Northeastern University
Electrical and Computer Engineering Department
Augmented Cognition Laboratory (ACLab)
https://web.northeastern.edu/ostadabbas/

Multi-Object Tracking in Complex Scenes: Extending Human Perception

Bishoy Galoaa

Electrical and Computer Engineering

Northeastern University

Northeastern University
Electrical and Computer Engineering Department
Augmented Cognition Laboratory (ACLab)
https://web.northeastern.edu/ostadabbas/

Multi-Object Tracking in Complex Scenes: Extending Human Perception

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes." - Marcel Proust

Bishoy Galoaa MS Thesis April 2025

With Deepest Gratitude

- **Prof. Sarah Ostadabbas**
  - For her exceptional **guidance** and unwavering **support** throughout my research journey
  - For creating an environment that fosters **innovation** and **critical** thinking
  - For believing in my **vision** and helping me transform ideas into **impactful** research
  - For the opportunity to continue my academic journey as **a PhD student at ACLab**

With Deepest Gratitude

- **Dr. Somaieh Amraee**
  - For her invaluable **mentorship** and technical guidance
  - For countless hours of collaboration on **all my** research projects
  - For sharing her **expertise** and insights that shaped my approach to computer vision

With Deepest Gratitude

- **Prof.** Stratis Ioannidis
  - For his foundational teaching in **Pattern Recognition**
  - For serving on my thesis committee

With Deepest Gratitude

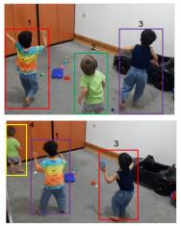
- **Prof. Michael Everett**
  - For his **exceptional** mentorship during the **exoskeleton** control project
  - For being on my thesis committee

MS Thesis Committee Members




Thank You!

My Journey at Northeastern University

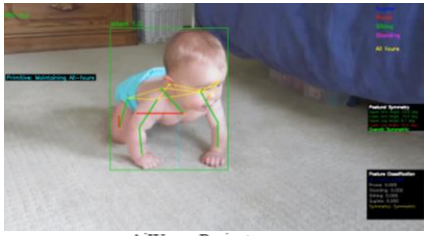


Joined ACLab

Sep 2023

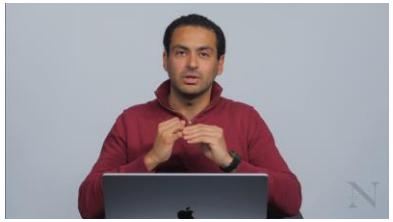
Exoskeleton Control

Dec 2023



AiWover Project

Summer 2024



Course Development

Jan 2025

NEURAL INFORMATION PROCESSING SYSTEMS

MS Thesis Defense

March 2025

Fall 2023

TA for Verifiable ML

Jan-May 2024

DragonTrack

Fall 2024

Publications

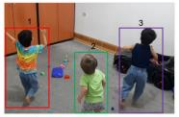
ICML

ICCV HONOLULU HAWAII OCT 19-23, 2025

Future

PhD Continuation

My Journey at Northeastern University



Joined ACLab

Sep 2023

Fall 2023

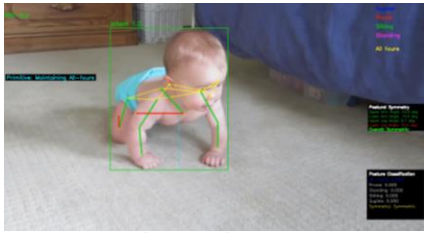
TA for Verifiable ML

Exoskeleton Control

Dec 2023

Jan-May 2024

DragonTrack

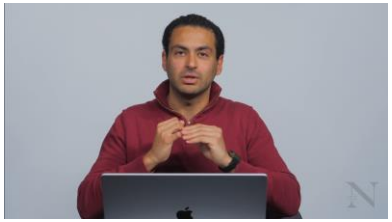


AiWover Project

Summer 2024

Fall 2024

Publications



Course Development

Jan 2025

Future

PhD Continuation

NEURAL INFORMATION PROCESSING SYSTEMS

March 2025

MS Thesis Defense

Bishoy Galoaa MS Thesis April 202

Publications

[1] Multiple Toddler Tracking in Indoor Videos
S. Amraee, **B. Galoaa**, M. Goodwin, E. Hatamimajoumerd, S. Ostadabbas
IEEE/CVF Winter Conference on Applications of Computer Vision ***(WACV) Workshops, 2024***

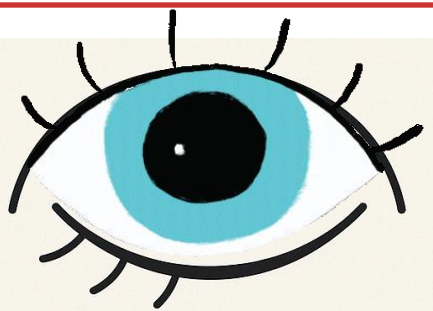
[2] DragonTrack
Transformer-Enhanced Graphical Multi-Person Tracking in Complex Scenarios
B. Galoaa, S. Amraee, S. Ostadabbas
Winter Conference on Applications of Computer Vision ***(WACV), 2025***

[3] MOTE: More Than Meets the Eye
Optical Flow-Based Multi-Object Tracking with Prolonged Occlusion Handling
B. Galoaa, S. Amraee, S. Ostadabbas
Under Review, International Conference on Machine Learning ***(ICML), 2025***

[4] UniTrack
A Differentiable Graph-Based Loss for Robust Multi-Object Tracking
B. Galoaa, U Nandi , V Dhir, S. Amraee, S. Ostadabbas
Under Review, International Conference on Computer Vision ***(ICCV), 2025***

The Hidden Hours We Don't See

- We blink 15–20 times per minute
- Each blink lasts ~0.3 seconds
- That's 6 seconds every minute
- 6 minutes every hour
- 1.5 hours a day — eyes closed, yet we never notice



Per minute:	Per hour:	Per day:
15-20 blinks ×1/3 seconds =	6 seconds/ minute × 60 =	6 minutes /hour ×16 hours =
6 seconds	**6 minutes**	**96 minutes**

The Hidden Hours We Don't See

- Our brain fills in the gaps. It sees even when we're blind for a moment.

But what about machines? Can we teach them to see through their own *blind spots, too*?

My Research Mission

- To extend human perception through intelligent **visual tracking systems** —
  enabling machines to see, reason, and remain aware where human attention cannot.

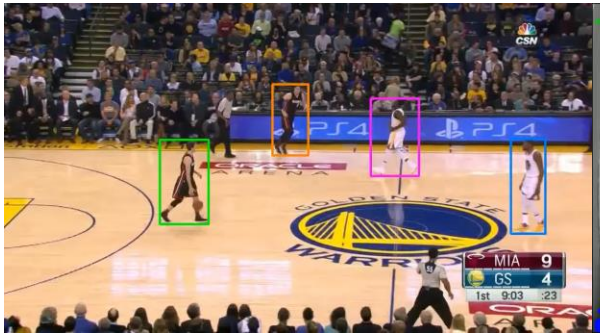
My Research Vision

- Advancing **visual tracking** to enable reliable machine perception in dynamic, complex, and attention-limited environments.
    - ➢ Multi-Object Tracking in Complex Scenes

This Thesis Vision: Solving Through Perception

What if machines could:

- Follow a child through a busy daycare, even when out in the crib
- Track a pedestrian through complex scenes
- Distinguish athletes on the same team
- Learn tracking from just a few examples



SportsMOT [ICCV 2023]

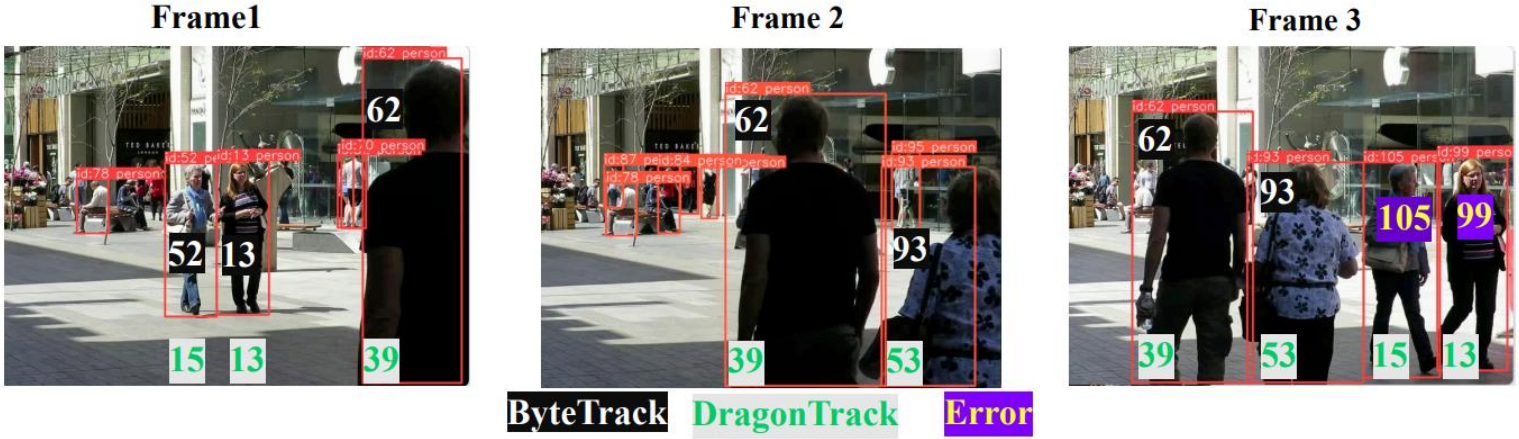
Moezzi S [WACVW 2025]

MOTE [ICML 2025]

Persistent Challenges in Tracking

Why Tracking Remains Challenging

- **Occlusions:** Objects disappearing temporarily behind others
  - Human visual system can track through occlusions
  - Most algorithms struggle when objects are not visible

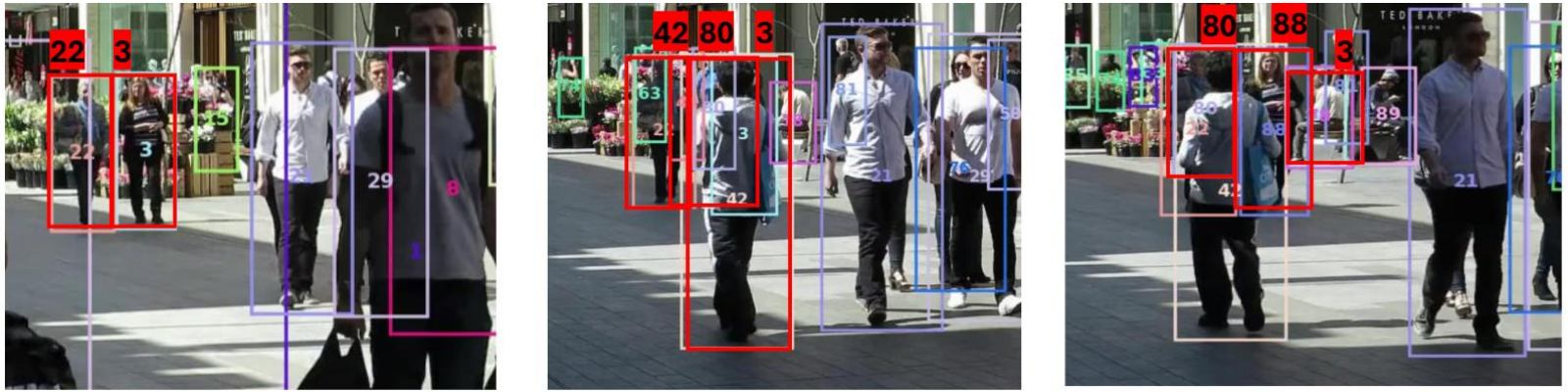


Bishoy Galoaa MS Thesis April 2025

DragonTrack [WACV 2025]

Why Tracking Remains Challenging

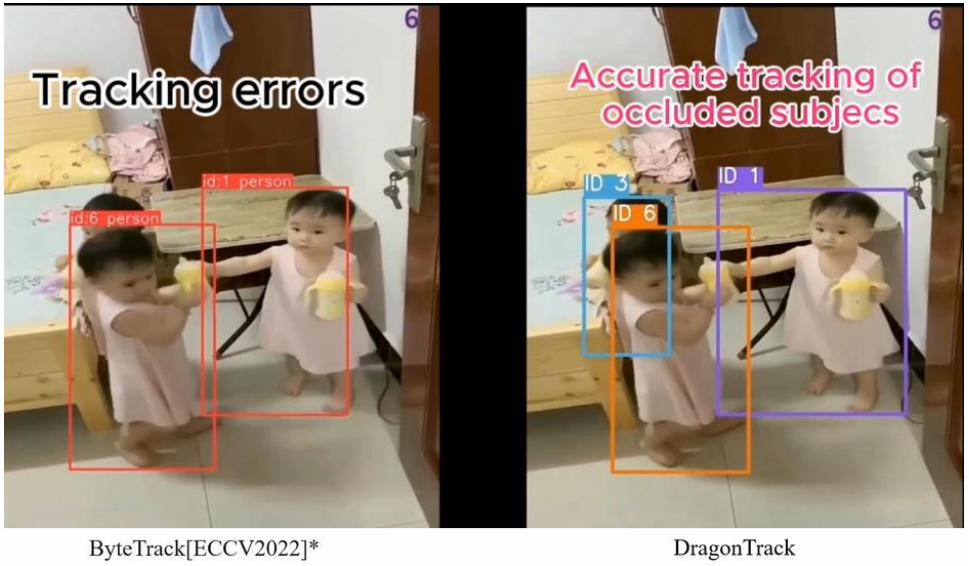
- **ID Switches:** Losing track of identities after occlusion
- Example: When two similar people cross paths and are later confused



UniTrack [ICCV 2025]

Why Tracking Remains Challenging

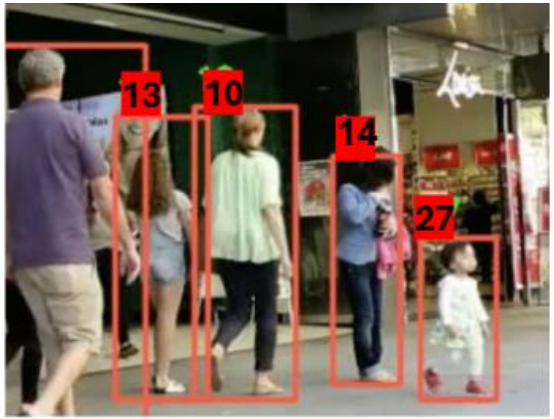
- **Similar Appearances**: Distinguishing between visually similar objects
- Particularly challenging in uniform scenarios (e.g., sports teams)



DragonTrack [WACV 2025]

Why Tracking Remains Challenging

- **Crowded Scenes**: Complex interactions between multiple objects Computationally demanding
- Dense crowds create frequent occlusions

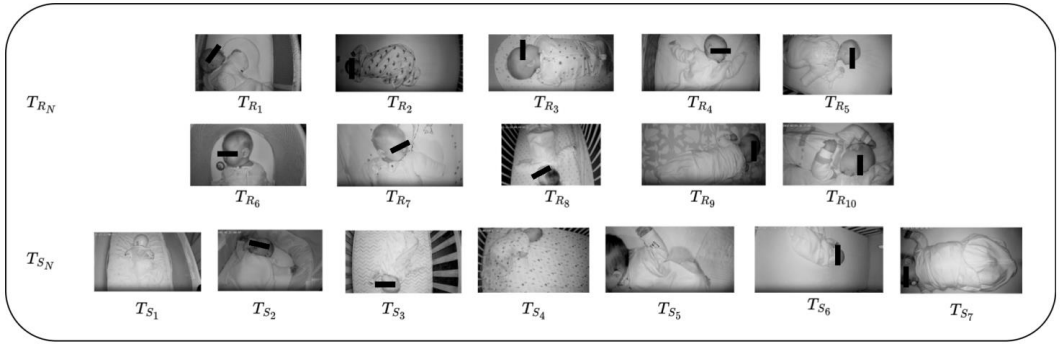


Bishoy Galoaa MS Thesis April 2025

UniTrack [ICCV 2025]

Why Tracking Remains Challenging

- **Small Data Domains**: Limited training examples in specialized areas
- Medical applications, infant monitoring
- Requires methods that can learn from sparse data



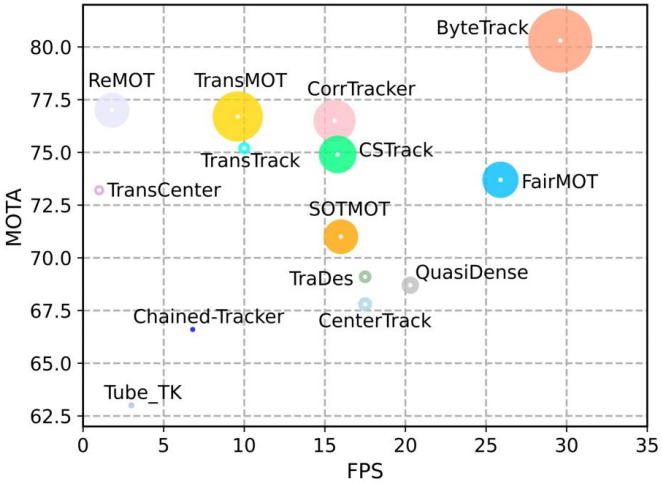
Moezzi S [WACVW 2025]

The Evolution of Multi-Object Tracking

Bishoy Galoaa MS Thesis April 2025

From Detection to Understanding

- Multi-Object tracking has evolved through three major paradigms:
  - Traditional Methods (2016-2018)
    - **SORT** (Simple Online Realtime Tracking)
    - **DeepSORT** (Integration with appearance features)
    - Limitations: They struggled with occlusions and similar appearances
  - Transform-based Methods (2021-Present)
    - **Trackformer**, **MOTR**, **TransTrack**
    - End-to-end learning
    - Better context understanding but room for improvement

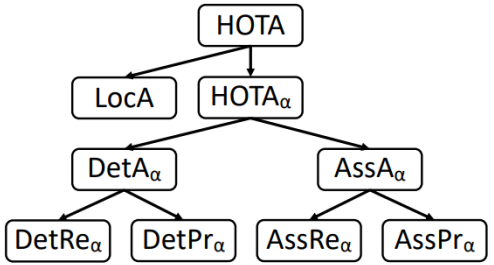


ByteTrack [ECCV 2022]

From Detection to Understanding

- **Key Evaluation Metrics**:
- HOTA (Higher Order Tracking Accuracy): Balances detection and association
- MOTA (Multiple Object Tracking Accuracy): Focuses on detection errors
- IDF1: Measures identity consistency throughout the tracking

Metric	Purpose	Equation
MOTA ↑	Overall Accuracy	$1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t GT_t}$
MOTP ↑	Localization	$\frac{\sum_{t,i} d_{t,i}}{\sum_t c_t}$
IDF1 ↑	Identity	$\frac{\lvert IDTP \rvert}{\lvert IDTP \rvert + 0.5\lvert IDFN \rvert + 0.5\lvert IDFP \rvert}$
HOTA ↑	Detection & Association	$\left(\prod_{i=1}^{n} \sqrt{DetA(i) \times AssA(i)} \right)^{\frac{1}{n}}$

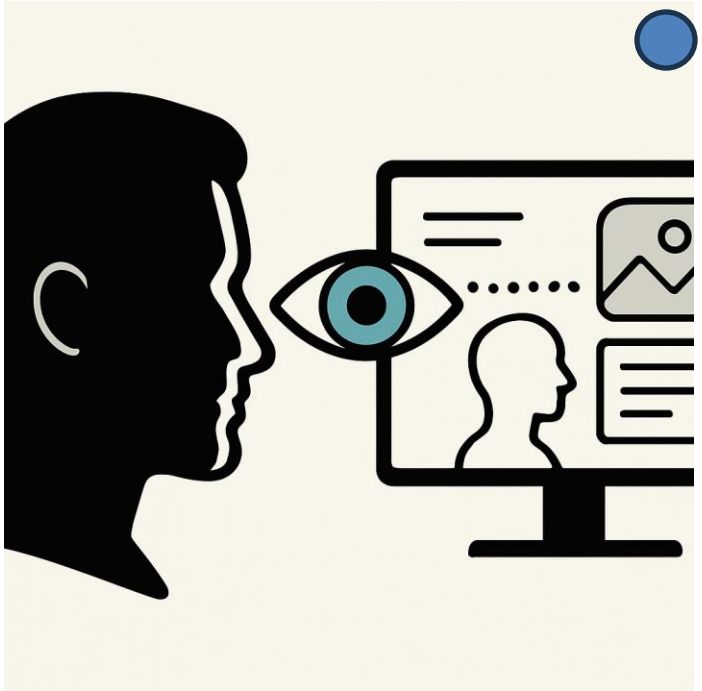


HOTA [IJCV 2020]

Research Framework Overview

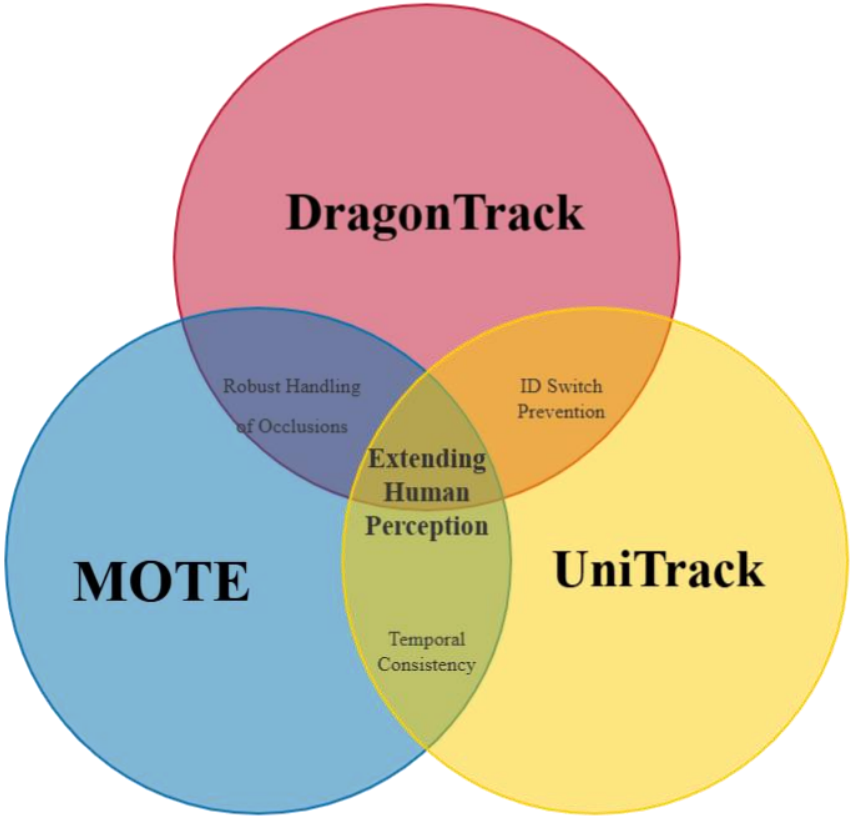
Our Approach to Complex Tracking

- **Research Philosophy**: "Extending human perception rather than replacing it“
- **Three main solutions**:
  - DragonTrack
  - MOTE
  - UniTrack



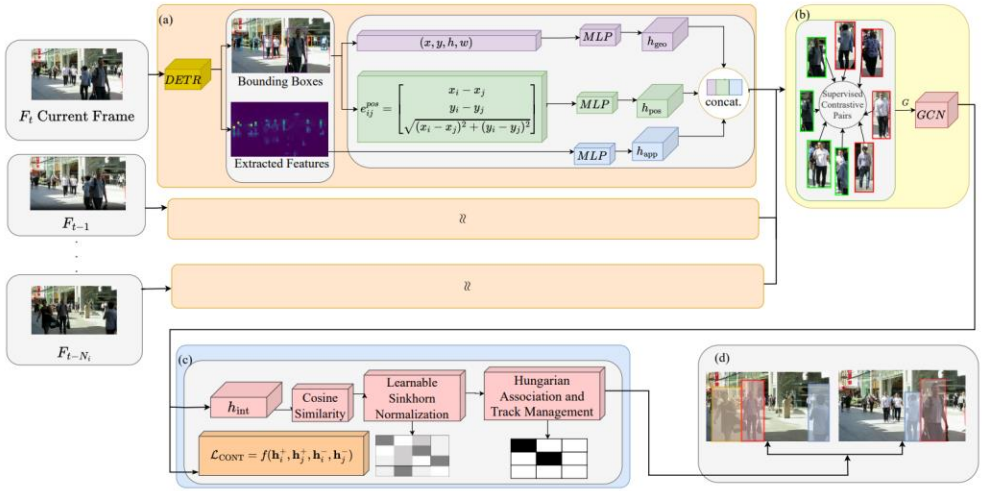
Unifying Theme

- Creating systems that maintain awareness where human attention falters
- **5 Major Barriers to Robust Tracking**
  1. **Occlusions** – temporary disappearance of objects behind others
  2. **ID Switches** – losing track of identities after occlusions
  3. **Similar Appearances** – visually indistinguishable subjects (e.g., uniforms)
  4. **Crowded Scenes** – dense interactions and visual overlap
  5. **Small Data Domains** – limited training data in healthcare/child monitoring



DragonTrack: Overview

- **DragonTrack**: Transformer-enhanced graphical tracking Addresses occlusion and ID switches
- Combines transformer detection with graph modeling
- **Published** in WACV 2025

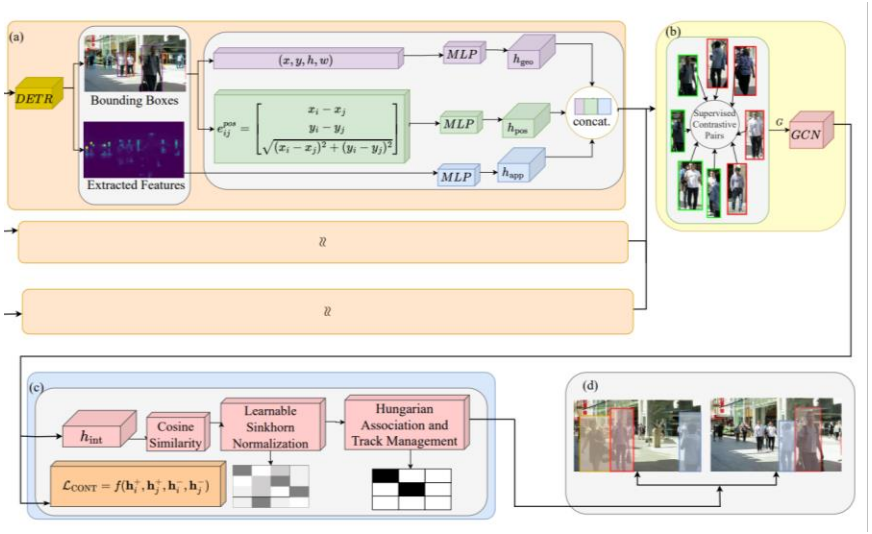


DragonTrack [WACV 2025]

DragonTrack: Method

Key Innovations:

- **Transformer-based Detection** *(DETR)* for spatial + appearance features
- **Graph Convolutional Network (GCN)** to model inter-object relationships
- **Learnable Sinkhorn Assignment** for optimal track-detection matching
- **Contrastive Loss + Weighted BCE** for identity-preserving embeddings



$$e_{ij}^{pos} = \left[\frac{x_i - x_j}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}, \frac{y_i - y_j}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}\right]$$

DragonTrack: Results

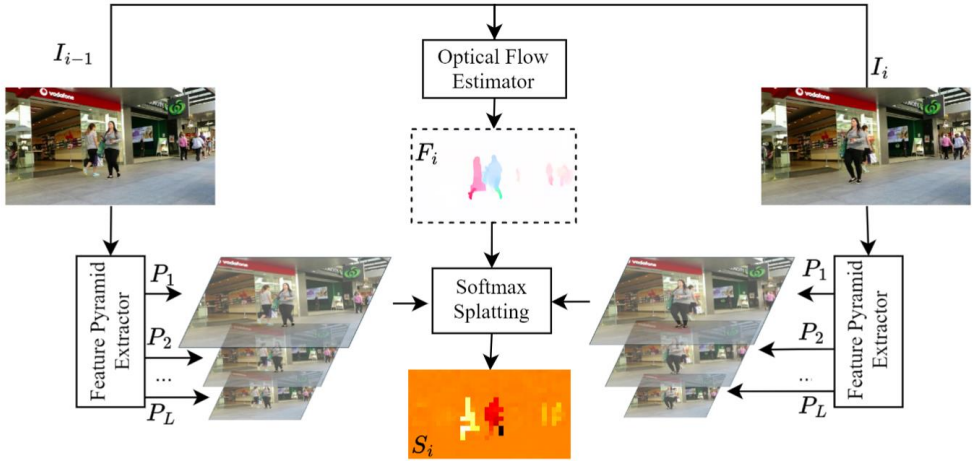
- We compared DragonTrack against both SOTA CNN-based and Transformer-based methods
- DragonTrack was evaluated over MOT17 and DanceTrack

Methods	HOTA↑	AssA↑	DetA↑	IDF1↑	MOTA↑	IDS↓
CNN-based:						
Tracktor++ [3]	44.8	45.1	44.9	52.3	53.5	2072
MPNTrack [6]	49.0	51.1	47.3	61.7	58.8	**1185**
CenterTrack [39]	52.2	51.0	53.8	64.7	67.8	3039
TraDeS [19]	52.7	50.8	55.2	63.9	69.1	3555
QDTrack [19]	53.9	52.7	55.6	66.3	68.7	3378
GSDT [28]	55.5	54.8	56.4	68.7	66.2	3318
FairMOT [37]	59.3	58.0	60.9	72.3	73.7	3303
CorrTracker [26]	60.7	58.9	62.9	73.6	76.5	3369
GRTU [27]	62.0	62.1	62.1	75.0	74.9	1812
MAATrack [22]	62.0	60.2	64.2	75.9	79.4	1452
StrongSORT [10]	63.5	63.7	63.6	78.5	78.3	1446
ByteTrack [36]	63.1	62.0	64.5	77.3	80.3	2196
Transformer-based:						
TrackFormer [17]	-	-	-	63.9	65.0	3528
TransTrack [24]	54.1	47.9	61.6	63.9	74.5	3663
MOTR [35]	57.8	55.7	60.3	68.6	73.4	2439
MOTRv2 [38]	62.0	60.6	63.8	75.0	78.6	-
DragonTrack (Ours)	**65.3**	**66.2**	**65.8**	**79.2**	**82.0**	1313

DragonTrack [WACV 2025]

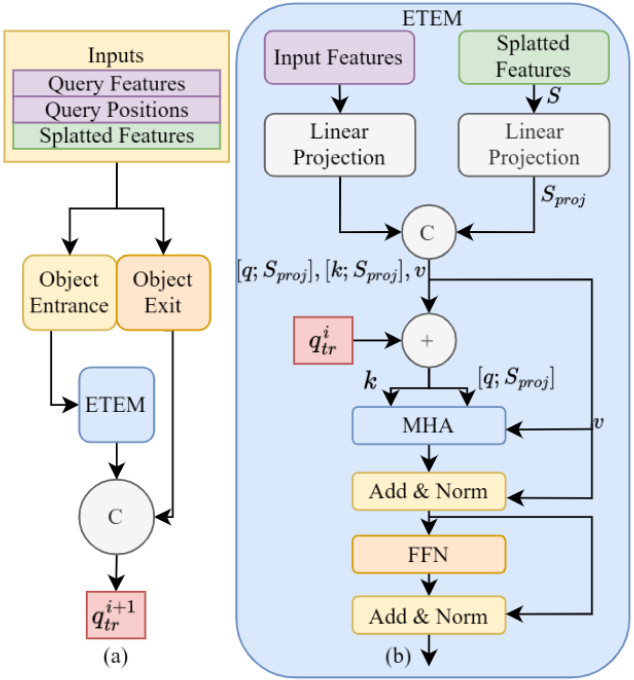
MOTE: Overview

- **MOTE** (More Than Meets the Eye):
- Optical flow-based tracking Maintains tracking during **prolonged occlusions**
- Uses **Softmax splatting** for disocclusion features
- **Under review** at ICML 2025



MOTE: Method

- **Optical Flow Estimation**
  - Computes dense motion fields between frames to model object movement—even when partially or fully occluded.
- **Softmax Splatting Module**
  - Warps feature maps along flow vectors to create **disocclusion-aware** representations.
- **ETEM (Enhanced Track Embedding Module)**
  - Inputs: Query features, positions, and splatted motion features
  - Uses **multi-head attention** to fuse temporal and appearance information
  - Applies **linear projection**, **add & norm**, and **feed-forward layers** for robust embedding updates
  - Outputs: Enhanced track query q_{tr}^{i+1} with improved identity continuity



MOTE [ICML 2025]

MOTE: Results

- MOTE was evaluated on the MOT15, MOT17, MOT20 and DanceTrack datasets

Methods	HOTA↑	AssA↑	DetA↑	MOTA↑	IDF1↑	IDS↓
CNN-based:						
Tracktor++(Bergmann et al., 2019)	44.8	45.1	44.9	53.5	52.3	2072
CenterTrack(Zhou et al., 2020)	52.2	51.0	53.8	67.8	64.7	3039
TraDeS (Pang et al., 2021)	52.7	50.8	55.2	69.1	63.9	3555
QDTrack (Pang et al., 2021)	53.9	52.7	55.6	68.7	66.3	3378
GSDT (Wang et al., 2021c)	55.5	54.8	56.4	66.2	68.7	3318
FairMOT(Zhang et al., 2021)	59.3	58.0	60.9	73.7	72.3	3303
CorrTracker (Wang et al., 2021a)	60.7	58.9	62.9	76.5	73.6	3369
GRTU (Wang et al., 2021b)	62.0	62.1	62.1	74.9	75.0	1812
MAATrack (Stadler & Beyerer, 2022)	62.0	60.2	64.2	79.4	75.9	1452
StrongSORT (Du et al., 2023)	63.5	63.7	63.6	78.3	78.5	1446
ByteTrack (Zhang et al., 2022b)	63.1	62.0	64.5	80.3	77.3	2196
BoostTrack (Zhang et al., 2023a)	65.4	64.2	64.8	80.5	80.2	1104
Transformer-based:						
TrackFormer (Meinhardt et al., 2021)	/	/	/	65.0	63.9	3528
TransTrack(Sun et al., 2020)	54.1	47.9	61.6	74.5	63.9	3663
MOTR(Zeng et al., 2022)	57.8	55.7	60.3	73.4	68.6	2439
MOTRv2(Zhang et al., 2023b)	62.0	60.6	63.8	78.6	75.0	/
MOTE (Ours)	**66.3**	**67.8**	**65.4**	**82.0**	**80.3**	**1412**

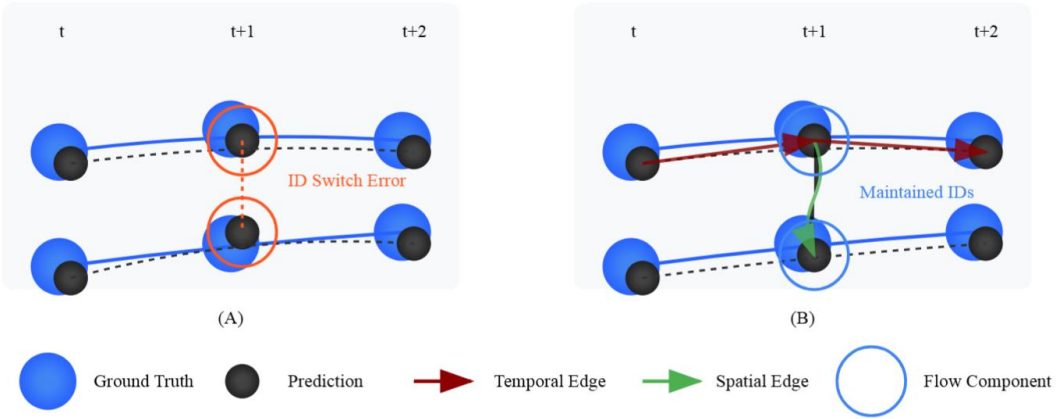
MOTE [ICML 2025]

MOTE: Visual Results

MOTE [ICML 2025]

UniTrack: Overview

- **UniTrack (Under Review** ICCV 2025) : Unified loss function for tracking optimization Graph-based differentiable loss
- Integration with existing architectures
- Addresses post-occlusion, temporal, and cross-subject errors



UniTrack [ICCV 2025]

UniTrack: Method

- Graph Definition:
  - Vt: Nodes representing tracked objects at time t
  - Et: Edges encoding spatial and temporal relationships
  - Wt: Edge weights derived from learned flow variables
- **Unified Loss Formulation**:
- **Temporal Edges** (Type 2 errors: drift)
- **Spatial Edges** (Type 3 errors: similar subjects)
- **Flow Conservation Constraint** (Track life cycle)

$$\mathcal{G} = \{G_t = (V_t, E_t, W_t)\}_{t=1}^{T}$$

$$\mathcal{L} = \mathcal{L}_{flow} + \lambda_s \mathcal{L}_{spatial} + \lambda_t \mathcal{L}_{temporal}$$

$$\mathcal{L}_{temporal} = \sum_{(i,j)\in E_t} \|v_t^i - v_{t+1}^i\|^2 f_{t,t+1}^i$$

$$\mathcal{L}_{spatial} = \sum_{(i,j)\in E_t} w_{ij} \cdot d(p_t^i, p_t^j) \cdot f_{t,t+1}^i$$

$$\sum_{j\in N^+(i)} f_{t,t+1}^i - \sum_{j\in N^-(i)} f_{t-1,t}^j = b_i$$

UniTrack: Results

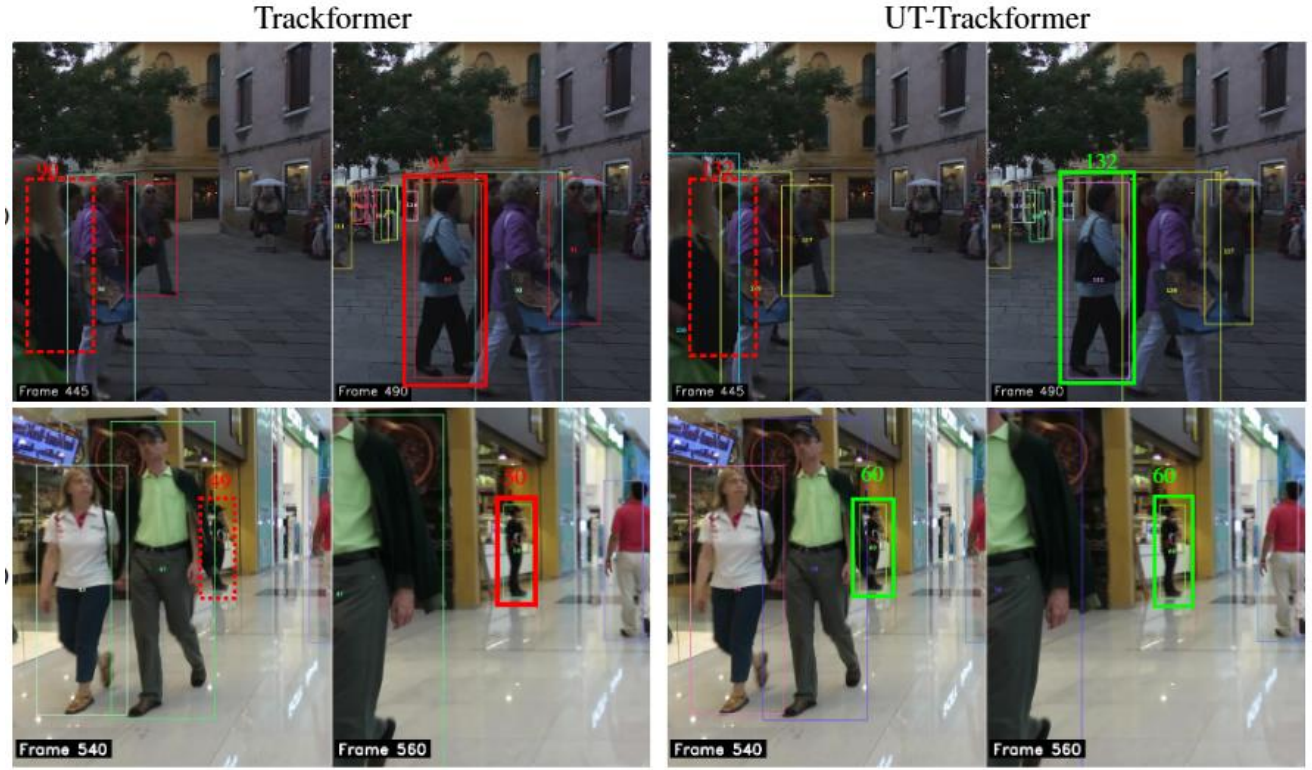
- UniTrack was tested over 3 tracker: MOTR, Trackformer and FairMOT
- The UT-proposed algorithm were evaluated over the MOT17 and MOT20 datasets

Method	MOTA↑	IDF1↑	HOTA↑	FP↓	FN↓	IDs↓
MOTR [14]	62.1	61.3	53.2	1843	21034	289
UT-MOTR	**64.8**	**63.9**	**55.7**	**1562**	**19845**	**356**
Trackformer [8]	62.3	57.6	52.8	1965	21893	643
UT-Trackformer	65.9	66.4	56.2	**1039**	16667	705
UT-Trackformer*	**69.05**	**69.05**	**62.13**	1299	**16605**	**396**
FairMOT [16]	61.7	61.5	52.9	1902	20456	388
UT-FairMOT	**64.5**	**64.2**	**55.3**	**1623**	**19234**	**482**

Method	MOTA↑	IDF1↑	HOTA↑	FP↓	FN↓	IDs↓
MOTR [14]	53.2	57.9	51.8	1843	25034	389
UT-MOTR	**55.8**	**60.4**	**54.2**	**1562**	**23845**	**356**
Trackformer [8]	54.1	56.2	50.9	1965	25893	643
UT-Trackformer	**56.16**	**64.14**	**57.66**	**1374**	**22004**	**314**
FairMOT [16]	53.5	58.3	52.4	1902	25456	488
UT-FairMOT	**55.2**	**61.5**	**55.8**	**1723**	**23234**	**402**

UniTrack [ICCV 2025]

UniTrack: Visual Results

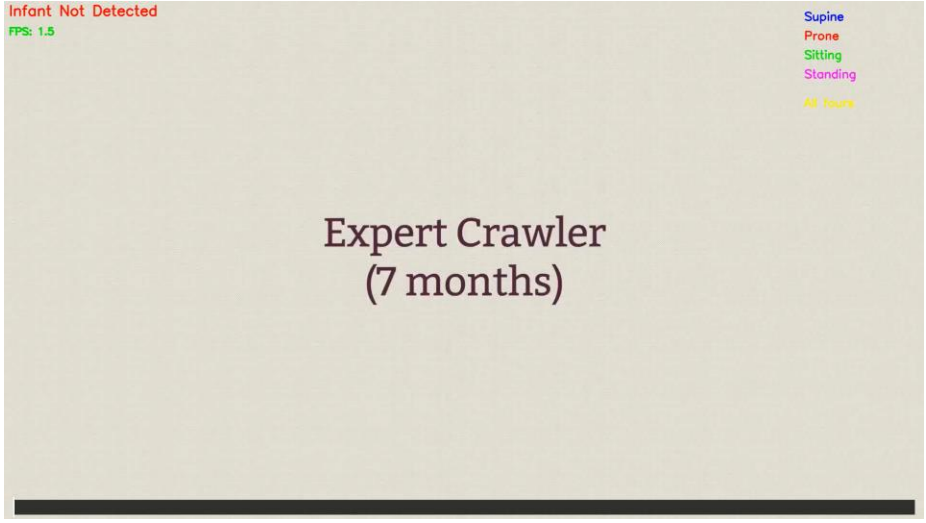


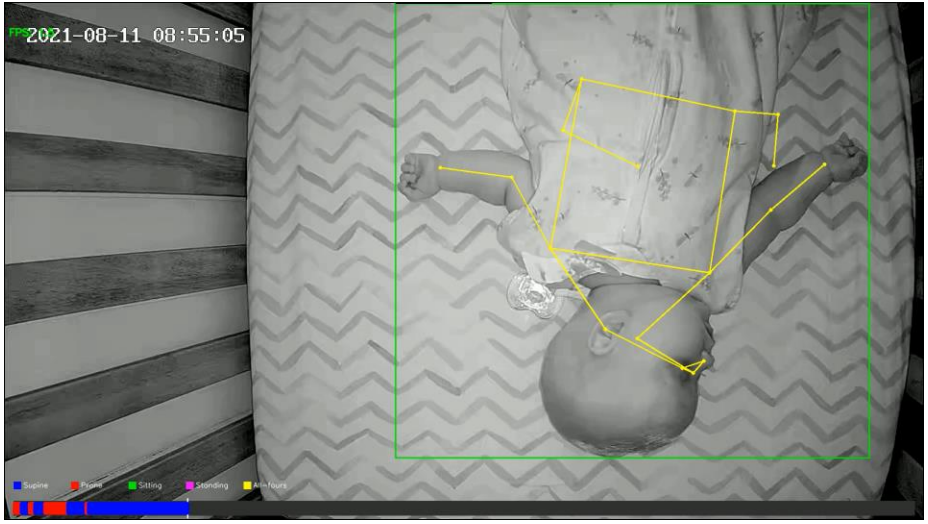
Bishoy Galoaa MS Thesis April 2025

UniTrack [ICCV 2025]

Real World Applications

- Tracking isn't for academia only
- Healthcare:
  - Infant Tracking
  - Sleep Monitoring
- Safety
  - Cars/Pedestrians interactions





Bishoy Galoaa MS Thesis April 2025

✅ Solved vs 🔧 Remaining Challenges

- **Occlusion Handling**
  - *MOTE* preserves tracking during prolonged occlusions using optical flow + splatting
- **Crowded Scene Robustness**
  - *DragonTrack* uses graph modeling for better discrimination in dense environments
- **Temporal Drift Reduction**
  - *UniTrack* enforces consistency across frames via graph flow constraints
- **Plug-and-Play Optimization**
  - *UniTrack* enhances identity tracking across multiple SOTA architectures

Remaining Challenges (Future Work)

- **Residual ID Switches**
  - Still occur in ambiguous re-identification cases despite *UniTrack*'s gains
- **Long-Term Temporal Memory**
  - Need for models to better remember identities across long video gaps
- **Extreme Appearance Similarity**
  - Challenging in uniforms/siblings—requires more than visual features
- **Data Scarcity Adaptability**
  - Real-world deployment needs robust performance with minimal fine-tuning

Future Research Direction

- **Multi-Camera Point Tracking:**
  - Novel approach to simultaneously track points across multiple calibrated cameras
  - Gain physical understanding of human motion and interaction
- **Infant Monitoring and Tracking:**
  - Solve data scarcity problem using, using minimal labeling techniques
- **Multi-Platform Multi Sensor Dataset Collection:**
  - Enabling broader research opportunities for multi-modal tracking
- **Long-Term Vision: Chain of Reasoning Events (CoRE)**
  - Integrating linguistic-cognitive intelligence into tracking systems
  - Creating tracking systems that can explain and reason about observed behaviors

Thank You!

To the entire ACLab team for their **collaboration** and **support!**

Thank You!

Questions?

UniTrack: Graph Components

where:
- Vt are **vertices** (detected objects at time t)
- Et are **edges** (spatial or temporal links)
- Wt are **weights** encoding association strength (flow)

- **Temporal Edges**
  - Represented by **red horizontal lines**
  - Connect same object across time
  - Enforce **motion smoothness**
  - 🔧 Mitigates **Type 2 errors**: identity drift over time
- **Spatial Edges**
  - Represented by **green vertical lines**
  - Link co-occurring objects within the same frame
  - Enforce **appearance and geometry-based consistency**
  - 🔧 Mitigates **Type 3 errors**: confusion between similar subjects
- **Flow Weights ft,t+1if_{t,t+1}^ift,t+1i**
  - **Blue circles** encode confidence in track continuity
  - Used in loss terms + flow conservation constraint
  - 🔧 Addresses **Type 1 errors**: false starts or ends of a track
- **Track State bi∈{−1,0,1}**
  - Indicates whether a track is starting, continuing, or ending

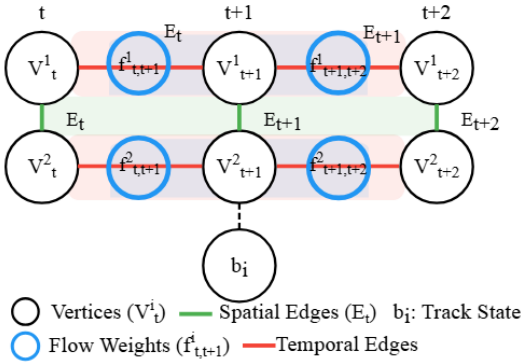
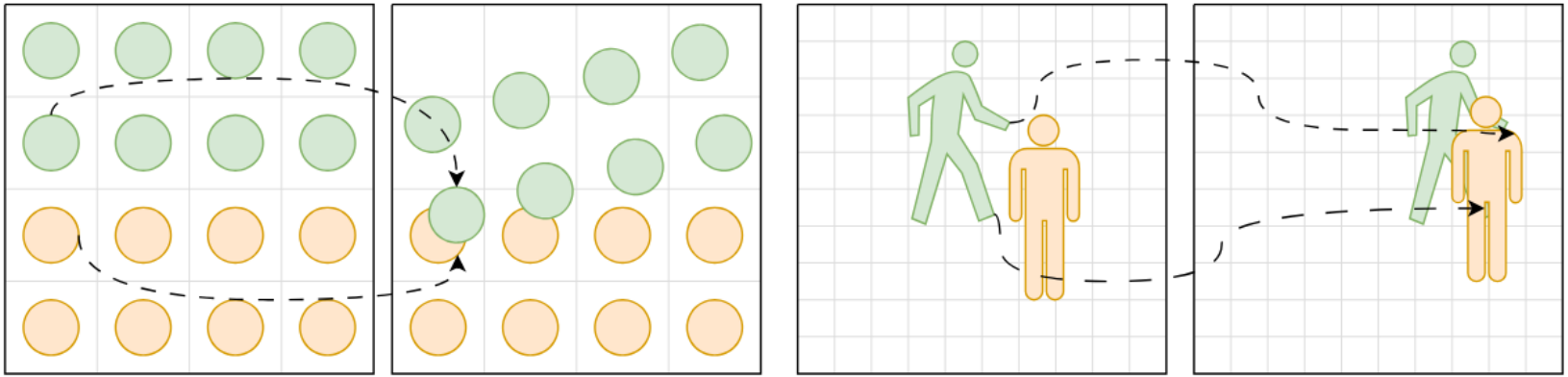


Figure 3. Graph-theoretic formulation of the tracking problem. We represent a multi-object tracking sequence as a temporal sequence of weighted directed graphs $\mathcal{G} = \{G_t = (V_t, E_t, W_t)\}_{t=1}^{T}$. Tracked objects are represented as vertices V_t^i where superscript i indicates object identity and subscript t denotes time. The figure shows three key components: (1) temporal edges (red horizontal lines) modeling motion consistency across frames to address Type 2 errors, (2) spatial edges (green vertical lines) capturing inter-object relationships within frames to mitigate Type 3 errors, and (3) flow components (blue circles) encoding association strengths $f_{t,t+1}^i$ to prevent Type 1 errors. Light red background highlights the temporal connections, light green background highlights the spatial relationships, and light blue highlights the flow components. The track state variable b_i represents initialization, continuation, or termination of tracks through flow conservation constraints.

MOTE: Method (cont.)



MOTE: Method (cont.)