# SPLATNet: Sparse Lattice Networks for Point Cloud Processing

Su, Hang, et al.

Otakar Jašek

May 30, 2019

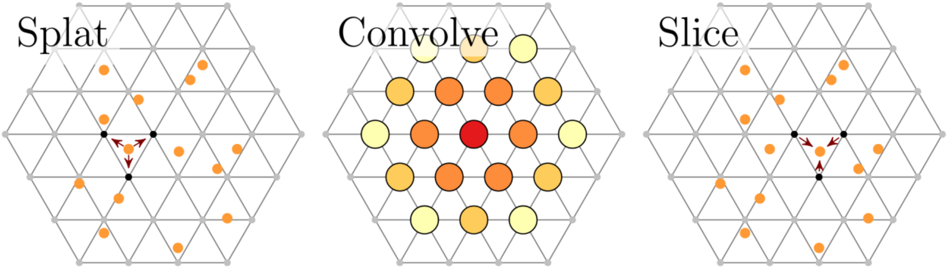
# Point cloud learning – history

- Hand-crafted features
- View-based methods
- Voxel representation
- PointNet
- Geometric deep learning

# Point cloud characteristics

- **Unordered** set of points
- Varying number of points
- Invariant to rigid transformations
- Often no features, location only

# Bilateral Convolution Filter (BCL)[1]




- Permutohedral lattice (can be regular though)
- Barycentric interpolation
- Output not necessarily the same as input
- Hash tables due to sparsity

[1] Jampani, Varun, Martin Kiefel, and Peter V. Gehler. "Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

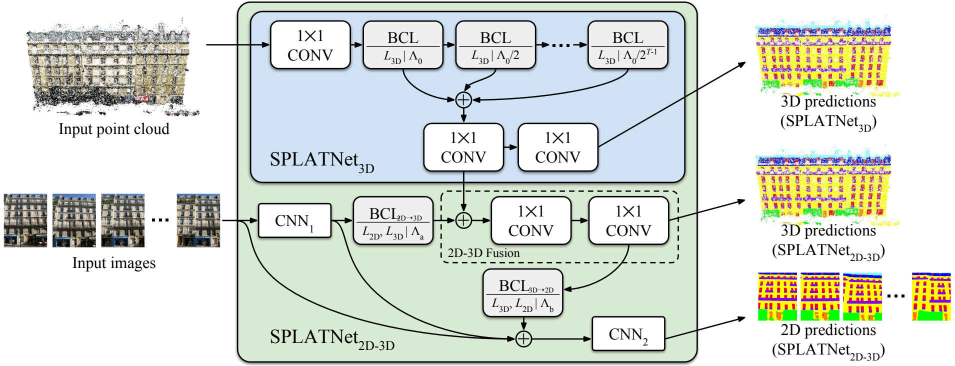
# BCL advantages for point clouds

- Points can be unordered
- Input an output points can be different
- Separating "what" and "where"
- Fully differentiable
- Scaling of features to simulate receptive field size (less parameters)
- Density normalization

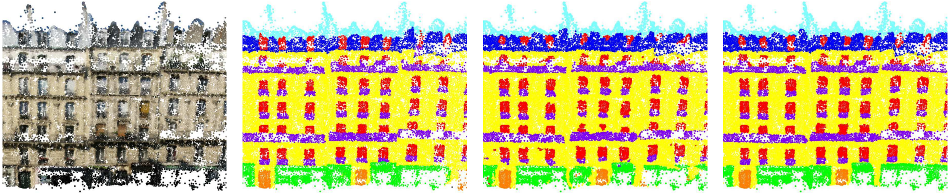
# Scaling and selecting features

$(x, y, z)$ $(8x, 8y, 8z)$ $(n_x, n_y, n_z)$

# Architecture



# Facade segmentation

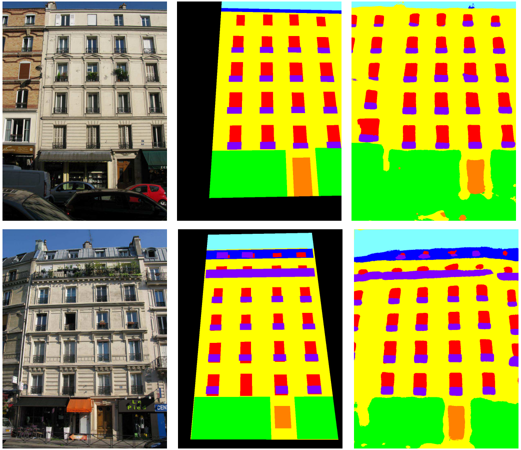


Input Point Cloud | Ground truth | $\text{SPLATNet}_{3D}$ | $\text{SPLATNet}_{2D\text{-}3D}$

| Method | Average IoU | Runtime (min) |
|---|---|---|
| *With only 3D data* | | |
| OctNet [35] | 59.2 | - |
| $\text{Autocontext}_{3D}$ [14] | 54.4 | 16 |
| $\text{SPLATNet}_{3D}$ (Ours) | **65.4** | 0.06 |
| *With both 2D and 3D data* | | |
| $\text{Autocontext}_{2D\text{-}3D}$ [14] | 62.9 | 87 |
| $\text{SPLATNet}_{2D\text{-}3D}$ (Ours) | **69.8** | 1.20 |

(a) Point cloud labeling

# Facade segmentation

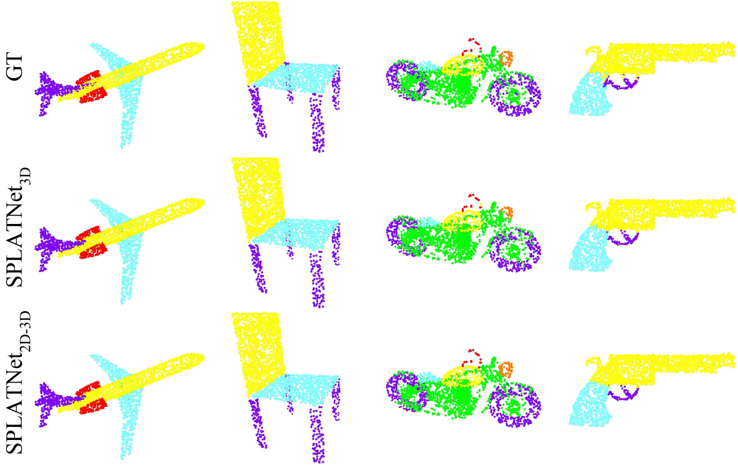


Input    Ground truth    SPLATNet$_{2D\text{-}3D}$

| Method | Average IoU | Runtime (min) |
|---|---|---|
| Autocontext$_{2D}$ [14] | 60.5 | 117 |
| Autocontext$_{2D\text{-}3D}$ [14] | 62.7 | 146 |
| DeepLab$_{2D}$ [10] | 69.3 | 0.84 |
| SPLATNet$_{2D\text{-}3D}$ (Ours) | **70.6** | 4.34 |

(b) Multi-view image labeling

# Shapenet



# Shapenet results

| #instances | | | 2690 | 76 | 55 | 898 | 3758 | 69 | 787 | 392 | 1547 | 451 | 202 | 184 | 283 | 66 | 152 | 5271 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | class avg. | instance avg. | air-plane | bag | cap | car | chair | ear-phone | guitar | knife | lamp | laptop | motor-bike | mug | pistol | rocket | skate-board | table |
| Yi *et al.* [44] | 79.0 | 81.4 | 81.0 | 78.4 | 77.7 | 75.7 | 87.6 | 61.9 | 92.0 | 85.4 | 82.5 | 95.7 | 70.6 | 91.9 | **85.9** | 53.1 | 69.8 | 75.3 |
| 3DCNN [31] | 74.9 | 79.4 | 75.1 | 72.8 | 73.3 | 70.0 | 87.2 | 63.5 | 88.4 | 79.6 | 74.4 | 93.9 | 58.7 | 91.8 | 76.4 | 51.2 | 65.3 | 77.1 |
| Kd-network [25] | 77.4 | 82.3 | 80.1 | 74.6 | 74.3 | 70.3 | 88.6 | 73.5 | 90.2 | **87.2** | 81.0 | 94.9 | 57.4 | 86.7 | 78.1 | 51.8 | 69.9 | 80.3 |
| PointNet [31] | 80.4 | 83.7 | **83.4** | 78.7 | 82.5 | 74.9 | 89.6 | 73.0 | 91.5 | 85.9 | 80.8 | 95.3 | 65.2 | 93.0 | 81.2 | 57.9 | 72.8 | 80.6 |
| PointNet++ [33] | 81.9 | 85.1 | 82.4 | 79.0 | 87.7 | 77.3 | **90.8** | 71.8 | 91.0 | 85.9 | 83.7 | 95.3 | 71.6 | 94.1 | 81.3 | 58.7 | 76.4 | **82.6** |
| SyncSpecCNN [45] | 82.0 | 84.7 | 81.6 | 81.7 | 81.9 | 75.2 | 90.2 | 74.9 | **93.0** | 86.1 | **84.7** | 95.6 | 66.7 | 92.7 | 81.6 | 60.6 | **82.9** | 82.1 |
| SPLATNet$_{3D}$ | 82.0 | 84.6 | 81.9 | 83.9 | 88.6 | 79.5 | 90.1 | 73.5 | 91.3 | 84.7 | 84.5 | **96.3** | 69.7 | 95.0 | 81.7 | 59.2 | 70.4 | 81.3 |
| SPLATNet$_{2D\text{-}3D}$ | **83.7** | **85.4** | 83.2 | **84.3** | **89.1** | **80.3** | 90.7 | **75.5** | 92.1 | 87.1 | 83.9 | **96.3** | **75.6** | **95.8** | 83.8 | **64.0** | 75.5 | 81.8 |

| | **MEAN** | AREO | BAG | CAP | CAR | CHAIR | EAR PHONE | GUITAR | KNIFE | LAMP | LAPTOP | MOTOR | MUG | PISTOL | ROCKET | SKATE BOARD | TABLE | **WINNING CATEGORIES** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # SHAPES | | 2690 | 76 | 55 | 898 | 3758 | 69 | 787 | 392 | 1547 | 451 | 202 | 184 | 283 | 66 | 152 | 5271 | |
| POINTNET | 83.7 | 83.4 | 78.7 | 82.5 | 74.9 | 89.6 | 73.0 | **91.5** | 85.9 | 80.8 | 95.3 | 65.2 | 93.0 | 81.2 | 57.9 | 72.8 | 80.6 | 1 |
| POINTNET++ | **85.1** | 82.4 | 79.0 | **87.7** | 77.3 | 90.8 | 71.8 | 91.0 | 85.9 | **83.7** | 95.3 | **71.6** | **94.1** | 81.3 | 58.7 | **76.4** | 82.6 | 5 |
| KD-NET | 82.3 | 80.1 | 74.6 | 74.3 | 70.3 | 88.6 | 73.5 | 90.2 | 87.2 | 81.0 | 94.9 | 57.4 | 86.7 | 78.1 | 51.8 | 69.9 | 80.3 | 0 |
| LOCALFEATURENET | 84.3 | **86.1** | 73.0 | 54.9 | **77.4** | 88.8 | 55.0 | 90.6 | 86.5 | 75.2 | **96.1** | 57.3 | 91.7 | **83.1** | 53.9 | 72.5 | **83.8** | 5 |
| OURS | **85.1** | 84.2 | **83.7** | 84.4 | 77.1 | **90.9** | **78.5** | **91.5** | **87.3** | 82.9 | 96.0 | 67.8 | 93.3 | 82.6 | **59.7** | 75.5 | 82.0 | **6** |

# Conclusion

- Hierarchical convolution on pointclouds
- Scalable receptive field of view
- Locally aware
- Order invariant
- Sparse and efficient learnable filters
- Ability to map to different output points