Adversarial Tracking: Attacking 3D Multi-Object Tracking in Real Time

Anonymous CVPR submission

Paper ID *****

Abstract

We plan to focus on attacking 3D Multi-Object Tracking (MOT) models for Vision-Based vehicle tracking in real time, without using 3D Lidar Data.

1. Introduction

The first tracking system used radar and sonar to track objects for military applications, and the algorithm consists of four parts: data association, state update, state prediction, and track management. Later, the employment of deep learning yields significant performance increases during data association for visual tracking.

Before the wide adoption of deep learning methods, it is popular to solve the visual tracking problem using low-level features and statistical learning techniques. In [16], Vo et al. reviewed three major techniques used to find the correspondence between detections and tracklet: The Joint Probabilistic Data Association Filter (JPDAF), Multi Hypothesis Tracking (MHT), and Random Finite Sets (RFS). They also introduced several nonlinear filtering methods: Bayesian Estimation, the Kalman Filter, the Particle Filter, and the Gaussian Sum Filter.

As research in deep learning advances, Krebs et al. reviewed the application of deep neural networks for object tracking: feature extraction, data association, and end-to-end tracking [8]. Most recent research focuses on tracking pedestrians, and Tracking-By-Detection (TBD) has become the most popular multi-object tracking framework. In [15], Sun et al. thoroughly analyzed the current developments in TBD-based MOT algorithms that includes four major components: localization, feature extraction, data association, and tack management.

There is also a growing trend of employing Joint Detection and Tracking (JDT), which is also described as end-to-end tracking in some literature [14]. More recently, a review on deep-learning-based visual multi-object tracking algorithms also includes Transformer-based methods [6].

1.1. Object Tracking

Recent research interests are shifting from Single Object Tracking (SOT) to Multi-Object Tracking (MOT). For MOT, there are three most popular frameworks:

- Tracking by Detection (TBD): A modular framework that heavily relies on the accuracy of the detector, such as SiamRPN++ [9] for SOT.
- Joint Detection and Tracking (JDT): The end-to-end tracking model is more efficient.
- Transformer: More accurate, but computationally expensive [13] for 2D MOT [12].

For autonomous driving, it is important to estimate vehicles' pose via 3D Object Tracking, and we plan to focus on vision-only multi-object trackers:

- Multi-Modality Tracker: Associate detection results with 3D Lidar data [18].
- Monocular Tracker: Vision-only methods [20] [7].

In [4], Ciaparrone et al. listed the most popular evaluation dataset (MOTChallenge, KITTI, nuScenes) and evaluation metrics (MOTA, MOTP, MT, ML, FP, FN).

1.2. Adversarial Attacks

The first adversarial attack against MOT attacks a TBD method that uses YOLOv3 as the object detection model, the Hungarian matching for the data association, and the Kalman filter for noisy measurement.

Most following work attacks SiameseRPN-based trackers for SOD [21] [10] [5] [1] [22]. Another research attacks the GOTURN model, which is an efficient end-to-end SOD model [19]. We only find one research paper that attacks FairMOT (JDT) [11]. There are many adversarial attacks against 3D Lidar Point Cloud [2] [3] [17].

1.3. Summary

In summary, we find several research papers that attack 2D SOT (caemera) and 3D MOT (Lidar). But we do not find attacks against 3D monocular trackers [20] [7]. Neither do we find research that attacks Transformers for 2D MOT.

References

[1] Xuesong Chen, Xiyu Yan, Feng Zheng, Yong Jiang, Shu-Tao Xia, Yong Zhao, and Rongrong Ji. One-shot adversarial attacks on visual tracking with dual attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10176–10185, 2020. 1

[2] Riran Cheng, Nan Sang, Yinyuan Zhou, and Xupeng Wang. Universal adversarial attack against 3d object tracking. In *2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, pages 34–40. IEEE, 2021. 1

[3] Riran Cheng, Nan Sang, Yinyuan Zhou, and Xupeng Wang. Non-rigid transformation based adversarial attack against 3d object tracking. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2744–2748. IEEE, 2022. 1

[4] Gioele Ciaparrone, Francisco Luque Sánchez, Siham Tabik, Luigi Troiano, Roberto Tagliaferri, and Francisco Herrera. Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381:61–88, 2020. 1

[5] Qing Guo, Xiaofei Xie, Felix Juefei-Xu, Lei Ma, Zhongguo Li, Wanli Xue, Wei Feng, and Yang Liu. Spark: Spatial-aware online incremental attack against visual tracking. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 202–219. Springer, 2020. 1

[6] Shuman Guo, Shichang Wang, Zhenzhong Yang, Lijun Wang, Huawei Zhang, Pengyan Guo, Yuguo Gao, and Junkai Guo. A review of deep learning-based visual multi-object tracking algorithms for autonomous driving. *Applied Sciences*, 12(21):10741, 2022. 1

[7] Hou-Ning Hu, Yung-Hsu Yang, Tobias Fischer, Trevor Darrell, Fisher Yu, and Min Sun. Monocular quasi-dense 3d object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1992–2008, 2022. 1

[8] Sebastian Krebs, Bharanidhar Duraisamy, and Fabian Flohr. A survey on leveraging deep neural networks for object tracking. In *2017 IEEE 20th international conference on Intelligent Transportation Systems (ITSC)*, pages 411–418. IEEE, 2017. 1

[9] Bo Li, Wei Wu, Qiang Wang, Fangyi Zhang, Junliang Xing, and Junjie Yan. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4282–4291, 2019. 1

[10] Siyuan Liang, Xingxing Wei, Siyuan Yao, and Xiaochun Cao. Efficient adversarial attacks for visual object tracking. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16*, pages 34–50. Springer, 2020. 1

[11] Delv Lin, Qi Chen, Chengyu Zhou, and Kun He. Trasw: Tracklet-switch adversarial attacks against multi-object tracking. *arXiv preprint arXiv:2111.08954*, 2021. 1

[12] Liting Lin, Heng Fan, Yong Xu, and Haibin Ling. Swintrack: A simple and strong baseline for transformer tracking. *arXiv preprint arXiv:2112.00995*, 2021. 1

[13] Tim Meinhardt, Alexander Kirillov, Laura Leal-Taixe, and Christoph Feichtenhofer. Trackformer: Multi-object tracking with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8844–8854, 2022. 1

[14] Sankar K Pal, Anima Pramanik, Jhareswar Maiti, and Pabitra Mitra. Deep learning in multi-object detection and tracking: state of the art. *Applied Intelligence*, 51:6400–6429, 2021. 1

[15] Zhihong Sun, Jun Chen, Liang Chao, Weijian Ruan, and Mithun Mukherjee. A survey of multiple pedestrian tracking based on tracking-by-detection framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5):1819–1833, 2020. 1

[16] Ba-ngu Vo, Mahendra Mallick, Yaakov Bar-Shalom, Stefano Coraluppi, Ronald Mahler, and Ba-tuong Vo. Multitarget tracking. 2015. 1

[17] Zhengyi Wang, Xupeng Wang, Ferdous Sohel, Mohammed Bennamoun, Yong Liao, and Jiali Yu. Adversary distillation for one-shot attacks on 3d target tracking. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2749–2453. IEEE, 2022. 1

[18] Xinshuo Weng, Jianren Wang, David Held, and Kris Kitani. 3d multi-object tracking: A baseline and new evaluation metrics. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10359–10366. IEEE, 2020. 1

[19] Rey Reza Wiyatno and Anqi Xu. Physical adversarial textures that fool visual object tracking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4822–4831, 2019. 1

[20] Jialian Wu, Jiale Cao, Liangchen Song, Yu Wang, Ming Yang, and Junsong Yuan. Track to detect and segment: An online multi-object tracker. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12352–12361, 2021. 1

[21] Xugang Wu, Xiaoping Wang, Xu Zhou, and Songlei Jian. Sta: Adversarial attacks on siamese trackers. *arXiv preprint arXiv:1909.03413*, 2019. 1

[22] Bin Yan, Dong Wang, Huchuan Lu, and Xiaoyun Yang. Cooling-shrinking attack: Blinding the tracker with imperceptible noises. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 990–999, 2020. 1