SCENE-Net V2: Interpretable Multiclass 3D Scene Understanding with Geometric Priors

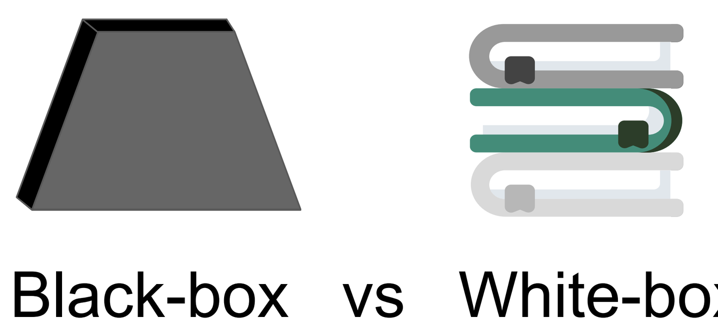
Diogo Lavado, Cláudia Soares and Alessandra Micheletti

UNIVERSITÀ DEGLI STUDI DI MILANO

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

ICML International Conference On Machine Learning

Introduction



Increasing the size of black-boxes to boost performance leads to diminishing returns, whereas white-boxes have domain constraints and simple designs.

Harnessing **geometric information** in 3D point clouds is the key to drive innovation.

We present an **interpretable** 3D semantic segmentation model that leverages **geometric priors** for feature extraction, has a **compact design** and **competitive performance**.

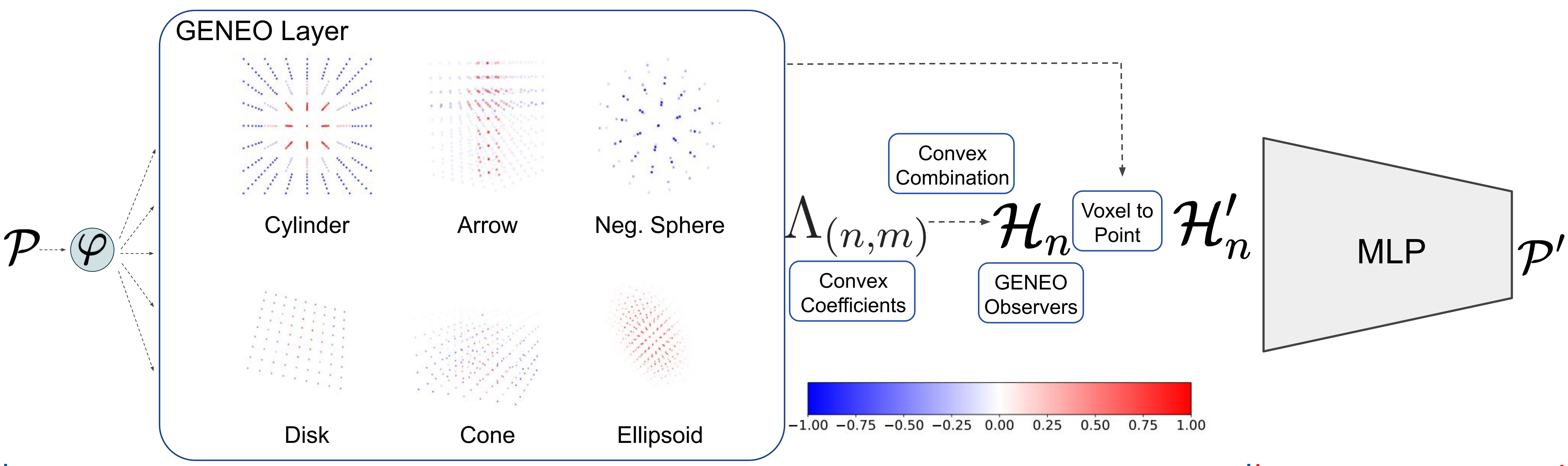
GENEOs

To build geometric priors, we leverage Group Equivariant Non-expansive Operators (GENEOs) [1].

These operators **provide a measure of the world**, analogous to patterns learnt by CNN kernels.

Unlike convolutional kernels, our GENEO-kernels are not blind to the underlying geometry of 3D scenes. **They are parameterized with meaningful features.**

SCENE-Net V2 is a gray-box model that pairs geometric interpretability and general application



White-box feature extraction phase with **540** meaningful shape parameters

Black-box classifier with **240K** parameters

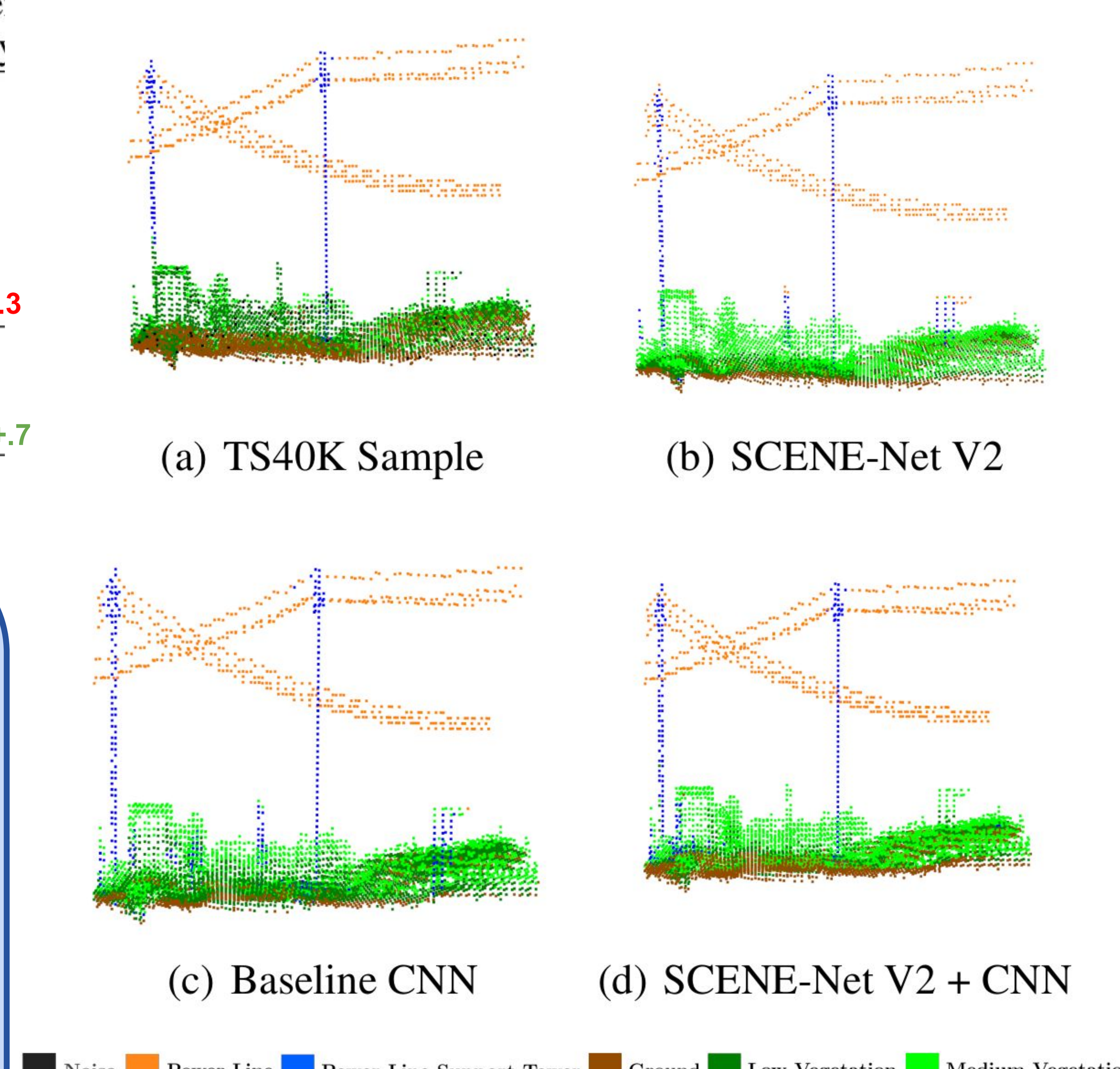
In the GENEO Layer, we instantiate *m* GENEO-kernels from 6 families of geometric priors. Such families are defined by meaningful shape parameters, such as the radius of a cylinder. They are then combined into *n* **observers** through convex combinations, creating more complex feature extraction outputs.
A CNN-based feature extraction process with an analogous architecture contains **21.4K parameters**.

Experiments

The Performance of SCENE-Net V2

Method	mIoU	#Parameters (M)	Parameter Efficiency
PointNet (Qi et al., 2017a)	44.58	0.40	7.96
PointNet++ (Qi et al., 2017b)	46.90	1.48	7.60
KPConv (Thomas et al., 2019)	57.58	14.9	8.03
RandLA-Net (Hu et al., 2020)	16.76	1.24	2.75
Point Transformer V1 (Wu et al., 2022)	62.67	12.8	8.81
Point Transformer V2 (Wu et al., 2023a)	**65.58**	46.2 x3	8.56 -.3
CNN Baseline	41.69	0.26	7.69
SCENE-Net V2 (Ours)	45.54 +3.8	**0.24** -.02	8.46
SCENE-Net V2 + CNN (Ours)	**50.21** +8.5	0.26	**9.27** +.7

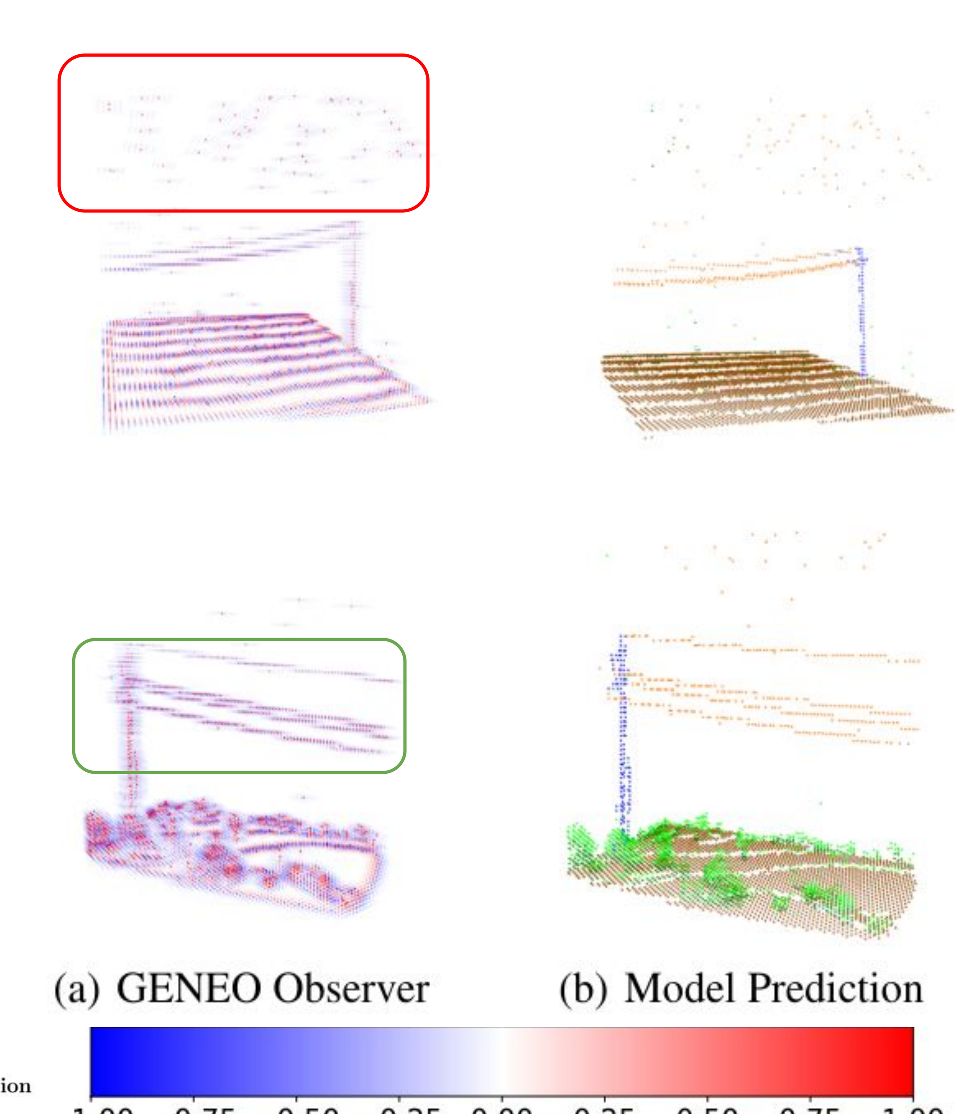
- Our model **surpasses** a comparable CNN, where its feature extraction phase includes 40x more parameters.
- In SOTA methods, PTV2 has triple the parameters of PTV1 and achieves a **less than 3% increase** in mIoU.
- Using our model as a geometric feature extractor **increases the CNN performance by more than 8% with virtually no increase in model size.**



(a) TS40K Sample (b) SCENE-Net V2

(c) Baseline CNN (d) SCENE-Net V2 + CNN

Interpretability Analysis



(a) GENEO Observer (b) Model Prediction

Future Work

- Using SCENE-Net V2 as a feature extraction tool for SOTA benchmarks;
- Applying GENEOs directly onto raw 3D point clouds;

SCAN ME

Acknowledgements: This research was funded by DM 351-2022 of the Italian Ministry of University and Research and with partial support from the Italian INDAM –GNAMPA group, by FCT I.P., through the strategic project NOVA LINCS (UIDB/04516/2020), “Sustainable Stone by Portugal” Nº C644943391-00000051/40 co-funded by Recovery and Resilience Plan and NextGeneration EU, and TaRDIS Horizon Europe (101093006).

[1] Bergomi, M. G., Frosini, P., Giorgi, D., and Quercioli, N. Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning. Nature Machine Intelligence, 1(9):423–433, 2019

PRR Plano de Recuperação e Resiliência
REPÚBLICA PORTUGUESA
Financiado pela União Europeia NextGenerationEU