3D Semantic Novelty Detection

Hossein Kakavand
Samaneh Gharedagh Sani

s308581@studenti.polito.it, s309100@studenti.polito.it

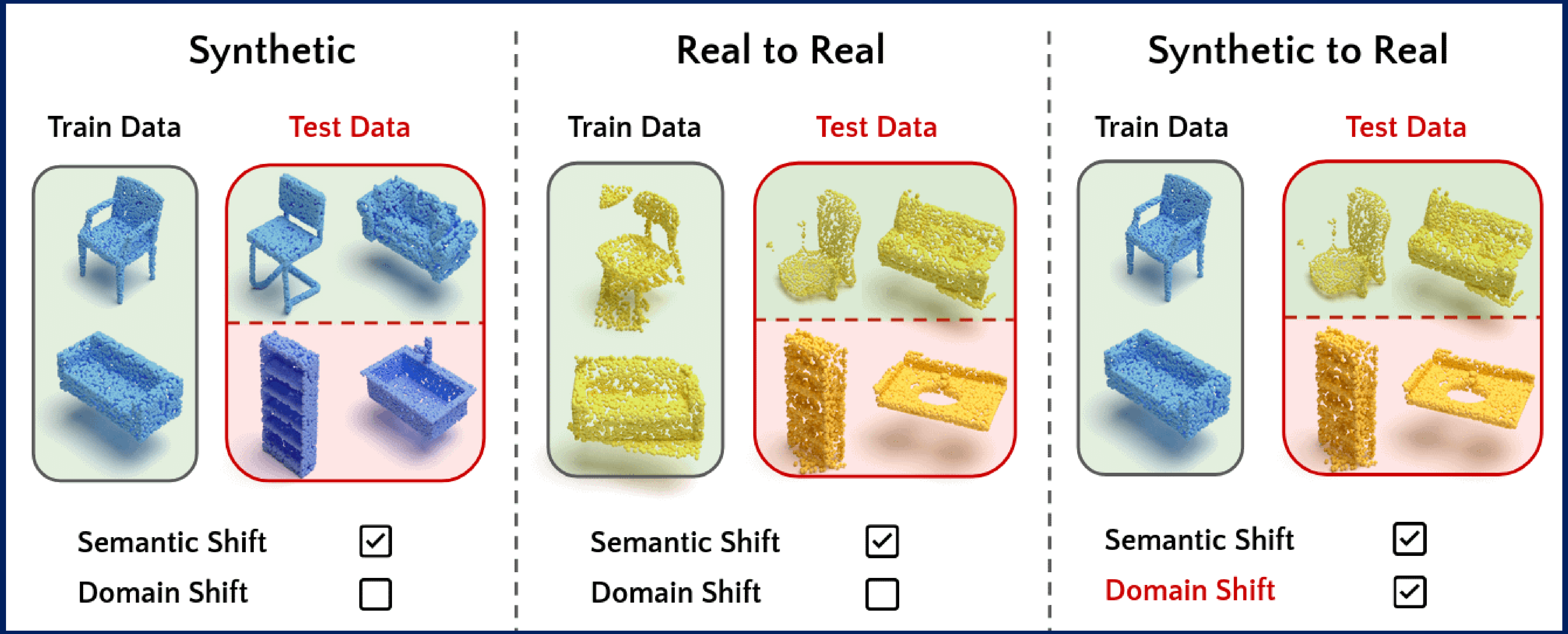
Politecnico di Torino

Introduction to Semantic Novelty Detection

Semantic Novelty Detection is essential in data analysis and machine learning. It involves finding new, unseen patterns or information in datasets, which is crucial for applications like anomaly detection and discovering new trends.

- Semantic: Focuses on understanding the meaning and relationships within data, going beyond just structure. This helps reveal important trends and patterns that might be missed otherwise.
- Novelty: Identifies new, original, or significantly different patterns within data.

Combining semantic analysis with novelty detection helps uncover valuable insights and respond proactively to data changes.

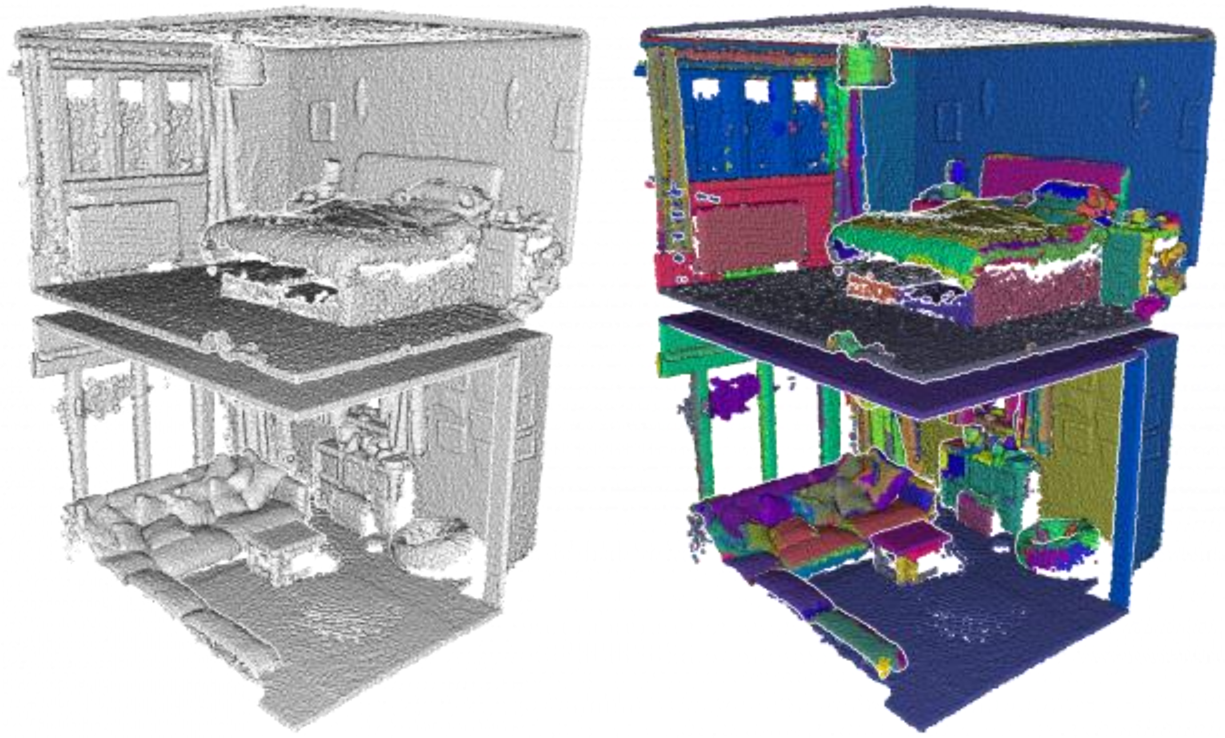


3DOS: Towards 3D Open Set Learning - Benchmarking and Understanding Semantic Novelty Detection on Point Clouds, Alliegro et al in NeurIPS 2022

Understanding 3D Point Clouds

3D Point Clouds consist of numerous individual points in 3D space, captured using technologies like LiDAR and stereo cameras. They are used in 3D modeling, computer vision, and GIS.

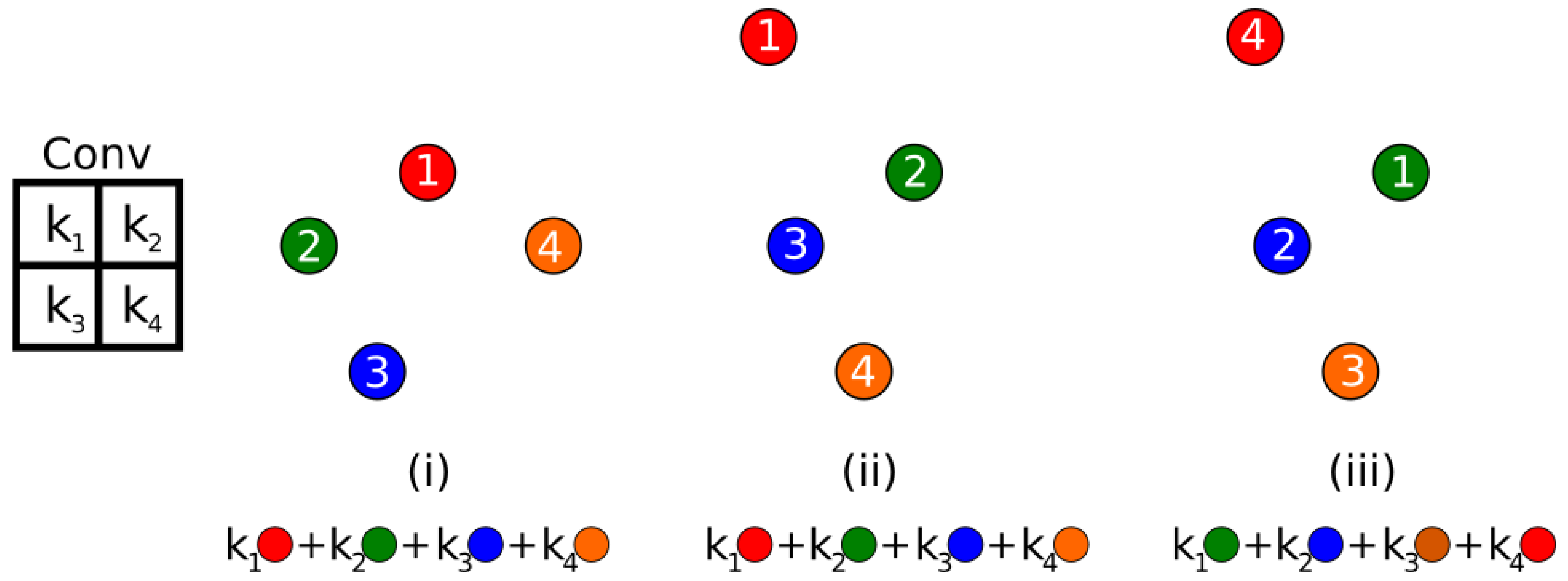
- **Applications:**
  - **3D Modeling**: Used in architecture, engineering, and entertainment.
  - **Computer Vision**: Helps in object recognition and scene reconstruction.
  - **GIS**: Creates accurate geographic representations.
- **Challenges:**
  - Lack of inherent structure.
  - No clear ordering.
  - Need for detailed manual annotation.



Deep Learning for 3D Point Clouds

Deep learning has shown significant promise in tackling the challenges of processing and understanding 3D shapes within point clouds. For a deep learning model to be effective in processing 3D point clouds, it must satisfy the following key requirements:

- **Geometric Transformations Invariance:** The model's output for downstream tasks must remain consistent even when rigid transformations (like rotation, translation, or scaling) are applied to the input data.
- **Permutation Invariance:** The model must be invariant to the N! possible permutations of the input points.

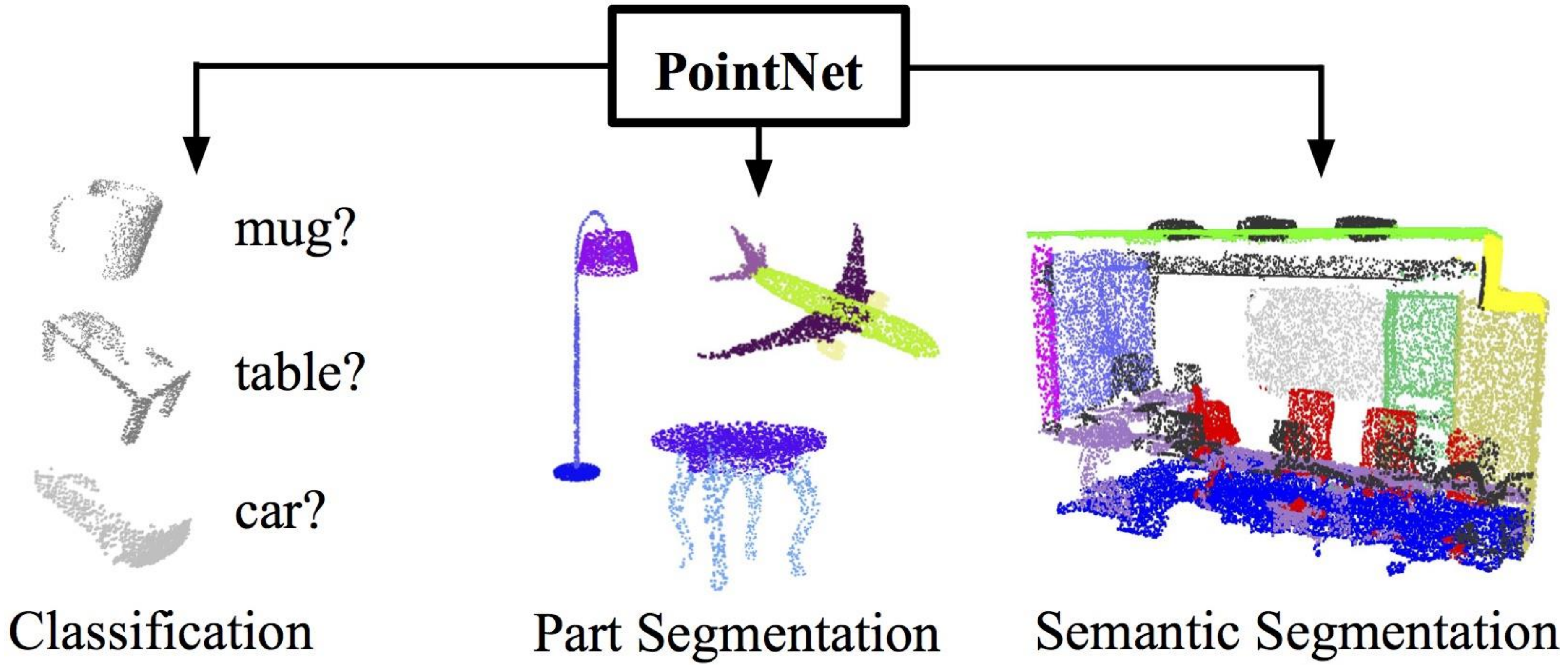


Point cloud problem statement

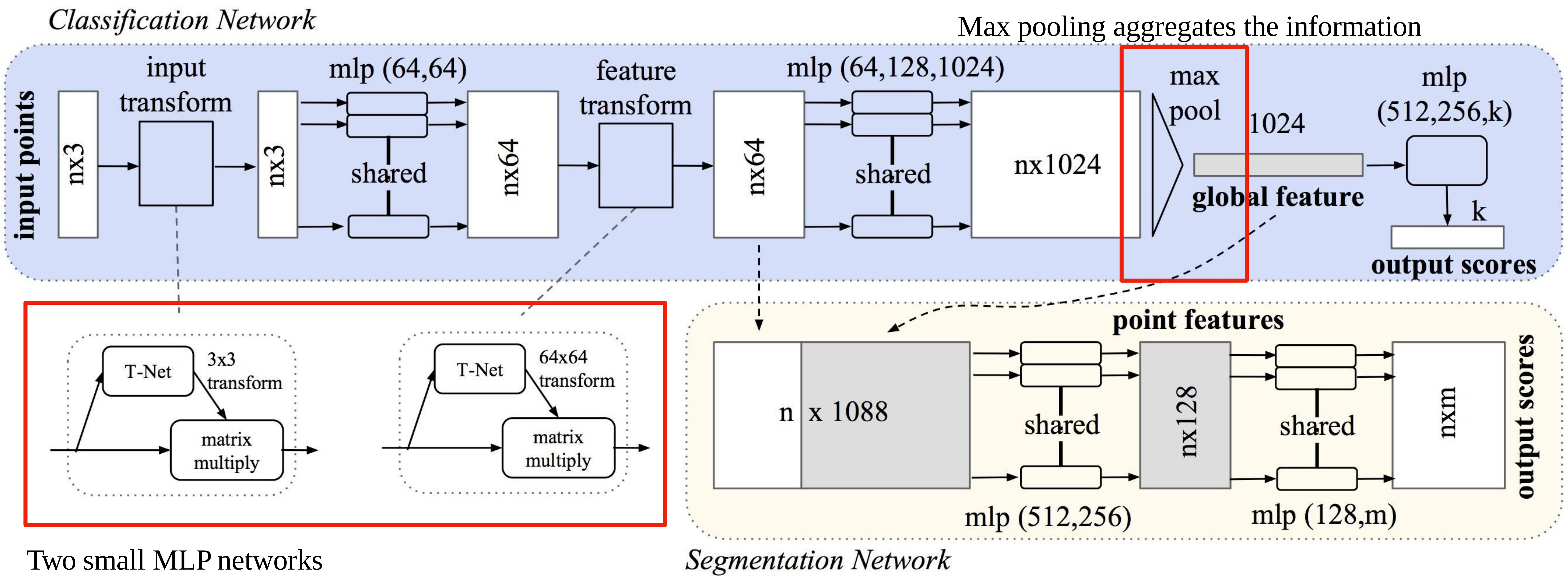
PointNet: Deep Learning for Point Clouds

PointNet is a deep learning architecture for direct point cloud processing, providing class or segment labels.

- **Key Innovation:**
  PointNet employs a single symmetric function, max pooling, to aggregate information from all points in the input data.
- **Insights:**
  - PointNet effectively summarizes complex 3D shapes by identifying a sparse set of key points.
  - Resembles the skeleton of objects based on visualization.



Max pooling aggregates the information



Two small MLP networks

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

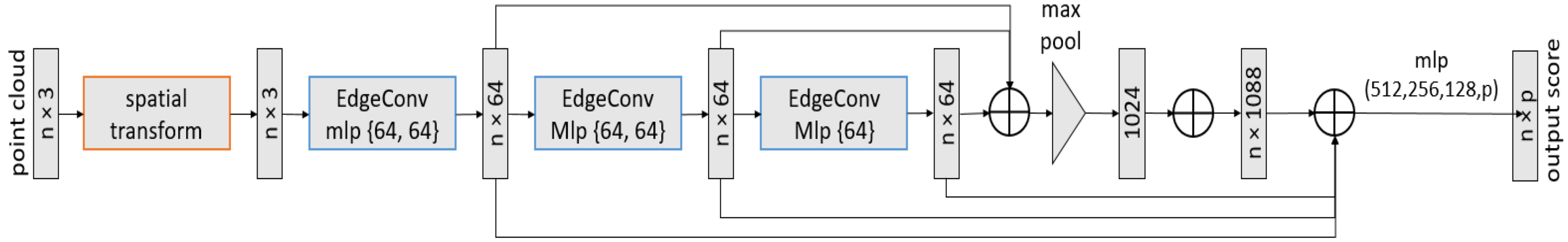
PointNet: Deep Learning for Point Clouds

DGCNN: Enhancing 3D Point Cloud Analysis

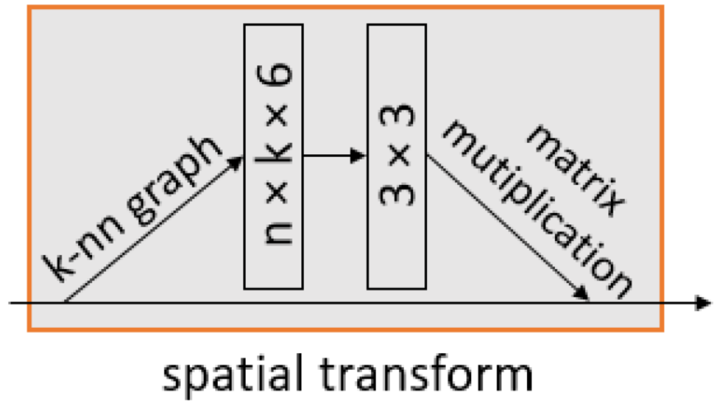
Dynamic Graph CNN (DGCNN) addresses the irregularity of point clouds using EdgeConv.

- Key Features:
  - **EdgeConv Operation:** Directly processes raw point cloud data.
  - **Permutation Invariance:** Maintains performance consistency by ensuring invariance to input point permutations.
  - **Geometric Structure Capture:** Learns edge embeddings to understand spatial relationships and geometric context.
- Advantages:
  - **Robust Local Feature Learning:** Groups points in both Euclidean and semantic spaces, crucial for tasks like segmentation.
  - **Reliable Performance:** Consistently accurate regardless of point order.
  - **Versatility in Applications:** Suitable for object recognition, scene understanding, and robotic perception.

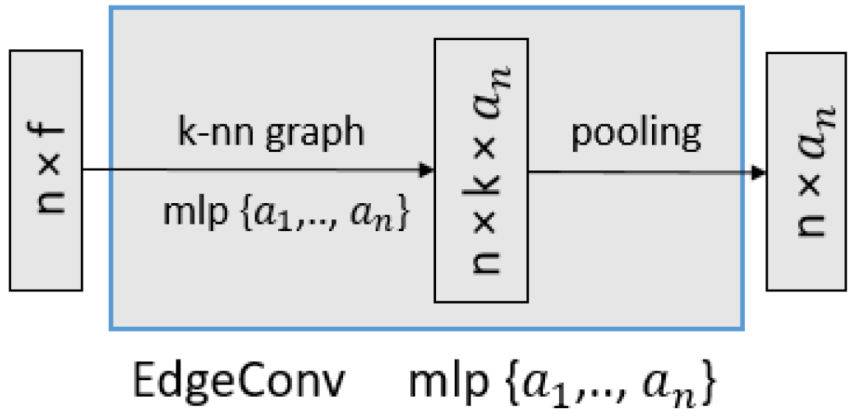
(**Dynamic Graph CNN for Learning on Point Clouds, YUE WANG et al**)



(a) The DGCNN semantic segmentation network architecture



(**b**) Spatial transformation block



(**c**) EdgeConv block

EdgeConv block: The EdgeConv block processes a tensor input by computing edge features for each point using a multi-layer perceptron (MLP).

Dynamic Graph CNN

3DOS: Towards 3D Open Set Learning

The 3DOS Benchmark provides a comprehensive framework for evaluating 3D Open Set learning methods, focusing on semantic novelty detection in point clouds. It covers various difficulty levels and evaluates both synthetic and real-world scenarios.

- **Benchmark Tracks:**
  - **Synthetic:** Evaluates models using synthetic data for controlled comparisons.
  - **Real to Real:** Tests models with real-world data to assess generalization.
  - **Synthetic to Real:** Examines how well models trained on synthetic data perform on real-world data, simulating real-world deployment.
- **Evaluation Metrics:**
  - **AUROC (Area Under the Receiver Operating Characteristic Curve):** Measures the ability to distinguish between known and unknown samples.
  - **FPR95 (False Positive Rate at 95% True Positive Rate):** Assesses false positive rate at 95% true positive rate, indicating precision and reliability.
- **Datasets:**
  Uses ShapeNetCore, ModelNet40, and ScanObjectNN to provide diverse 3D objects for thorough model evaluation.

3DOS: Synthetic To Real Benchmark

Objective: Assess model performance in a cross-domain context.

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

- **Dataset Categories:**
  - **SR1 and SR2:** Include matching classes from both ModelNet40 and ScanObjectNN.
  - **SR3:** Comprises ScanObjectNN classes that do not have direct counterparts in ModelNet40 .

- **Methodology:**
  Models are trained on known classes from ModelNet40 and evaluated on both known and unknown classes from ScanObjectNN.

Figure 4: Visualization of the object categories in each of the sets of the Synthetic to Real Benchmark. **SR1**: chair, shelf, door, sink, sofa. **SR2**: bed, toilet, desk, table, display. **SR3**: bag, bin, box, pillow, cabinet.

3DOS: Evaluation Methods

The evaluation of 3D Open Set learning methods in the context of semantic novelty detection involves several diverse approaches. Here are the primary families of methods used in the evaluation:

- **Discriminative Methods:** These methods involve standard classifiers trained with cross-entropy loss, which are typically used in closed set classification. Examples include MSP, MLS, ODIN, Energy, GradNorm, and ReAct.
- **Representation and Distance Based Methods:** Involve learning feature embeddings for identifying novel categories. (ARPL+CS Cosine proto CE (L2) SupCon and SubArcFace).
- **Density and Reconstruction-Based Methods:** Unsupervised approaches such as Variational Autoencoders (VAE) and Normalizing Flow (NF) models fall under this category. These methods focus on reconstructing the input data and use reconstruction error as a metric for detecting novelty.
- **Outlier Exposure with OOD Generated Data:** This approach assesses the performance of models trained with Outlier Exposure (OE), where the models are exposed to artificially generated out-of-distribution (OOD) data during training.

3DOS Baselines and DGCNN Failure Cases

DGCNN - SR2 (TAR1+TAR2)

Method	AUROC	FPR95
MSP	0.6352	0.8801
MLS	0.6739	0.8497
Entropy	0.6413	0.8459
Distance	0.632	0.9101
Distance(Prototypes)	0.5979	0.9462
Cosine	0.6856	0.8716
Cosine(Prototypes)	0.6352	0.9115
ODIN	0.675	0.8492
Energy	0.6755	0.852
GranNorm	0.6405	0.8535
React(+Energy)	0.6761	0.852

DGCNN SR1

DGCNN - SR1 (TAR1+TAR2)

Method	AUROC	FPR95
MSP	0.7212	0.9028
MLS	0.6997	0.9107
Entropy	0.7203	0.9028
Distance	0.6863	0.8575
Distance(Prototypes)	0.6712	0.8826
Cosine	0.6829	0.8789
Cosine(Prototypes)	0.626	0.9193
ODIN	0.6998	0.9101
Energy	0.6984	0.9089
GranNorm	0.6846	0.9272
React(+Energy)	0.6896	0.9028

DGCNN SR2

MSP FAILURE CASES SR1

Predict	Actual
Door	Chair
Door	Chair
Door	Bookshelf
Chair	Bookshelf
Chair	Bookshelf

MSP FAILURE CASES SR2

Predict	Actual
Bed	Desk
Bed	Desk
Bed	Desk
Bed	Desk
Bed	Desk

MSP Metric distinct failure instances

EUCLIDEAN DISTANCE FAILURE SR1

Predict	Actual
Door	Chair
Sink	Chair
Door	Chair
Chair	Door
Door	Bookshelf

EUCLIDEAN DISTANCE FAILURE SR2

Predict	Actual
Bed	Desk
Bed	Desk
Bed	Desk
Bed	Desk
Bed	Desk

Euclidean Distance distinct failure instances

	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++

OpenShape is a technique designed to develop joint representations across text, images, and point clouds, with a focus on improving 3D shape comprehension in open-world contexts.

This method combines several 3D datasets, removes noisy text data, and investigates strategies for scaling 3D backbone networks.

Performance evaluations on zero-shot 3D classification benchmarks reveal that **OpenShape** achieves excellent accuracy on the Objaverse-LVIS and ModelNet40 datasets.

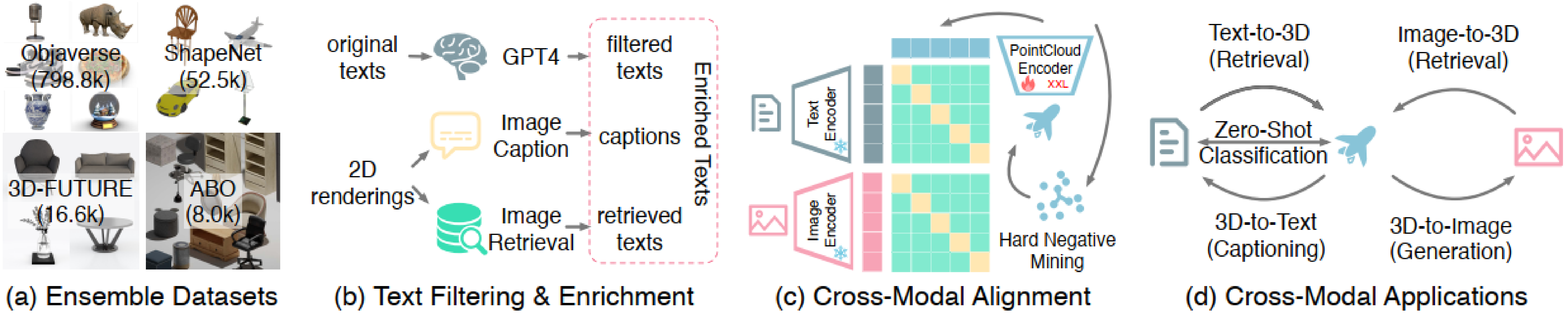


Figure 2: (a) We ensemble four public 3D shape datasets, resulting in 876k shapes that encompass diverse categories and concepts. (b) We propose three strategies to automatically filter and enrich the noisy texts in the original datasets. (c) We train a 3D point cloud encoder to align the 3D shape embedding space with the CLIP's text and image embedding spaces. We perform cross-modal contrastive learning with scaled 3D backbones and hard negative mining. (d) OpenShape embeddings can be easily integrated with other CLIP-based models, enabling various cross-modality tasks.

OpenShape

OPENSHAPE - B32 - SR1 (TAR1+TAR2)

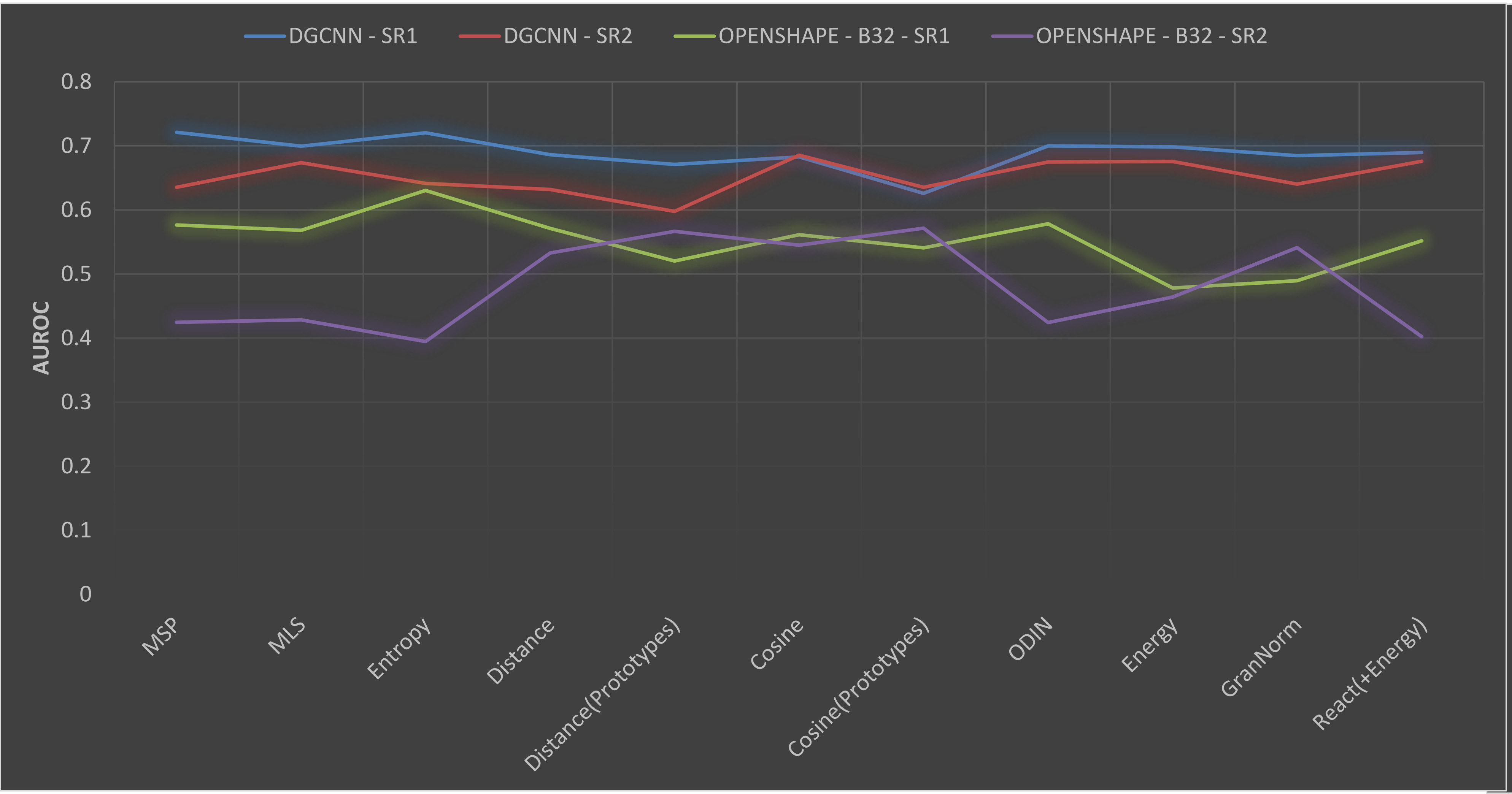
Method	AUROC	FPR95
MSP	0.5766	0.9743
MLS	0.5681	0.9547
Entropy	0.6305	0.8941
Distance	0.5712	0.9492
Distance(Prototypes)	0.5205	0.967
Cosine	0.5614	0.945
Cosine(Prototypes)	0.5408	0.9486
ODIN	0.5783	0.9768
Energy	0.4782	0.9725
GranNorm	0.4896	0.9743
React(+Energy)	0.5519	0.9596

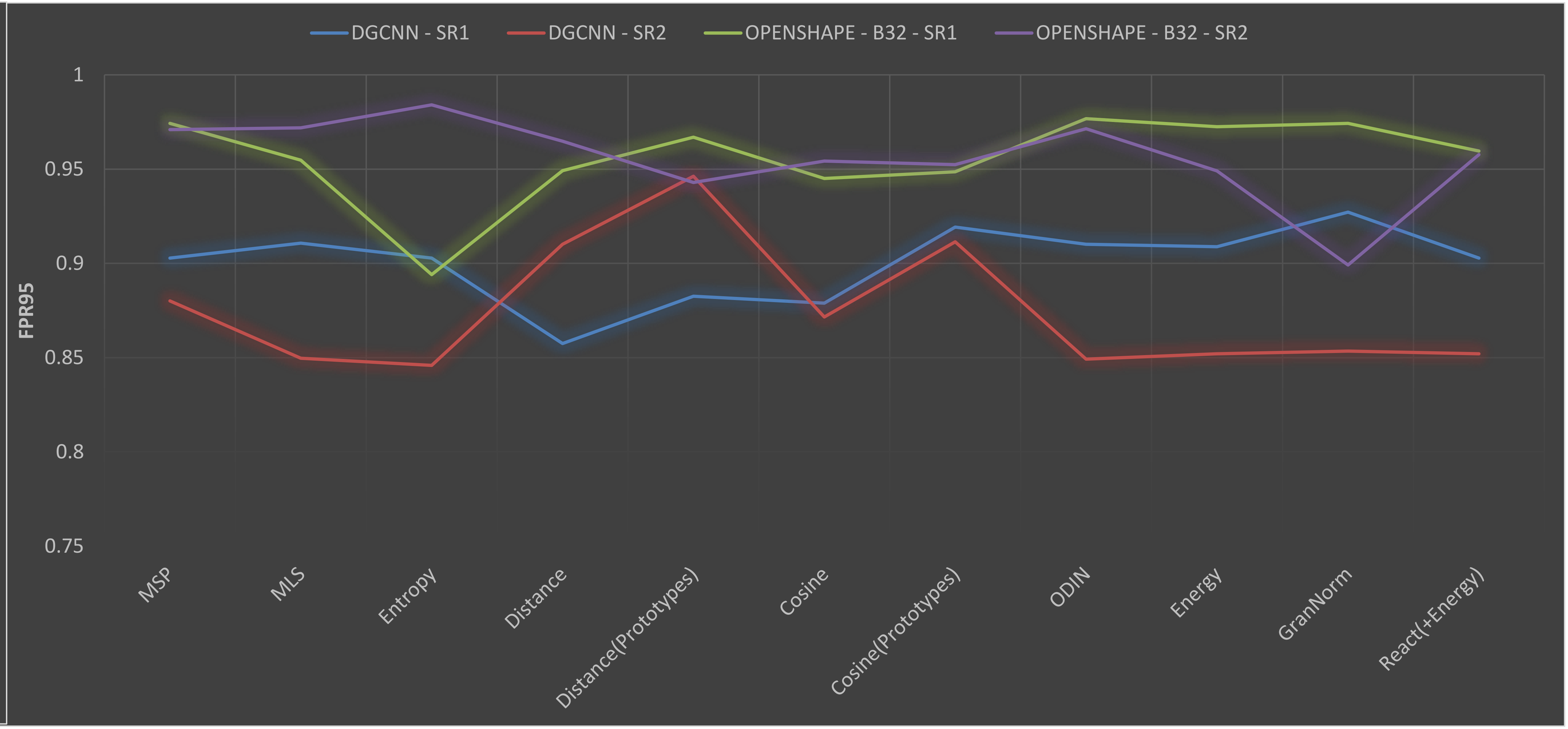
OPENSHAPE - B32 - SR2 (TAR1+TAR2)

Method	AUROC	FPR95
MSP	0.4247	0.971
MLS	0.4284	0.9719
Entropy	0.3946	0.9841
Distance	0.5327	0.9648
Distance(Prototypes)	0.5666	0.9429
Cosine	0.5451	0.9543
Cosine(Prototypes)	0.5716	0.9524
ODIN	0.4242	0.9714
Energy	0.4641	0.9491
GranNorm	0.5413	0.8991
React(+Energy)	0.4021	0.9577

OpenShape encounters difficulties in generalizing to real-world test data, resulting in variable performance as reflected in AUROC and FPR95 metrics across different scenarios.

OpenShape Evaluation with 3DOS

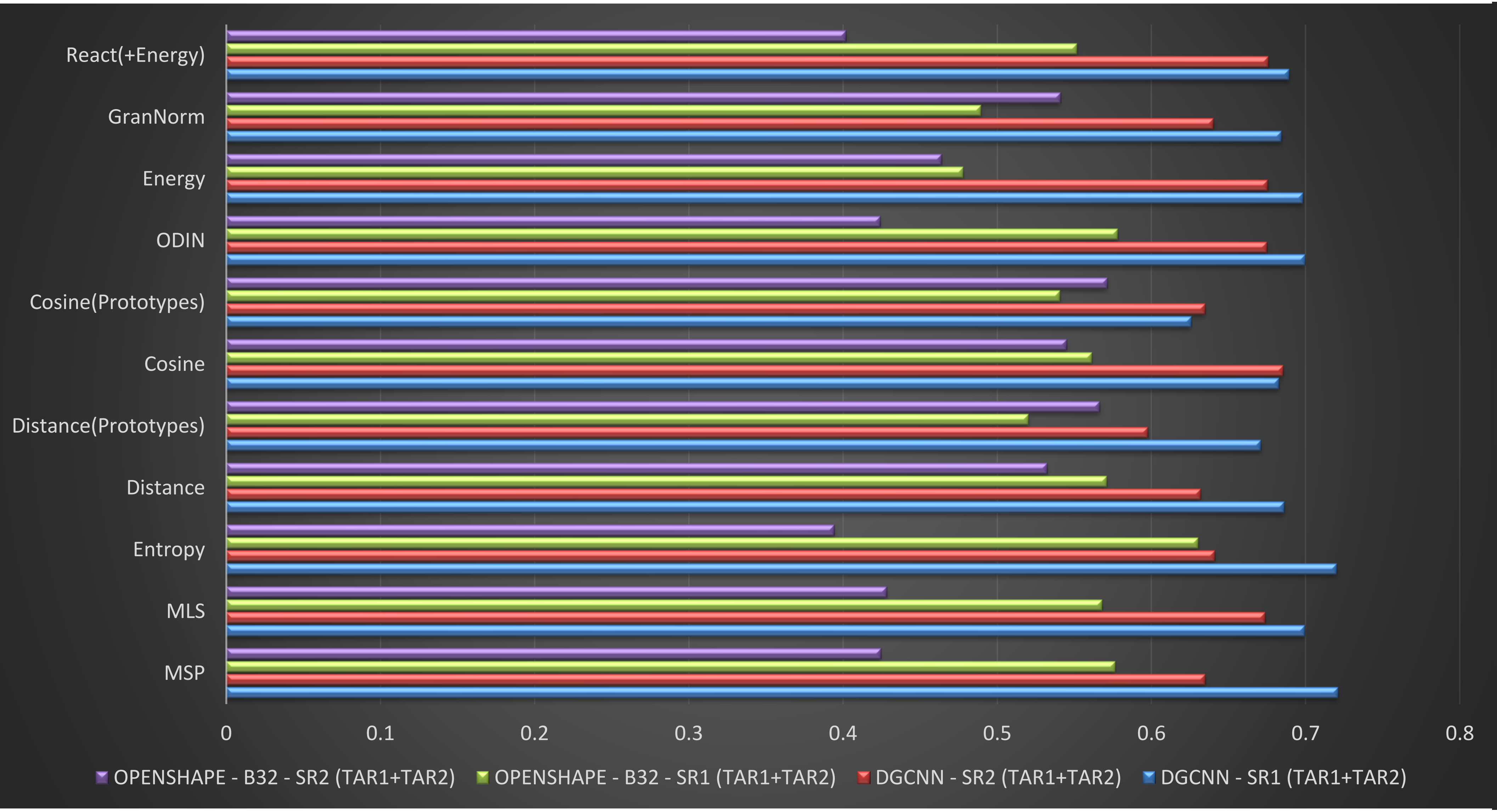




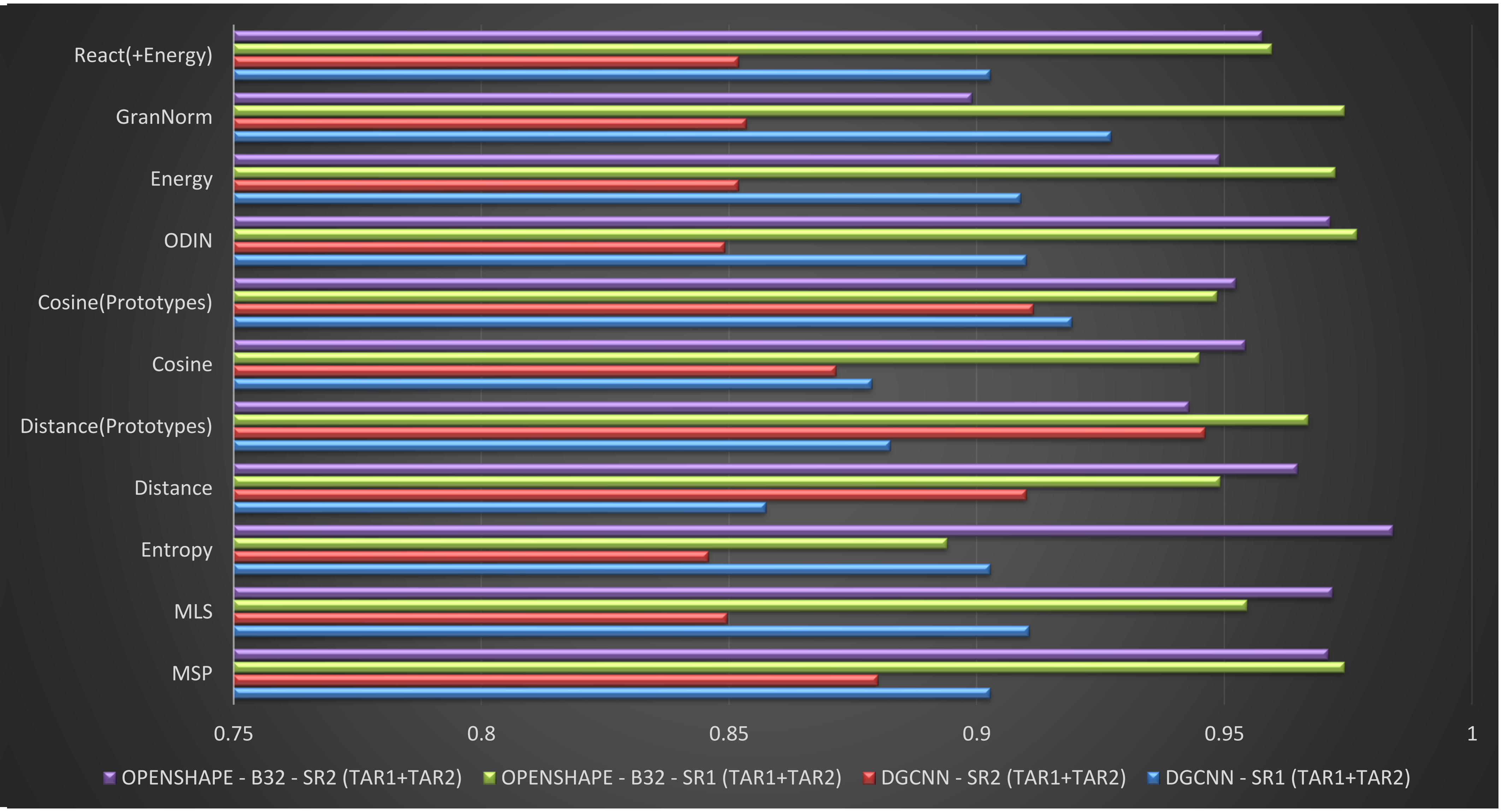
AUROC

FPR95

Final results



AUROC



FPR95

While the results for OpenShape are promising, DGCNN still outperforms it in many instances.

Final results

Thank You for your attention

Hossein Kakavand, Samaneh Gharedagh Sani

s308581@studenti.polito.it, s309100@studenti.polito.it