3D Semantic Novelty Detection

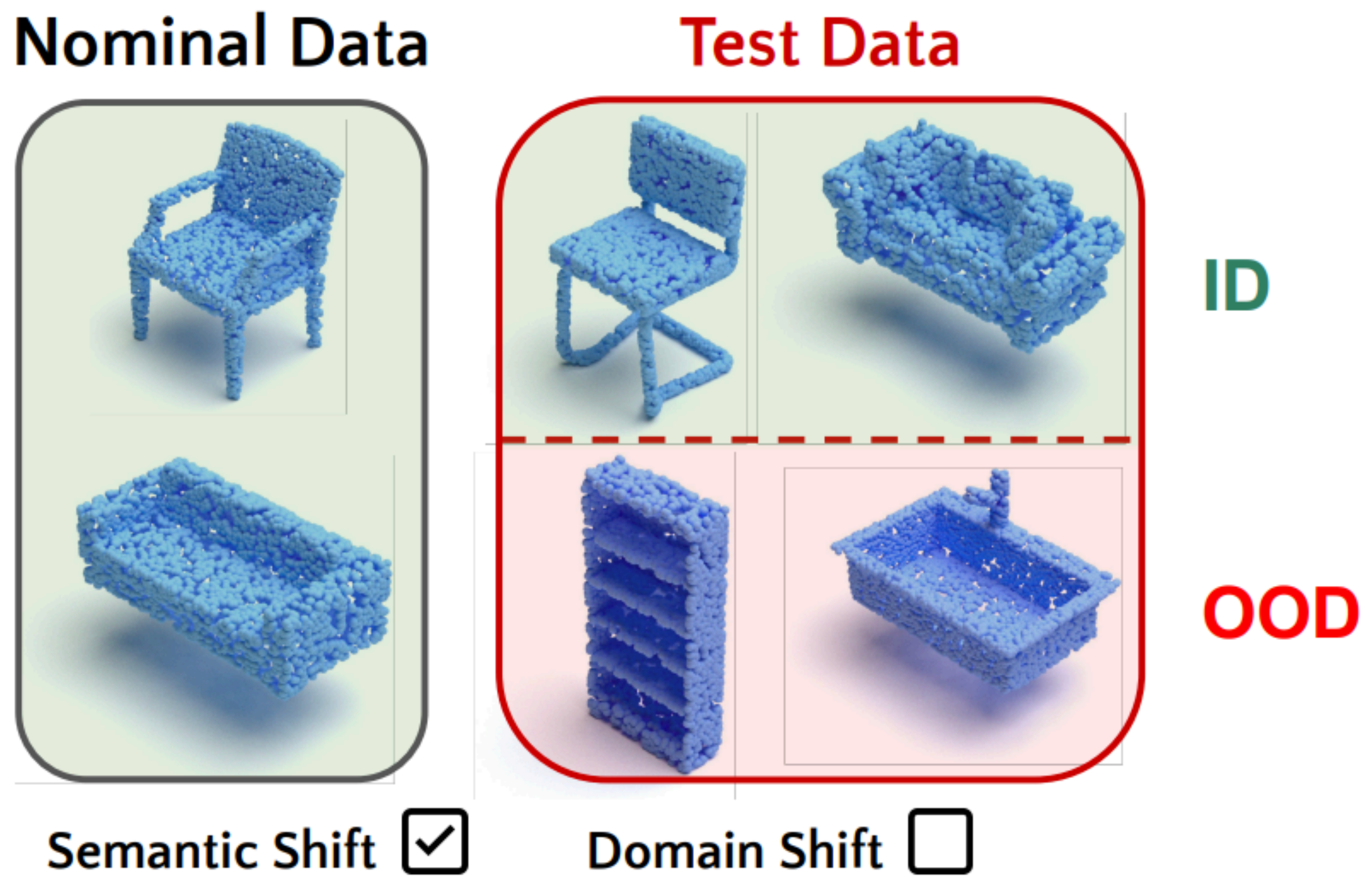
Gabriele Quaranta, Giulio Maselli, George Florin Eftime

s318944@studenti.polito.it, s306125@studenti.polito.it, s303483@studenti.polito.it

What is Semantic Novelty Detection

Semantic novelty detection is the process of identifying new or previously unseen patterns, concepts, or information within a dataset or data stream.

- Semantic: refers to the meaning or interpretation of data, focusing on understanding relationships and associations between information rather than just its structure.
- Novelty: involves identifying new, original, or significantly different patterns or concepts within data.



3DOS: Towards 3D Open Set Learning - Benchmarking and Understanding Semantic Novelty Detection on Point Clouds, Alliegro et al in NeurIPS 2022

3D Point Clouds

Point clouds are three-dimensional datasets composed of individual points.

They are generated using technologies like LiDAR or stereo cameras and find applications in 3D modeling, computer vision, and GIS.

As technology advances and new applications are found, processing point clouds still presents several challenges:

- Lack of structure
- Absence of clear ordering
- Requirement for meticulous manual annotation

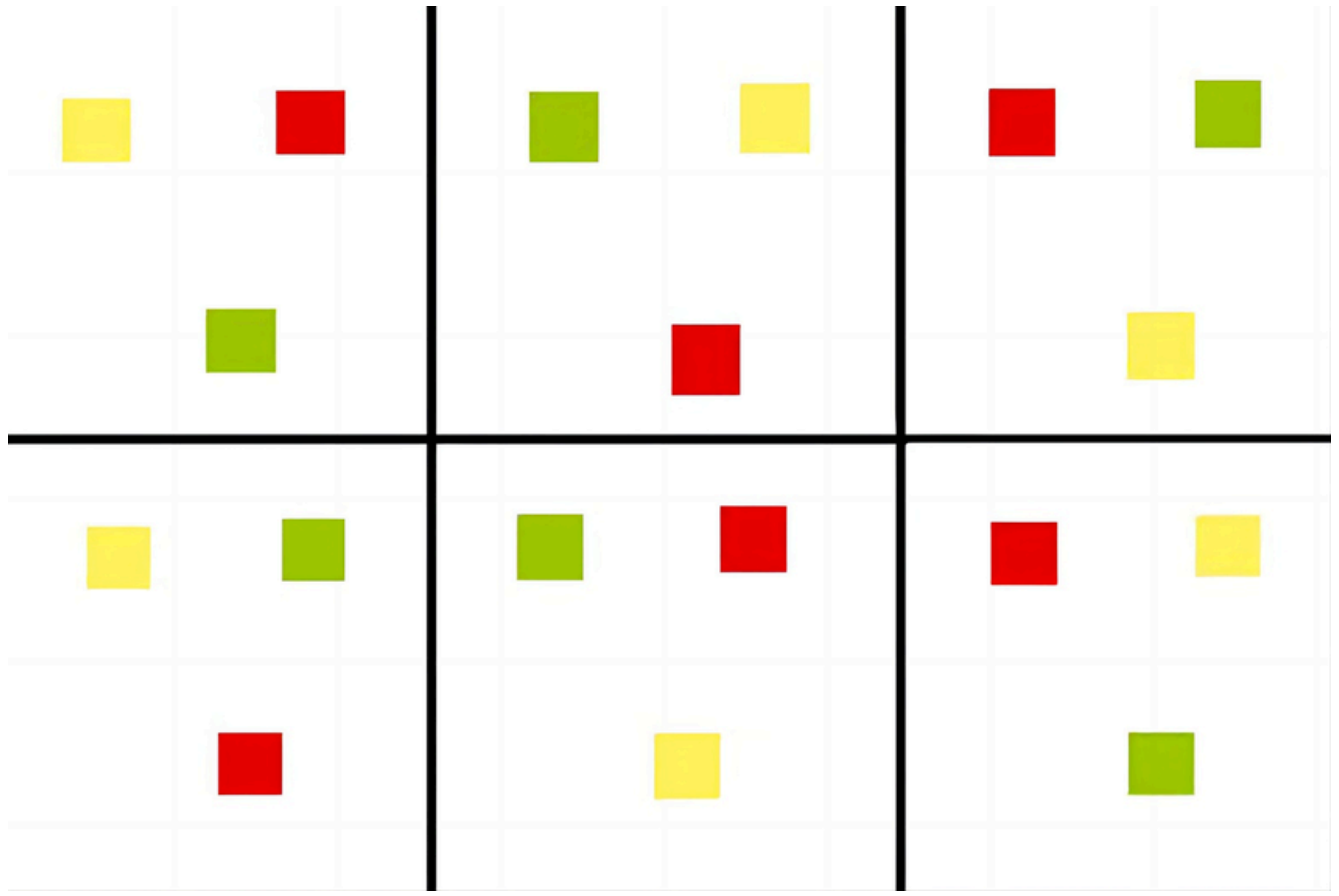
https://www.open3d.org/docs/release/tutorial/geometry/pointcloud.html

Deep Learning for PC

Point cloud processing and understanding 3D shapes pose significant challenges, with deep learning showing promising potential.

Given the challenges a deep learning model for 3D point clouds must ensure:

- Permutation Invariance: It should be invariant to the N! possible permutations of the input.
- Geometric Transformations Invariance: The results of downstream tasks should remain consistent despite any rigid transformations applied.



Permutation invariance of a point cloud. For the three points in the figure, six different permutations exist, but the shapes they express are the same.

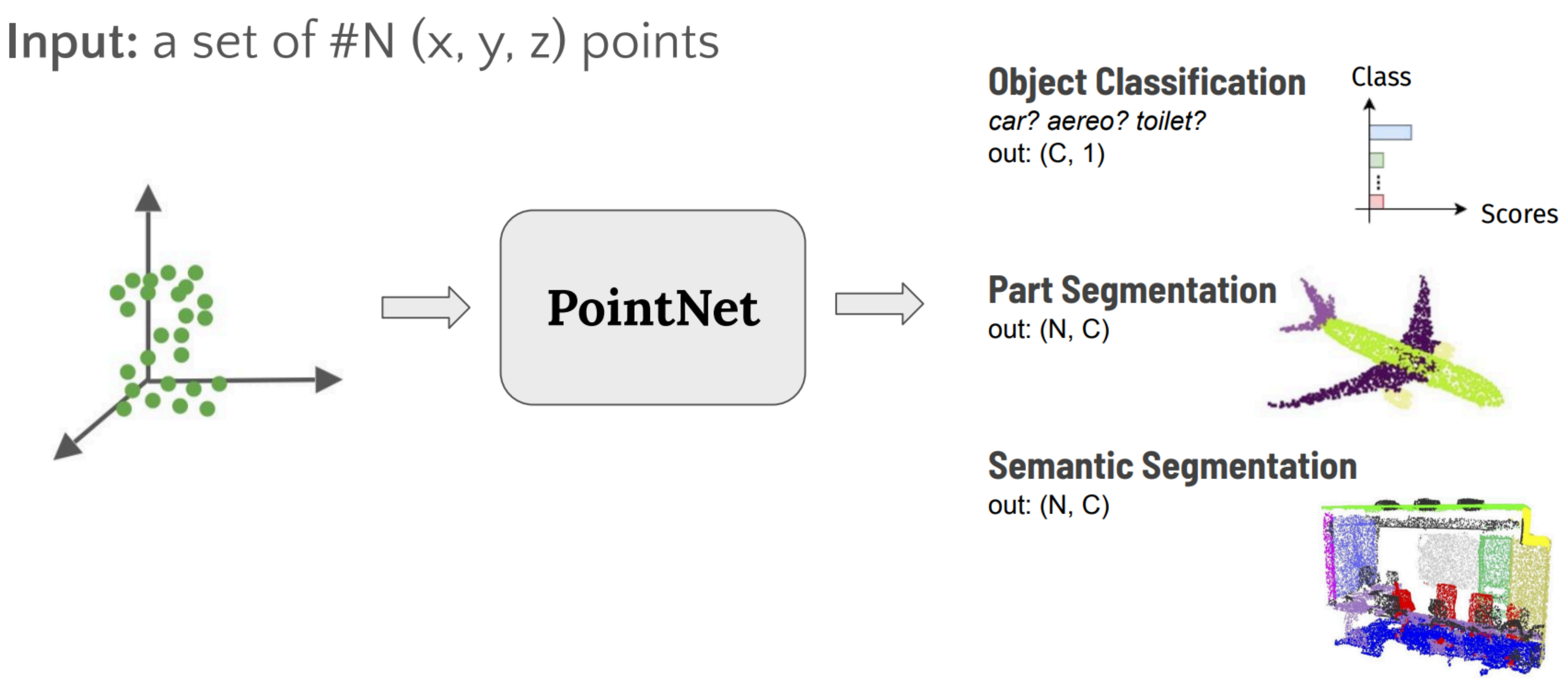
Zhang, Ansi & Li, Song & Wu, Jie & Li, Shaobo & Zhang, Bao. (2023). Exploring Semantic Information Extraction From Different Data Forms in 3D Point Cloud Semantic Segmentation.

PointNet: Deep Learning for Point Clouds

PointNet is a unified architecture designed to process point clouds directly, producing either class labels for the entire input or segment/part labels for individual points.

Key Innovation:
- Utilization of a single symmetric function: max pooling.
- Enables learning of optimization criteria for selecting significant points.

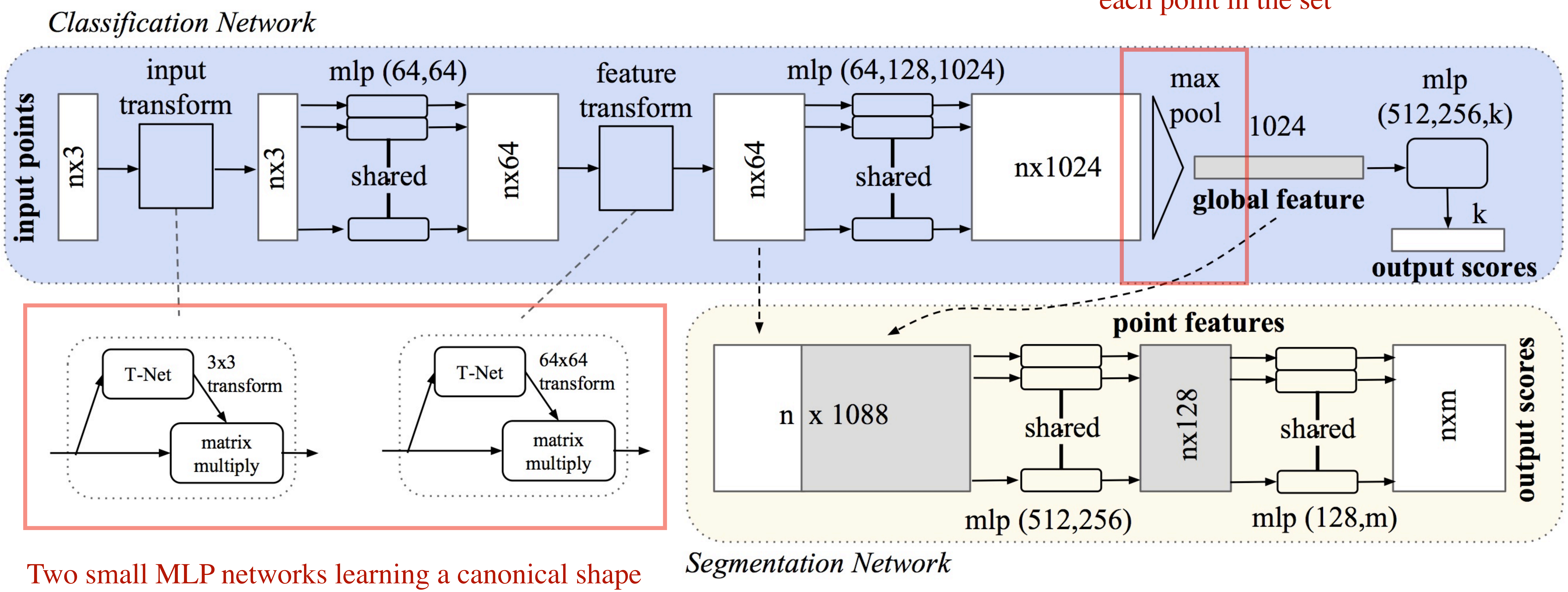
Insights:
- Learns to summarize input point clouds by a sparse set of key points.
- Resembles the skeleton of objects based on visualization.



PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al in CVPR 2017

PointNet: Deep Learning for Point Clouds

Max pooling aggregates the information from each point in the set



Two small MLP networks learning a canonical shape transformation

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al in CVPR 2017

Dynamic Graph CNN

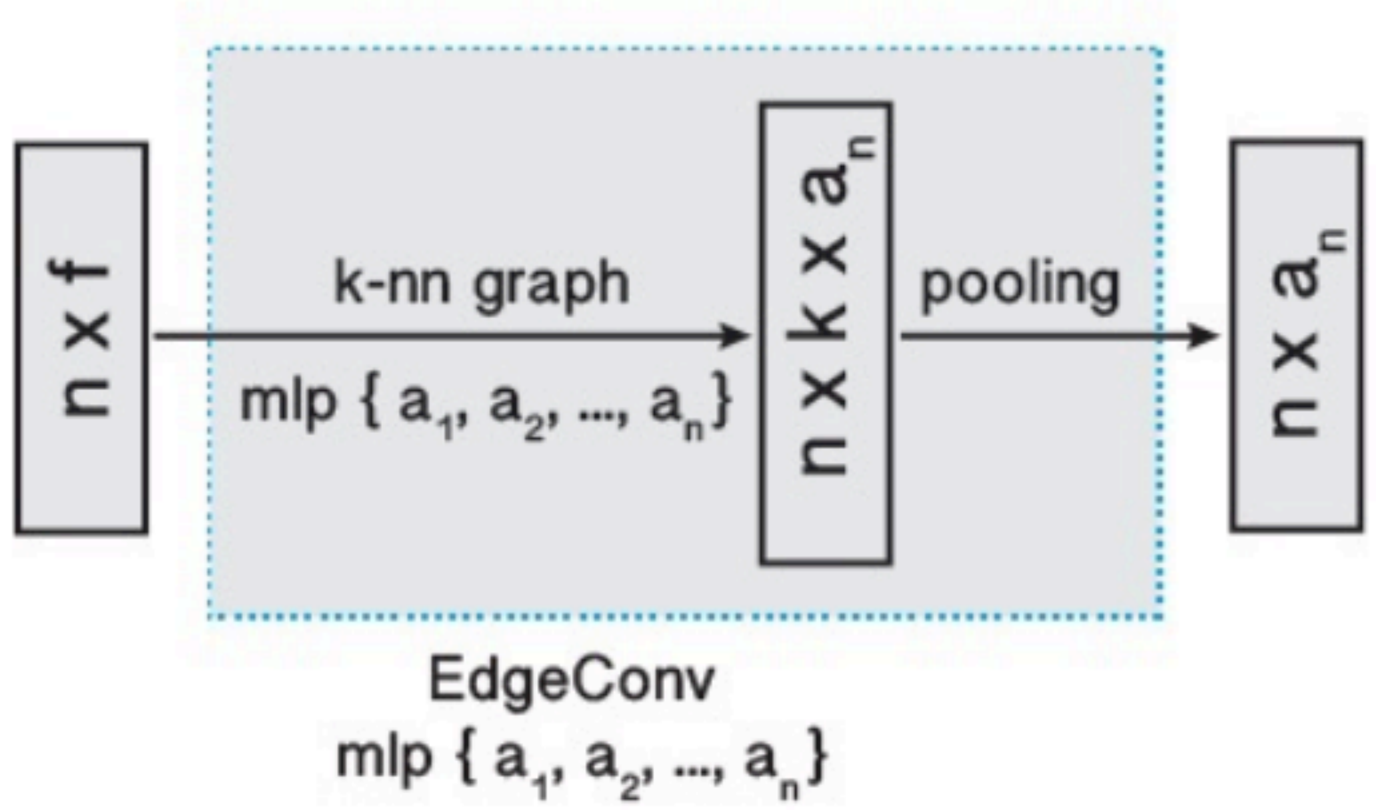
Deep neural networks struggle with the irregularity of point clouds, limiting their ability to capture local geometric relationships.

Solution: EdgeConv, a novel operation designed to address the limitations of existing techniques.

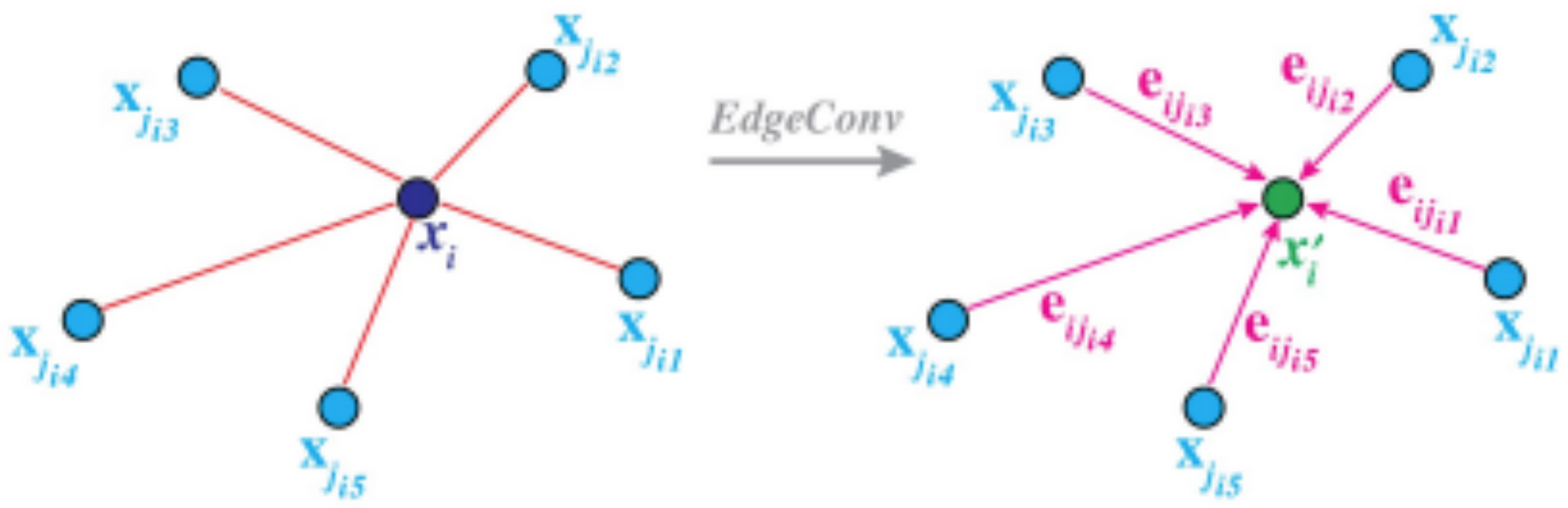
Key Features:
- Operates directly on raw point cloud data.
- Captures local geometric structure while maintaining permutation invariance.

Advantages:
- Maintains permutation invariance.
- Constructs a local graph to learn embeddings for edges.
- Enables grouping of points in both Euclidean and semantic space.

Dynamic Graph CNN



EdgeConv block. Takes as input a tensor, computes edge features for each point by applying a multi-layer perceptron (mlp) and generates a tensor of shape after pooling among neighboring edge features.



The EdgeConv operation. The output of EdgeConv is calculated by aggregating the edge features associated with all the edges emanating from each connected vertex.

Dynamic Graph CNN for Learning on Point Clouds, YUE WANG et al.

3DOS: Towards 3D Open Set Learning

The first benchmark for 3D Open Set learning, considering several settings with increasing levels of difficulty.

It includes three main tracks, the first is meant to investigate the behavior of existing Open Set methods on 3D data, the other two are designed to simulate real-world deployment conditions:

- Synthetic
- Real to Real
- Synthetic to Real (the focus of our studies)

The evaluation of the ability to detect unknown samples in test data is done using two metrics: AUROC and FPR95.

3DOS Benchmark is based on top of three 3D objects datasets: ShapeNetCore, ModelNet40 and ScanObjectNN.

3DOS: Towards 3D Open Set Learning - Benchmarking and Understanding Semantic Novelty Detection on Point Clouds, Alliegro et al in NeurIPS 2022

3DOS: Synthetic To Real Benchmark

Objective: Evaluate model performance in cross-domain scenario.

Training Data: Synthetic point clouds from ModelNet40.
Testing Data: Real-world point clouds from ScanObjectNN.

The datasets are divided in three categories: SR1, SR2 and SR3. SR1 or SR2 treated as known; others as unknown.

- SR1 and SR2: Matching classes from ModelNet40 and ScanObjectNN.
- SR3: ScanObjectNN classes without direct counterparts in ModelNet40.

The models are trained on known classes from ModelNet40 then evaluated on both known and unknown classes from ScanObjectNN.

3DOS: Towards 3D Open Set Learning - Benchmarking and Understanding Semantic Novelty Detection on Point Clouds, Alliegro et al in NeurIPS 2022

3DOS: Evaluation Methods

The evaluation encompasses several families of methods:

Discriminative Methods: Standard closed set classifier trained with cross-entropy. (MSP, MLS, ODIN, Energy, GradNorm, and ReAct)

Density and Reconstruction Based Methods: Unsupervised approaches such as VAE with reconstruction-based scoring and Normalizing Flow (NF) models.

Outlier Exposure with OOD Generated Data: Assessing the performance of the OE approach using fake OOD data produced via point cloud mixup.

Representation and Distance Based Methods: Involve learning feature embeddings for identifying novel categories. (ARPL+CS, Cosine proto, CE (L2), SupCon, and SubArcFace).

3DOS: Towards 3D Open Set Learning - Benchmarking and Understanding Semantic Novelty Detection on Point Clouds, Alliegro et al in NeurIPS 2022

3DOS Baselines and DGCNN Failure Cases

PointNet++

	Synth to Real Benchmark - PointNet++ [34]					
	SR 1 (easy)		SR 2 (hard)		Avg	
Method	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓
MSP [18]	81.0	79.6	70.3	86.7	75.6	83.2
MLS	82.1	76.6	67.6	86.8	74.8	81.7
ODIN [27]	81.7	77.3	70.2	84.4	76.0	80.8
Energy [28]	81.9	77.5	67.7	87.3	74.8	82.4
GradNorm [21]	77.6	80.1	68.4	86.3	73.0	83.2
ReAct [41]	81.7	75.6	67.6	87.2	74.6	81.4
VAE [31]	-	-	-	-	-	-
NF	78.0	84.4	74.7	84.2	76.4	84.3
OE+mixup [19]	71.2	89.7	60.3	93.5	65.7	91.6
ARPL+CS [7]	**82.8**	74.9	68.0	89.3	75.4	82.1
Cosine proto	79.9	**74.5**	**76.5**	**77.8**	**78.2**	**76.1**
CE (L^2)	79.7	84.5	75.7	80.2	77.7	82.3
SubArcFace [11]	78.7	84.3	75.1	83.4	76.9	83.8

PointNet++ Results from original paper

DGCNN

DGCNN - SR1		
Method	AUROC	FPR95
MSP[3]	0.7234	0.9175
MLS	0.6964	0.9632
Entropy	0.7181	0.9327
Distance[2]	0.6228	0.9391
Distance (Prototypes)	0.6039	0.9454
Cosine[11]	0.6629	0.9048
Cosine (Prototypes)	0.5652	0.9086
ODIN[5]	0.6965	0.9619
Energy[7]	0.6940	0.9708
GradNorm[4]	0.6908	0.9543
React (+Energy)[9]	0.6799	0.9721

DGCNN - SR2		
Method	AUROC	FPR95
MSP	0.6409	0.8855
MLS	0.6100	0.8985
Entropy	0.6420	0.8890
Distance	0.6631	0.8536
Distance (Prototypes)	0.5622	0.9575
Cosine	0.6679	0.8796
Cosine (Prototypes)	0.6485	0.9103
ODIN	0.6108	0.8973
Energy	0.6069	0.8973
GradNorm	0.5614	0.9244
React (+Energy)	0.6088	0.8949

DGCNN Failure Cases

Confidence	Predicted	Actual
1.0000	desk	chair
0.9998	bed	chair
1.0000	display	door
1.0000	bed	shelf
1.0000	bed	sink
0.9999	desk	shelf
0.9999	desk	sink
1.0000	bed	sofa

MSP Metric distinct failure instances

Distance	Predicted	Actual
0.1534	desk	door
0.1649	display	sink
0.1942	toilet	shelf
0.1726	bed	chair

Euclidean Distance distinct failure instances

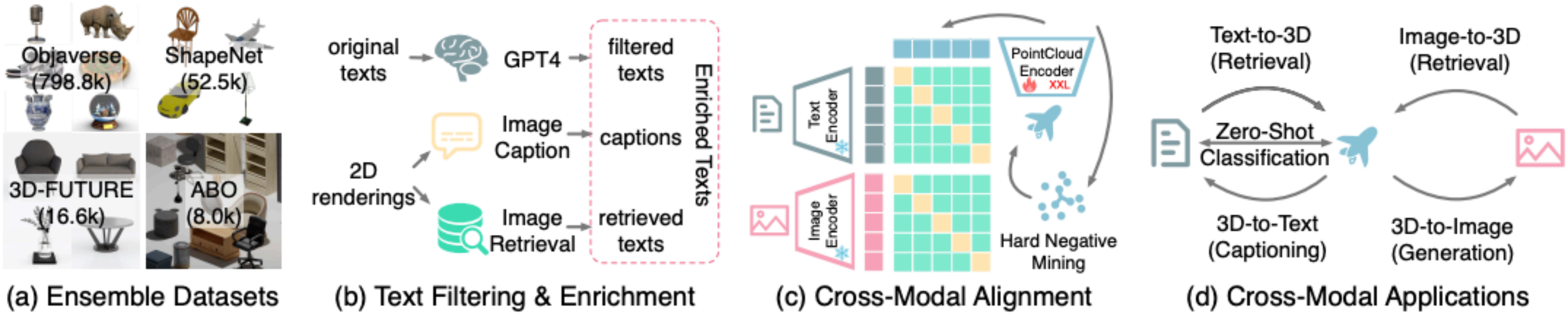
3DOS: Towards 3D Open Set Learning - Benchmarking and Understanding Semantic Novelty Detection on Point Clouds, Alliegro et al in NeurIPS 2022

OpenShape

OpenShape is a method for learning multi-modal joint representations of text, image, and point clouds, focusing on scaling up 3D representations for open-world 3D shape understanding.

It ensembles multiple 3D datasets, filters noisy text descriptions and explores strategies for scaling 3D backbone networks.

Evaluation on zero-shot 3D classification benchmarks demonstrates OpenShape's high accuracy on Objaverse-LVIS and ModelNet40 datasets.



Liu et al. (2023). OpenShape: Scaling Up 3D Shape Representation Towards Open-World Understanding.

OpenShape Evaluation with 3DOS

OpenShape G14 - SR1		
Method	AUROC	FPR95
MSP	0.5386	0.9505
MLS	0.5309	0.9581
Entropy	0.5375	0.9505
Distance	0.5307	0.9454
Distance (Prototypes)	0.5012	0.9581
Cosine	0.5564	0.9277
Cosine (Prototypes)	0.5302	0.9594
ODIN	0.5404	0.9492
Energy	0.5284	0.9492
GradNorm	0.5427	0.9645
React (+Energy)	0.5003	0.9645

OpenShape G14 - SR2		
Method	AUROC	FPR95
MSP	0.4498	0.9514
MLS	0.4676	0.9371
Entropy	0.4438	0.9530
Distance	0.5370	0.9076
Distance (Prototypes)	0.5132	0.9251
Cosine	0.5413	0.9203
Cosine (Prototypes)	0.5167	0.9514
ODIN	0.4465	0.9562
Energy	0.4802	0.9355
GradNorm	0.4443	0.9530
React (+Energy)	0.5132	0.9227

OpenShape seems to struggle to generalize on real-world test data, leading to variable performance AUROC and FPR95 metrics across different scenarios.

Final results

	Synth to Real Benchmark - PointNet++ [34]					
	SR 1 (easy)		SR 2 (hard)		Avg	
Method	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓
MSP [18]	81.0	79.6	70.3	86.7	75.6	83.2
MLS	82.1	76.6	67.6	86.8	74.8	81.7
ODIN [27]	81.7	77.3	70.2	84.4	76.0	80.8
Energy [28]	81.9	77.5	67.7	87.3	74.8	82.4
GradNorm [21]	77.6	80.1	68.4	86.3	73.0	83.2
ReAct [41]	81.7	75.6	67.6	87.2	74.6	81.4
VAE [31]	-	-	-	-	-	-
NF	78.0	84.4	74.7	84.2	76.4	84.3
OE+mixup [19]	71.2	89.7	60.3	93.5	65.7	91.6
ARPL+CS [7]	**82.8**	74.9	68.0	89.3	75.4	82.1
Cosine proto	79.9	**74.5**	**76.5**	**77.8**	**78.2**	**76.1**
CE (L^2)	79.7	84.5	75.7	80.2	77.7	82.3
SubArcFace [11]	78.7	84.3	75.1	83.4	76.9	83.8

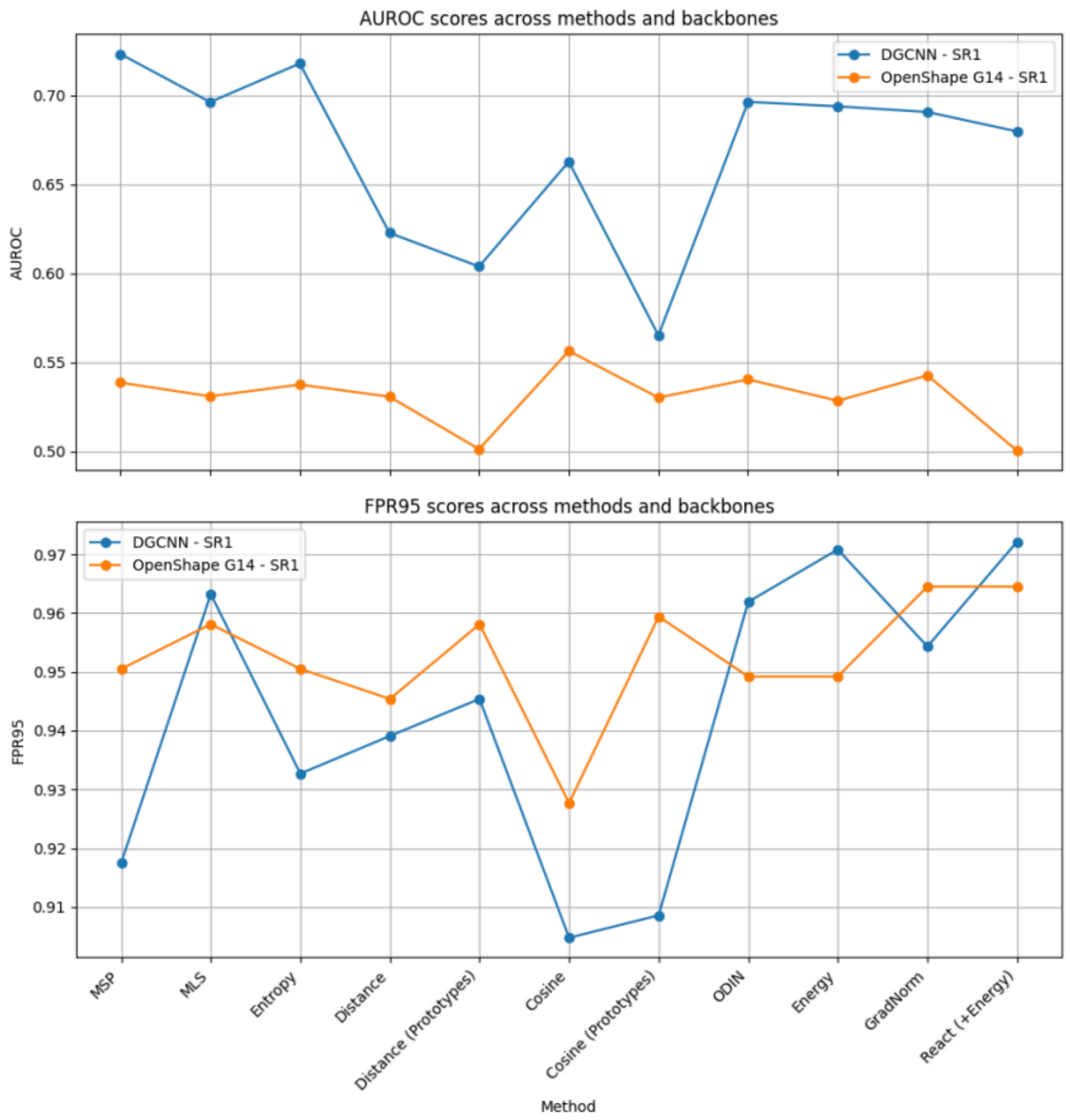
OpenShape G14 - SR1		
Method	AUROC	FPR95
MSP	0.5386	0.9505
MLS	0.5309	0.9581
Entropy	0.5375	0.9505
Distance	0.5307	0.9454
Distance (Prototypes)	0.5012	0.9581
Cosine	0.5564	0.9277
Cosine (Prototypes)	0.5302	0.9594
ODIN	0.5404	0.9492
Energy	0.5284	0.9492
GradNorm	0.5427	0.9645
React (+Energy)	0.5003	0.9645

DGCNN - SR1		
Method	AUROC	FPR95
MSP[3]	0.7234	0.9175
MLS	0.6964	0.9632
Entropy	0.7181	0.9327
Distance[2]	0.6228	0.9391
Distance (Prototypes)	0.6039	0.9454
Cosine[11]	0.6629	0.9048
Cosine (Prototypes)	0.5652	0.9086
ODIN[5]	0.6965	0.9619
Energy[7]	0.6940	0.9708
GradNorm[4]	0.6908	0.9543
React (+Energy)[9]	0.6799	0.9721

OpenShape G14 - SR2		
Method	AUROC	FPR95
MSP	0.4498	0.9514
MLS	0.4676	0.9371
Entropy	0.4438	0.9530
Distance	0.5370	0.9076
Distance (Prototypes)	0.5132	0.9251
Cosine	0.5413	0.9203
Cosine (Prototypes)	0.5167	0.9514
ODIN	0.4465	0.9562
Energy	0.4802	0.9355
GradNorm	0.4443	0.9530
React (+Energy)	0.5132	0.9227

DGCNN - SR2		
Method	AUROC	FPR95
MSP	0.6409	0.8855
MLS	0.6100	0.8985
Entropy	0.6420	0.8890
Distance	0.6631	0.8536
Distance (Prototypes)	0.5622	0.9575
Cosine	0.6679	0.8796
Cosine (Prototypes)	0.6485	0.9103
ODIN	0.6108	0.8973
Energy	0.6069	0.8973
GradNorm	0.5614	0.9244
React (+Energy)	0.6088	0.8949

Final results



The results are promising but DGCNN still outperforms OpenShape in many cases.

Detecting semantic novelty in unseen data with OpenShape requires further refinement, possibly fine-tuning on top of the benchmark datasets.

Thank You for your attention!

Gabriele Quaranta, Giulio Maselli, George Florin Eftime

s318944@studenti.polito.it, s306125@studenti.polito.it, s303483@studenti.polito.it