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 that have labels for 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.
- To generate the ground annotations, we developed an automatic ground labeler that uses a 3D plane-fitting technique [3] (Fig.1).
- 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 if up-sampling would improve prediction maps without undermining the prediction time per scan.
- To generate the ground annotations, we developed an automatic ground labeler that uses a 3D plane-fitting technique [3] (Fig.1).
- 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.

**Results**
- Precision, recall, IoU and prediction time.

**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.

**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.

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

**References**

**Table 1:** Prediction time per point cloud tested on a GeForce GTX 1060 MaxQ GPU.

<table>
<thead>
<tr>
<th>Method</th>
<th>Prediction time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down-sampling</td>
<td></td>
</tr>
<tr>
<td>w/ CRF</td>
<td>0.0424</td>
</tr>
<tr>
<td>w/o CRF</td>
<td>0.0261</td>
</tr>
<tr>
<td>Up-sampling</td>
<td></td>
</tr>
<tr>
<td>w/ CRF</td>
<td>0.0306</td>
</tr>
<tr>
<td>w/o CRF</td>
<td>0.0085</td>
</tr>
<tr>
<td>Original Approach</td>
<td></td>
</tr>
<tr>
<td>w/ CRF</td>
<td>0.0424</td>
</tr>
<tr>
<td>w/o CRF</td>
<td>0.0261</td>
</tr>
</tbody>
</table>