Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans

Romain Loiseau[1, 2] Elliot Vincent[1, 3] Mathieu Aubry[1] Loic Landrieu[1, 2]




[1]LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, France

[2]Univ Gustave Eiffel, IGN, ENSG, LASTIG, France [3]INRIA Paris, France

1

2

1

École des Ponts
ParisTech

2

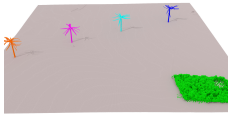
3

1, 2

Contributions

Large aerial 3D scan

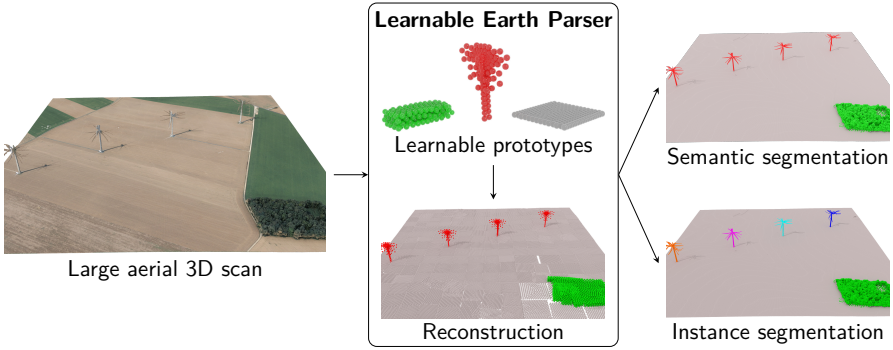
Semantic segmentation



Instance segmentation

Prior works ✗ need large annotations ✗ has low interpretability ✗ on toy scenes / 1 object

Contributions

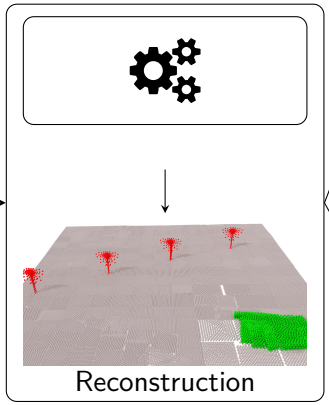


Prior works	✗ need large annotations	✗ has low interpretability	✗ on toy scenes / 1 object
Ours	✓ is purely unsupervised	✓ produces visual prototype	✓ on real world data

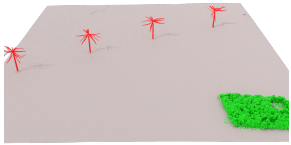
- **Learnable Earth Parser:** unsupervised framework to parse large real-world 3D scans into interpretable parts
- **Earth Parser Dataset:** new aerial LiDAR dataset with annotations in various environments

Discovering 3D Shapes in Aerial Scans

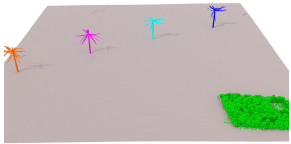
Large aerial 3D scan



Reconstruction



Semantic segmentation

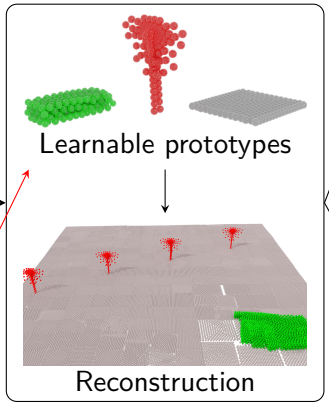


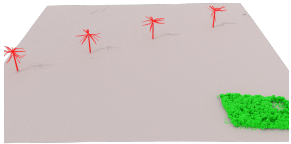
Instance segmentation

Discovering 3D Shapes in Aerial Scans

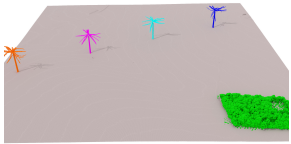
Large aerial 3D scan

Point clouds !



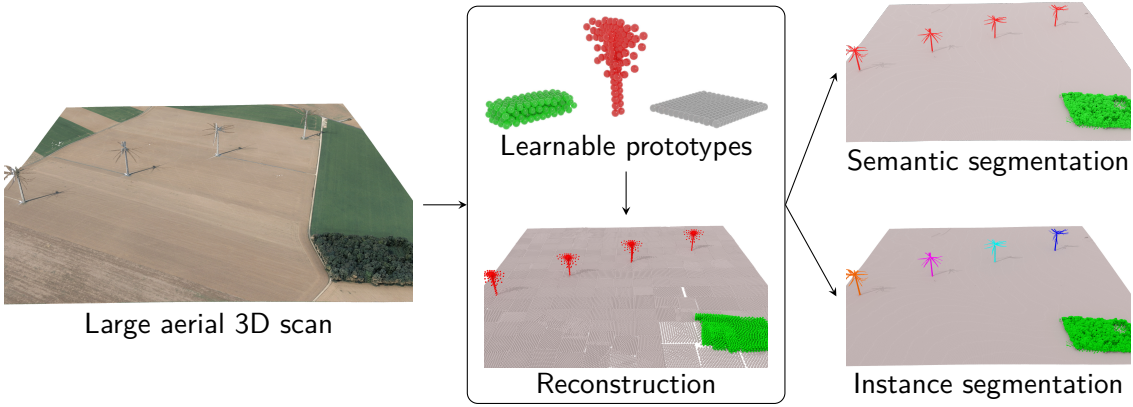


Semantic segmentation



Instance segmentation

Discovering 3D Shapes in Aerial Scans

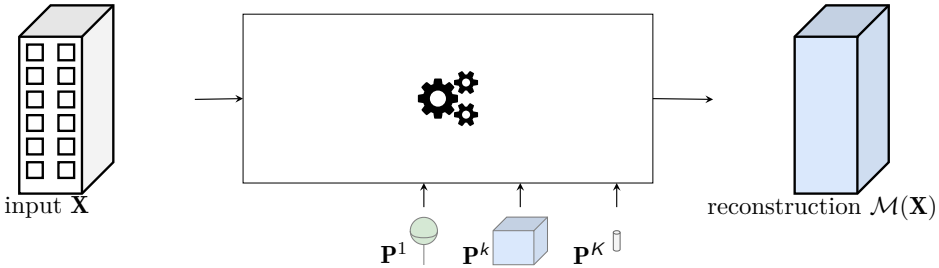


Challenges

? Diversity of objects

? Ambiguous objects separation, varying number of objects

Diversity of Objects

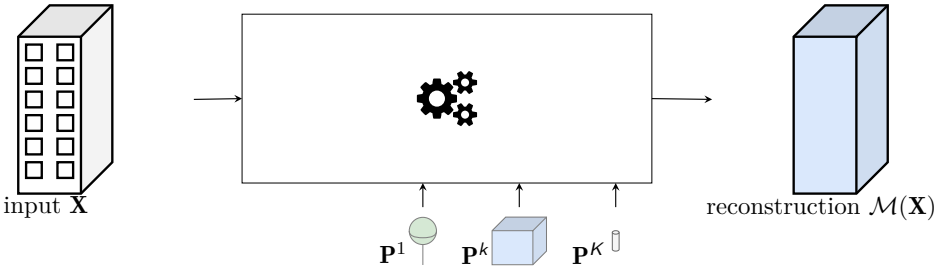


We can discover 3D prototypes in large point cloud collections [3]

- **Unsupervised** method to represent large point cloud collections
- **Viewable** representation in input space using 3D deformable prototypes
- *State-of-the-art* **few-shot semantic segmentation** performance

[3] R. Loiseau, T. Monnier, M. Aubry, and L. Landrieu, Representing Shape Collections With Alignment-Aware Linear Models, 3DV, 2021

Slot Modeling

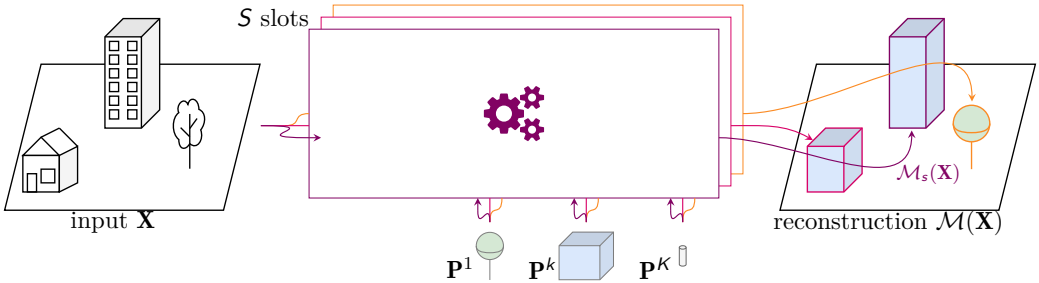


Challenges

✓ Diversity of objects

? Ambiguous objects separation, varying number of objects

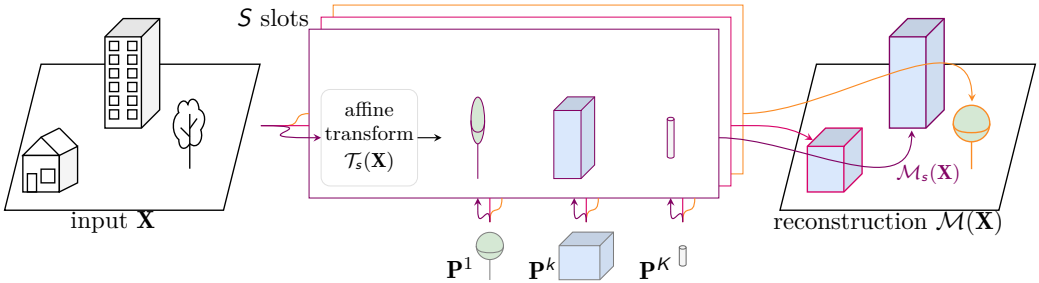
Slot Modeling



- S slot models $\mathcal{M}_s$:

$$\mathcal{M}(\mathbf{X}) = \bigcup_{s=1\cdots S} \mathcal{M}_s(\mathbf{X})$$

Slot Modeling



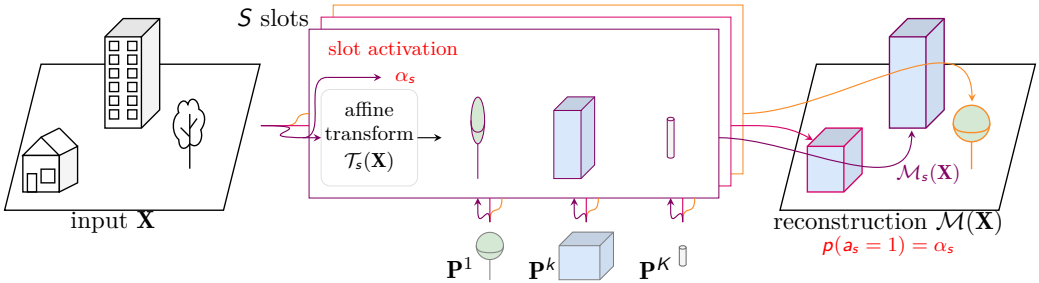
- S slot models $\mathcal{M}_s$:

$$\mathcal{M}(\mathbf{X}) = \bigcup_{s=1\cdots S} \mathcal{M}_s(\mathbf{X})$$

- The k-th output of $\mathcal{M}_s(\mathbf{X})$ is $\mathbf{Y}_s^k$:

$$\mathbf{Y}_s^k = \mathcal{T}_s(\mathbf{X})[\mathbf{P}^k]$$

Probabilistic modeling



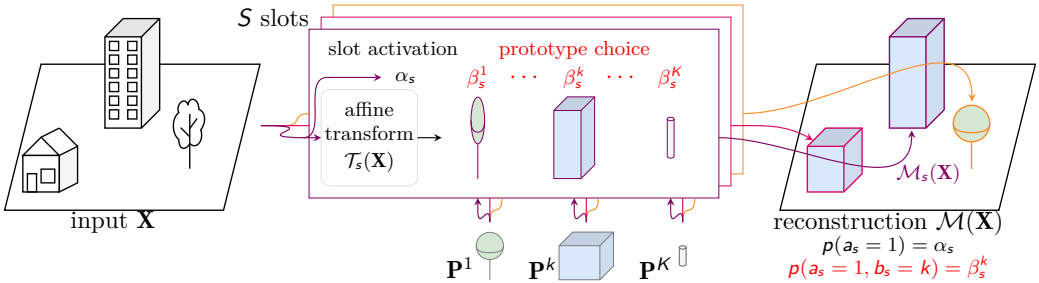
- S slot models $\mathcal{M}_s$ with activations determined by a binary variable a_s:

$$\mathcal{M}(\mathbf{X}) = \bigcup_{\substack{s=1\cdots S \\ a_s=1}} \mathcal{M}_s(\mathbf{X})$$

- The k-th output of $\mathcal{M}_s(\mathbf{X})$ is $\mathbf{Y}_s^k$:

$$\mathbf{Y}_s^k = \mathcal{T}_s(\mathbf{X})[\mathbf{P}^k]$$

Probabilistic modeling



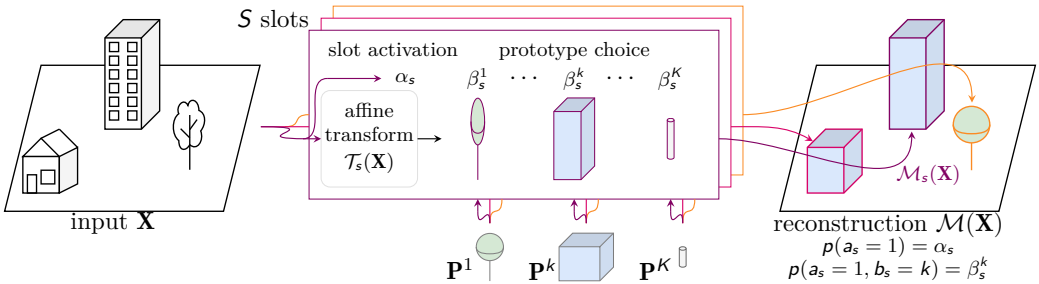
- S slot models $\mathcal{M}_s$ with activations determined by a binary variable a_s:

$$\mathcal{M}(\mathbf{X}) = \bigcup_{\substack{s=1\cdots S \\ a_s=1}} \mathcal{M}_s(\mathbf{X})$$

- $\mathcal{M}_s(\mathbf{X})$ determined by a variable $b_s \in \{1, \cdots, K\}$. If $b_s = k$, the output of $\mathcal{M}_s(\mathbf{X})$ is $\mathbf{Y}_s^k$:

$$\mathbf{Y}_s^k = \mathcal{T}_s(\mathbf{X})[\mathbf{P}^k]$$

Learnable Earth Parser

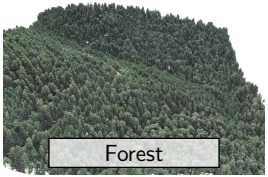


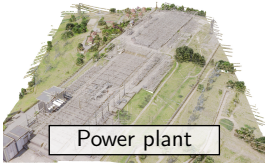
- We learn alongside the **prototypes** $\mathbf{P}$, **deformations** $\mathcal{T}$ and **activation probabilities** α, β!

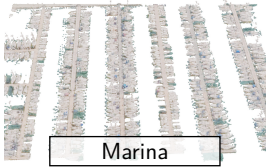
$$\mathcal{L}(\mathcal{M}) = \mathbb{E}_{\mathbf{X}}\left[\mathcal{L}_{\text{acc}}\left(\mathcal{M}, \mathbf{X}\right) + \mathcal{L}_{\text{cov}}\left(\mathcal{M}, \mathbf{X}\right)\right] + \lambda_{\text{act}}\mathcal{L}_{\text{act}}(\mathcal{M}) + \lambda_{\text{slot}}\mathcal{L}_{\text{slot}}(\mathcal{M})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

Earth Parser Dataset

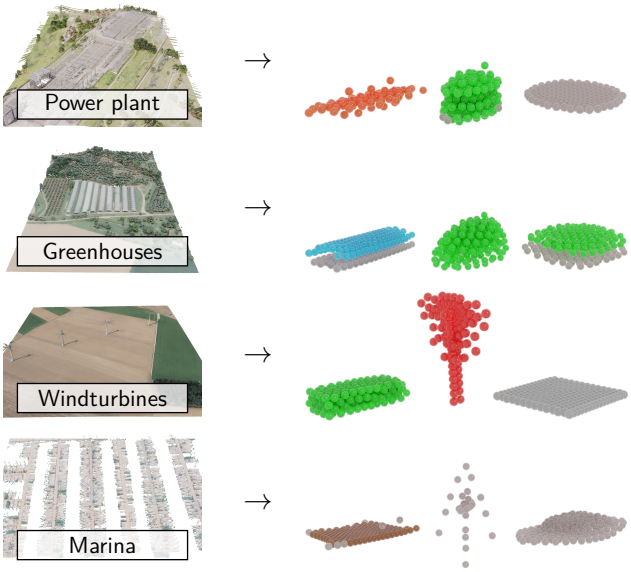
- Annotated aerial LiDAR scans in diverse environments (IGN program LiDAR-HD [2])



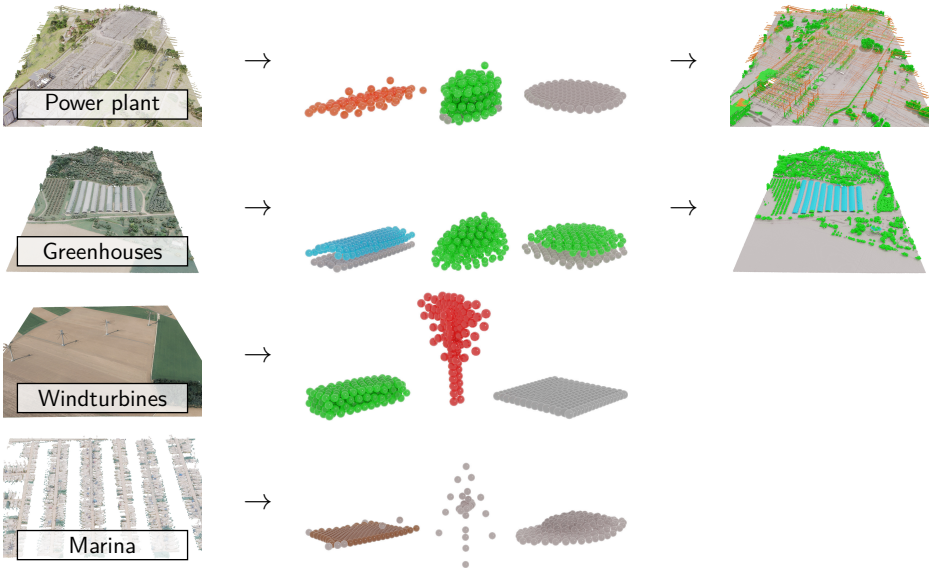




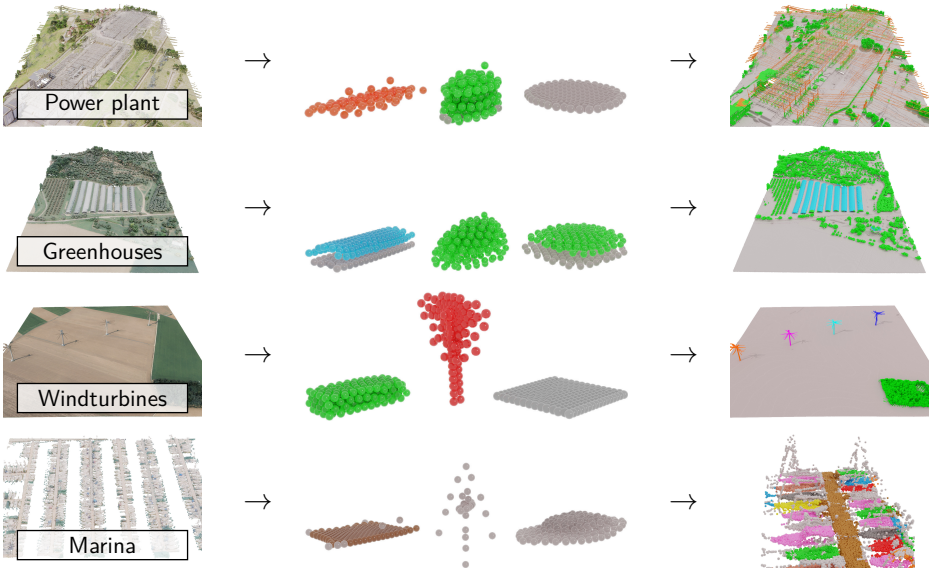
Learned Prototypes



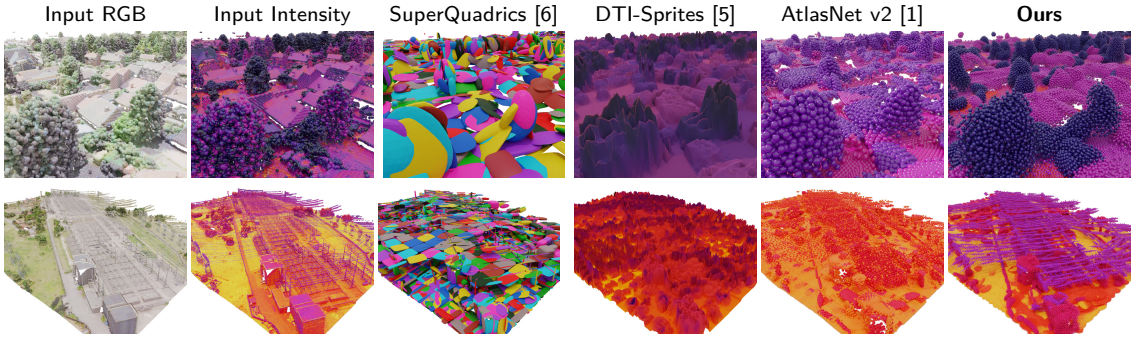
Learned Prototypes → Semantic Segmentation



Learned Prototypes → Semantic and Instance Segmentation



Quantitative results

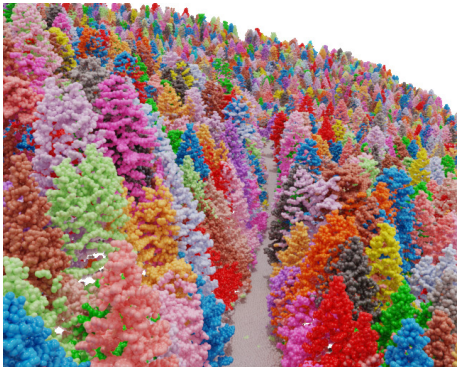


	Rec.	Semantic	Crop fields		Forest		Greenhouses		Marina		Power plant		Urban		Windturbines	
			Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.	mIoU	Cham.	mIoU
k-means (i,z) [4]	✗	✓	—	93.8	—	71.5	—	39.3	—	41.4	—	42.8	—	56.5	—	87.6
SuperQuadrics [6]	3D	✗	0.86	—	1.04	—	0.60	—	0.93	—	0.58	—	0.40	—	13.50	—
DTI-Sprites [5]	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	25.9
AtlasNet v2 [1]	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	9.47	48.1
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**	**93.4**

Conclusion

Contributions

- **Unsupervised** method for parsing complex real-world aerial scans into simple parts
- **Viewable** representation in input space
- *State-of-the-art* **semantic segmentation** performance



R. Loiseau, E. Vincent, M. Aubry, and L. Landrieu, Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, CVPR, 2024

References I

[1] T. Deprelle, T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry. Learning elementary structures for 3D shape generation and matching. NeurIPS, 2019.

[2] IGN. LiDAR-HD: a 3D mapping of France's soil and subsoil.

[3] R. Loiseau, T. Monnier, M. Aubry, and L. Landrieu. Representing shape collections with alignment-aware linear models. 3DV, 2021.

[4] J. MacQueen. Classification and analysis of multivariate observations. 5th Berkeley Symp. Math. Statist. Probability, 1967.

[5] T. Monnier, E. Vincent, J. Ponce, and M. Aubry. Unsupervised layered image decomposition into object prototypes. ICCV, 2021.

[6] D. Paschalidou, A. O. Ulusoy, and A. Geiger. Superquadrics revisited: Learning 3D shape parsing beyond cuboids. CVPR, 2019.