# Exploring the Depths of 3D Semantic Novelty Detection

Samaneh Gharehdagh Sani
Politecnico di Torino
s309100@studenti.polito.it

Hossein Kakavand
Politecnico di Torino
s308581@studenti.polito.it

## Abstract

*Deep learning achieved a breakthrough with some of the challenges that existed in the processing of point clouds and the understanding of 3D shapes. The point clouds' unstructured nature somewhat challenges conventional CNNs. A pioneering approach, PointNet, processes each point independently and then aggregates all the information without local structure. Recent methods derive inspiration from PointNet, where each point is enriched from local neighborhood processing, hierarchical feature extraction, or graph-convolutional modules. This paper discusses those challenges and the improvements in a short introduction on the rise of deep-learning techniques for point cloud management.*

## 1. Introduction

The project deals with the problem of 3D semantic novelty detection, i.e. making the distinction between whether a given data sample belongs to a nominal distribution of known semantic classes. Modern work in machine learning mostly assumes samples at training and test time come i.i.d from the same distribution. This is hardly ever true in practice. A shift in the data distribution can be as drastic as new object categories or entirely new data domains, which creates a significant risk factor, especially for self-driving cars. The construction of a model that can simultaneously distinguish well among known classes and detect unknown categories is a crucial but very challenging problem.

This work evaluates the performance of different models and backbones in such a setup, therefore providing a fair comparison of their generalization when exposed to OOD data, which is presented in ”3DOS: Towards 3D Open Set Learning”. The benchmark is divided into three tracks, namely, Synthetic, Real to Real, and Synthetic to Real—all implemented in a way that the same conditions present in a real-world deployment are replicated.

In consideration of the 3DOS benchmark, and more specifically the Synthetic to Real track, in which nominal data samples are obtained through the generation of point clouds of a synthetic scene of CAD model obtained, and test samples are real-world point cloud samples obtained from sensor scans, we want to distinguish between samples that belong to the in-distribution, that is, the nominal data, and the out-distribution, which represents unknown semantic classes.

In this work, we focus on one out of the three tracks provided, which is the Synthetic to Real Scenario. In that track, the general behavior of two backbones, that is, DG CNN and PointNet, is re-evaluated. In addition, failure case analysis for DGCNN is carried out, and performance of a pre-trained model, OpenShape, is analyzed with its large scale to extract feature representations of both nominal and test data distributions.

## 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 of the critical contributions in this direction, namely PointNet and DGCNN, have paved the way for further research, including that of 3D semantic novelty detection. And it is in this vein, and under the background of these pioneering works mentioned above, that comes OpenShape, as a complementary approach set to scale up 3D shape representation learning for open-world understanding by multi-modal joint representation of text, image, and point clouds.

### 2.1. PointNet

The revolutionary work of PointNet by Qi et al. offers a direct point cloud processing method—a fresh solution to the spatial invariance problem caused by the disorder of 3D data. The operation processes every point individually and aggregates global features, pushing the bars of efficiency and effectiveness in 3D data analysis forward. Of course, though, it has limitations in catching local geometric structures that are highly important in detailed 3D understanding.

### 2.2. Dynamic Graph CNN

Following the reasoning by PointNet, Dynamic Graph CNN advances a meaningful step further and proposes the

EdgeConv operation to aggregate the local structures that PointNet misses. The essence of DGCNN is that the structural graph can be dynamically adjusted, empowering the integration of local geometric details and global shape understanding. This greatly enhances the capability of the model applied to a detailed 3D shape analysis, closely following the ultimate goals of the semantic novelty detection in point clouds.

### 2.3. OpenShape

OpenShape continues the discussion of point cloud processing and, in particular, offers a new way to scale up 3D representations for open-world shape understanding. Multi-modal contrastive learning and open-world adaptation through the use of multiple 3D datasets in OpenShape tackle the challenges of limited 3D training scale data and poor generalization to unseen shape categories. More broadly, this paper's focus on multi-modal representation alignment with hard negative mining and integration into enriched text descriptions with 3D shape and image data makes a step toward comprehensive 3D shape understanding—open-world context.

## 3. 3DOS: Towards Open Set 3D Learning

The benchmark 3D Open Set (3DOS) is paramount in showing a strict frame of reference for the estimation of model-wise performance in different scenarios, imitating challenges that exist in real-life deployments. In this regard, the result points toward the requirement for models that can classify known objects accurately while, at the same time, showing that they can identify and control novel objects in a way that allows for confident performance in dynamic scenarios.

This benchmark estimates the performance of the Open Set approaches on the task of detecting unknown samples in the test data. Here, the Open Set methods are judged on two important metrics: AUROC and FPR95 and Classification Accuracy (ACC) to judge the ability of the Open Set methods to accurately classify known data.

### 3.1. Datasets

ShapeNetCore, ModelNet40, and ScanObjectNN shape the three most popular 3D object datasets. ShapeNetCore: It is a 55 object category dataset with a collection of instances that is synthetic in nature. ModelNet40: This dataset encapsulates 40 categories of 3D CAD models of man-made object data. ScanObjectNN: This dataset contains 15 categories which have 3D scans of real-world objects and the samples are already present in the form of point clouds.

### 3.2. Benchmark Description(Synthetic to Real Benchmark)

In the cross-domain scenario of synthetic to real-world, the source domain consisted of synthetic point clouds. For training purposes, 3D CAD models were taken from ModelNet40 in order to generate synthetic point clouds. Three category sets are defined for evaluation:

SR1: Contains classes present in both ModelNet40 and ScanObjectNN SR2: Contains classes from ScanObjectNN that have a one-to-one mapping with ModelNet40 classes SR3: Contains classes from ScanObjectNN not having a one-to-one mapping with ModelNet40 The models are trained with ModelNet40 samples from the known classes and evaluated on ScanObjectNN samples from both known and unknown classes, with the possibility of one-to-one mapping, i.e., in the SR1 and SR2 scenarios respectively.

## 4. Experimental Results

More importantly, 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. In particular, the methodologies behind PointNet and DGCNN lay a strong foundation for the challenges the project raises regarding the analysis of 3D data. In that light, OpenShape can scale in 3D shape representation learning, while the multi-modal data integrations form promising paths toward better separability between unknown and known semantic classes in point clouds. Related work and comprehensive background set forth the objectives of the project, furthering the developments in 3D semantic analysis and novelty detection.

### 4.1. Simple Baselines: DGCNN and PointNet++

The backbone used makes an enormous difference in the performance of these distance-based methods. Though Cosine proto performs very well on PN2, it performs poorly on DGCNN, likely due to DGCNN prototypes not representing real-world test data well. The same observations hold for the CE (L2) case, pointing out the fact that PointNet++ is of great robustness to domain shifts. SubArc-Face proves the consistency by giving good results over different backbones and is best in overall performance on average.

### 4.2. DGCNN Failure Case Analysis

The following presented will be an analysis of the attributes of misclassification cases retrieved from a fail-case examination of a Convolutional Neural Network (CNN) applied to the SR2 (hard) benchmark. The two methodologies used for this analysis are Maximum Softmax Probability (MSP) and Distance-based. One works by taking into account the confidence levels of predictions, the other, known as the Distance-based methodology, makes use of the Euclidean distance that lies in between feature vectors.

| **DGCNN - SR1** | **AUROC** | **FPR95** |
|---|---|---|
| MSP [3] | 0.7212 | 0.9028 |
| MLS | 0.6997 | 0.9107 |
| Entropy | 0.7203 | 0.9028 |
| Distance [2] | 0.6863 | 0.8575 |
| Distance (Prototypes) | 0.6712 | 0.8826 |
| Cosine [11] | 0.6829 | 0.8789 |
| Cosine (Prototypes) | 0.626 | 0.9193 |
| ODIN [5] | 0.6998 | 0.0.9101 |
| Energy [7] | 0.6984 | 0.9089 |
| GradNorm [4] | 0.6846 | 0.9272 |
| React (+Energy) [9] | 0.6896 | 0.9028 |

Table 1. DGCNN - SR1 Benchmark Results

| **DGCNN - SR2** | **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 |
| GradNorm | 0.6405 | 0.8535 |
| React (+Energy) | 0.6761 | 0.852 |

Table 2. DGCNN - SR2 Benchmark Results

| 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 | 80.4 |
| 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 |

Table 3. Synth to Real Benchmark - PointNet++ [8]

#### 4.2.1 MSP Metric - SR1

- Average ID Score: 0.9295
- Average OOD Score: 0.8508
- Threshold: 0.99
- Total Misclassified OOD: 392 out of 1255

The predicted classes have a high confidence level close to the ideal classification confidence level of 1.0 for the classes. Notable instances are the confusion of 'chair' with 'door' and 'Bookshelf' with 'door' and 'chair'. Such observations would suggest the model is highly confident in its wrong predictions; therefore, that can cause real-world big errors.

| **Predict** | **Actual** |
|---|---|
| Door | Chair |
| Door | Chair |
| Door | Bookshelf |
| Chair | Bookshelf |
| Chair | Bookshelf |

Table 4. Misclassifications in the SR1 with predicted labels, and actual labels.

#### 4.2.2 Euclidean Distance Based Metric - SR1

- Average ID Distance: 0.3509
- Average OOD Distance: 0.3086
- Threshold: 0.3859
- Total Misclassified OOD: 240 out of 1255

The distance-based approach, on the other hand, simulates low confidence in the misclassified samples with generally larger distances over the threshold. For instance, some of the misclassifications of interest include 'chair' → 'door' and 'chair' → 'sink'. Such relatively low confidence level connotes the difficulty of the model in separating some classes and may thus lead to broader misclassification patterns.

| **Predicted** | **Actual** |
|---|---|
| desk | door |
| display | sink |
| toilet | shelf |
| bed | chair |

Table 5. Misclassifications in the SR2 (hard) benchmark with distance values, predicted labels, and actual labels.

#### 4.2.3 MSP Metric - SR2

- Average ID Score: 0.8822
- Average OOD Score: 0.8331
- Threshold: 0.99
- Total Misclassified OOD: 515 out of 1255

The predicted classes have a high confidence level close to the ideal classification confidence level of 1.0 for the classes. Notable instances are the confusion of 'desk' with 'bed'. Such observations would suggest the model is highly confident in its wrong predictions; therefore, that can cause real-world big errors.

| Predict | Actual |
|---|---|
| Bed | Desk |
| Bed | Desk |
| Bed | Desk |
| Bed | Desk |
| Bed | Desk |

Table 6. Misclassifications in the SR2 with predicted labels, and actual labels.

#### 4.2.4 Euclidean Distance Based Metric - SR2

- Average ID Distance: 0.3894
- Average OOD Distance: 0.3425
- Threshold: 0.4283
- Total Misclassified OOD: 265 out of 1255

The distance-based approach, on the other hand, simulates low confidence in the misclassified samples with generally larger distances over the threshold. For instance, some of the misclassifications of interest include 'chair' $\rightarrow$ 'door' and 'chair' $\rightarrow$ 'sink'. Such relatively low confidence level connotes the difficulty of the model in separating some classes and may thus lead to broader misclassification patterns. This gives a definite affirmation on the importance of both the levels of confidence, as well as the crucial point of the distances between the features, when determining the model performance. Very high confidences, even on misclassifications, could indicate critical problems. On the other hand, rather low confidences may indicate an element of ambiguity amongst the classes. This might be a strong indicator for future improvement in the highly radical new training paradigms toward higher classification accuracies and reliability in real-world scenarios.

| Predicted | Actual |
|---|---|
| Bed | Desk |
| Bed | Desk |
| Bed | Desk |
| Bed | Desk |
| Bed | Desk |

Table 7. Misclassifications in the SR2 (hard) benchmark with distance values, predicted labels, and actual labels.

### 4.3. Evaluation of large pre-trained models: OpenShape

Qualitative analysis will allow us to deduce the relevance of such large pre-trained models in the context of 3D semantic novelty detection. We compare the performance of OpenShape to DCGNN through the metrics AUROC and FPR95, looking to better understand responses to real-world data characteristics that usually differ drastically from synthetic data they were trained on.

The outcomes clearly are that the large representation capacity of OpenShape seriously lacks generalization for real-world test data. This alone makes a case for the variable performances in the AUROC and FPR95 metrics between the SR1 and SR2 scenarios. With such a finding, the approach of OpenShape is promising because it draws from large datasets and from multimodal representations of learning. Practical utility of OpenShape in being able to detect semantic novelty in unseen data requires further refinement.

This can include transfer-learning techniques, data augmentation, or a variant of semi-supervised or unsupervised learning techniques to ensure that the model is maximized to the best sensitivity of new classes of objects. We want to stress the importance of developing training and fine-tuning strategies that allow pre-trained models to better align with the specifics of real-world data.

| OpenShape B32 - SR1 | 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 |
| GradNorm | 0.4896 | 0.9743 |
| React (+Energy) | 0.5519 | 0.9596 |

Table 8. OpenShape B32 - SR1 Benchmark Results

| OpenShape B32 - SR2 | 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 |
| GradNorm | 0.5413 | 0.8991 |
| React (+Energy) | 0.4021 | 0.9577 |

Table 9. OpenShape B32 - SR2 Benchmark Results

#### 4.3.1 DGCNN vs OpenShape Performance

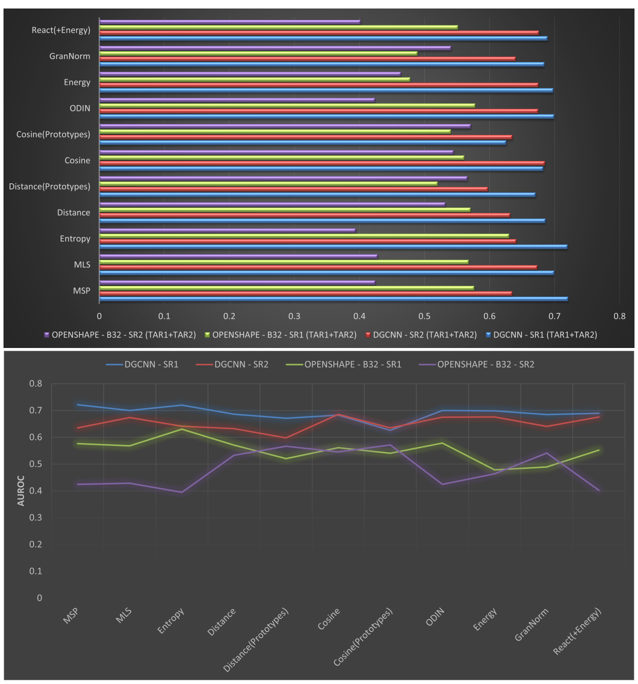


Figure 1. DGCNN - OpenShape AUROC Metric

## 5. Possible real-world application

3D Semantic novelty detection can be described as an attribute that deals with the identification and categorization of novel or uncommon elements in 3D in an environment with exposure to a system. It is critical in very many fields, for it allows systems to recognize and respond to novel situations or objects. Below we discuss possible practical robotic, medical imaging, and augmented reality applications and coping strategies of the inherent imperfections of 3D scans.

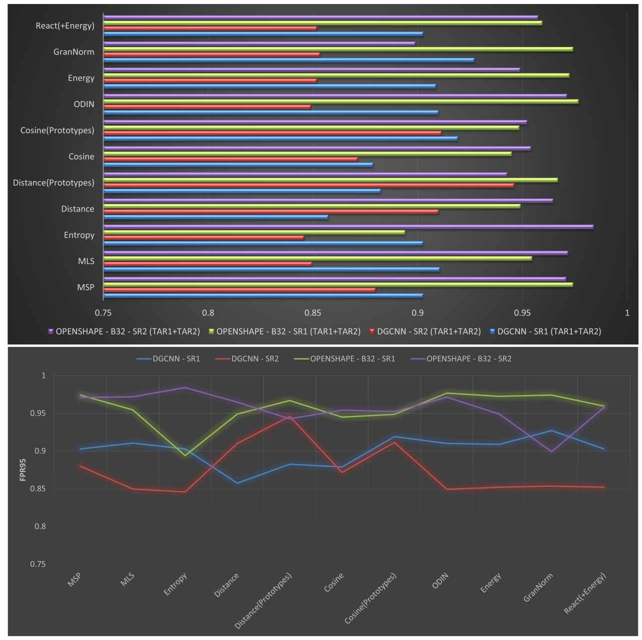


Figure 2. DGCNN - OpenShape FPR95 Metric

### 5.1. Augmented Reality

Richness in the AR user's experience is achieved through such things as 3D semantic novelty detection, where applications might understand and interact with the world. For example, it can use novelty detection to identify an object unknown in a user's environment to offer context or overlay of digital interaction. This ability to have digital content conceptually mixed and applied in real physical surroundings for engaging and interactive experiences is a key feature in seamlessly applied education, entertainment, and retail.

### 5.2. Medical Imaging

Some applications in which 3D semantic novelty detection is used include those in medical imaging to monitor abnormal structures, such as tumors and other pathologies, where the usual anatomical structures in the scans are not observed. Through the recognition of novelties, it can mark the scans for the human experts to revise them further, possibly leading to an early diagnosis and, finally, the elaboration of the treatment plan. Such an application would have to be very precise and sensitive, as the costs paid in the case of false negatives might be high.

### 5.3. Robotics

Thus, the ultimate job of 3D semantic novelty detection for a robot is to enable the robots to traverse and manipulate scenes in new environments while potentially facing the presence of new and unknown objects, obstacles, or even other entities within the surrounding environment. For example, a disaster site requires the robot to negotiate through this site for search and rescue missions, to find its path and a void, to classify victims, and to interact with things that it has never seen before. In other words, inclusion of 3D semantic novelty detection in the robot enables the robot to classify the unseen objects as obstacles and items of interest, thus making navigational and interaction strategies based on the object.

## 5.4. Tackling Inherent Imperfections

The 3D scans are usually incomplete and noisy, and often corrupted by artifacts, imparting imperfection to them, which in turn makes the task of semantic novelty detection hard. Now, these imperfections were corrected through the following strategies:

### 5.4.1 Data Refinement

The implemented data refinements denoise, remove outliers, and fill holes to reduce noise and artifacts, which enhance the 3D data quality prior to processing as much as possible in order to make detection less sensitive to imperfections.

### 5.4.2 Robust Feature Extraction

The extractions are robust with respect to the changes in scale, rotation, occlusions, and the other kinds of typical deformations introduced by processes of 3D scans; this allows the system to perform well on recognizing objects despite imperfections in the scans.

### 5.4.3 Managing Incomplete Data

This could then be applied in robotics, where it would be necessary to design a system properly working with incomplete data. It might be able to infer even the presence of an object from incomplete scans by using the context and prior knowledge of some attributes that are important of the object being partially observed. For example, if the robot feels something that is part of the chair, it will be able to infer the presence of the chair.

# 6. Specific Application Integration: Medical Imaging

3D semantic novelty detection has significant applications in medical imaging, particularly for monitoring abnormal structures such as tumors and other pathologies. In these applications, the system identifies novelties in scans where usual anatomical structures are not observed, marking them for further review by human experts. This process can lead to early diagnosis and the development of effective treatment plans.

## 6.1. Continuous Model Updating

The system continuously updates a model of the anatomical environment with every new scan. This model combines supervised learning for identifying known anatomical structures with unsupervised or semi-supervised learning to detect novel or abnormal findings. This integration may also involve multimodal fusion, incorporating 3D data with other types of medical imaging data, such as MRI, CT, and ultrasound, to provide a comprehensive understanding of the patient's condition.

## 6.2. Probabilistic Framework

A medical imaging system can operate within a probabilistic framework to manage imperfections in the data, ensuring that the confidence level of its detections corresponds to the quality of the scan data. When partial or noisy data are detected, which might create uncertainty about the presence of an abnormality, the system can adapt its behavior to ensure accurate diagnostics. This adaptation could involve taking additional scans, using alternative imaging modalities, or seeking expert human intervention.

## 6.3. Precision and Sensitivity

For such an application to be effective, it must be extremely precise and sensitive, as the cost of false negatives—failing to detect an abnormality—can be very high. Therefore, the system must ensure high sensitivity to detect even subtle novelties in the scans, while also maintaining a low false-positive rate to avoid unnecessary alarm and additional testing.

By implementing these strategies, a medical imaging application can effectively handle the imperfections of 3D scans, enhancing the detection of abnormalities in clinical settings. This approach balances the need for accurate novelty detection with the practical limitations of current 3D scanning technologies, paving the way for the incorporation of 3D semantic novelty detection in medical diagnostics and treatment planning.

# 7. Conclusions

This paper pushes the envelope of 3D point cloud processing further to evaluate deep learning methodologies, such as PointNet, against DGCNN, over the novel 3DOS benchmark for synthetic-to-real scenarios. Very important is the development of methodologies that allow a clear distinction between categories that are known and novel, which is important in any application facing real-world variabilities. The 3DOS benchmark is one big step toward the evaluation of approaches for measurement and improvement

of robustness in out-of-distribution data detection. Our research advocates the development not only in novel aspects but also in robust and flexible models in this area, which is relatively important for future research in 3D semantic novelty detection. This work builds on the excellent improvement in 3D point cloud analysis and can be used in a number of practical applications to achieve improved safety and accuracy.
The implementation of the codes can be found on GitHub: https://github.com/Hosseinkakavand1376/AML_3DOS

## 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

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

[4] Rui Huang, Andrew Geng, and Yixuan Li. "On the importance of gradients for detecting distributional shifts in the wild". In: NeurIPS. 2021. 3

[5] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. "Enhancing the reliability of out-of-distribution image detection in neural networks". In: ICLR. 2018. 3

[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. 3

[8] Charles R Qi et al. "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space". In: NeurIPS. 2017. 3

[9] Yiyou Sun, Chuan Guo, and Yixuan Li. "ReAct: Out-of-distribution Detection With Rectified Activations". In: NeurIPS. 2021. 3

[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. 3

[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: CVPR. 2015.