3D Semantic Novelty Detection

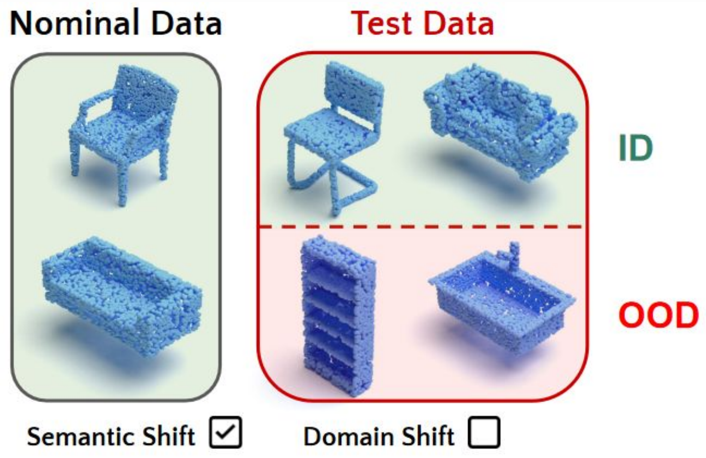
Margherita Lavena, Mattia Pecce, Marcello Vitaggio

What is Semantic Novelty Detection?

Identify if a 3D point cloud fits into known categories (in-distribution) or represents something new and unfamiliar (out-of-distribution).

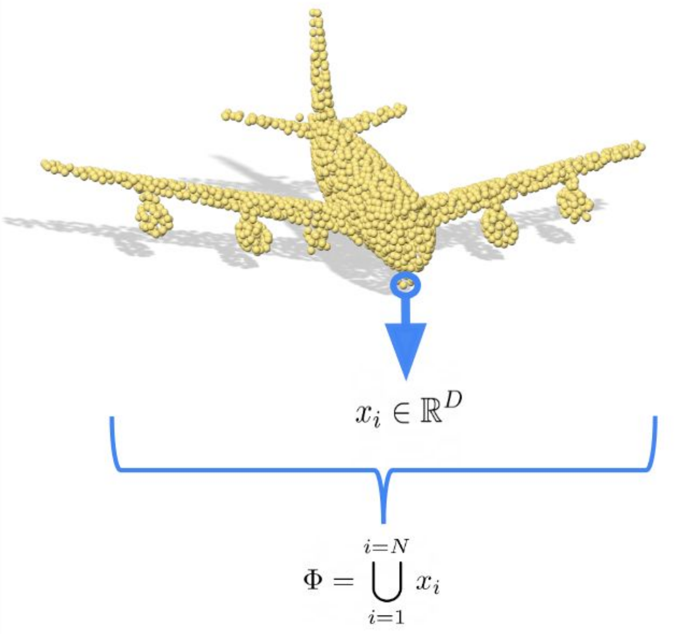
Semantic: Understanding the meaning and relationships in the data.

Novelty: Detecting new or different patterns.



Point Clouds

- Represent objects as sets of points in 3D space, each with (x, y, z) coordinates.
- Can include extra data like color or surface details.
- Created by sensors like LiDAR or TOF, now found in devices like smartphones.
- Efficient for capturing 3D shapes but are unstructured (no fixed order).
- Require specialized techniques for analysis, as traditional CNNs don't work well on them.

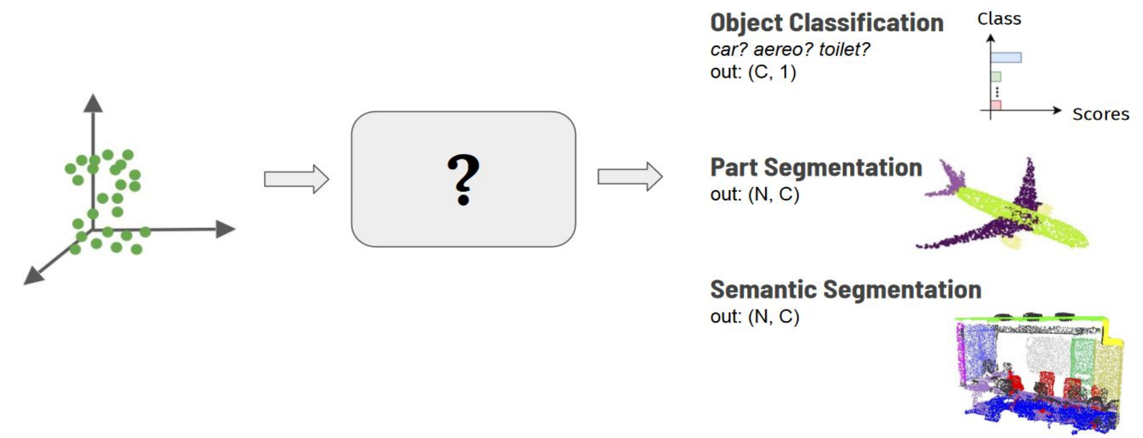


Learning from Point Clouds

Deep learning helps models learn patterns directly from raw 3D point cloud data.

These models must guarantee:

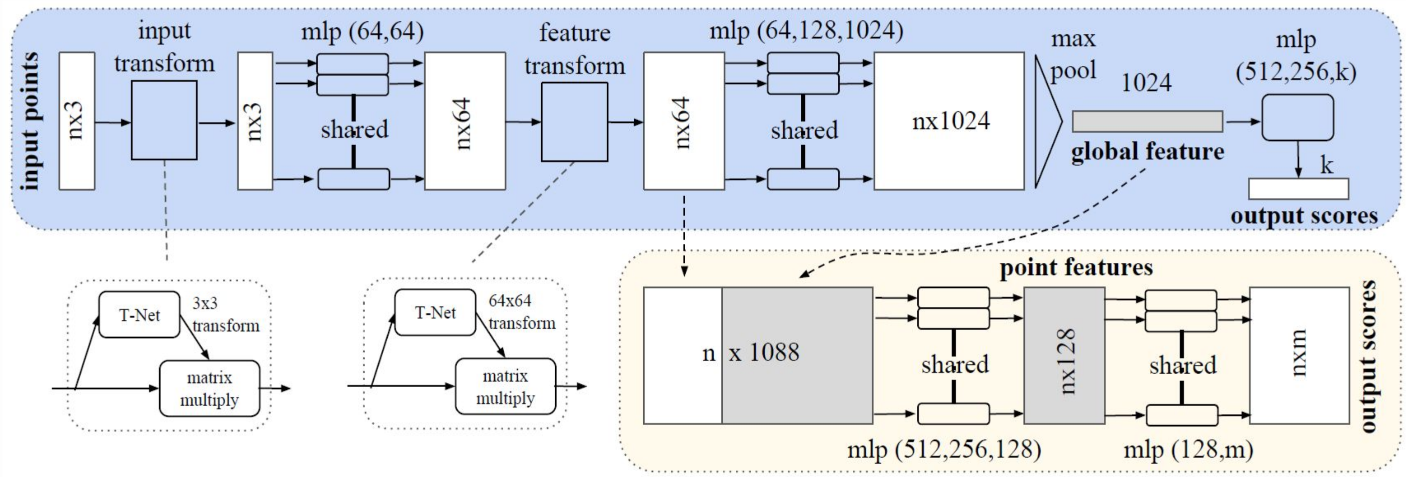
- **Permutation Invariance**: Output remains the same regardless of point order in input.
- **Geometric Transformations Invariance**: Consistent performance even with rotations or shifts in the point cloud.



PointNet

First deep learning model to directly process 3D point clouds by treating each point independently.

- **Max Pooling**: Symmetric Function that select the maximum value for each feature across all points, the results is a global descriptor that represent the entire shape.



PointNet’s Limitation and PointNet++

Limitations:

- Doesn't capture local structures at multiple scales.
- May overlook important local geometric relationships in complex scenes.

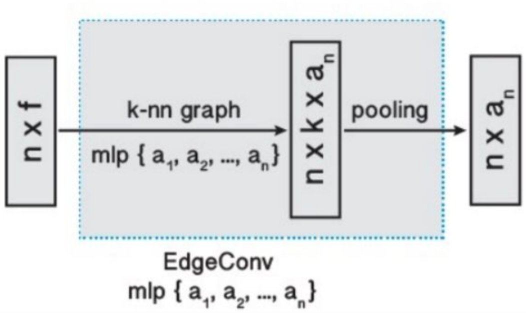
How PN2+ solve these problems?

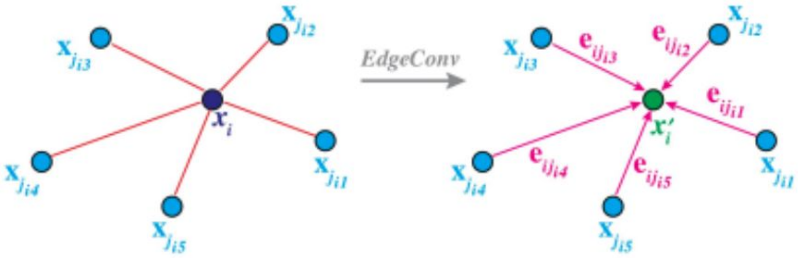
- **Hierarchical Learning**: Builds on PointNet by adding a hierarchical structure, capturing local context at various scales.
- **Multi-scale Feature Learning**: Learns fine details and global shapes, similar to how CNNs operate on images.
- **Robust to Non-Uniformity**: Adapts to varying point densities, making it more reliable with real-world data.

Dynamic Graph CNN

EdgeConv: a novel operation designed to overcome the limitations of existing techniques when dealing with point clouds.

- Operates directly on raw point cloud data
- Builds local neighborhood graphs to capture local geometric structures
- Maintains permutation invariance





3DOS: Towards 3D Open Set Learning

- This is the first benchmark focused on 3D Open Set learning, featuring multiple scenarios with increasing difficulty levels.
- It is organized into three main tracks, with the first examining the performance of existing Open Set methods on 3D data, while the other two replicate real-world deployment: **Synthetic, Real to Rea**l, **Synthetic to Real** (which is the primary focus of this project)
- Model performance is evaluated by measuring its ability to detect unknown samples, using two metrics: **AUROC** and **FPR95**.
- The 3DOS Benchmark is built upon three 3D object datasets: **ShapeNetCore**, **ModelNet40**, and **ScanObjectNN**.

3DOS: Synthetic To Real Benchmark

- **Goal:** Assess model performance in a cross-domain setting.
- **Training Data:** Synthetic point clouds sourced from ModelNet40.
- **Testing Data:** Real-world point clouds taken from ScanObjectNN.
- The datasets are split into three groups: SR1, SR2, and SR3. SR1 or SR2 are considered known, while the remaining groups are treated as unknown.
- **SR1 and SR2:** Corresponding classes from both ModelNet40 and ScanObjectNN.
- **SR3:** ScanObjectNN classes that have no direct equivalents in ModelNet40.
- The models are trained on the known classes from ModelNet40 and then tested on both known and unknown classes from ScanObjectNN.

3DOS: Evaluation Methods

- **Discriminative Approaches:** These methods involve training a typical classifier on a closed set using cross-entropy. Examples include MSP, MLS, ODIN, Energy, GradNorm, and ReAct.
- **Density and Reconstruction Approaches:** These are unsupervised techniques, such as VAE for reconstruction-based scoring, and models using Normalizing Flow (NF).
- **Outlier Exposure Using Generated OOD Data:** This evaluates the OE method's performance by utilizing synthetic OOD data created through point cloud mixup.
- **Representation and Distance-Based Approaches:** These focus on learning feature embeddings to detect novel categories, including methods like ARPL+CS, Cosine Proto, CE (L2), SupCon, and SubArcFace.

3DOS: Evaluation Methods

- **MSP (Maximum Softmax Probability):** the logits go through a softmax function before selecting the maximum of the obtained class probabilities as the score.
- **CE (L2) :** use the inverse of the distance from the nearest training sample as each test sample normality score.

3DOS Baselines Results

Method	PointNet++				DGCNN			
	SR 1		SR 2		SR 1		SR 2	
	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95
MSP	0.8078	0.8122	0.6794	0.8882	0.7096	0.9101	0.6026	0.9068
MLS	0.8189	0.7578	0.6598	0.8711	0.6804	0.9315	0.6196	0.8877
Entropy	0.8102	0.7841	0.6817	0.8744	0.7066	0.9205	0.6055	0.8972
Odin	0.8145	0.7529	0.6836	0.8278	0.6805	0.9315	0.6203	0.8873
Energy	0.8179	0.7731	0.6580	0.8525	0.6788	0.9309	0.6203	0.8972
GradNorm	0.7731	0.8202	0.6748	0.8492	0.6538	0.9474	0.6013	0.8844
ReAct	0.8178	0.7768	0.6615	0.8421	0.6703	0.9242	0.6304	0.8972
CE L^2	0.8131	0.7713	0.7306	0.8206	0.6836	0.8765	0.6182	0.9605
Cosine (Proto)	0.7079	0.9370	0.6905	0.8844	0.6333	0.9272	0.5993	0.9353

Table 1. Performance comparison of different methods on PointNet and DGCNN across SR1 and SR2 metrics for AUROC and FPR95.

PointNet++ Failure Case Results

	Bed	Toilet	Table+Desk	Display	OOD
Chair	7	34	44	1	86
Shelf	2	0	36	39	28
Door	0	0	0	22	44
Sink	11	0	4	0	9
Sofa	0	2	4	0	19

Table 8. Misclassifications on SR2 with Euclidean distance (threshold = 0.2723)

	Bed	Toilet	Table+Desk	Display	OOD
Chair	1	24	32	0	47
Shelf	1	0	24	20	17
Door	0	0	1	11	53
Sink	1	5	11	0	47
Sofa	8	0	3	0	4

Table 6. Misclassifications on SR2 with MSP (threshold = 0.9989)

Large scale pre-trained models

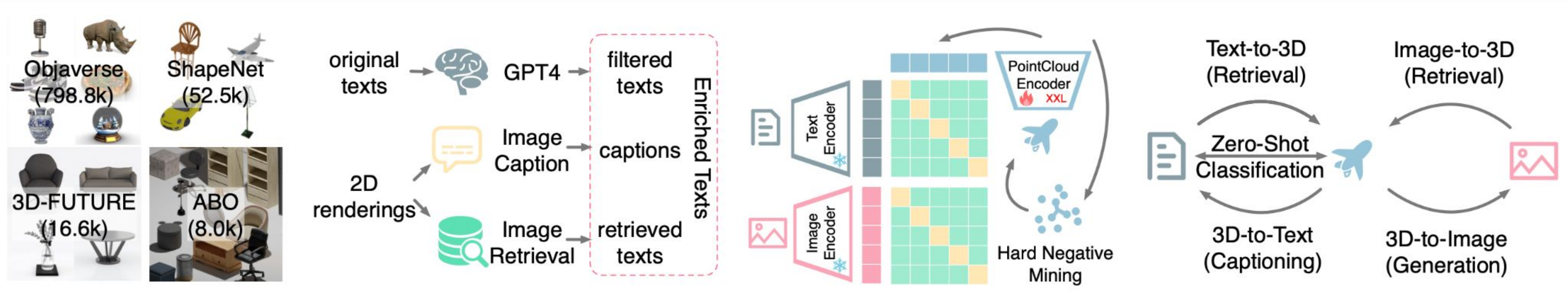
Instead of training specialized models, can we leverage a large pre-trained model for the task of 3D Semantic Novelty Detection?

- **Reduced computational cost**
- **Scalability** and **generalization**
- **Robustness**
- Improved performance on **low-data scenarios**

Openshape

Method introduced in 2023 in the paper "OpenShape: Scaling Up 3D Shape Representation Towards Open-World Understanding".

- **Multi-modal contrastive learning** across text, images and point clouds.
- **Text enrichment** and **hard negative mining** to develop robust and generalizable 3D shape embeddings



Ensamble Dataset   Text FIltering & Enrichment   Cross-Modal Alignment   Cross-Modal Applications

Large scale pre-trained models

	PointNet++		DGCNN		PointBERT	
	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95
CE L^2	0.7718	0.7959	0.6509	0.9185	0.5404	0.9645
	+23	-17	+11	-5		

Further analysis and interventions are needed for this model to be competitive in identifying semantic novelty in unseen data.

Future study

Involve other models proposed in Openshape

Compare PointBERT-ViTg14 with other checkpoints

Fine-tune the model

Thank you for your attention!

Do you have any questions?

Margherita Lavena, s313642@studenti.polito.it
Mattia Pecce, s309825@studenti.polito.it
Marcello Vitaggio, s318904@studenti.polito.it