NON-RIGID TRANSFORMATION BASED ADVERSARIAL ATTACK AGAINST 3D OBJECT TRACKING

*Riran Cheng, Nan Sang, Yinyuan Zhou, Xupeng Wang**

University of Electronic Science and Technology of China
School of Information and Software Engineering
No.4, Section 2, North Jianshe Road

ABSTRACT

It is well-recognized that 3D visual tasks based on deep neural networks are vulnerable to adversarial attacks. Existing methods to generate adversarial examples are mainly developed from injecting imperceptible perturbations into the inputs. However, aggressive characteristic of geometric transformations, which are common in 3D objects, are rarely investigated. In this paper, we propose the non-rigid transformation based adversarial attack method against 3D object tracking. The adversarial example is generated by deforming parts of the tracking template, leading to deviation of the tracking predictions from the ground truth. Specifically, a clustering-based region segmentation module is designed to divide the tracking template into local regions. Furthermore, an objective function, which combines IoU loss, confidence loss and distance loss, is leveraged to update the poses of local regions. Experiments conducted on an efficient 3D tracker demonstrate that 3D trackers are extremely vulnerable to non-rigid deformation.

Index Terms— 3D object tracking, adversarial attack,non-rigid transformation, region segmentation

1. INTRODUCTION

3D object tracking plays an important role in real-world applications, such as autonomous driving [1, 2, 3, 4]. The introduction of deep neural networks prompts 3D object tracking to achieve superior performance, which however has the disadvantage of vulnerability to adversarial attacks [5, 6, 7, 8]. Considering that security is regarded as a top priority by safety-critical tasks, adversarial attack against deep 3D trackers emerges as a great real-world concern. This allows to figure out vulnerabilities of existing deep neural networks, leading to the improvement on their performance.

In recent years, adversarial attacks against 3D classifiers have been widely researched. The method to generate adversarial examples were proposed in [9]. By injecting imperceptible perturbation into the point cloud, the generated adversarial example leads the classifier to produce wrong predictions. In [10], statistical operations were leveraged to produce robust perturbations that make the adversarial examples aggressive. [11] first proposed an adversarial network named LG-GAN to craft flexible adversarial examples based on perturbation. In addition, adversarial attacks against 3D object detection were investigated in [12, 13].

Existing methods to generate adversarial examples are mostly based on injecting perturbations, which are extended from 2D images. However, geometric transformations of 3D objects are ignored. [14] first showed existing 3D classifier are vulnerable to rigid deformations. A novel block-box attack framework based on Thompson Sampling was designed to generate adversarial examples by rotating 3D samples. Up to now, the robustness of 3D trackers under non-rigid deformation has not been evaluated.

In this paper, we propose a non-rigid transformation [15] based adversarial attack method against 3D object tracking. To achieve non-rigid transformation of point cloud, we adopt the strategy of rotating local regions of the tracking template. Hence, a region segmentation module based on clustering is developed to divide tracking template into local regions, following which part of them are selected to rotate. An objective function, which consists of IoU loss, confidence loss and distance loss, is utilized to update the rotation angles of selected regions. IoU loss leads the 3D tracker to generate deviated target proposals. Confidence loss is designed to make the corresponding target scores become fuzzy. Distance loss constrains the non-rigid deformation with human imperception. An adversarial example is produced by aggregating all local regions, which is able to confuse 3D tracker.

In summary, the main contributions of this paper are listed as follows:

- We propose a non-rigid transformation based adversarial attack method against 3D trackers.
- The non-rigid transformation based adversarial attack is achieved from rotating local regions of the tracking template, which is implemented by a region segmentation module.
- An objective function is designed to update the poses of local regions, which are further aggregated to generate

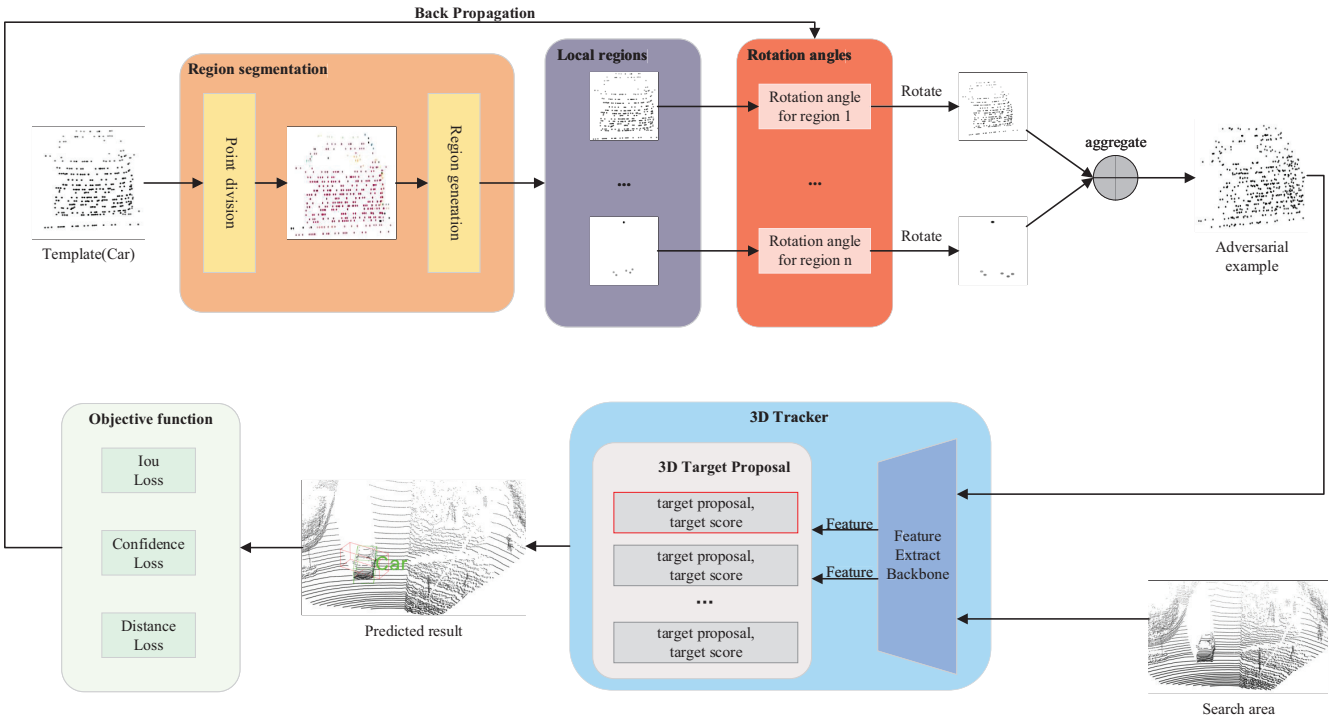

Fig. 1. Illustration of the proposed non-rigid transformation based adversarial attack method. Given the tracking template, region segmentation module divides the template into local regions, from which several of them are selected to rotate. Afterwards, all local regions are aggregated to generate the adversarial example. The 3D tracker is led to create the fuzzy target scores and the offset target proposal by the adversarial example so that the predicted result (red box) deviates from the ground truth (green box). In addition, the poses of selected regions are updated by optimizing the objective function.

adversarial examples.

- Experiments on popular 3D object tracking models show a superior attack performance of our method.

2. METHOD

Given tracking template P_{tmp} and corresponding search area P_{sea}, a 3D tracker T takes them as input to predict the position of the tracking object. Denote $R(P_{tmp}, \omega)$ as the adversarial example generated by rotating local regions of P_{tmp}.The goal of our method is to obtain the proper rotation angles ω to generate adversarial example, which declines the performance of T and keeps visual imperceptible at the same time.

The pipeline of non-rigid transformation based adversarial attack method is illustrated in Fig.1. The afferent template are first divided into local regions by region segmentation module. Local regions are selected for rotation and the rotation angels are updated by the objective function. The objective function consists of IoU loss, confidence loss and distance loss. Specifically, IoU loss leads the 3D tracker to generate deviated target proposals, with their corresponding target scores being fuzzy due to the confidence loss. The combination of IoU loss and confidence loss renders the tracking results deviated from the ground truth. Distance loss is used to constrain the rotation angles to ensure visual imperceptibility of the modification. An adversarial example is generated by aggregating all local regions, which can tamper the output of 3D target proposal leading to a degraded performance of 3D tracker.

2.1. Region Segmentation

For the purpose of non-rigid deformation of the point cloud, we adopt the strategy of rotating local regions of tracking template. Hence, Region segmentation module, which consists of point division and region generation, is proposed to divide the template into local regions. Clustering-based point division traverses each point in the template, dividing them into different regions and independent points. Afterwards, region generation divides the tracking template into local regions according to the results of point division, while all the independent points are gathered to generate a special region. In order to maintain attack performance when the 3D tracker has defensive method such as eliminating independent points, we remove the special region from the range of rotation available regions. Finally, region generation produces a set of local re-

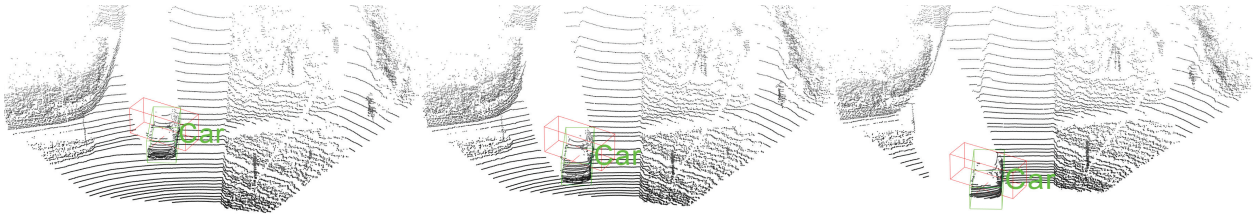

Fig. 2. Visualization of the attack performance.

gions ordered by the size of the points in the region.

2.2. Attack Details

A 3D tracker generates target proposals associated with corresponding target scores. Target proposals contains potential location of the object, while corresponding target scores represent the probability of the tracked target. The 3D tracker chooses the target proposal with the highest target score as the predicted result. This indicates that both offsetting target proposals and tampering the target scores are able to attack 3D tracker. Hence, confidence attack aimed at target scores and IoU attack aimed at target proposals are proposed.

For IoU attack, we calculate IoU scores between all target proposals generated by the 3D tracker and the ground truth. By degrading these scores, a 3D tracker is led to generate offset target proposals. Compared to the task of classification, tracking has temporal coherence [16]. Hence, temporal motions among sequential frames of the target object are considered. Extra IoU scores, which represent temporal motion of previous frame are computed to enhance attack effect. IoU loss function is defined as follow:

$$\mathcal{L}_{IoU} = -\sum_{i=0}^{n} \left(IoU\left(G_t, p_t^i\right) + IoU\left(G_{t-1}, p_t^i\right)\right) log(1-s_t^i) \quad (1)$$

G_t, G_{t-1} represent the ground truth at frame t and previous frame. p_t^i, s_t^i denote i^{th} target proposal at frame t and corresponding target score.

For confidence attack, the difference in target scores between high probability target proposals and low probability target proposals is reduced. Therefore, the 3D tracker is confused to generate fuzzy target scores. Target proposals with similar target scores are also located similarly in the search area. Therefore, an attack against a single target score may not be significantly effective [17]. In this case, we change multiple target scores to improve the attack performance. Confidence loss is defined as follow:

$$\mathcal{L}_{confidence} = \sum_{i=0}^{q} S_i - \sum_{i=p}^{r} S_i \quad (2)$$

S_i represents i^{th} target score of the ranked target scores and p, q, r denote index of target scores.

In order to make the attack visually imperceptible, L_2 distance is adopted as metric to limit the modification. Distance loss is defined as follow:

$$\mathcal{L}_{distance} = \left(\sum_i \left(x_i - y_i\right)^2\right)^{\frac{1}{2}} \quad (3)$$

x_i, y_i denote points of tracking template and adversarial example, respectively.

To sum up, the rotation angles are updated by optimizing objective function $\mathcal{L}$, which is formulated as follows:

$$\mathcal{L} = \alpha\mathcal{L}_{IoU} + \beta\mathcal{L}_{confidence} + \gamma\mathcal{L}_{distance} \quad (4)$$

α, β, γ are parameters to balance these loss functions. By minimizing $\mathcal{L}$, an adversarial example is generated, which leads to the deviation of the predictions from the ground truth.

3. EXPERIMENT

In this section, a series of experiments on KITTI[18] benchmark were conducted to verify the effectiveness of the proposed attack method. The P2B[19] was adopted as the victim 3D tracker.

3.1. Dataset

The KITTI[18] dataset, which consists of 3D lidar data and 2D images, is designed to evaluate object detection and object tracking. The KITTI dataset contains 50 video sequences, 21 of which are used for training and the rest are used for testing. Since the ground truth of test dataset is inaccessible offline, only 19-20 scenes of the training dataset were used for our experiment, following [19].

3.2. Implementation Details

For the rotation in X-Y plane, we iterated to update the rotation angles until the maximum number of iterations was reached. The maximum number of iterations was set to 60. The parameters of the objective function were set as α = 10, β = 1 and γ = 1, which achieves the best performance in our experiment. In addition, we adopted DBSCAN[20] as clustering algorithm, which is used for region segmentation. Local regions with the maximum number of points and the least

points were selected to rotated. A number of search area generation methods were adopted by the P2B tracker, resulting in distinct performances. We chose the method with the best performance throughout the experiment.

3.3. Evaluation Metrics

Success and Precision are used to measure the tracking performance of 3D tracker. "Success" denotes the proportion of scenarios with value larger than 0.6. "Precision" denotes the proportion of scenarios with error smaller than 0.8, and error represents the distance between the center of the ground truth and the predicted results.

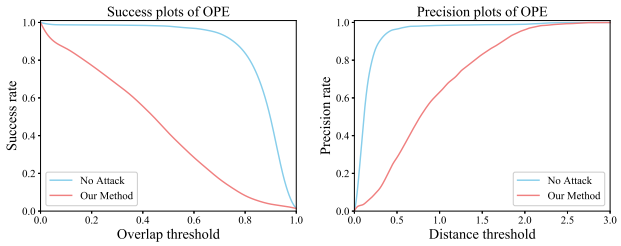

Fig. 3. Evaluation results of the tracker with or without adversarial attack.

3.4. Quantitative Study

To verify the effectiveness of the non-rigid transformation based adversarial attack, the performance of attack was evaluated by comparing with benign 3D tracker. Comparison results are shown in Table 1. Our method reduced Success and Precision significantly by 42.9% and 36%, indicating that our attack method can invalidate the 3D tracker.

Extra evaluation results can be learned from Fig. 3. In success plots, our method reduce the Success by 76% when overlap threshold is 0.8. In precision plots, our method reduced the precision by up to 70%. These results mean that more predictions were far away the ground truth under the adversarial attack.

To evaluate the contribution of non-rigid transformation, a baseline method was designed for comparison, which generates adversarial examples by adding perturbations. To be specific, the perturbations are generated by the object function proposed by our method. The comparison results are shown in Table 2. In the case of perturbation, the success and precision of the tracker are reduced by 21.7% and 17.1%

Table 1. The tracking performance with and without adversarial attack.

Method	Success(%)	Precision(%)
Original P2B	86.6	91.7
Our method	**43.7**	**55.7**

Table 2. Comparison of the performance using different attack method.

Attack Method	Success(%)	Precision(%)
Non-rigid transformation	43.7	55.7
Perturbation	64.9	74.6

Table 3. Comparison of performance using different objective functions to update rotation angles.

Objective Function	Success(%)	Precision(%)
Our default setting	43.7	55.7
Without IoU loss	51.7	65.4
Without confidence loss	49.6	61.3

respectively. Hence, Our method outperforms the baseline method in success and precision by 21.2% and 18.9% respectively, which proves non-rigid transformation is a more effective strategy of adversarial attack than adding perturbations.

The objective function consists of IoU loss, confidence loss and distance loss. Distance loss is used to limit modification caused by the non-rigid transformation, hence an ablation study was performed to evaluate the contribution of IoU loss and confidence loss, which are designed for attacking the 3D tracker. We compared the baseline objective function with objective function under situations as: without IoU loss and without confidence loss.Results are shown in Table 3. The success and precision are reduced by 34.9% and 26.3% respectively without IoU loss, and by 37% and 30.4% respectively without confidence loss. In the case of our default setting, the success and precision are further reduced to the best results. The results show that the combination of IoU loss and confidence loss can further enhance the attack effectiveness.

3.5. Qualitative Study

Attack performance on some frames of a video sequence are demonstrated in Fig. 2, which shows that the tracking results(red box) of the 3D tracker attacked by our method were significant deviated from the ground truth(green box), which proves the effectiveness of the proposed non-rigid deformation based adversarial attack.

4. CONCLUSION

In this paper, we propose a non-rigid deformation based adversarial attack method against 3D object tracking, which generates adversarial examples by rotating local regions of the tracking template. A clustering-based region segmentation module is proposed to divide the template into local regions. An objective optimization function, which combines IoU loss, confidence loss, and distance loss, are leveraged to update the rotation angles. The results of experiment prove the 3D trackers are vulnerable to non-rigid deformation.

5. REFERENCES

[1] Andrew I Comport, Éric Marchand, and François Chaumette, "Robust model-based tracking for robot vision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2004, vol. 1, pp. 692–697.

[2] Wenjie Luo, Bin Yang, and Raquel Urtasun, "Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 3569–3577.

[3] Eiji Machida, Meifen Cao, Toshiyuki Murao, and Hiroshi Hashimoto, "Human motion tracking of mobile robot with kinect 3d sensor," in *Proceedings of SICE Annual Conference*. IEEE, 2012, pp. 2207–2211.

[4] Ankith Manjunath, Ying Liu, Bernardo Henriques, and Armin Engstle, "Radar based object detection and tracking for autonomous driving," in *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility*. IEEE, 2018, pp. 1–4.

[5] Naveed Akhtar and Ajmal Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *Ieee Access*, vol. 6, pp. 14410–14430, 2018.

[6] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[7] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard, "Deepfool: a simple and accurate method to fool deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2574–2582.

[8] Wei Jiang, Xiangyu Wen, Jinyu Zhan, Xupeng Wang, and Ziwei Song, "Interpretability-guided defense against backdoor attacks to deep neural networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2021.

[9] Chong Xiang, Charles R Qi, and Bo Li, "Generating 3d adversarial point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9136–9144.

[10] Chengcheng Ma, Weiliang Meng, Baoyuan Wu, Shibiao Xu, and Xiaopeng Zhang, "Towards effective adversarial attack against 3d point cloud classification," in *IEEE International Conference on Multimedia and Expo*. IEEE, 2021, pp. 1–6.

[11] Hang Zhou, Dongdong Chen, Jing Liao, Kejiang Chen, Xiaoyi Dong, Kunlin Liu, Weiming Zhang, Gang Hua, and Nenghai Yu, "Lg-gan: Label guided adversarial network for flexible targeted attack of point cloud based deep networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10356–10365.

[12] Chang Chen and Teng Huang, "Camdar-adv: Generating adversarial patches on 3d object," *International Journal of Intelligent Systems*, vol. 36, no. 3, pp. 1441–1453, 2021.

[13] Xupeng Wang, Mumuxin Cai, Ferdous Sohel, Nan Sang, and Zhengwei Chang, "Adversarial point cloud perturbations against 3d object detection in autonomous driving systems," *Neurocomputing*, 2021.

[14] Yue Zhao, Yuwei Wu, Caihua Chen, and Andrew Lim, "On isometry robustness of deep 3d point cloud models under adversarial attacks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1201–1210.

[15] Xupeng Wang, Ferdous Sohel, Mohammed Bennamoun, Yulan Guo, and Hang Lei, "Scale space clustering evolution for salient region detection on 3d deformable shapes," *Pattern Recognition*, vol. 71, pp. 414–427, 2017.

[16] Shuai Jia, Yibing Song, Chao Ma, and Xiaokang Yang, "Iou attack: Towards temporally coherent black-box adversarial attack for visual object tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6709–6718.

[17] 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 *Proc. IEEE conference on Computer Vision and Pattern Recognition*, 2020, pp. 10176–10185.

[18] Andreas Geiger, Philip Lenz, and Raquel Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.

[19] Haozhe Qi, Chen Feng, Zhiguo Cao, Feng Zhao, and Yang Xiao, "P2b: Point-to-box network for 3d object tracking in point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6329–6338.

[20] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *kdd*, 1996, vol. 96, pp. 226–231.