Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans

Romain Loiseau1, 2
romain.loiseau@enpc.fr
Elliot Vincent1, 3
elliot.vincent@enpc.fr
Mathieu Aubry1
mathieu.aubry@enpc.fr
Loic Landrieu1, 2
loic.landrieu@enpc.fr
1LIGM, ENPC 2LASTIG, ENSIG/IGN 3Inria and DIENS

Overview

Goal: provide a practical tool for analyzing 3D scenes without relying on application-specific user annotations.

Approach: a probabilistic reconstruction model that decomposes inputs into a small set of learned prototypical shapes.

Earth Parser Dataset: aerial scans in diverse environments.

Results: outperforms state-of-the-art unsupervised methods, visually interpretable, does not require any manual annotations.

Method

Scene reconstruction model:

\[ M(X) = \bigcup_{s=1}^{S} M_s(X), \text{ with } M_s(X) = Y_s = T_s(X) [P^k] \text{ if } b_s = k. \]

Proportional modeling:

\( a \) and \( b \) as random variables following (multi-)Bernoulli distributions:

\[ p(a_s = 1) = \alpha_s : \text{ proba. the slot } s \text{ is activated} \]

\[ p(a_s = 1, b_s = k) = \beta_s^k : \text{ proba. it is activated and selects prototype } k \]

Training losses:

- Slots average of the expected distance between \( M_s(X) \) and \( X \):

\[ \mathcal{L}_{\text{acc}}(M, X) = \frac{1}{S} \sum_{s=1}^{S} E_{a,b} \left[ d(M_s(X), X) \right]. \]

- Average over all points \( x \) of the expected distance between \( x \) and its closest point in the reconstruction:

\[ \mathcal{L}_{\text{cov}}(M, X) = \frac{1}{|X|} \sum_{x \in X} E_{a,b} \left[ \min_{b_s=1} d(x, M_s(X)) \right]. \]

The final loss is the sum of reconstruction losses and regularization:

\[ \mathcal{E}_{X} [\mathcal{L}_{\text{acc}}(M, X) + \mathcal{L}_{\text{cov}}(M, X)] + \lambda_{\text{act}} \mathcal{L}_{\text{act}} + \lambda_{\text{slot}} \mathcal{L}_{\text{slot}} + \lambda_{\text{proto}} \mathcal{L}_{\text{proto}}. \]

Quantitative Results

<table>
<thead>
<tr>
<th></th>
<th>Ref.</th>
<th>Semi-sup.</th>
<th>Crop fields</th>
<th>Forest</th>
<th>Greenhouses</th>
<th>Marina</th>
<th>Power plant</th>
<th>Urban</th>
<th>Windturbines</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-means</td>
<td>(i,z)</td>
<td>×</td>
<td>Cham. mIoU</td>
<td>Cham. mIoU</td>
<td>Cham. mIoU</td>
<td>Cham. mIoU</td>
<td>Cham. mIoU</td>
<td>Cham. mIoU</td>
<td>Cham. mIoU</td>
</tr>
<tr>
<td>SuperQuadrics [1]</td>
<td>3D×</td>
<td></td>
<td>93.8</td>
<td>71.5</td>
<td>39.3</td>
<td>41.4</td>
<td>42.8</td>
<td>56.5</td>
<td>—</td>
</tr>
<tr>
<td>DTI-Sprites [2]</td>
<td>2.5D+i</td>
<td>✓</td>
<td>6.10</td>
<td>83.2</td>
<td>40.2</td>
<td>42.0</td>
<td>53.6</td>
<td>29.0</td>
<td>47.3</td>
</tr>
<tr>
<td>AtlasNet v2 [3]</td>
<td>3D+i</td>
<td>✓</td>
<td>1.07</td>
<td>43.1</td>
<td>71.4</td>
<td>5.56</td>
<td>49.1</td>
<td>41.6</td>
<td>63.8</td>
</tr>
<tr>
<td>Ours</td>
<td>3D+i</td>
<td>✓</td>
<td>0.72</td>
<td>96.9</td>
<td>83.7</td>
<td>4.04</td>
<td>91.3</td>
<td>78.7</td>
<td>52.2</td>
</tr>
</tbody>
</table>

Meaningful and Interpretable Prototypes

Learned Prototypes. Selected learned prototypes on different scenes. We show three prototypes among those selected by our post-processing selection.

Bibliography