Real-Time Semantic Segmentation of Sparse LIDAR Point Clouds using SqueezeSeg and Recurrent CRF

Ingrid Navarro Anaya, *ITESM*, Dr. Luis E. Navarro-Serment, *CMU*

Motivation

- The use of **Deep Learning** approaches for semantic segmentation of sparse LIDAR Point Clouds has **not** been fully explored.
  - High levels of sparsity in the data makes it difficult to interpret object structure
- Our goal is to develop a real-time semantic segmentation system for highly sparse 3D point clouds using a lightweight CNN called **SqueezeSeg** [1] with **Recurrent Conditional Random Fields (CRF).**

Approach

- Our focus is to address semantic segmentation in point clouds collected from LIDAR scans with **sparse vertical density.**
  - Classes of interest: **car**, **pedestrian**, **cyclist** and **ground.**
  - Sensor: **Velodyne VLP-16**
- Gathering training data is a labour intensive task. Thus, we leverage on the **KITTI dataset [2],** which
  - has labels for **car**, **pedestrian** and **cyclist** but **not** for **ground**.
  - contains LIDAR scans that have **high vertical density**
- To generate the **ground annotations**, we developed an automatic ground labeler that uses a **3D plane-fitting** technique [3] (Fig. 1).

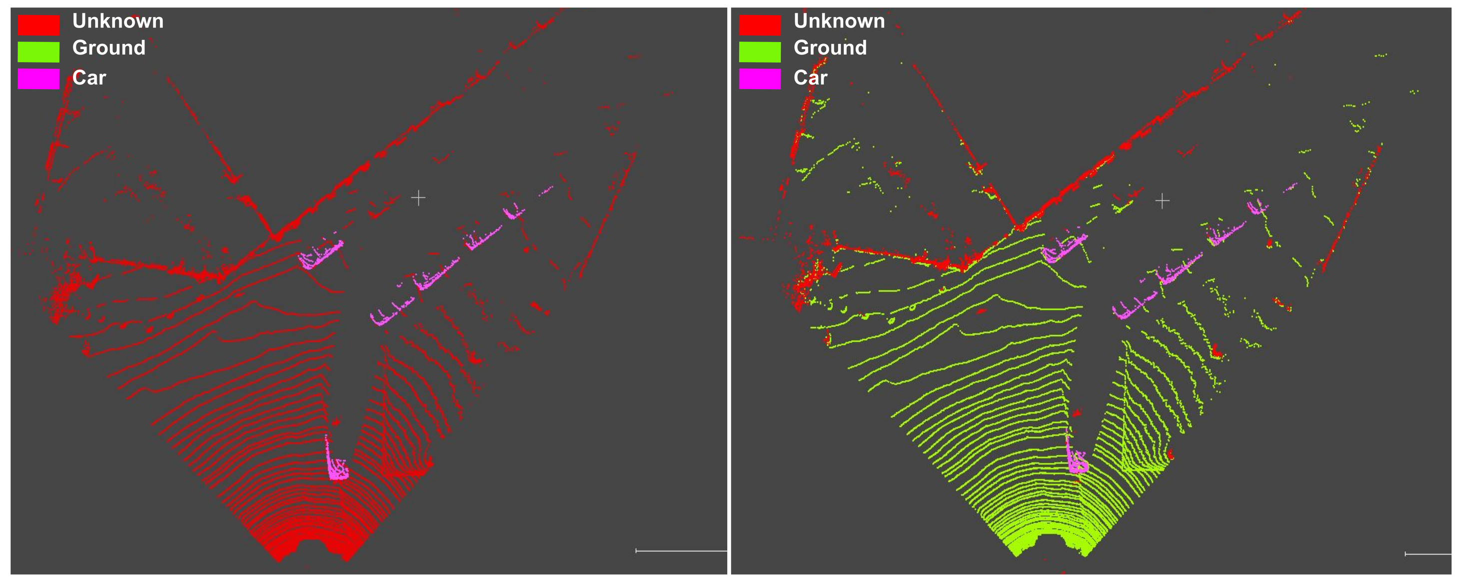

Fig.1: Original point cloud (left). Point cloud after estimating the ground (right).

- To obtain **sparse** point clouds from [2], we explored two methods: the **down-sampling** and the **up-sampling** method.
- **Down-sampling (Fig. 2):** reduces the vertical density of the point cloud by generating **16-ring** point clouds from **64-ring** scans.
  - Analyze the performance of the model if the network operates directly with sparse scans.

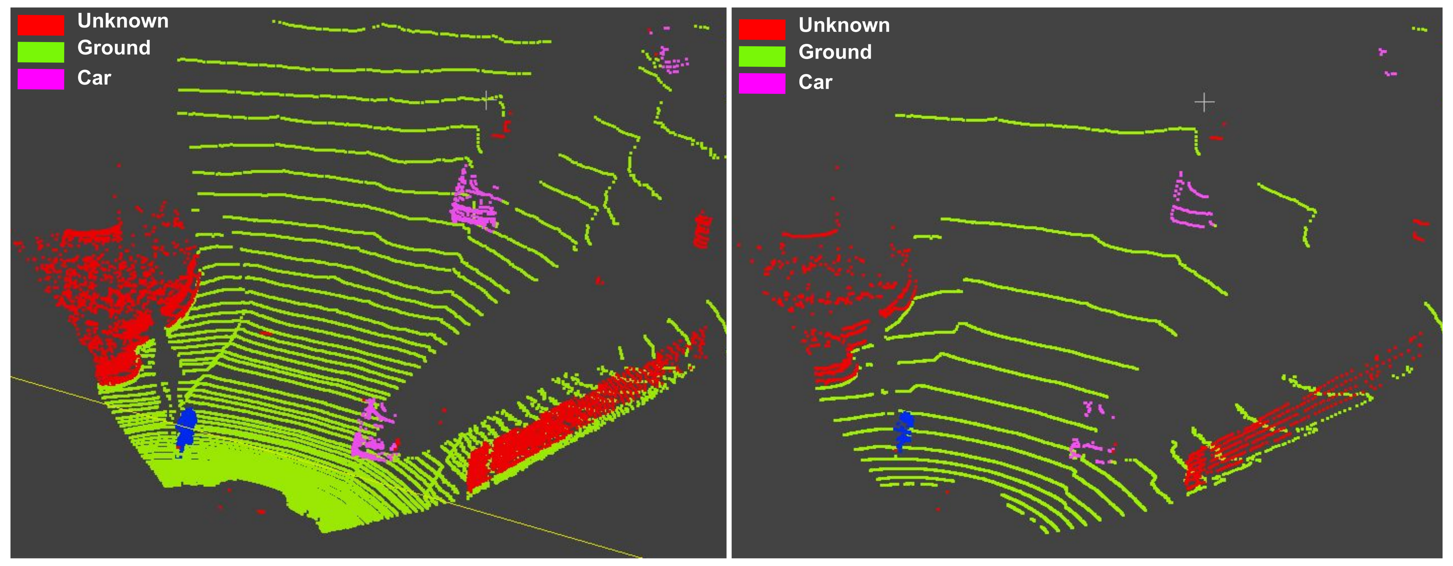

Fig. 2: Original 64-ring LIDAR scan (left) vs down-sampled 16-ring scan (right)

- **Up-sampling (Fig. 3).** Increase vertical density of a sparse point cloud by generating **32-ring** point clouds from **16-ring** using interpolation.
  - Analyze if up-sampling would **improve prediction maps** without undermining the prediction time per scan.
  - Analyze if the model is able to **respond robustly against noise** that might be introduced through the interpolations.

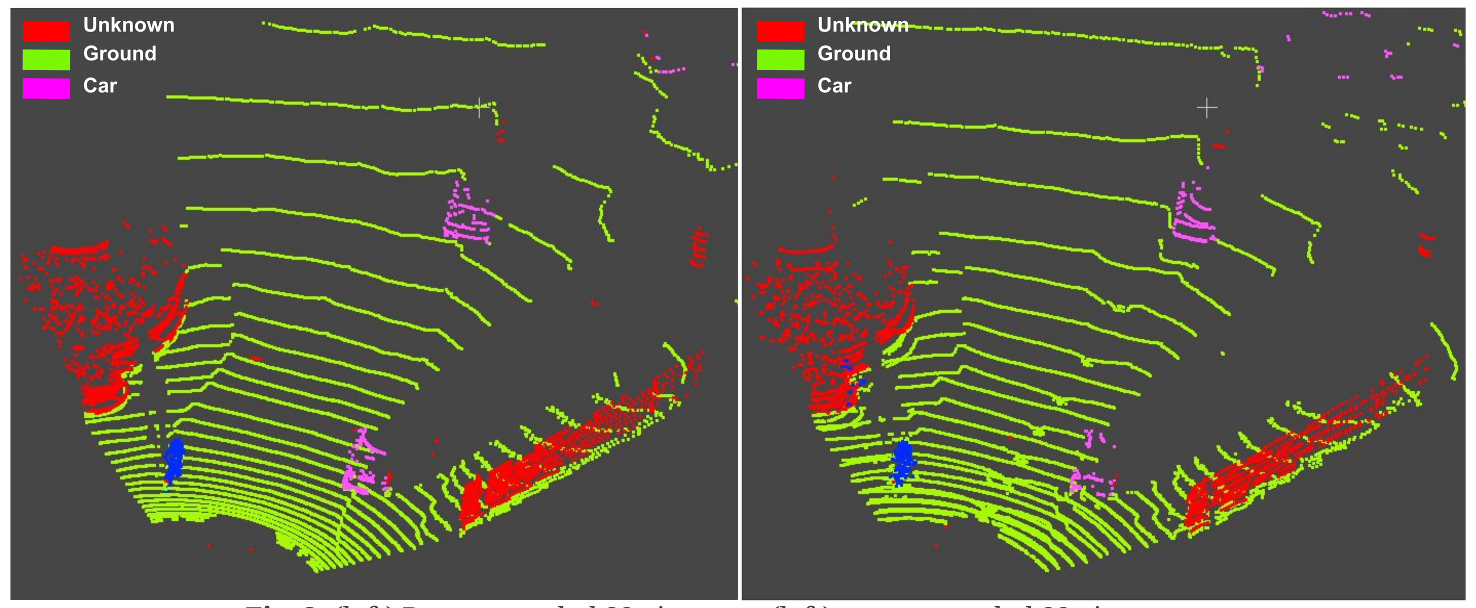

Fig. 3: (left) Down-sampled 32-ring scan (left) vs up-sampled 32-ring scan.

- Finally, to perform the our experiments, we modified the **SqueezeSeg** architecture to operate with **16-ring** and **32-ring** point clouds, since it was originally designed for 64-ring point clouds.

Experiments

- For each experiment,
  - we perform **end-to-end** training,
  - we evaluate if using **Recurrent CRF** allows to refine predictions.
  - we compare **precision, recall, IoU** and **prediction time.**

Results

Down-sampling method	Car			Ground		
	Precision	Recall	IoU	Precision	Recall	IoU
w/ CRF	0.5206	**0.9124**	0.4969	**0.9649**	0.8903	0.8624
w/o CRF	**0.5858**	0.8607	**0.5351**	0.9595	**0.9054**	**0.8720**

Table 1: **Down-sampling** method performance on the validation set.

Up-sampling method	Car			Ground		
	Precision	Recall	IoU	Precision	Recall	IoU
w/ CRF	**0.6485**	**0.9056**	**0.6074**	0.9622	**0.9081**	0.8731
w/o CRF	0.6089	0.8785	0.5616	**0.9655**	0.9053	**0.8769**

Table 2: **Up-sampling** method performance on the validation set.

Original method	Car			Ground		
	Precision	Recall	IoU	Precision	Recall	IoU
w/ CRF	0.7726	0.9736	0.7567	0.9745	**0.9127**	0.8913
w/o CRF	**0.7830**	**0.9792**	**0.7702**	**0.9770**	0.9108	**0.8916**

Table 3: Performance on the **original dataset**.

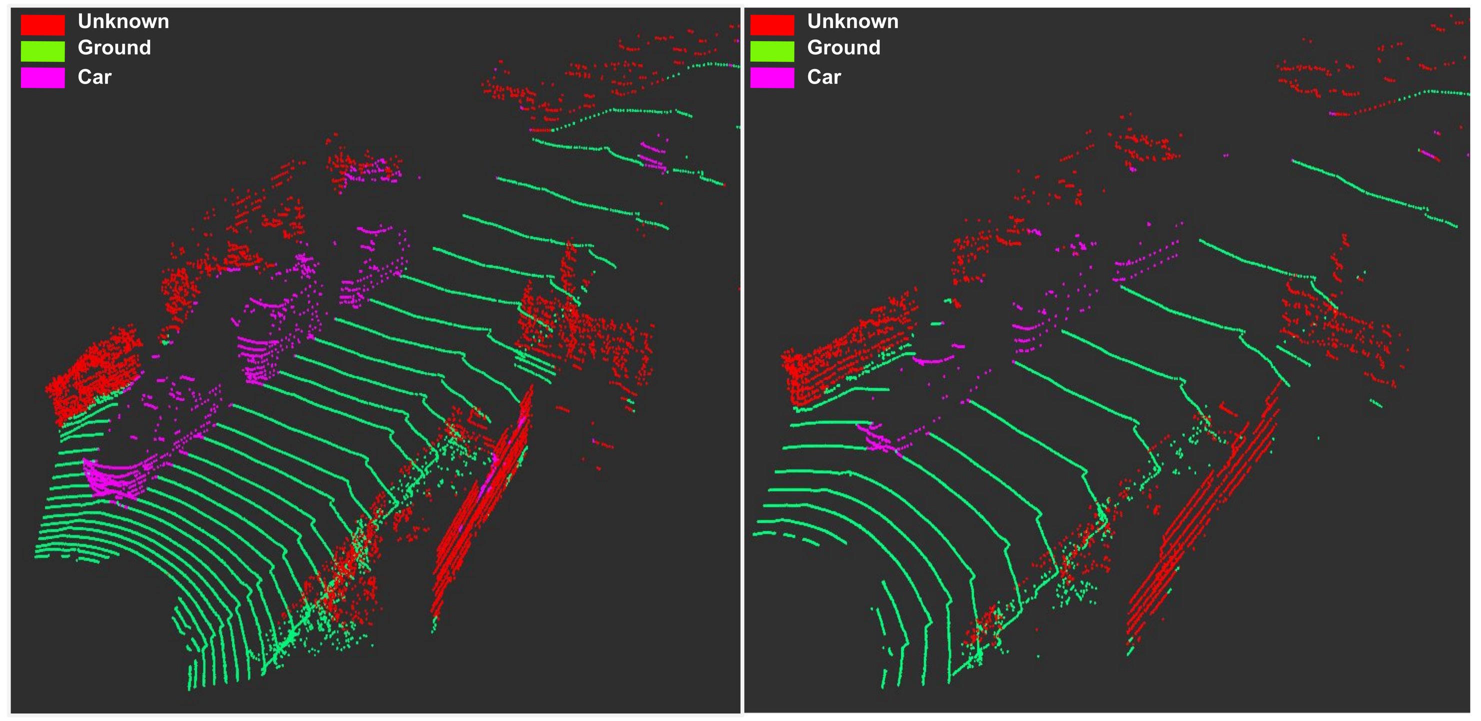

Fig 5: Predictions on 32-ring point clouds and 16-ring point clouds

Prediction time (s)			
	Down-sampling (8, 192 points per point cloud)	Up-sampling (16,384 points per point cloud)	Original Approach (32,768 point per point cloud)
w/ CRF	0.02241	0.03696	0.06186
w/o CRF	0.01615	0.02812	0.06038

Table 3: Prediction time per point cloud tested on a GeForce GTX 1060 MaxQ GPU

Conclusion

- We proposed two methods to perform fast and accurate semantic segmentation of highly sparse LIDAR point clouds for instances **car** and **ground** using **Deep Learning.**
  - In general, we achieve high ground-classification performance.
  - Our results for the car class are comparable to the results in **[1]**.
  - In all of our results, we achieve **high recall** scores.
  - Finally, there is room for improvement concerning the results using **Conditional Random Fields.**

Future Work

- **Extend** the capabilities of the network to perform classification for **pedestrians** and **cyclists.**
- **Integration** of our system using **NVIDIA Jetson TX2**.
- **Explore** the capabilities of the network in **unstructured environments.**

References

[1] B. Wu, A. Wan, X. Yue, and K. Keutzer. *SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for real-time road-object segmentation from 3D LiDAR Point Clouds.* arXiv:1710.07368, 2017.
[2] A. Geiger, P. Lenz, C. Stiller, and R. Urtasum. *Vision meets robotics: the kitti dataset.* 2013
[3] D. Zermas, I. Izzat, and N. Papanikolopoulou. *Fast segmentation of 3D Point Clouds: A paradigm on LiDAR data for autonomous vehicle applications.* In 2017 IEEE International Conference on Robotics and Automation (ICRA), pages 5067-6073, May 2017.

Acknowledgments

- Thanks to **Tecnologico de Monterrey** and **The Robotics Institute** for funding this work, and **Dr. Luis Navarro-Serment** for his guidance and support.

TECNOLÓGICO DE MONTERREY®