Exploring the Depths of 3D Semantic Novelty Detection

Gabriele Quaranta, Giulio Maselli, George Florin Eftime
Politecnico di Torino, Italy
s318944@studenti.polito.it, s306125@studenti.polito.it, s303483@studenti.polito.it

Abstract

Deep learning has shown promise in addressing challenges in point cloud processing and 3D shape understanding. However, the unstructured nature of point clouds presents obstacles for traditional CNNs. PointNet[8] has introduced an innovative approach by independently processing each point and aggregating information, albeit with limitations in capturing local structures. Inspired by PointNet, recent methods enrich point cloud representations through hierarchical feature extraction or graph-convolutional modules. This paper provides a brief overview of these challenges and advancements, highlighting the evolution of deep learning techniques in managing point clouds.

1. Introduction

The project centers on 3D semantic novelty detection, aiming to differentiate whether a given data sample belongs to a predefined nominal distribution of known semantic classes.

Traditional machine learning models often assume that training and test data stem from the same distribution, which may not hold true in real-world scenarios characterized by distributional shifts. These shifts can involve new object categories or data from new domains, leading to out-of-distribution (OOD) data that pose risks, particularly for autonomous agents like self-driving cars. Developing robust models capable of effectively discriminating among known classes while detecting unknown categories is essential yet challenging.

The research conducted in ”3DOS: Towards 3D Open Set Learning” enables us to assess the performance of different models and backbones in this context, offering a comprehensive analysis of their effectiveness when confronted with data falling outside the training distribution. The benchmark includes tracks for Synthetic, Real to Real, and Synthetic to Real scenarios, mimicking real-world deployment conditions.

Leveraging the 3DOS benchmark[1], particularly in the Synthetic to Real scenario, where nominal data samples are synthetic (CAD model-derived) and test samples are real-world point clouds from sensor scans, the objective is to distinguish between in-distribution (known semantic classes) and out-of-distribution (unknown) samples.

In this project, our focus is on a subset of this benchmark, specifically the Synthetic to Real Scenario. We re-evaluate the behavior of two backbones, DGCNN[12] and PointNet, in this track. Additionally, we conduct failure case analysis for DGCNN and explore the performance of a pre-trained model, OpenShape[6], leveraging its large scale to extract feature representations of both nominal and test data distributions.

Our code (developed on top of the 3DOS benchmark) is available on github: https://github.com/gabriquaranta/3D-semantic-novelty-detection

2. Related Works

The development of deep learning methods for point cloud processing has significantly advanced the field of 3D shape understanding and representation. Two pivotal contributions in this area, PointNet[8] and Dynamic Graph CNN (DGCNN)[12], have laid the groundwork for subsequent research, including our project's focus on 3D semantic novelty detection. Alongside these foundational works, OpenShape[6] emerges as a complementary approach, aiming to scale up 3D shape representation learning towards open-world understanding by learning multi-modal joint representations of text, image, and point clouds.

2.1. PointNet

Qi et al.'s PointNet introduces a groundbreaking method for directly processing point clouds, offering a novel solution to the unstructured nature of 3D data. By independently processing each point and aggregating global features, PointNet sets a precedent for efficiency and effectiveness in 3D data analysis. However, it encounters limitations in capturing the local geometric structures of data, which are crucial for detailed 3D understanding.

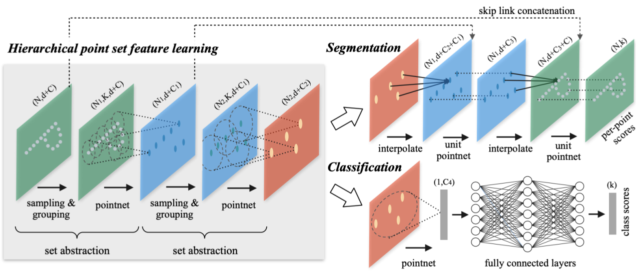

Figure 1. PointNet++

2.2. Dynamic Graph CNN (DGCNN)

Building on the insights from PointNet, Wang et al. propose the DGCNN, incorporating the EdgeConv operation to overcome PointNet's limitations in local structure capture. DGCNN dynamically updates graph structures, allowing for the integration of local geometric details with global shape understanding. This method significantly enhances the model's capacity for detailed 3D shape analysis, aligning closely with the objectives of semantic novelty detection in point clouds.

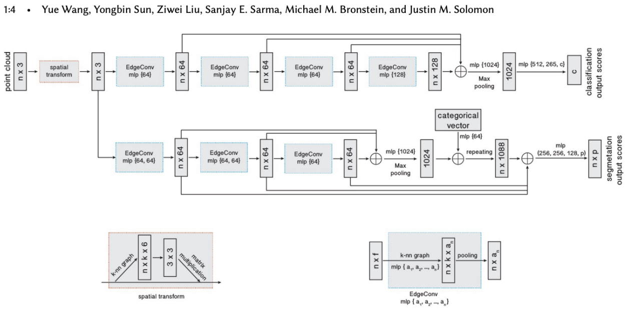
1:4 • Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon
point cloud
spatial transform
EdgeConv
Max pooling
categorical vector
repeating
classification output scores
segmentation output scores
spatial transform
k-nn graph
matrix multiplication
pooling
EdgeConv

Figure 2. DGCNN

2.3. OpenShape

The OpenShape framework extends the conversation on point cloud processing by focusing on scaling up 3D representations for open-world shape understanding. By leveraging multi-modal contrastive learning and ensembling multiple 3D datasets, OpenShape addresses the challenge of limited 3D training data scales and poor generalization to unseen shape categories. The method's emphasis on multi-modal representation alignment, hard negative mining, and the integration of enriched text descriptions with 3D shape and image data marks a significant step towards comprehensive 3D shape understanding in open-world settings.

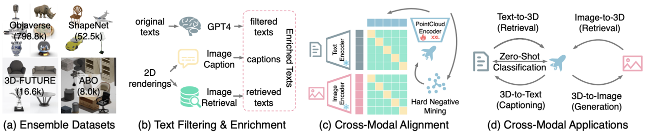

Figure 3. OpenShape Method

3. 3DOS: Towards Open Set 3D Learning

The 3D Open Set (3DOS) benchmark is integral to our research, providing a rigorous framework for assessing model performance across a variety of scenarios that mirror the challenges encountered in real-world deployments. By focusing on the Synthetic to Real scenario, our work leverages 3DOS to critically evaluate the adaptability of the Dynamic Graph CNN (DGCNN) and PointNet models when faced with point cloud data that transitions from synthetic origins to real-world complexities. This evaluation underscores the necessity for models that not only accurately classify known objects but also effectively identify and manage novel objects, thereby ensuring robust performance in dynamic environments.

It distinguishes itself by offering a comprehensive suite of tests that encompass Synthetic, Real to Real, and Synthetic to Real scenarios. This diversity allows for a nuanced analysis of model behaviors under conditions that range from controlled, synthetic environments to the unpredictability of real-world data. Our focus on the Synthetic to Real track is motivated by its relevance to practical applications, where models must often bridge the gap between neatly structured, synthetic training data and the messier, less predictable data captured from real environments.

This benchmark assesses the performance of Open Set methods in detecting unknown samples in test data using two primary metrics: AUROC and FPR95. AUROC represents the Area Under the Receiver Operating Characteristic Curve and indicates the probability that a known test sample has a greater normality score than an unknown one, while FPR95 measures the False Positive Rate at a True Positive Rate of 95%. Additionally, classification accuracy (ACC) is computed to evaluate the ability of Open Set methods to correctly classify known data.

Datasets: The benchmark utilizes three well-known 3D object datasets: ShapeNetCore, ModelNet40, and

ScanObjectNN. ShapeNetCore comprises synthetic instances of 55 object categories, while ModelNet40 contains 3D CAD models from 40 man-made object categories. ScanObjectNN includes 3D scans of real-world objects from 15 categories, with data samples already in the form of point clouds.

Synthetic to Real Benchmark: In the Synthetic to Real-World cross-domain scenario, synthetic point clouds from ModelNet40[13] are used for training, while real-world point clouds from ScanObjectNN[10] are used for testing. Three category sets, SR1, SR2, and SR3, are defined for evaluation, with SR1 and SR2 consisting of matching classes from ModelNet40 and ScanObjectNN, and SR3 comprising ScanObjectNN classes without a one-to-one mapping with ModelNet40. Models are trained on ModelNet40 samples of the known classes and evaluated on ScanObjectNN samples of both known and unknown classes in either SR1 or SR2 scenarios.

4. Experimental Results

Our project on 3D semantic novelty detection intersects with these related works through its focus on leveraging deep learning techniques for enhanced point cloud processing and understanding. Specifically, the methodologies pioneered by PointNet and DGCNN provide a solid foundation for addressing the challenges of 3D data analysis in our project. OpenShape's approach to scaling up 3D shape representation learning and its emphasis on multimodal data integration offer promising avenues for improving our project's ability to distinguish between known and unknown semantic classes in point clouds. The combination of these related works forms a comprehensive backdrop against which our project's objectives are set, driving forward the advancements in 3D semantic analysis and novelty detection.

4.1. Simple Baselines: DGCNN and PointNet++

The performance of distance-based methods is greatly influenced by the backbone used. While Cosine proto excels on PN2, it performs poorly on DGCNN likely due to DGCNN prototypes not accurately representing real-world test data. Similar observations hold for CE (L2), underscoring PointNet++'s resilience to domain shifts. SubArcFace proves its consistency, delivering commendable results across different backbones and achieving the best overall performance on average.

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

Synth to Real Benchmark - PointNet++ [8]					
SR 1 (easy)		SR 2 (hard)		Avg	
AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓	AUROC ↑	FPR95 ↓
81.0	79.6	70.3	86.7	75.6	83.2
82.1	76.6	67.6	86.8	74.8	81.7
81.7	77.3	70.2	84.4	76.0	80.8
81.9	77.5	67.7	87.3	74.8	82.4
77.6	80.1	68.4	86.3	73.0	83.2
81.7	75.6	67.6	87.2	74.6	81.4
-	-	-	-	-	-
78.0	84.4	74.7	84.2	76.4	84.3
71.2	89.7	60.3	93.5	65.7	91.6
82.8	74.9	68.0	89.3	75.4	82.1
79.9	**74.5**	**76.5**	**77.8**	**78.2**	**76.1**
79.7	84.5	75.7	80.2	77.7	82.3
78.7	84.3	75.1	83.4	76.9	83.8

4.2. DGCNN Failure Case Analysis

The following analysis presents the misclassification cases and their characteristics from a failcase examination of a Convolutional Neural Network (CNN) applied to the SR2 (hard) benchmark. The examination involves two methodologies: MSP (Maximum Softmax Probability) and

Distance-based. The MSP method focuses on the confidence level of predictions, while the Distance-based method considers the Euclidean distance between feature vectors.

4.2.1 MSP Metric

- Average ID Score: 0.9211
- Average OOD Score: 0.8722
- Threshold: 0.9998
- Total Misclassified OOD: 183 out of 1255

The misclassification instances are predominantly characterized by high confidence levels (close to 1.0) in the predicted classes, with notable instances including the misclassification of ’chair’ as ’desk’ and ’door’ as ’display’. These occurrences suggest that the model’s confidence in incorrect predictions is often high, potentially leading to significant errors in real-world applications.

Below are the ***distinct*** misclassification 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

Table 1. Misclassifications in the SR2 (hard) benchmark with confidence scores, predicted labels, and actual labels.

4.2.2 Distance Based Metric

- Average ID Distance: 0.1197
- Average OOD Distance: 0.1053
- Threshold: 0.1461
- Total Misclassified OOD: 85 out of 1255

In contrast to the MSP method, the distance-based approach indicates lower confidence levels in misclassified samples, with distances generally exceeding the established threshold. Notable instances include misclassifications such as ’door’ as ’desk’ and ’sink’ as ’display’. The relatively low confidence levels suggest that the model may struggle to differentiate between certain classes, leading to more varied misclassification patterns.

Below are the ***distinct*** misclassification cases:

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

Table 2. Misclassifications in the SR2 (hard) benchmark with distance values, predicted labels, and actual labels.

4.2.3 Conclusion

The analysis highlights the importance of considering both confidence levels and feature distances in evaluating model performance. While high confidence levels in misclassifications can indicate potentially critical errors, lower confidence levels may signify ambiguity in distinguishing between classes. Future improvements in model training and architecture design could focus on addressing these challenges to enhance classification accuracy and reliability in real-world scenarios.

4.3. Evaluation of large pre-trained models: OpenShape

We now focus on analyzing OpenShape, a large pre-trained model, to gain a different perspective on the applicability of such models in the context of 3D semantic novelty detection. By comparing OpenShape with DGCNN through the metrics AUROC and FPR95, we aim to understand how large-scale models can be optimized to handle real-world data, which often significantly differs from the synthetic data on which they were originally trained.

The results indicate that despite OpenShape’s vast representation capacity, the model struggles to generalize to real-world test data. This is highlighted by variable performances in the AUROC and FPR95 metrics across the SR1 and SR2 scenarios. Such findings suggest that while OpenShape’s approach is promising for its ability to learn from large datasets and through multi-modal representations, its practical utility in detecting semantic novelty in unseen data requires further refinements.

We want to highlight the importance of developing training and fine-tuning strategies that can better adapt pre-trained models to the specifics of real-world data. Adaptation might include transfer learning techniques, data augmentation, or the use of semi-supervised or unsupervised learning approaches to maximize the model’s sensitivity to new object classes.

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

4.3.1 DGCNN vs Openshape Performance

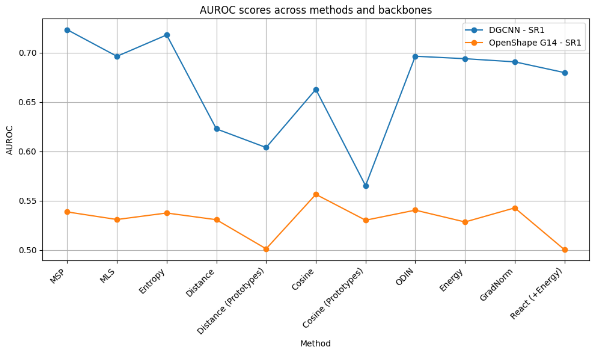

Figure 4. DGCNN - OpenShape AUROC Metric

5. Possible real-world application

3D semantic novelty detection involves identifying and categorizing new or unusual elements in 3D data that a system has not encountered before. This capability is crucial in various fields, allowing systems to recognize and respond to novel situations or objects. Below, we explore practical applications in robotics, medical imaging, and augmented reality, and discuss strategies for dealing with the inherent imperfections of 3D scans.

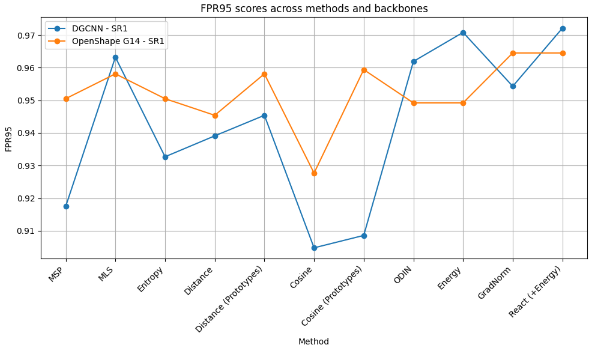

Figure 5. DGCNN - OpenShape FPR95 Metric

Robotics

In robotics, 3D semantic novelty detection is essential for autonomous navigation and manipulation in unknown or dynamic environments. For example, a robot navigating a disaster site for search and rescue missions must identify obstacles and paths, recognize victims, and interact with objects it has never seen before. Integrating 3D semantic novelty detection allows the robot to categorize unknown objects as potential obstacles or items of interest, enabling it to make informed decisions about navigation and interaction strategies.

Medical Imaging

In medical imaging, 3D semantic novelty detection can be used to identify anomalous structures in scans, such as tumors or other pathologies that deviate from typical anatomical structures. By recognizing these novelties, the system can flag scans for further review by a medical professional, potentially aiding in early diagnosis and treatment planning. This application requires high precision and sensitivity, as the cost of false negatives (missed anomalies) can be very high.

Augmented Reality (AR)

In AR, 3D semantic novelty detection can enhance user experiences by allowing applications to understand and interact with the real world in more complex ways. For example, an AR application could use novelty detection to identify unknown objects in a user's environment and provide contextual information or overlay digital interactions. This capability could be used in education, entertainment, or even retail, blending digital content with the physical world in a seamless and interactive manner.

5.1. Addressing Inherent Imperfections

3D scans are often incomplete, noisy, and affected by artifacts, which can challenge semantic novelty detection. Addressing these imperfections involves:

Data Preprocessing
Techniques like denoising, outlier removal, and hole filling can improve the quality of 3D data before processing. This step reduces the impact of noise and artifacts on the detection process.

Robust Feature Extraction
Using robust feature descriptors that are invariant to common distortions in 3D scans, such as changes in scale, rotation, or partial occlusions, can help ensure that the system accurately recognizes objects despite scan imperfections.

Partial Data Handling
The system should be designed to handle partial data effectively. Instead of classifying partially scanned objects as 'unknown' or 'out-of-distribution' (OOD) by default, the system could use contextual clues and prior knowledge to make educated guesses about the nature of partially visible objects. For example, in a robotics application, if a robot sees part of a chair, it can infer the presence of the chair even if the scan is incomplete.

6. Specific Application Integration: Robotics

Integrating 3D semantic novelty detection in a robotics application could involve creating a system that continuously updates its model of the environment with each new scan, using a combination of supervised learning for known objects and unsupervised or semi-supervised learning for novelty detection. The system could employ a multi-modal approach, combining 3D data with other sensor inputs (e.g., visual, infrared) to improve its understanding of the environment.

For handling imperfections, the robot could use a probabilistic framework to assess the confidence level of its detections, considering the quality of the scan data. When encountering partial or noisy data that might indicate a novel object, the system could adjust its behavior—slowing down, taking additional scans from different angles, or even asking for human assistance—to ensure safety and accuracy.

By integrating these strategies, the robotics application can effectively address the challenges posed by 3D scan imperfections, enabling the robot to operate safely and efficiently in unknown or dynamic environments. This approach balances the need for novelty detection with the practical limitations of current 3D scanning technologies, offering a path forward for incorporating 3D semantic novelty detection into real-world applications.

7. Conclusions

This study advances the field of 3D point cloud processing by evaluating deep learning approaches like PointNet and DGCNN against the novel 3DOS benchmark for synthetic-to-real scenarios. It highlights the critical need for models to discern between known and novel categories, a task increasingly vital for applications facing real-world variabilities. The introduction of the 3DOS benchmark represents a significant stride towards assessing and enhancing model robustness in identifying out-of-distribution data. Our research stands for the development of models that are not only innovative but also resilient and versatile, paving the way for future research in 3D semantic novelty detection. This work sets a foundation for advancing 3D point cloud analysis, promising improved safety and accuracy in practical applications.

References

[1] Antonio Alliegro, Francesco Cappio Borlino, and Tatiana Tommasi. “Towards Open Set 3D Learning: Benchmarking and Understanding Semantic Novelty Detection on Pointclouds”. In: *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. 2022. URL: https://openreview.net/forum?id=X2dHozbd1at.

[2] Dario Fontanel et al. “Detecting Anomalies in Semantic Segmentation with Prototypes”. In: *CVPR-W*. 2021.

[3] Dan Hendrycks and Kevin Gimpel. “A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks”. In: *ICLR* (2017).

[4] Rui Huang, Andrew Geng, and Yixuan Li. “On the importance of gradients for detecting distributional shifts in the wild”. In: *NeurIPS*. 2021.

[5] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. “Enhancing the reliability of out-of-distribution image detection in neural networks”. In: *ICLR*. 2018.

[6] Minghua Liu et al. *OpenShape: Scaling Up 3D Shape Representation Towards Open-World Understanding*. 2023. arXiv: 2305.10764 [cs.CV].

[7] Weitang Liu et al. “Energy-based Out-of-distribution Detection”. In: *NeurIPS*. 2020.

[8] Charles R Qi et al. “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”. In: *NeurIPS*. 2017.

[9] Yiyou Sun, Chuan Guo, and Yixuan Li. “ReAct: Out-of-distribution Detection With Rectified Activations”. In: *NeurIPS*. 2021.

[10] Mikaela Angelina Uy et al. “Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data”. In: *ICCV*. 2019.

[11] Hao Wang et al. “CosFace: Large Margin Cosine Loss for Deep Face Recognition”. In: *CVPR*. 2018.

[12] Yue Wang et al. “Dynamic Graph CNN for Learning on Point Clouds”. In: *ACM Transactions on Graphics (TOG)* 38.5 (2019), pp. 1–12.

[13] Z. Wu et al. “3D ShapeNets: A Deep Representation for Volumetric Shapes”. In: (2015).