Center for Research in Computer Vision
UNIVERSITY OF CENTRAL FLORIDA
THE UNIVERSITY OF WESTERN AUSTRALIA

3DMODT: Attention-Guided Affinities for Joint Detection & Tracking in 3D Point Clouds

Jyoti Kini, Ajmal Mian, Mubarak Shah

ICRA LONDON 2023

Motivation

- Traditionally, tracking multiple objects in 3D point clouds
  - Object detection
  - Data association

  } 2-step process
- Sub-par detections using off-the-shelf detectors affect overall tracking
- LiDAR point clouds are characterized by challenges
  - Data sparsity, uneven sampling density, lack of texture
  - Simple affinity estimation leads to noisy correspondences

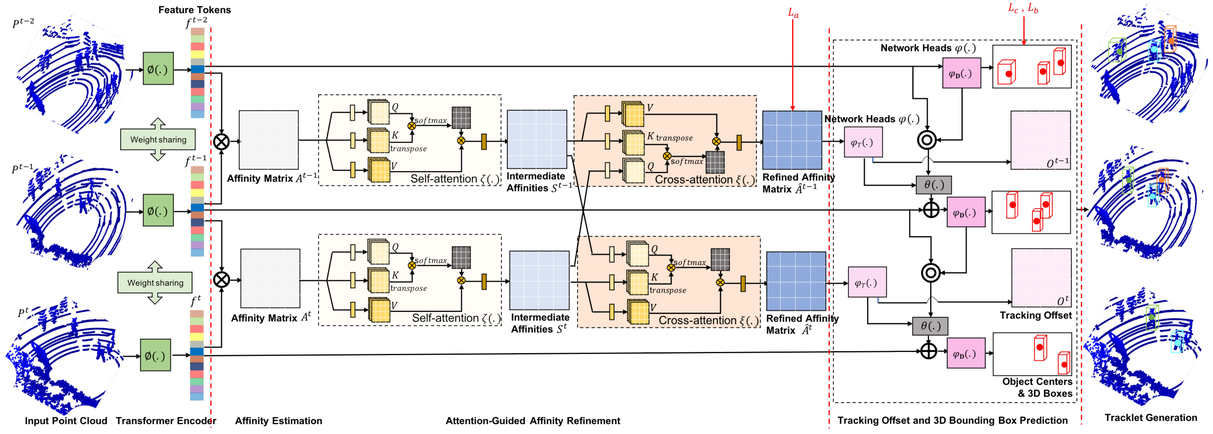
Contributions

- End-to-end 3D point cloud-based joint detection and tracking method
  - Eliminates the reliance on external object detector
  - No complex post-processing algorithms for data association needed
- Attention-guided affinity refinement module
  - Captures global context across affinities
  - Suppresses erroneous correspondences by seeking consensus among the affinities



- Competitive performance on tracking benchmarks
  - Waymo
  - KITTI
  - JRDB

Architecture



Transformer encoder generates feature tokens → Cosine similarity across feature tokens from consecutive scans used to construct affinity matrices → **Attention-mechanism on affinity matrices** refines correspondences → Network heads compute 3D bounding box information and tracking offsets → Network head predictions are collectively used to track the objects across timesteps

Attention-Guided Affinity Refinement

Self-attention Applied on individual affinity matrix

$$\zeta(A^{t-1}) = \text{softmax}(\bar{Q}^{t-1}\bar{K}^{t-1\mathsf{T}})\bar{V}^{t-1}$$

$$\zeta(A^{t}) = \text{softmax}(\bar{Q}^{t}\bar{K}^{t\mathsf{T}})\bar{V}^{t}$$

Cross-attention Applied across affinity matrices

$$\xi(A^{t-1}) = \text{softmax}(\bar{Q}^{t}\bar{K}^{t-1\mathsf{T}})\bar{V}^{t-1}$$

$$\xi(A^{t}) = \text{softmax}(\bar{Q}^{t-1}\bar{K}^{t\mathsf{T}})\bar{V}^{t}$$

Losses

La Forward Association Loss on Affinities $L_a = \frac{\sum(G \odot (-\log \hat{A}^t))}{\sum G}$

Lc $\ell 1$ Center Offset Loss $L_c = \frac{1}{R}\sum_{r=1}^{R}\left|\hat{C}_r - C_r\right|$

Lb $\ell 1$ Dimension Size Loss on Bounding Boxes $L_b = \frac{1}{R}\sum_{r=1}^{R}|\hat{\gamma}_r - \gamma_r|$

Objective Function $L = L_a + \lambda_c L_c + \lambda_b L_b$

Qualitative Results



Long Tracks | Poor Lighting | Adverse Weather

Quantitative Results

JRDB Test Set

Method	Input Data	External Detector	Complex Post-Processing	*MOTA*↑	*MOTP*↑	IDS↓
AB3DMOT [17]	*L*	✓	✓	19.34	42.01	6,179
JRMOT [5]	*L* + *I*	×	✓	20.15	42.44	4,216
PC-DAN [16]	*L*	✓	✓	**22.56**	6.01	26,022
3DMODT	*L*	×	×	21.74	**47.61**	**3,805**

Quantitative Results

Waymo Vehicle Validation Set

Method	Input Data	External Detector	Complex Post-Processing	*MOTA*↑	*MOTP*↑	Miss↓	Miss Match↓	*FP*↓
Sun et al. [46]	*L*	✓	✓	42.5 / 40.1	18.6 / 18.6	40.0 / 43.4	0.14 / 0.13	17.3 / 16.4
CenterPoint [36]	*L*	×	×	51.4 / 47.9	17.6 / 17.6	47.7 / 41.4	0.19 / 0.18	10.7 / 10.6
SimTrack [19]	*L*	×	×	53.1 / 49.6	17.4 / 17.4	35.5 / 39.8	0.20 / 0.19	11.2 / 10.5
3DMODT	*L*	×	×	**55.9 / 51.2**	**18.9 / 18.9**	**33.1 / 36.7**	**0.13 / 0.12**	**10.3 / 10.2**

KITTI Cars Validation Set

Method	Input Data	External Detector	Complex Post-Processing	*sAMOTA*↑	*AMOTA*↑	*AMOTP*↑	*MOTA*↑	*MOTP*↑
mmMOT [6]	*L* + *I*	✓	✓	70.61	33.08	72.45	74.07	78.16
FANTrack [47]	*L* + *I*	✓	×	82.97	40.03	75.01	74.30	75.24
AB3DMOT [17]	*L*	✓	✓	93.28	45.43	77.41	86.24	78.43
PC-TCNN [48]	*L*	×	×	95.44	47.64	–	–	–
3D DetecTrack [49]	*L* + *I*	×	×	96.49	48.87	81.56	91.46	82.24
3DMODT	*L*	×	×	**96.91**	**49.01**	**82.14**	**92.33**	**83.52**

Access:
https://tinyurl.com/2p8bv25u