Classification of Point Cloud for Road Scene Understanding with Multiscale Voxel Deep Network

Xavier Roynard
Jean-Emmanuel Deschaud
François Goulette

xavier.roynard@mines-paristech.fr, jean-emmanuel.deschaud@mines-paristech.fr, francois.goulette@mines-paristech.fr

October 1, 2018
Presentation Outline

1. Context
2. State of the Art
3. Our Approach
4. Results
5. Work in progress
1 Context

2 State of the Art
   - Point-wise Classification
   - Region-wise Classification
   - Segmentation-based Classification

3 Our Approach
   - Training on 3D point cloud scenes
   - Multi-Scale Architecture

4 Results
   - Results on Public Benchmarks
   - Comparison Mono/Multi-scales

5 Work in progress
Autonomous vehicles require HD-Maps for navigation and decision-making process

A production pipeline of HD-Maps can be:

- 3D point cloud acquisition by Mobile Laser Scanning (MLS),
- Precise 3D localization of relevant objects (road signs and ground markings),
- Extraction of mobile objects,
- Detection of navigation area and buildings.
1 Context

2 State of the Art
- Point-wise Classification
- Region-wise Classification
- Segmentation-based Classification

3 Our Approach
- Training on 3D point cloud scenes
- Multi-Scale Architecture

4 Results
- Results on Public Benchmarks
- Comparison Mono/Multi-scales

5 Work in progress
Point-wise Classification

- Hand-Made Features (dimensionality attributes, multi-scale) \(^a\),
- Deep Learning on Voxel Grid Neighborhood \(^b\)


State of the Art

Region-wise Classification

- on images: Snapnet\(^a\)
- on voxel Grid: SEGCloud\(^b\)


State of the Art

Segmentation-based Classification

- SPGraph\(^a\)

1 Context

2 State of the Art
   - Point-wise Classification
   - Region-wise Classification
   - Segmentation-based Classification

3 Our Approach
   - Training on 3D point cloud scenes
   - Multi-Scale Architecture

4 Results
   - Results on Public Benchmarks
   - Comparison Mono/Multi-scales

5 Work in progress
Our Approach

Training a Deep Neural Network on fully annotated 3D point cloud scenes

- Some challenges:
  - very unbalanced classes,
  - most represented classes are also the least geometrically diversified (ground, buildings),
  - billion of samples.

- Using all samples (points) in one epoch would be infeasible.

Proposed solution

- randomly sample $N > 0$ points in each class of the training dataset,
- then one epoch is: pass randomly all sampled points in the network
Our Approach

Multi-Scale Architecture

Mono-Scale

Multi-Scale
1 Context

2 State of the Art
   - Point-wise Classification
   - Region-wise Classification
   - Segmentation-based Classification

3 Our Approach
   - Training on 3D point cloud scenes
   - Multi-Scale Architecture

4 Results
   - Results on Public Benchmarks
   - Comparison Mono/Multi-scales

5 Work in progress
## Results on Semantic3D

<table>
<thead>
<tr>
<th>Rank</th>
<th>Method</th>
<th>Averaged IoU</th>
<th>Overall Accuracy</th>
<th>man-made terrain</th>
<th>natural terrain</th>
<th>high vegetation</th>
<th>low vegetation</th>
<th>buildings</th>
<th>hard scape</th>
<th>scanning artefacts</th>
<th>cars</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SPGraph $^1$</td>
<td>73.2%</td>
<td>94.0%</td>
<td>97.4%</td>
<td>92.6%</td>
<td>87.9%</td>
<td>44.0%</td>
<td>93.2%</td>
<td>31.0%</td>
<td>63.5%</td>
<td>76.2%</td>
</tr>
<tr>
<td>2</td>
<td>MS3_DVS (Ours)</td>
<td>65.3%</td>
<td>88.4%</td>
<td>83.0%</td>
<td>67.2%</td>
<td>83.8%</td>
<td>36.7%</td>
<td>92.4%</td>
<td>31.3%</td>
<td>50.0%</td>
<td>78.2%</td>
</tr>
<tr>
<td>3</td>
<td>RF_MSSF $^2$</td>
<td>62.7%</td>
<td>90.3%</td>
<td>87.6%</td>
<td>80.3%</td>
<td>81.8%</td>
<td>36.4%</td>
<td>92.2%</td>
<td>24.1%</td>
<td>42.6%</td>
<td>56.6%</td>
</tr>
<tr>
<td>4</td>
<td>SegCloud $^3$</td>
<td>61.3%</td>
<td>88.1%</td>
<td>83.9%</td>
<td>66.0%</td>
<td>86.0%</td>
<td>40.5%</td>
<td>91.1%</td>
<td>30.9%</td>
<td>27.5%</td>
<td>64.3%</td>
</tr>
<tr>
<td>5</td>
<td>SnapNet $^4$</td>
<td>59.1%</td>
<td>88.6%</td>
<td>82.0%</td>
<td>77.3%</td>
<td>79.7%</td>
<td>22.9%</td>
<td>91.1%</td>
<td>18.4%</td>
<td>37.3%</td>
<td>64.4%</td>
</tr>
</tbody>
</table>

---


Results on Paris-Lille-3D

New Benchmark for Point Cloud Classification: Paris-Lille-3D\(^a\):

- Training set: 140 million manually annotated points, 50 classes, 2km, 2 cities
- Test set: 30 million points, 9 classes, 2 other cities


<table>
<thead>
<tr>
<th>Rank</th>
<th>Method</th>
<th>Averaged IoU</th>
<th>ground</th>
<th>building</th>
<th>pole</th>
<th>bollard</th>
<th>trash can</th>
<th>barrier</th>
<th>pedestrian</th>
<th>car</th>
<th>natural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MS3_DVS (Ours)</td>
<td>66.89%</td>
<td>99.03%</td>
<td>94.76%</td>
<td>52.40%</td>
<td>38.13%</td>
<td>36.02%</td>
<td>49.27%</td>
<td>52.56%</td>
<td>91.3%</td>
<td>88.58%</td>
</tr>
<tr>
<td>2</td>
<td>RF_MSSF(^5)</td>
<td>56.28%</td>
<td>99.25%</td>
<td>88.63%</td>
<td>47.75%</td>
<td>67.27%</td>
<td>2.31%</td>
<td>27.09%</td>
<td>20.61%</td>
<td>74.79%</td>
<td>78.83%</td>
</tr>
</tbody>
</table>

Comparison Mono/Multi-scales

Precision and Recall on Paris-Lille-3D

Improvement on most of the classes.

Mean F1 Score

The contribution of multi-scale network is obvious.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th></th>
<th>Recall</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MS3_DVS</td>
<td>MS1_DVS</td>
<td>MS3_DVS</td>
<td>MS1_DVS</td>
</tr>
<tr>
<td>ground</td>
<td>97.74%</td>
<td>97.08%</td>
<td>98.70%</td>
<td>98.28%</td>
</tr>
<tr>
<td>buildings</td>
<td>85.50%</td>
<td>84.28%</td>
<td>95.27%</td>
<td>90.65%</td>
</tr>
<tr>
<td>poles</td>
<td>93.30%</td>
<td>92.27%</td>
<td>92.69%</td>
<td>94.16%</td>
</tr>
<tr>
<td>bollards</td>
<td>98.60%</td>
<td>98.61%</td>
<td>93.93%</td>
<td>94.16%</td>
</tr>
<tr>
<td>trash cans</td>
<td>95.31%</td>
<td>93.52%</td>
<td>79.60%</td>
<td>80.91%</td>
</tr>
<tr>
<td>barriers</td>
<td>85.70%</td>
<td>81.56%</td>
<td>77.08%</td>
<td>73.85%</td>
</tr>
<tr>
<td>pedestrians</td>
<td>98.53%</td>
<td>93.62%</td>
<td>95.42%</td>
<td>92.89%</td>
</tr>
<tr>
<td>cars</td>
<td>93.51%</td>
<td>96.41%</td>
<td>98.38%</td>
<td>97.71%</td>
</tr>
<tr>
<td>natural</td>
<td>89.51%</td>
<td>88.23%</td>
<td>92.52%</td>
<td>91.53%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>MS3_DVS</th>
<th>MS1_DVS</th>
<th>VoxNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris-Lille-3D</td>
<td>89.29%</td>
<td>88.23%</td>
<td>86.59%</td>
<td></td>
</tr>
<tr>
<td>Semantic3D</td>
<td>79.36%</td>
<td>74.05%</td>
<td>71.66%</td>
<td></td>
</tr>
</tbody>
</table>

1 Context

2 State of the Art
   - Point-wise Classification
   - Region-wise Classification
   - Segmentation-based Classification

3 Our Approach
   - Training on 3D point cloud scenes
   - Multi-Scale Architecture

4 Results
   - Results on Public Benchmarks
   - Comparison Mono/Multi-scales

5 Work in progress
- Use network architectures closer to the state of the art (Inception/ResNet).
- Adapt the Multi-Scale architecture to U-Net networks for semantic segmentation.
- Get closer to real-time inference with an Octree structure.
- Ensemble on several networks or several orientations of input point cloud.
Thank you!

Questions?


