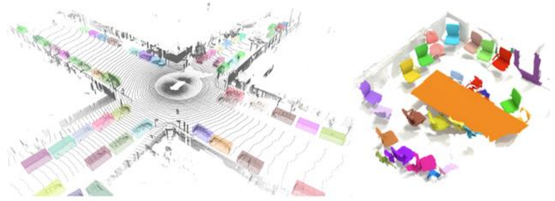
Semi-supervised 3D Object Detection via Temporal Graph Neural Networks

Jianren Wang[1], Haiming Gang[2], Siddharth Ancha[1], Yi-Ting Chen[3], David Held[1]
[1]Carnegie Mellon University, [2]Honda Research Institute, [3]National Chiao Tung University

3DV 2021 International Conference on 3D Vision Online, December 1-3, 2021

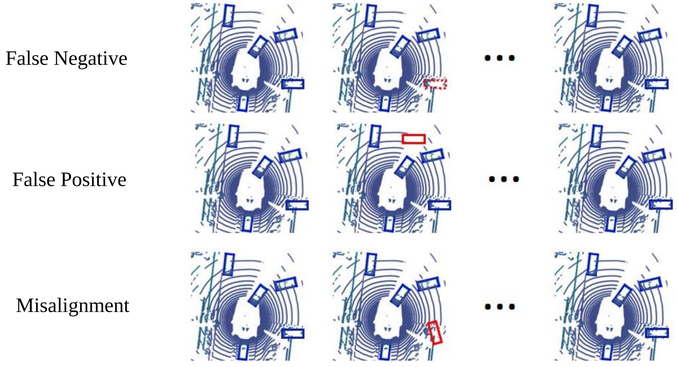
Problem

- 3D object detection is an essential and fundamental problem in autonomous driving.

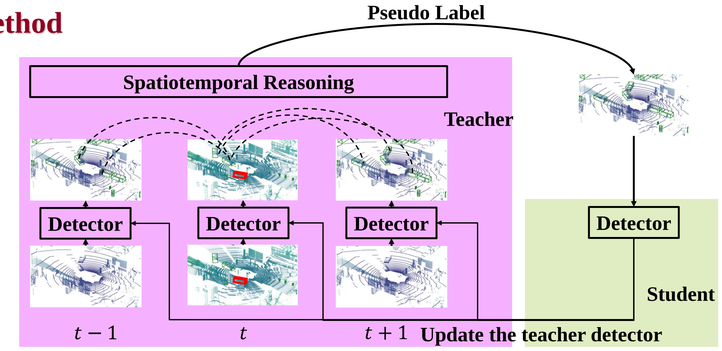


Motivation

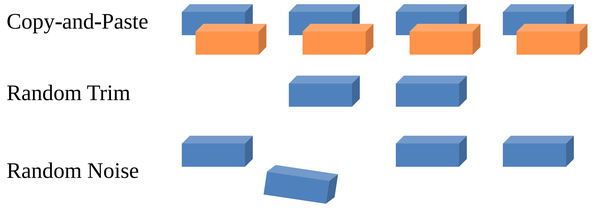
- 3D object detection is an essential and fundamental problem in autonomous driving.
- However, most of the existing approaches are strongly supervised and require the availability of a large amount of well-annotated 3D data.
- On the other hand, in most applications, point clouds are recorded over time as a data stream. A point cloud video contains richer spatio temporal information than a single frame.
- In this work, we propose to use this information to improve a single frame 3D object detector through semi-supervised training.



Method



Teacher Network: Graph Augmentation



Uncertainty-weighting

Uncertainty-weighting Loss

$$u = -s\log(s) - (1-s)\log(1-s)$$

$$l = \begin{cases} -(1-u)^k \log(p), & \text{if } s \geq 0.5 \\ -(1-u)^k \log(1-p), & \text{if } s < 0.5 \end{cases}$$

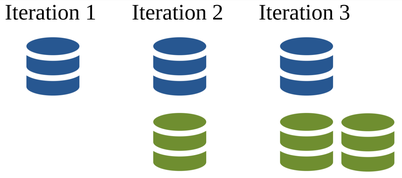
u: uncertainty

s: calibrated teacher's prediction

p: student's prediction

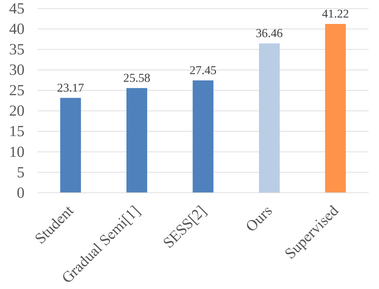
k: focusing parameter

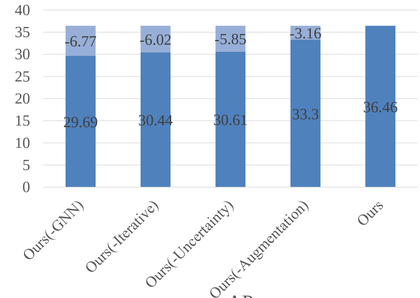
Gradually Semi-supervised Training

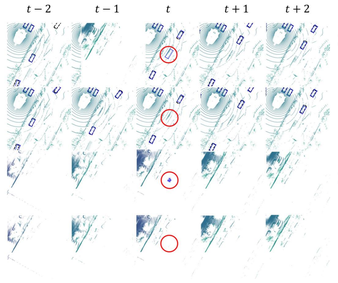


Experiments

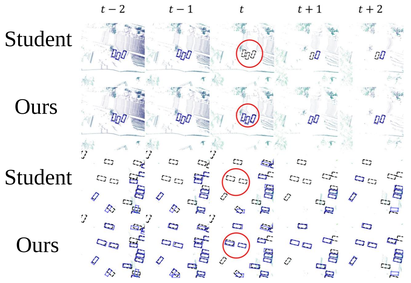
- 3D object detector pre-trained with 50 scenes on nuScenes train
- 500 scenes from nuScenes train for Semi-Supervised Training
- 150 scenes from nuScenes validation as test set







Correct False Negative



Correct False Negative

Conclusion

- We propose to leverage unlabeled point cloud videos by semi-supervised learning
- Our method incorporates uncertainty-aware semi-supervised training with a GNN for spatiotemporal reasoning
- Our method achieves state-of-the-art detection performance on the challenging nuScenes and H3D benchmarks
- Our method removes the need for excessive efforts in data annotation

[1] Teichman, Alex et al. "Tracking-based semi-supervised learning.", IJRR2012
[2] Zhao, Na et al. "Sess: Self-ensembling semi-supervised 3d object detection.", CVPR2020