Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans

Romain Loiseau[1, 2] romain.loiseau@enpc.fr
Elliot Vincent[1, 3] elliot.vincent@enpc.fr
Mathieu Aubry[1] mathieu.aubry@enpc.fr
Loic Landrieu[1, 2] loic.landrieu@enpc.fr

[1]LIGM, ENPC [2]LASTIG, ENSG/IGN [3]Inria and DIENS

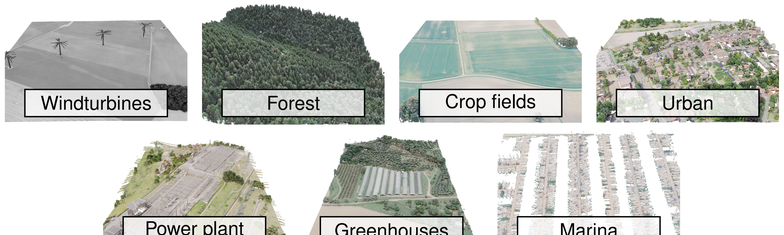
romainloiseau.fr/learnable-earth-parser/

Overview

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

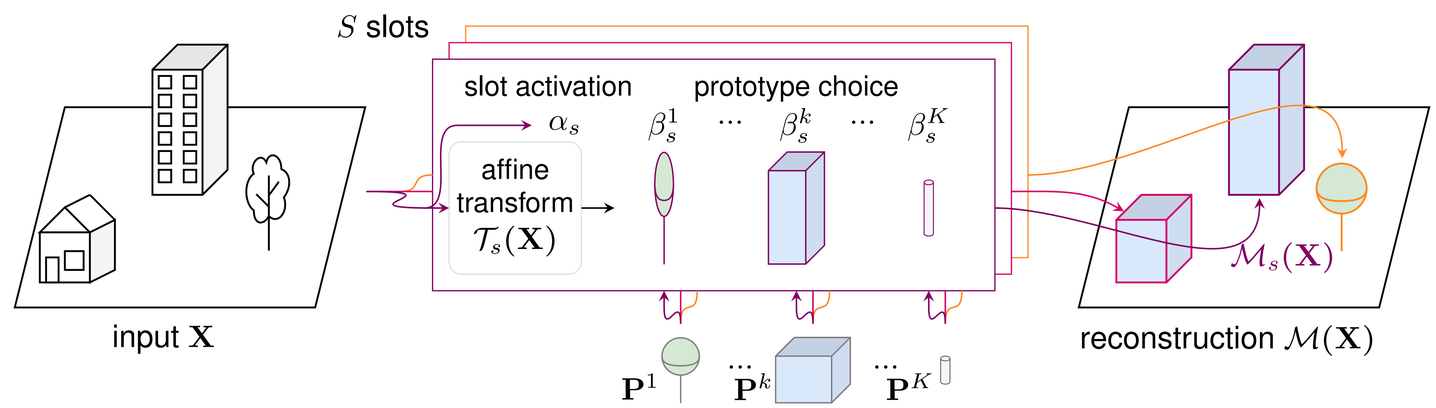
Approach: a **probabilistic reconstruction model** that decomposes inputs into a small set of **learned prototypical shapes**.

Earth Parser Dataset: aerial scans in diverse environments.



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

Learnable Earth Parser



Method Overview. Our model approximates an input point cloud $\mathbf{X}$ with S slot models. Each slot maps $\mathbf{X}$ to an affine 3D deformation $\mathcal{T}_s(\mathbf{X})$, a slot activation probability α_s, and the joint probabilities $\beta_s^1, \cdots, \beta_s^K$ of the slot being activated and choosing one of the K prototype point clouds $\mathbf{P}^1, \cdots, \mathbf{P}^K$. The output $\mathcal{M}_s(\mathbf{X})$ of an activated slot s is obtained by applying $\mathcal{T}_s(\mathbf{X})$ to its most likely prototype.

Method

Scene reconstruction model:

$$\mathcal{M}(\mathbf{X}) = \bigcup_{\substack{s=1\cdots S \\ a_s=1}} \mathcal{M}_s(\mathbf{X}) \text{ , with } \mathcal{M}_s(\mathbf{X}) = \mathbf{Y}_s^k = \mathcal{T}_s(\mathbf{X})[\mathbf{P}^k] \text{ if } b_s = k \text{ .}$$

Probabilistic modeling:

a and b as random variables following (multi)-Bernoulli distributions
$p(a_s = 1) = \alpha_s$: proba. the slot s is activated
$p(a_s = 1, b_s = k) = \beta_s^k$: proba. it is activated and selects prototype k

Training losses:

Slots average of the expected distance between $\mathcal{M}_s(\mathbf{X})$ and $\mathbf{X}$:

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

Average over all points x of $\mathbf{X}$ of the expected distance between x and its closest point in the reconstruction:

$$\mathcal{L}_{\text{cov}}(\mathcal{M}, \mathbf{X}) = \frac{1}{|\mathbf{X}|} \sum_{x \in \mathbf{X}} \mathbb{E}_{a,b} \left[\min_{s|a_s=1} d\left(x, \mathcal{M}_s(\mathbf{X})\right) \right] \text{ .}$$

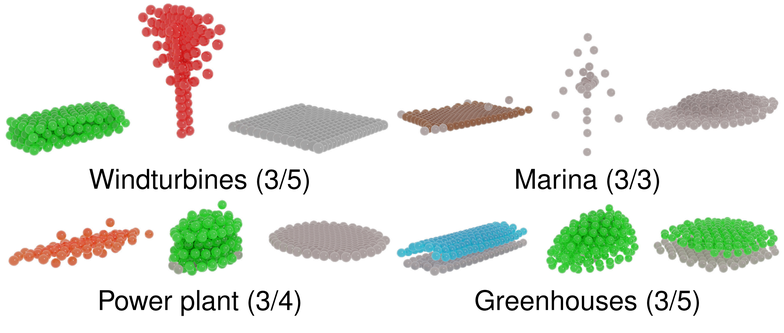
The final loss is the sum of reconstruction losses and regularization:

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

Quantitative Results

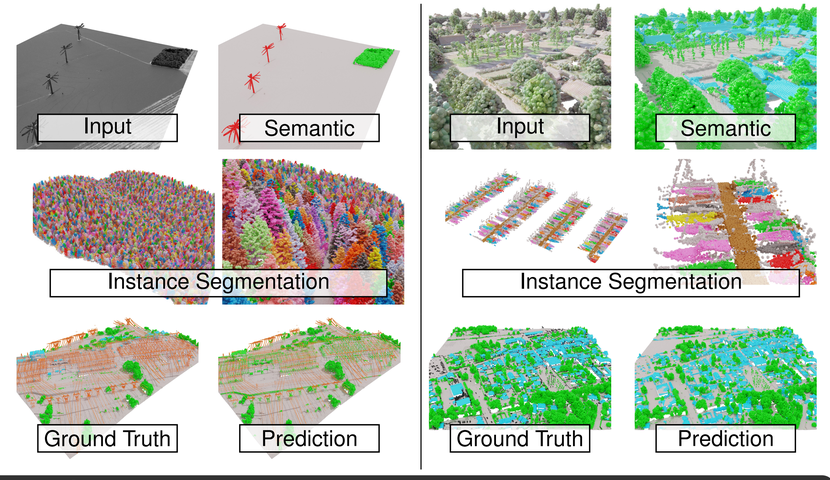
	Rec.	Semantic	Crop fields		Forest		Greenhouses		Marina		Power plant		Urban		Windturbines
			Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.
k-means (i,z)	✗	✓	—	93.8	—	71.5	—	39.3	—	41.4	—	42.8	—	56.5	—
SuperQuadrics [1]	3D	✗	0.86	—	1.04	—	0.60	—	0.93	—	0.58	—	0.40	—	13.5
DTI-Sprites [2]	2.5D+i	✓	6.10	83.2	14.59	40.2	5.36	42.0	6.16	41.4	5.36	29.0	2.99	47.3	36.19
AtlasNet v2 [3]	3D+i	✓	1.07	43.1	1.58	71.4	0.56	49.1	**0.73**	42.1	0.45	41.6	0.63	48.8	8.80
Ours	3D+i	✓	**0.72**	**96.9**	**0.88**	**83.7**	**0.40**	**91.3**	0.82	**78.7**	**0.44**	**52.2**	**0.29**	**83.2**	**6.65**

Meaningful and Interpretable Prototypes



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

Qualitative Results



Acknowledgements

This work was supported by ANR project READY3D ANR-19-CE23-0007. The work of MA was partly supported by the European Research Council (project DISCOVER, number 101076028).

Bibliography

[1] Paschalidou *et al*. Superquadrics revisited: Learning 3d shape parsing beyond cuboids. CVPR19. [2] Monnier *et al*. Unsupervised layered image decomposition into object prototypes. ICCV21. [3] Deprelle *et al*. Learning elementary structures for 3D shape generation and matching. NeurIPS19. [4] Loiseau *et al*. Representing Shape Collections with Alignment-Aware Linear Models. 3DV21.