Attentional PointNet for 3D-Object Detection in Point Clouds

Anshul Paigwar, Özgür Erkent, Christian Wolf, Christian Laugier

Attentional PointNet for 3D-Object Detection in Point Clouds

Anshul Paigwar, Ozgur Erkent, Christian Wolf, Christian Laugier
Inria – Team Chroma, Univ. Grenoble Alpes, INSA Lyon
Email: {first_name}.{second_name}@inria.fr

Inria inventeurs du monde numérique

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

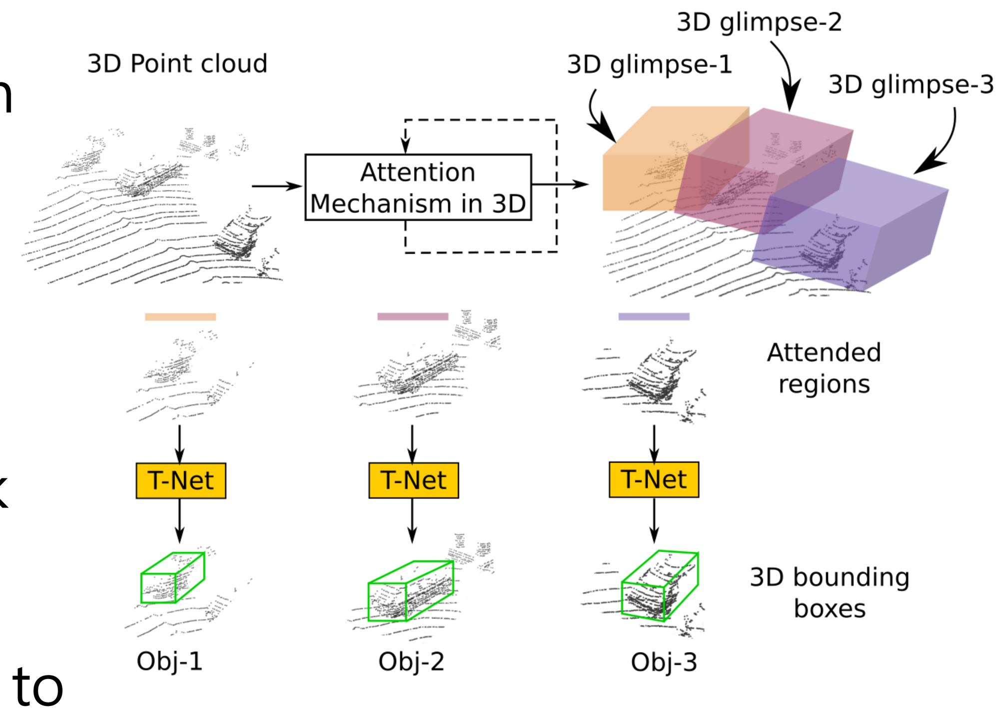
Introduction

Accurate detection of objects in 3D point clouds is a central problem for autonomous navigation. Most existing methods require data from multiple sensors for the detections. Such methods are prone to sensor failure.

We propose a novel deep architecture called Attentional PointNet for 3D object detection. The network directly operates on sparse 3D points.

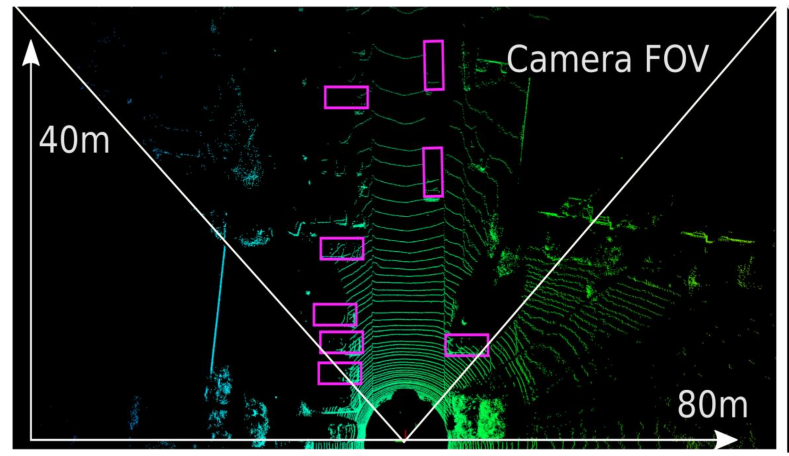
Attention- Mechanism in 3D

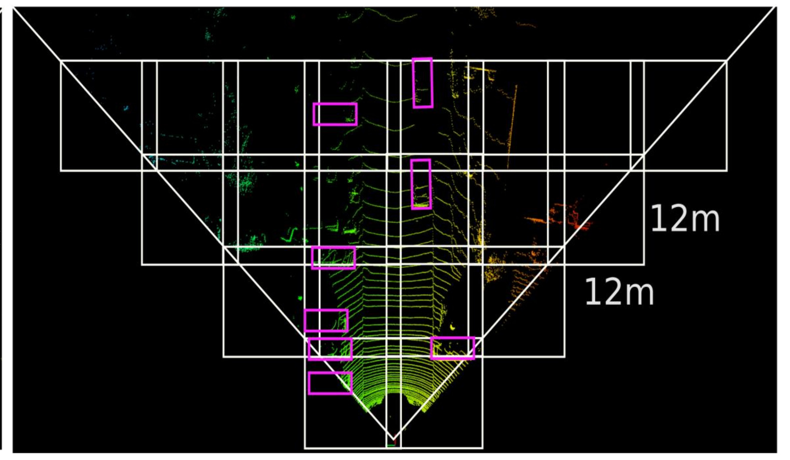
- We extend visual attention mechanisms to 3D point clouds for multiple object detection.
- Given a cluttered environment, The network learns to attend to the objects of interest, thus reducing the data needed to be processed.



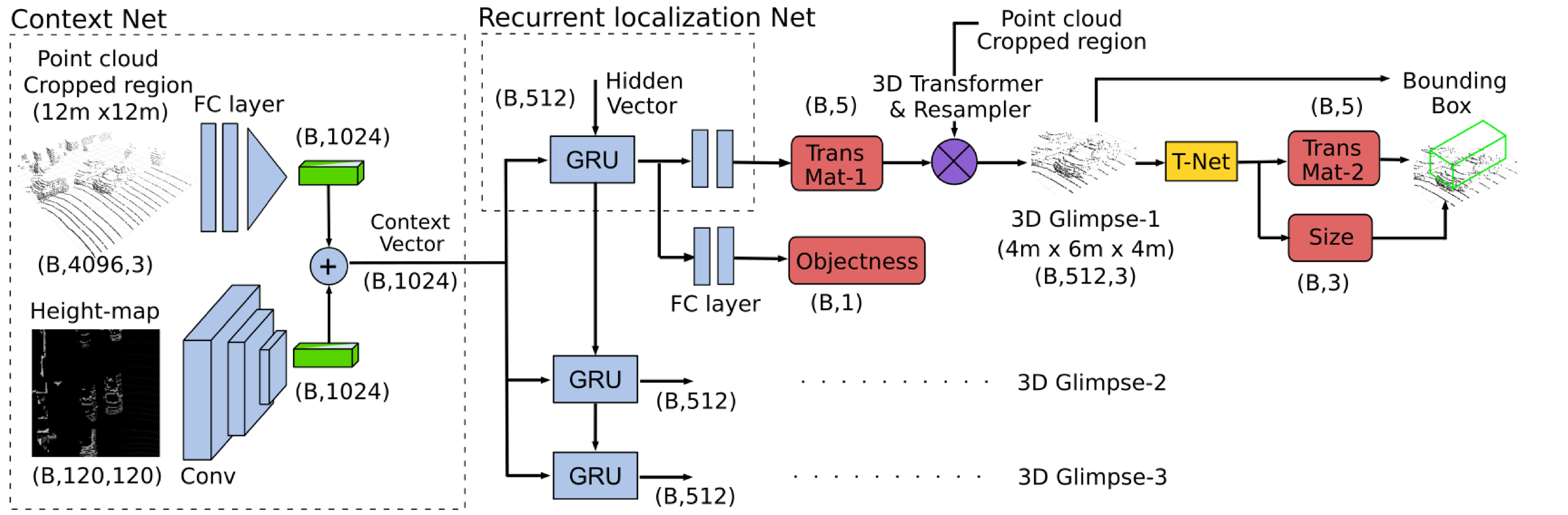
Data Augmentation

- We train the model on a custom KITTI dataset.
- We subdivide the FOV area from each scan into equally spaced cropped regions of 12m×12m with an overlap of 1m.
- Each cropped region of size 12m×12m is also converted into a grayscale image of size 120×120 pixels encoding height.





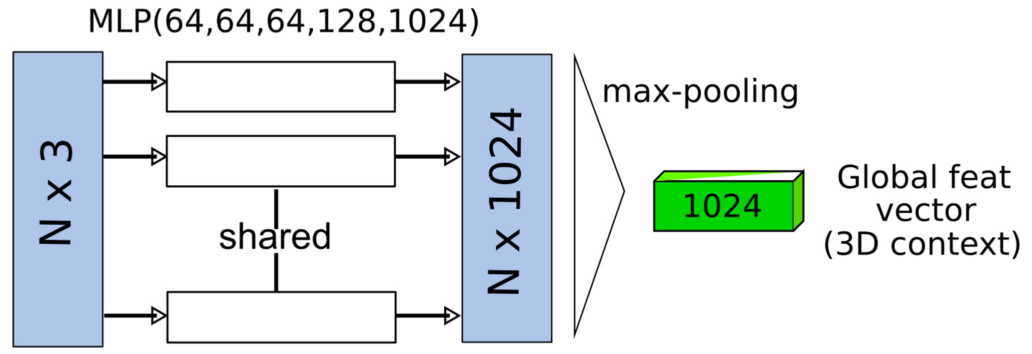
Architecture



Given the point cloud and the corresponding height map, network sequentially regress parameters of a 3D Transformation matrix representing pose of a fixed size 3D glimpse.

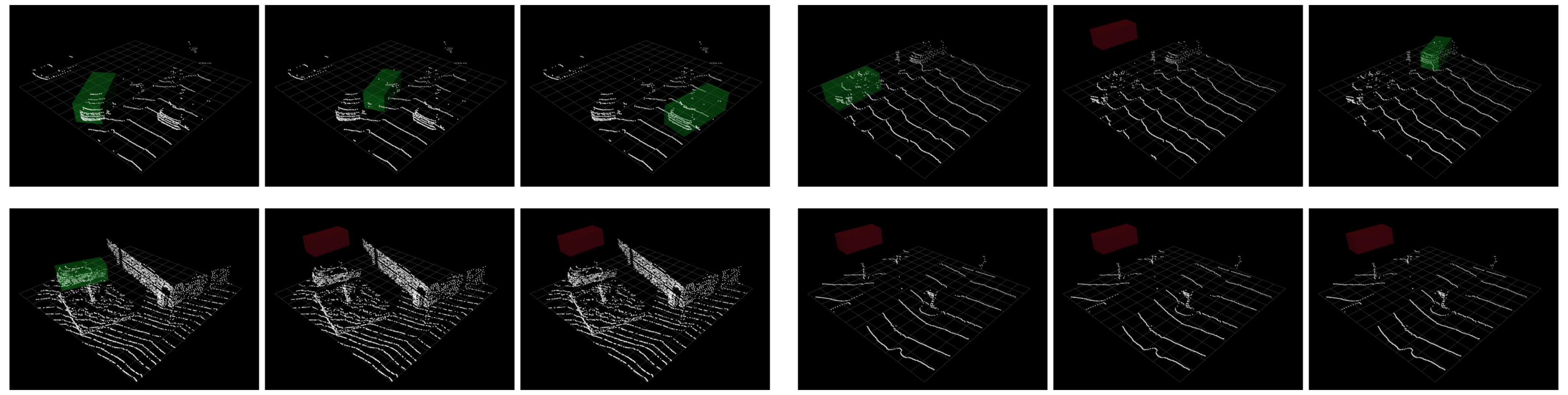
PointNet (T-Net)

A modified PointNet (T-Net) then estimates another 3D transformation matrix and size representing the 3D bounding box of the object inside the glimpse.



Sequential detection

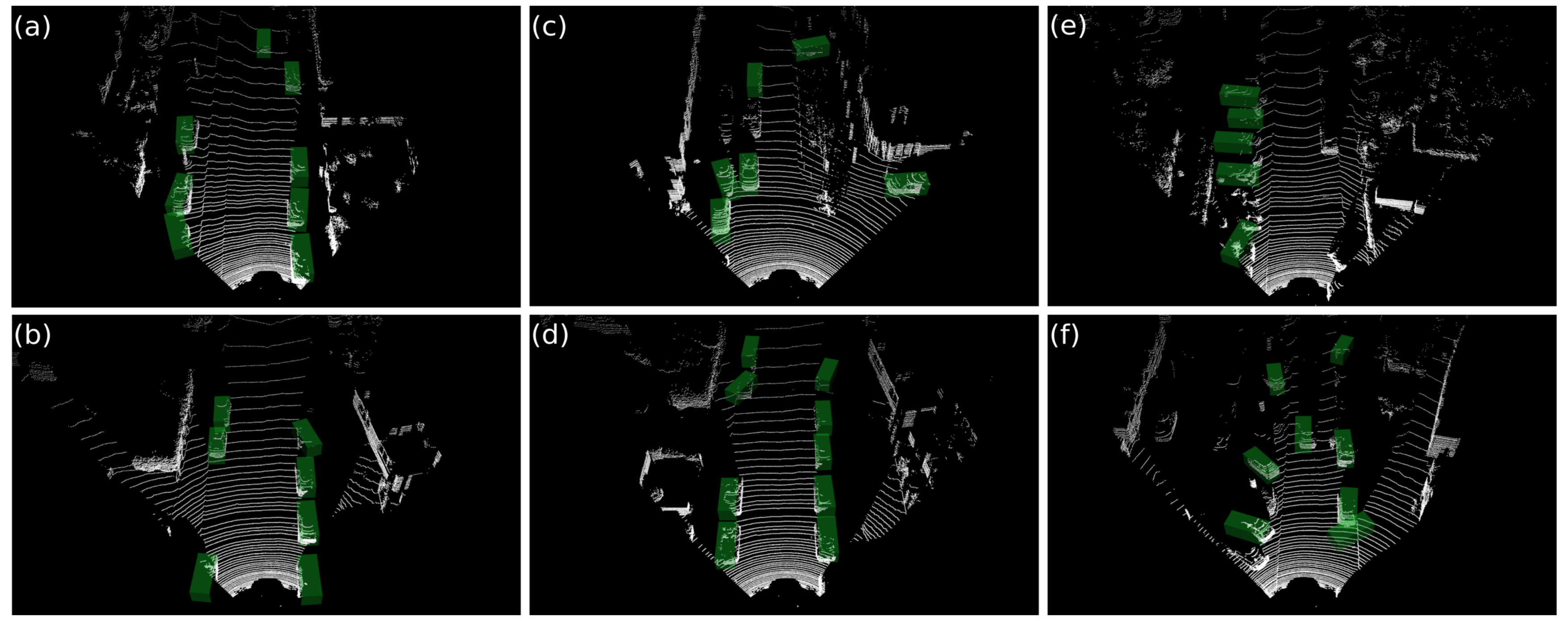
- For each cropped region network makes three predictions sequentially classifying and localizing the cars in the scene.
- When there are less than three cars in the scene, the network focuses outside the cropped region and appropriately classify them as negative detections.



Experiments & Results

Method	Modality	FPS	Car		
			Easy	Mod.	Hard
MV3D	Lidar+Mono	2.8	71.09	62.35	55.12
F-PointNet	Lidar+Mono	5.9	81.20	70.39	62.19
AVOD	Lidar+Mono	12.5	73.59	65.78	58.38
RoarNet	Lidar+Mono	10	83.95	75.79	67.88
VeloFCN	Lidar	-	15.20	13.66	15.98
RT3D	Lidar	11.23	23.49	21.27	19.81
VoxelNet	Lidar	4.3	67.27	52.87	46.62
Complex-YOLO	Lidar (BV)	16.6	55.63	49.44	44.13
A-PointNet (vanilla)	Lidar	12.5	49.47	44.64	41.71
Attentional-PointNet	Lidar	8.06	58.62	52.28	47.23

For car detection, Attentional PointNet achieves comparable AP of 52.28% among the architectures using LiDAR data only and surpasses many approaches in terms of inference time.



References

1. Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.

Contact

Code & Dataset

Acknowledgement: This work has been conducted within the ENABLE-S3 project that has received funding from the ECSEL joint undertaking under grant agreement NO 692455. The work was also partially funded by ANR grant Deepvision (ANR-15-CE23-0029, STPGP-479356-15).