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

ICML
International Conference
On Machine Learning

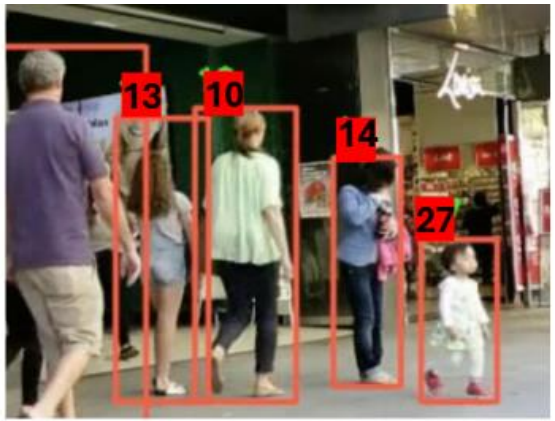
More Than Meets the Eye: Enhancing Multi-Object Tracking Even with Prolonged Occlusions

Bishoy Galoaa · Somaieh Amraee · Sarah Ostadabbas*



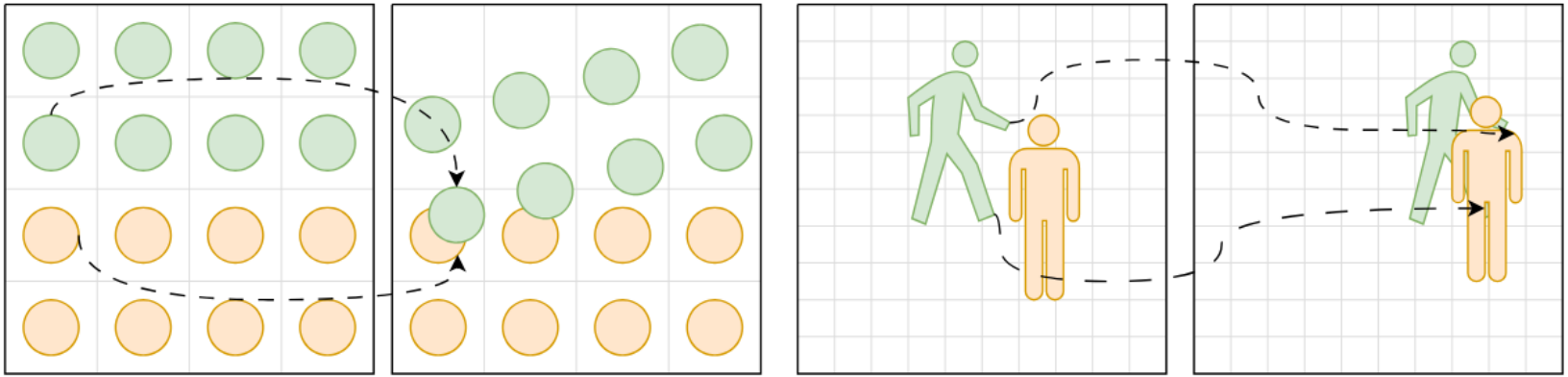
Why Prolonged Occlusion is Challenging

- **Prolonged Occlusions:** Extended periods where tracked objects are completely hidden
- **Why it matters:** Essential for crowd analysis, autonomous navigation, and human behavior understanding



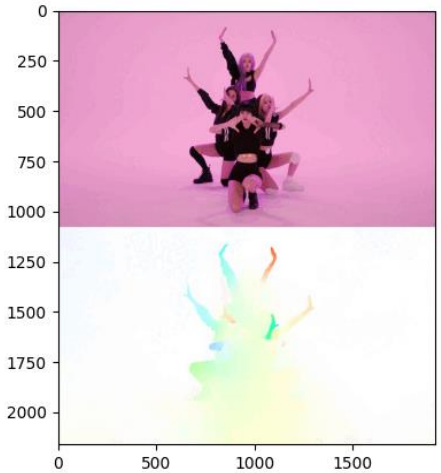
MOTE: Motivation

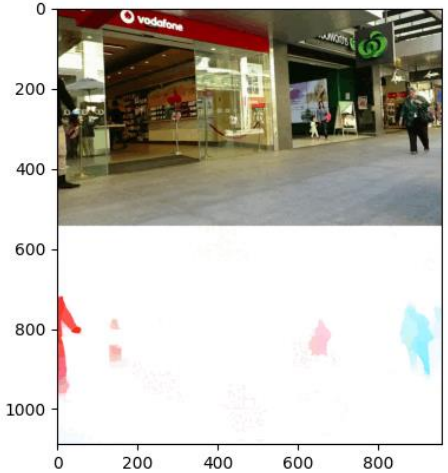
Applying Softmax Splatting to Occlusions: This technique propagates object features through occluded regions by combining motion-guided warping with visibility weighting, preserving identity information when direct observation is impossible.



Optical Flow as a Feature

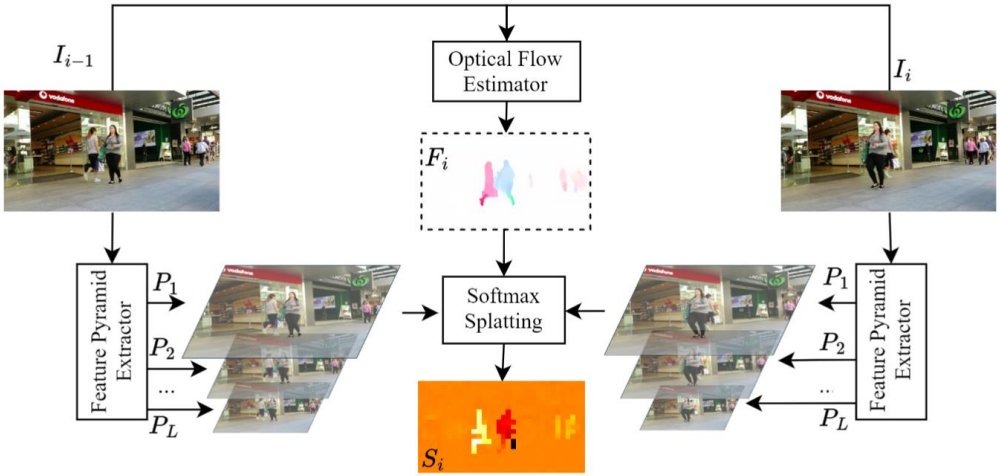
To capture motion features, we utilize optical flow—a natural complement to multi-object tracking that provides crucial temporal information between frames.





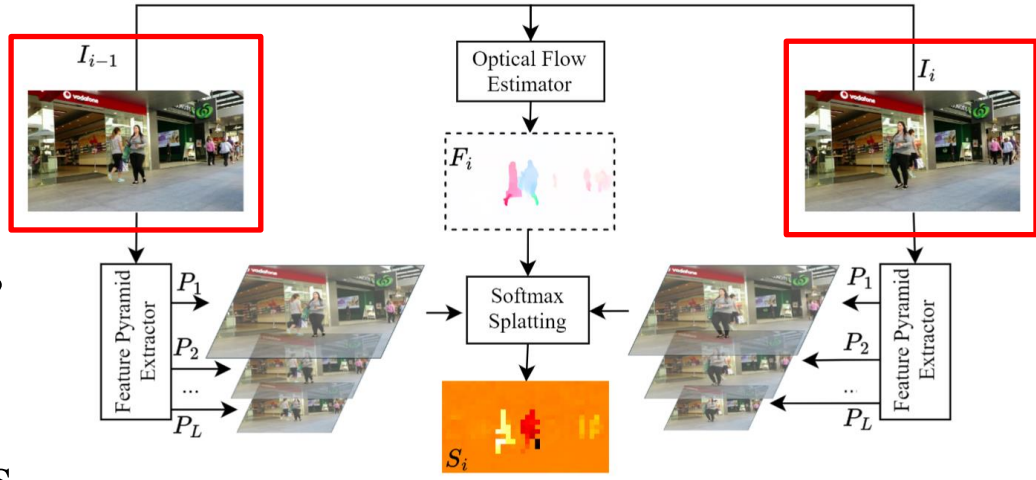
MOTE: Overview

- **MOTE** (More Than Meets the Eye): Our occlusion-aware tracking framework
- **Key strength**: Leverages optical flow to track objects through prolonged occlusions
- **Technical innovation**: Employs softmax splatting to generate disocclusion features



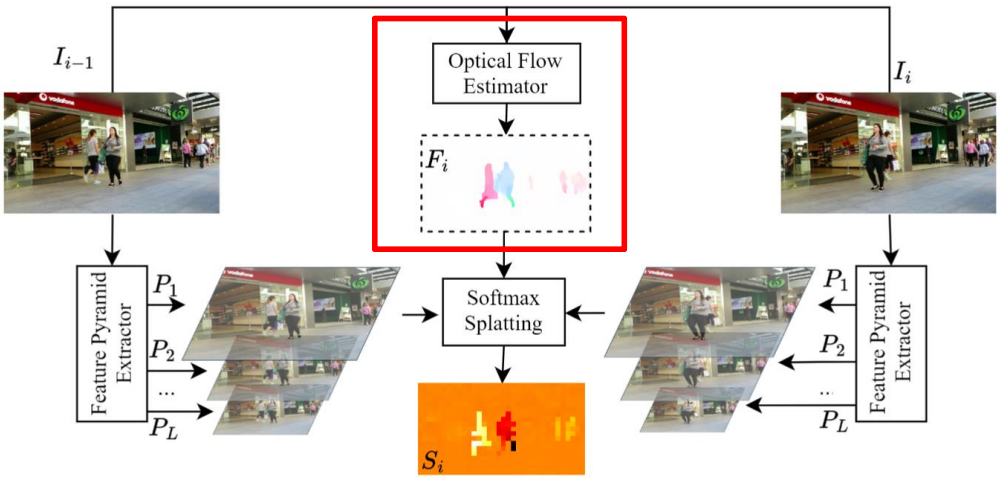
MOTE: Overview

- **MOTE** (More Than Meets the Eye): Our occlusion-aware tracking framework
- **Key strength**: Leverages optical flow to track objects through prolonged occlusions
- **Technical innovation**: Employs softmax splatting to generate disocclusion features



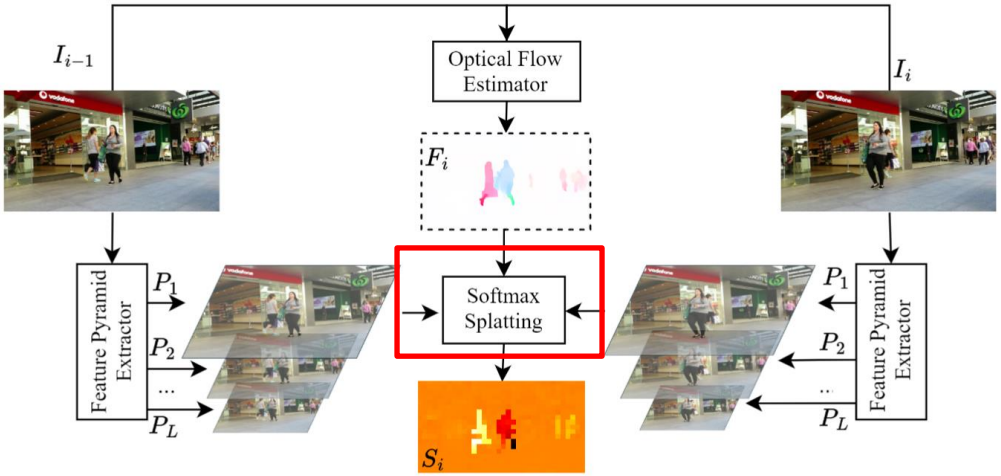
MOTE: Overview

- **MOTE** (More Than Meets the Eye): Our occlusion-aware tracking framework
- **Key strength**: Leverages optical flow to track objects through prolonged occlusions
- **Technical innovation**: Employs softmax splatting to generate disocclusion features



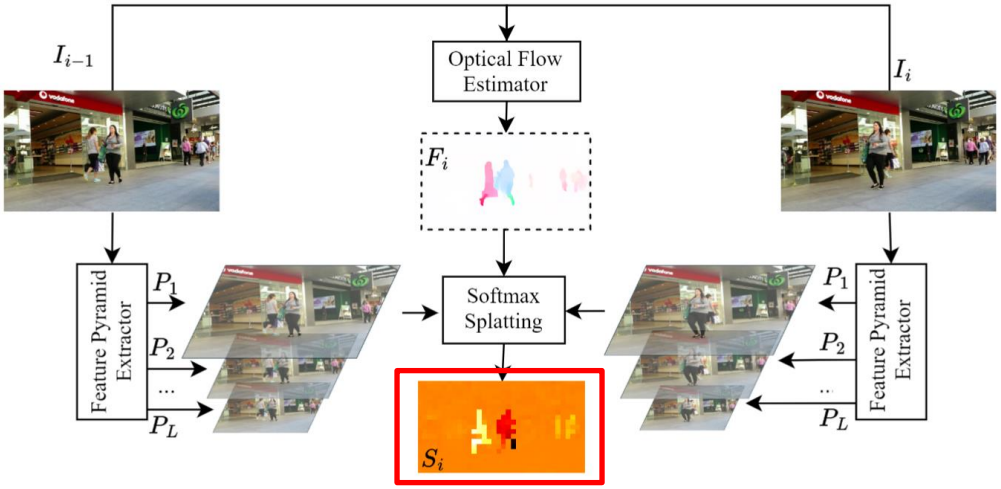
MOTE: Overview

- **MOTE** (More Than Meets the Eye): Our occlusion-aware tracking framework
- **Key strength**: Leverages optical flow to track objects through prolonged occlusions
- **Technical innovation**: Employs softmax splatting to generate disocclusion features



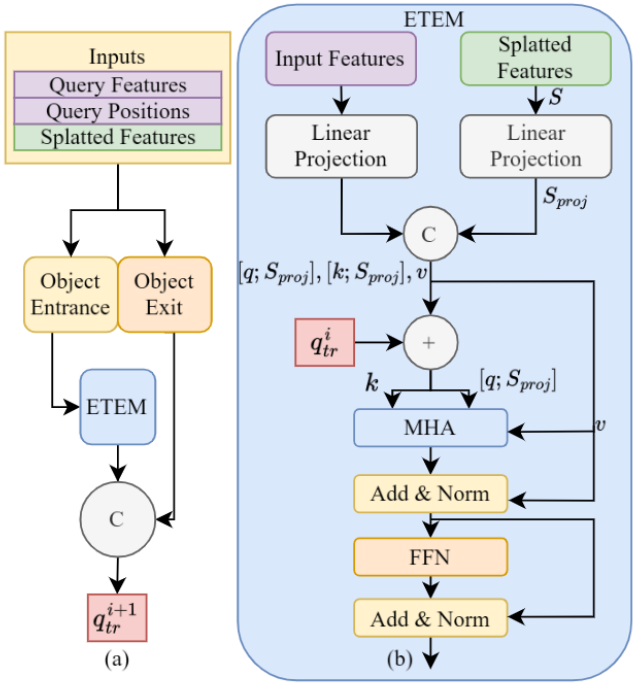
MOTE: Overview

- **MOTE** (More Than Meets the Eye): Our occlusion-aware tracking framework
- **Key strength**: Leverages optical flow to track objects through prolonged occlusions
- **Technical innovation**: Employs softmax splatting to generate disocclusion features



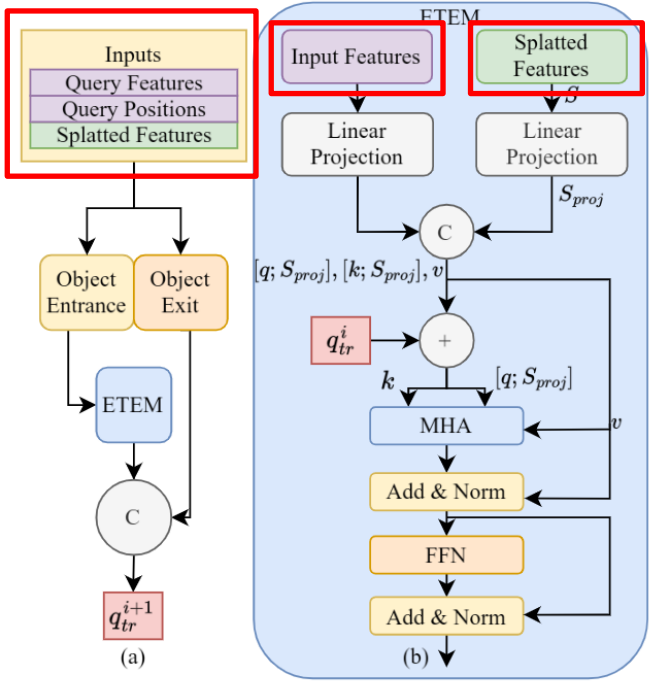
MOTE: Method

- **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



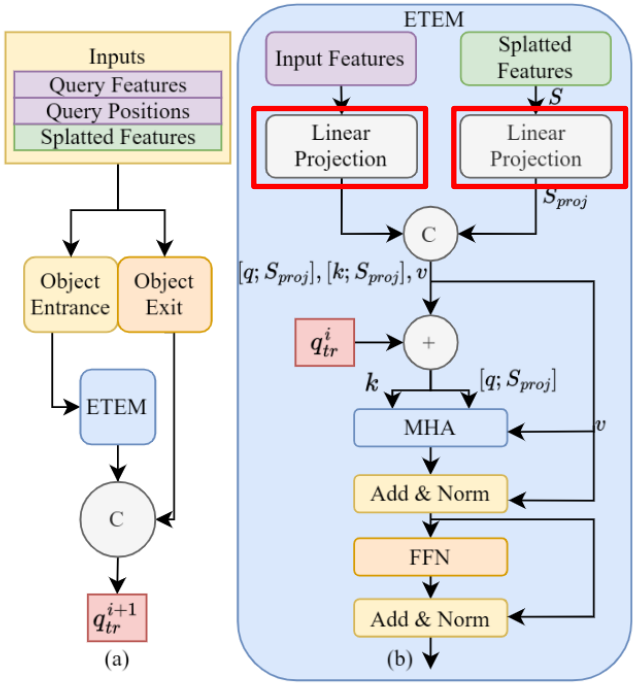
MOTE: Method

- **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



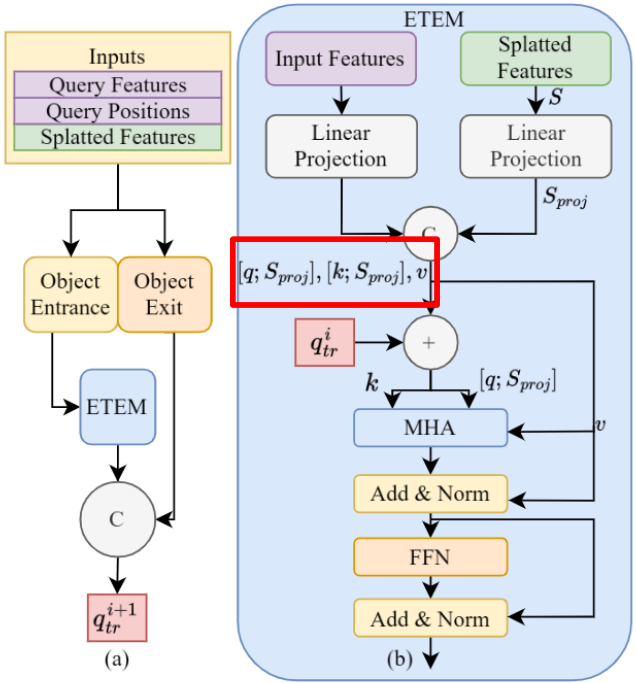
MOTE: Method

- **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



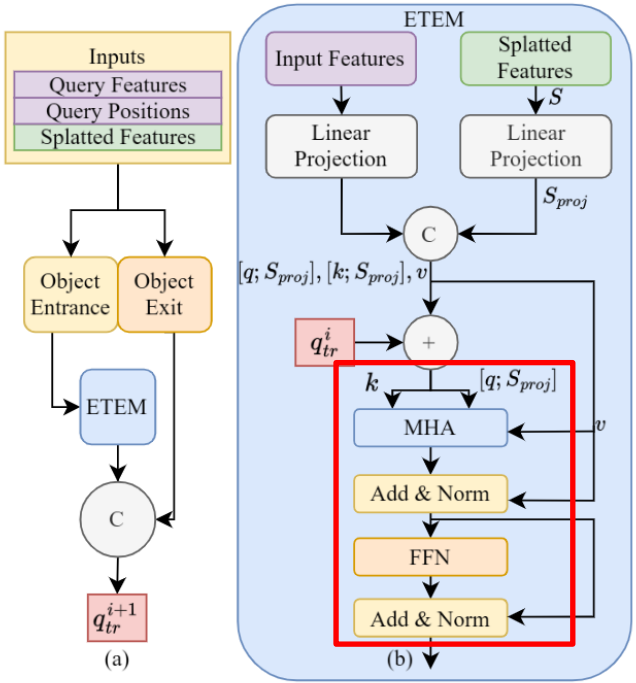
MOTE: Method

- **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



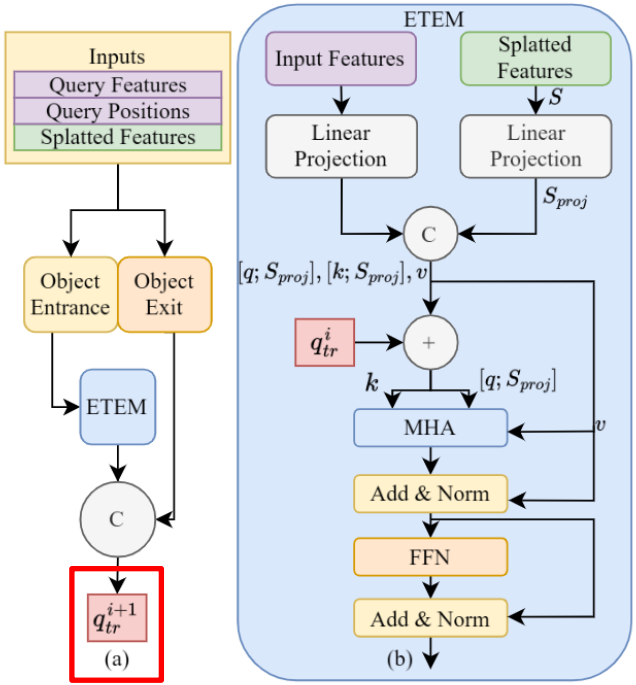
MOTE: Method

- **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: Method

- **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



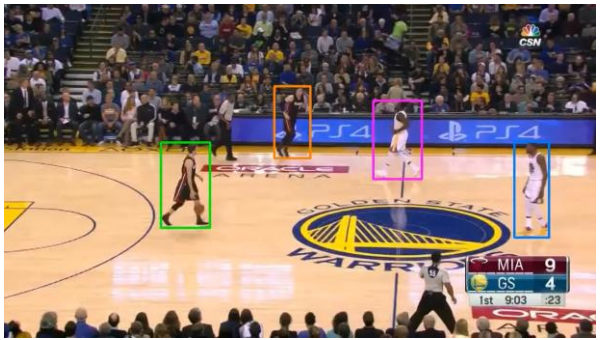
How MOTE Sees the World

- **MOTE** processes video frames (left), estimates optical flow (middle), and generates disocclusion matrices (right) to track objects through occlusions.

MOTE: Benchmark Used

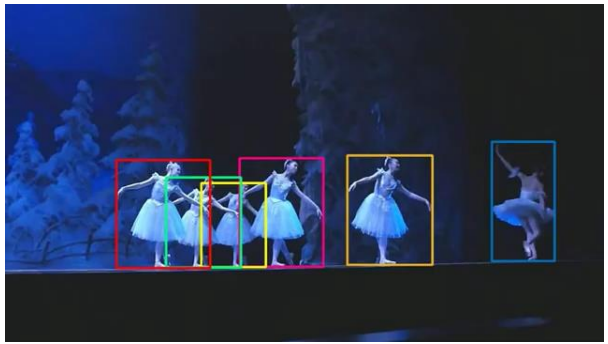
MOTE was evaluated on:

- MOT Challenge: MOT15 MOT17, MOT20
- DanceTrack
- SportsMOT



SportsMOT [ICCV 2023]

MOT Challenge [IJCV 2010]

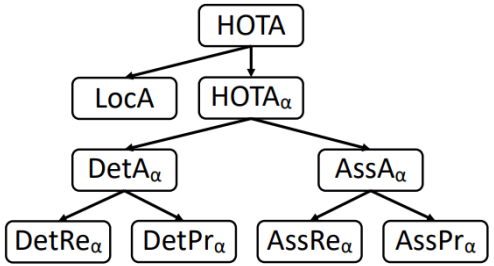


DanceTrack Challenge [2022]

MOTE: Metrics Used

- **We reported 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]

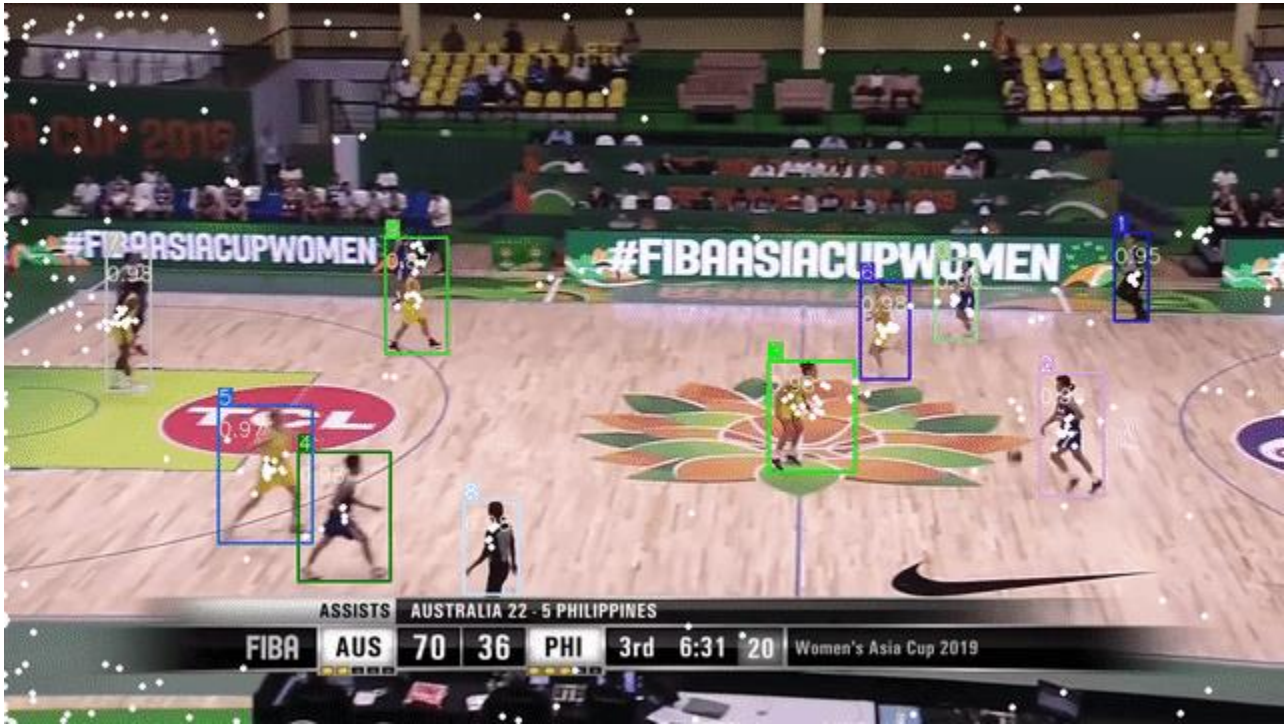
Prolonged Occlusions

ByteTrack[ECCV2022]*

*According to the results presented in the paper, ByteTrack ranks as the third-best method in terms of MOTA, following **MOTE** (ours).

Crowded Scene with Multiple Occlusions (MOT20)

SportsMOT Results



SportsMOT Results

SportsMOT Results

MOTE: Results

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

DanceTrack

Methods	HOTA↑	AssA↑	DetA↑	MOTA↑	IDF1↑
CNN-based:					
FairMOT (Zhang et al., 2021)	39.7	23.8	66.7	82.2	40.8
CenterTrack (Zhou et al., 2020)	41.8	22.6	78.1	86.8	35.7
TraDeS (Pang et al., 2021)	43.3	25.4	74.5	86.2	41.2
QDTrack (Pang et al., 2021)	54.2	38.7	81.0	87.7	50.4
ByteTrack (Zhang et al., 2022b)	47.7	31.0	71.0	91.5	48.8
OC-SORT (Cao et al., 2022)	55.1	38.3	80.3	92.0	54.6
Transformer-based:					
TransTrack (Sun et al., 2020)	45.5	27.5	75.9	88.4	45.2
GTR (Wang et al., 2021b)	48.0	31.9	72.5	89.7	50.3
MOTRv2 (Zhang et al., 2023b)	69.9	59.0	83.0	91.9	71.7
MOTRv2* (Zhang et al., 2023b)	73.4	64.4	**83.7**	92.1	**76.0**
MOTE (Ours)	**74.2**	**65.2**	82.6	**93.2**	75.2

MOT17

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**

SportsMOT

Methods	MOTA↑	IDF1↑	FPS↑
ByteTrack (Zhang et al., 2022b)	17.9	31.4	30.2
MOTR (Zeng et al., 2022)	44.1	48.7	7.5
MOTE (Ours)	**45.7**	**50.2**	**22.2**

MOTE: Results

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

DanceTrack

Methods	HOTA↑	AssA↑	DetA↑	MOTA↑	IDF1↑
CNN-based:					
FairMOT (Zhang et al., 2021)	39.7	23.8	66.7	82.2	40.8
CenterTrack (Zhou et al., 2020)	41.8	22.6	78.1	86.8	35.7
TraDeS (Pang et al., 2021)	43.3	25.4	74.5	86.2	41.2
QDTrack (Pang et al., 2021)	54.2	38.7	81.0	87.7	50.4
ByteTrack (Zhang et al., 2022b)	47.7	31.0	71.0	91.5	48.8
OC-SORT (Cao et al., 2022)	55.1	38.3	80.3	92.0	54.6
Transformer-based:					
TransTrack (Sun et al., 2020)	45.5	27.5	75.9	88.4	45.2
GTR (Wang et al., 2021b)	48.0	31.9	72.5	89.7	50.3
MOTRv2 (Zhang et al., 2023b)	69.9	59.0	83.0	91.9	71.7
MOTRv2* (Zhang et al., 2023b)	73.4	64.4	**83.7**	92.1	**76.0**
MOTE (Ours)	**74.2**	**65.2**	82.6	**93.2**	75.2

MOT17

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**

SportsMOT

Methods	MOTA↑	IDF1↑	FPS↑
ByteTrack (Zhang et al., 2022b)	17.9	31.4	30.2
MOTR (Zeng et al., 2022)	44.1	48.7	7.5
MOTE (Ours)	**45.7**	**50.2**	**22.2**

MOTE: Results

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

DanceTrack

Methods	HOTA↑	AssA↑	DetA↑	MOTA↑	IDF1↑
CNN-based:					
FairMOT (Zhang et al., 2021)	39.7	23.8	66.7	82.2	40.8
CenterTrack (Zhou et al., 2020)	41.8	22.6	78.1	86.8	35.7
TraDeS (Pang et al., 2021)	43.3	25.4	74.5	86.2	41.2
QDTrack (Pang et al., 2021)	54.2	38.7	81.0	87.7	50.4
ByteTrack (Zhang et al., 2022b)	47.7	31.0	71.0	91.5	48.8
OC-SORT (Cao et al., 2022)	55.1	38.3	80.3	92.0	54.6
Transformer-based:					
TransTrack (Sun et al., 2020)	45.5	27.5	75.9	88.4	45.2
GTR (Wang et al., 2021b)	48.0	31.9	72.5	89.7	50.3
MOTRv2 (Zhang et al., 2023b)	69.9	59.0	83.0	91.9	71.7
MOTRv2* (Zhang et al., 2023b)	73.4	64.4	**83.7**	92.1	**76.0**
MOTE (Ours)	**74.2**	**65.2**	82.6	**93.2**	75.2

MOT17

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**

SportsMOT

Methods	MOTA↑	IDF1↑	FPS↑
ByteTrack (Zhang et al., 2022b)	17.9	31.4	30.2
MOTR (Zeng et al., 2022)	44.1	48.7	7.5
MOTE (Ours)	**45.7**	**50.2**	**22.2**

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

Check out our paper and project page

Thank you!