DepthContrast: Learning Self-supervised 3D Features from Single-view Depth Scans
Zaiwei Zhang¹, Rohit Girdhar², Armand Joulin², Ishan Misra²
¹University of Texas at Austin  ²Facebook AI Research

Motivation
• Expensive 3D data labeling
• Availability of existing large collection of single-view depth scans
• More commercial 3D sensors will lead to more unlabeled single-view 3D data

Key Takeaways
• Works with single/multi-view depth scans acquired by varied sensors (Lidar or Kinect)
• Works on point cloud and voxel-based model architectures
• Improves label efficiency on downstream tasks

Method & Formulation
\[ l_i = -\log \left( \frac{\exp(v_{i,1}v_{i,2}/\tau)}{\sum_k \exp(v_{i,k}^Tv_{i,j}/\tau)} \right) \]

Extension to Multiple 3D Input Formats
- Point Network
- Voxel Network

Main Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Stats</th>
<th>Task</th>
<th>Gain of DepthContrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-supervised Pretraining</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ScanNet-vid (Dai et al., 2017)</td>
<td>190K single-view depth maps (Indoor)</td>
<td>Det. +3.6% mAP</td>
<td></td>
</tr>
<tr>
<td>Redwood-vid (Choi et al., 2016)</td>
<td>190K single-view depth maps (Indoor/Outdoor)</td>
<td>Seg. +6.9% mIOU</td>
<td></td>
</tr>
</tbody>
</table>

Benefits on Label Efficiency

Scaling on Model & Pretraining Data

Baseline Comparison

<table>
<thead>
<tr>
<th>Initialization</th>
<th>ScanNet</th>
<th>SUNRGBD</th>
<th>Matterport3D</th>
<th>S3DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scratch</td>
<td>58.6</td>
<td>57.4</td>
<td>38.8</td>
<td>31.2</td>
</tr>
<tr>
<td>Supervised</td>
<td>59.1 (+1.5)</td>
<td>41.7 (+29)</td>
<td>48.5 (+17.2)</td>
<td></td>
</tr>
<tr>
<td>DepthContrast (Ours)</td>
<td>61.3 (+2.2)</td>
<td>41.9 (+3.1)</td>
<td>43.6 (+2.4)</td>
<td></td>
</tr>
<tr>
<td>PointContrast (Xie et al., 2020)</td>
<td>59.2 (+2.4)</td>
<td>57.9 (+1.6)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

VoteNet model on ScanNet
VoteNet model on SUNRGBD

Same perf. ∼2× fewer labels

Final Loss:
\[ L_i = l_{i,1}^{ab} + l_{i,2}^{ba} + l_a^{aa} + l_b^{bb} \]

Rounding across format
\[ l_{i,1}^{ab} \]

Rounding within format
\[ l_a^{aa} \]

Code Link: https://github.com/facebookresearch/DepthContrast