Learning Object-level Point Augmentor for Semi-supervised 3D Object Detection
Cheng-Ju Ho\textsuperscript{1}  Chen-Hsuan Tai\textsuperscript{1}  Yi-Hsuan Tsai\textsuperscript{2}  Yen-Yu Lin\textsuperscript{1}  Ming-Hsuan Yang\textsuperscript{3}
\textsuperscript{1}National Yang Ming Chiao Tung University \textsuperscript{2}Phiar Technologies \textsuperscript{3}University of California at Merced

Observation

- Existing 3D semi-supervised methods usually employ only global augmentation, but it’s sub-optimal.
  - It ignores the object-level data variance, which is crucial for the instance-level object detection task.
- Apply augmentations to the point clouds within each object bounding box directly.
  - Its performance depends on proper augmentation settings.
- Compared with rotation, point displacement can enhance data variance while keeping object orientations.

Object-level point augmentor

- Adversarial learning strategy.
  - With jointly pre-train a detector with an augmentor. The augmentor is optimized to generate proper augmented scene $x_a$ while the detector is derived to localize and recognize the augmented data accurately.

Augmentation Objective.

- Augmented scene $x_a$ should be more challenging.
  - $\mathcal{L}_d(x_{i\ast}, y_{a\ast}) \geq \mathcal{L}_d(x_{i\ast}, y_{a\ast}')$
- $x_i$ and $x_a$ should be classified as the same class.

Teacher-Student Framework in SSL.

- We initialize the student and teacher models from the pre-trained detector, and apply our object-level augmentor and asymmetric data augmentations to make this framework effective.

Quantitative results on indoor datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>mAP @5%</th>
<th>mAP @10%</th>
<th>mAP @20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ScanNet [16]</td>
<td>OA</td>
<td>47.1</td>
<td>39.0</td>
<td>21.1</td>
</tr>
<tr>
<td>SUN RGB-D 5%</td>
<td>OA</td>
<td>42.7</td>
<td>24.9</td>
<td>13.6</td>
</tr>
<tr>
<td>PRE-DEF [5]</td>
<td>OA</td>
<td>40.4</td>
<td>40.6</td>
<td>20.4</td>
</tr>
<tr>
<td>PRE-DEF [5]</td>
<td>OA</td>
<td>40.0</td>
<td>40.5</td>
<td>20.9</td>
</tr>
<tr>
<td>PRE-DEF [5]</td>
<td>OA</td>
<td>47.3</td>
<td>39.5</td>
<td>20.5</td>
</tr>
</tbody>
</table>

Methodology

- Emphasize object instances rather than irrelevant backgrounds.
- Dynamically adjust the augmentation magnitude according to the detector’s ability.
- After augmenting, making the augmented data more useful for object detector training.

Qualitative results on the ScanNet

<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>OPA (ours)</th>
<th>3DLoUMatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source code:</td>
<td>![QR Code](QR Code)</td>
<td>![Source Code](Source Code)</td>
</tr>
</tbody>
</table>