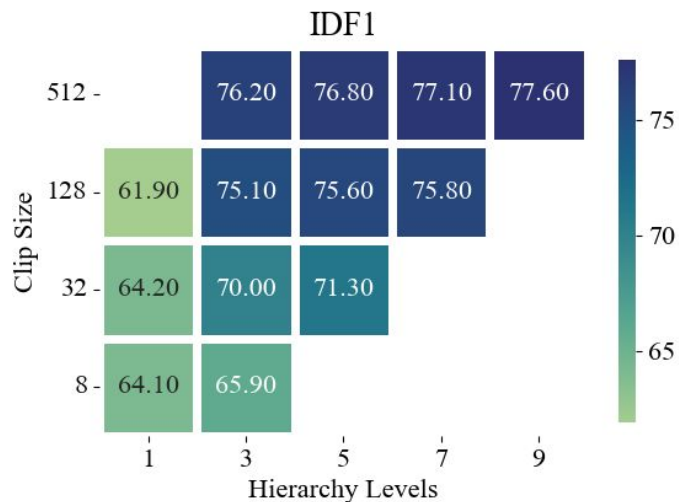


Ablation studies

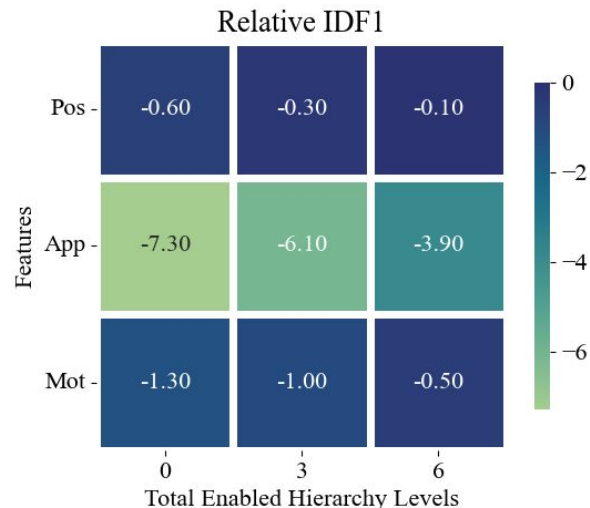


Increasing the number of hierarchy levels yields +13.4 IDF1 over a naive, i.e., single level baseline.

Ablation studies



Increasing the number of hierarchy levels yields +13.4 IDF1 over a naive, i.e., single level baseline.



Appearance is the most important feature while motion has a moderate impact. Position is only relevant for short-term association.

Qualitative results - per level

Level 1



Level 3



Level 6



Level 9



Qualitative comparison with SOTA

APLift



MPNTrack



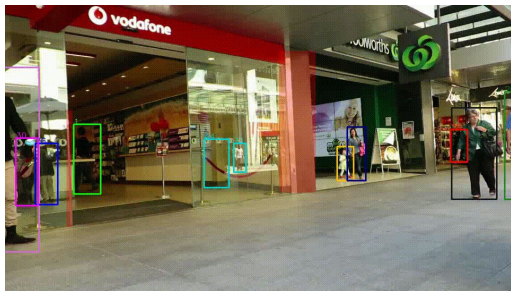
LPC_MOT



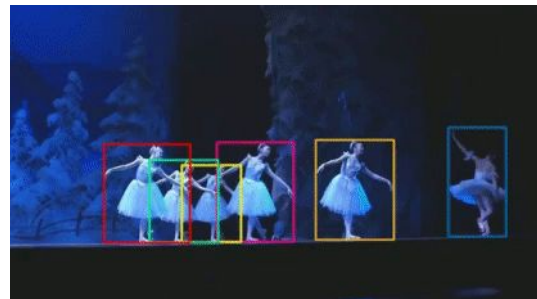
Ours



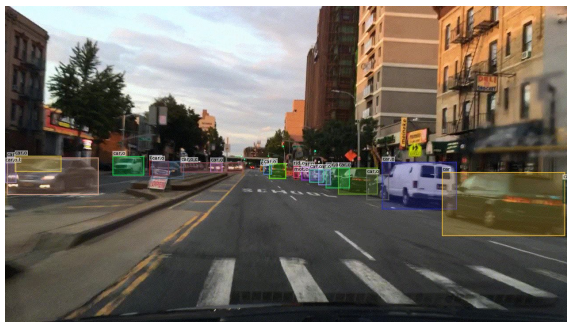
Datasets



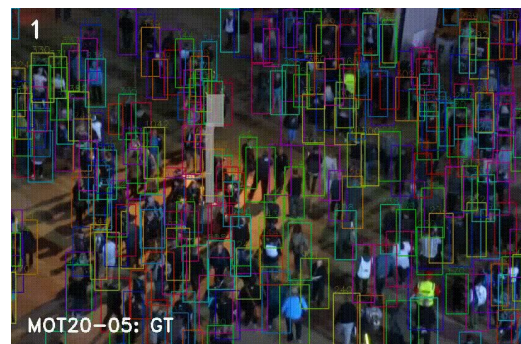
MOT17



DanceTrack



BDD



MOT20

Benchmark results

Method	Det Ref.	IDF1 ↑	HOTA ↑	MOTA ↑	ID Sw. ↓
MOT17 - Public					
Tracktor [2]	Tracktor	55.1	44.8	56.3	1987
LPT [24]	Tracktor	57.7	–	57.3	1424
MPNTrack [4]	Tracktor	61.7	49.0	58.8	1185
Lif_T [15]	Tracktor	65.6	51.3	60.5	1189
ApLift [16]	Tracktor	65.6	51.1	60.5	1709
GMT [14]	Tracktor	65.9	51.2	60.2	1675
LPC_MOT [8]	Tracktor	66.8	51.5	59.0	1122
+4.7 IDF1	SUSHI (Ours)	71.5	54.6	62.0	1041

Method	mIDF1 ↑	mMOTA ↑	IDF1 ↑	MOTA ↑	ID Sw. ↓
BDD - Test					
Yu <i>et al.</i> [60]	44.7	26.3	68.2	58.3	14674
QDTrack [30]	52.3	35.5	72.3	64.3	10790
TETer [23]	53.3	37.4	–	–	–
ByteTrack [65]	55.8	40.1	71.3	69.6	15466
+4.2 mIDF1	SUSHI (Ours)	60.0	40.2	76.2	13626

Method	IDF1 ↑	HOTA ↑	MOTA ↑	AssA ↑	DetA ↑
DanceTrack					
CenterTrack [67]	35.7	41.8	86.8	22.6	78.1
FairMOT [66]	40.8	39.7	82.2	23.8	66.7
TraDes [54]	41.2	43.3	86.2	25.4	74.5
GTR [68]	50.3	48.0	84.7	31.9	72.5
QDTrack [30]	50.4	54.2	87.7	36.8	80.1
MOTR [63]	51.5	54.2	79.7	40.2	73.5
ByteTrack [65]	53.9	47.7	89.6	32.1	71.0
+9.5 IDF1	SUSHI (Ours)	63.4	63.3	88.7	50.1

Method	Det Ref.	IDF1 ↑	HOTA ↑	MOTA ↑	ID Sw. ↓
MOT20 - Public					
Tracktor [2]	Tracktor	52.7	42.1	52.6	1648
LPT [24]	Tracktor	53.5	–	57.9	1827
ApLift [16]	Tracktor	56.5	46.6	58.9	2241
MPNTrack [4]	Tracktor	59.1	46.8	57.6	1210
LPC_MOT [8]	Tracktor	62.5	49.0	56.3	1562
+9.1 IDF1	SUSHI (Ours)	71.6	55.4	61.6	1053

Thank you for your attention!



Paper



Code