CASL: Curvature-Augmented Self-supervised Learning for 3D Anomaly Detection

Yaohua Zha[1,2], Xue Yuerong[1], Chunlin Fan[1], Yuansong Wang[1], Tao Dai[3*], Ke Chen[2], Shu-Tao Xia[1,2]

[1]Tsinghua Shenzhen International Graduate School, Tsinghua University
[2]Institute of Perceptual Intelligence, Pengcheng Laboratory
[3]College of Computer Science and Software Engineering, Shenzhen University

Abstract

Deep learning-based 3D anomaly detection methods have demonstrated significant potential in industrial manufacturing. However, many approaches are specifically designed for anomaly detection tasks, which limits their generalizability to other 3D tasks. In contrast, self-supervised point cloud models aim for general representation learning, yet our investigation reveals that these classical models are suboptimal at anomaly detection under the unified fine-tuning paradigm. This motivates us to develop a more generalizable 3D model that can effectively detect anomalies without relying on task-specific designs. Interestingly, we find that using only the curvature of each point as its anomaly score already outperforms several classical self-supervised and dedicated anomaly detection models, highlighting the critical role of **curvature** in 3D anomaly detection. In this paper, we propose a **C**urvature-**A**ugmented **S**elf-supervised **L**earning (**CASL**) framework based on a reconstruction paradigm. Built upon the classical U-Net architecture, our approach introduces multi-scale curvature prompts to guide the decoder in predicting the coordinates of each point. Without relying on any dedicated anomaly detection mechanisms, it achieves leading detection performance through straightforward anomaly classification fine-tuning. Moreover, the learned representations generalize well to standard 3D understanding tasks such as point cloud classification.

Code — https://github.com/zyh16143998882/CASL
Extended version —
https://www.arxiv.org/abs/2511.12909

Introduction

Deep learning-based anomaly detection algorithms (Pang et al. 2021; Zamanzadeh Darban et al. 2024; Bergmann et al. 2019; Xie et al. 2024) have played a significant role in industrial manufacturing due to their efficiency in quality control. Early research (Bergmann et al. 2018, 2020; Roth et al. 2022) primarily focused on the 2D domain and achieved notable progress. More recently, 3D point cloud anomaly detection (Liu et al. 2023; Li et al. 2024; Zhu et al. 2024) has emerged as a prominent research direction, as its ability to capture more comprehensive spatial structures leads to significantly improved detection performance compared to traditional 2D approaches.

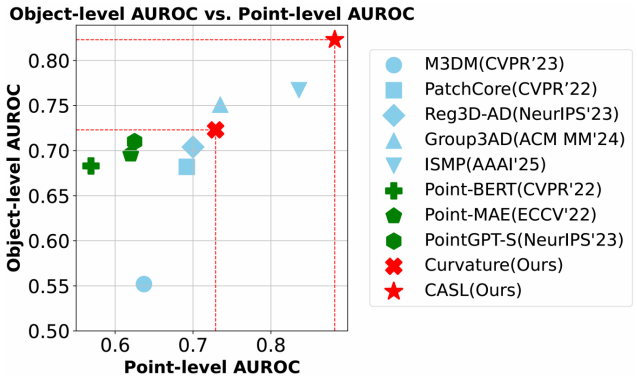

Figure 1: Anomaly detection performance of different methods on the Real3D-AD dataset. Blue markers represent task-specific anomaly detection methods, while green markers correspond to classical self-supervised models fine-tuned under the "pre-training and fine-tuning" paradigm. Red markers are our methods.

Existing 3D anomaly detection methods can be broadly categorized into feature-matching-based methods (Roth et al. 2022; Liu et al. 2023; Zhu et al. 2024) and reconstruction-based methods (Li et al. 2024; Zhou et al. 2024b). Feature-matching methods build a feature pool from normal samples during training and identify anomalies at inference by computing feature distances. Reconstruction-based methods learn to reconstruct normal samples from local point information during training and detect anomalies by comparing the reconstructed data with the original input. These methods are tailored specifically for anomaly detection tasks. Despite achieving impressive detection performance, their task-specific design inherently restricts their generalizability to broader 3D understanding tasks.

In contrast, self-supervised representation learning on point clouds (Pang et al. 2022; Zha et al. 2023b, 2025; Chen et al. 2024) aims to learn general representations through a unified "pre-training and fine-tuning" paradigm. For instance, methods such as Point-MAE (Pang et al. 2022),

*Corresponding author. (daitao.edu@gmail.com)

CrossPoint (Afham et al. 2022), and PointGPT (Chen et al. 2024) first pre-train models using self-supervised signals and then transfer them to downstream tasks such as classification and segmentation via simple task-head-based fine-tuning. This unified paradigm significantly enhances the generalizability of the models. Although some previous 3D anomaly detection methods (e.g., PatchCore (Roth et al. 2022), Reg3D-AD (Liu et al. 2023)) have employed pre-trained point cloud models, they primarily rely on feature matching mechanisms specifically designed for anomaly detection, rather than adhering to the unified "pre-training and fine-tuning" paradigm. Therefore, these methods offer limited evidence regarding the generalizability of pre-trained models to anomaly detection tasks.

To this end, we begin by fine-tuning existing pre-trained point cloud models using a simple pseudo-anomaly-based classification task, aiming to assess their generalizability to anomaly detection scenarios. As illustrated in Figure 1, our experimental results show that these classical models exhibit suboptimal performance when applied to anomaly detection via this straightforward fine-tuning approach. We attribute this suboptimal performance to the prevalent issue of geometric shortcuts (Zha et al. 2024d; Zhang, Zhang, and Yan 2024) in existing classical self-supervised methods. These approaches typically perform reconstruction directly from the coordinate space to the coordinate space, resulting in a complete overlap between the semantic and positional domains. Consequently, the learned representations overly rely on low-level spatial features, leading to fine-grained representational collapse in anomaly detection.

Furthermore, we make an interesting observation: ***a non-learning-based method that uses only the curvature of each point as its anomaly score outperforms several classical self-supervised and task-specific anomaly detection models***, as shown in Figure 1. This finding underscores the potential of curvature as a highly informative cue for identifying structural irregularities. As an intrinsic geometric property parallel to the coordinate domain, curvature can effectively alleviate the issue of geometric shortcuts when used as the semantic representation in self-supervised learning, offering a compelling alternative to traditional coordinate-based semantics.

In this paper, we propose a curvature-augmented representation learning framework grounded in a reconstruction paradigm. Built upon the classical U-Net (Ronneberger, Fischer, and Brox 2015) architecture, our method introduces multi-scale curvature prompts to guide the decoder in predicting the spatial coordinates of each point. In contrast to conventional reconstruction-based approaches, which typically mask a subset of coordinates and train the model to recover their coordinates, our framework adopts a fundamentally different strategy: we mask out all point coordinates and rely exclusively on curvature-based prompts to reconstruct the original coordinates. By embedding curvature, a geometric cue that is highly sensitive to anomalies, into the learning process, the model is encouraged to focus on local surface variations and fine-grained structural details. This design compels the network to learn rich geometric representations solely from curvature information, thereby not only avoiding geometric shortcuts but also deepening the understanding of 3D structural patterns. Remarkably, without relying on any dedicated anomaly detection mechanisms, the proposed framework achieves leading performance through simple classification-based fine-tuning, improving the average Object-level AUROC (O-AUROC) by 5.6% on the Real3D-AD (Liu et al. 2023) dataset and 4.8% on the Anomaly-ShapeNet (Li et al. 2024) dataset. Furthermore, the learned representations also exhibit strong generalization ability across standard 3D understanding tasks.

Our main contributions are summarized as follows:

- We find that a non-learning method based solely on point-wise curvature outperforms classical self-supervised and several anomaly-specific approaches on benchmark datasets, highlighting the importance of curvature in 3D anomaly detection.
- We develop a curvature-augmented representation learning framework (CASL) based on a reconstruction paradigm. We introduce a full coordinate masking strategy combined with multi-scale curvature prompts to guide the prediction of original coordinates, effectively avoiding geometric shortcuts and promoting a deeper understanding of 3D structural patterns.
- Our model enables effective anomaly detection without relying on any anomaly-specific detection mechanisms, requiring only a simple pseudo-anomaly-based classification fine-tuning. It significantly outperforms existing methods on the Real3D-AD and Anomaly-ShapeNet datasets, and demonstrates superior generalization across other 3D understanding tasks.

Related Work

Point Cloud Self-supervised Learning

Self-supervised learning (SSL) methods for 3D point clouds can be broadly categorized into discriminative and reconstruction paradigms. Discriminative methods such as PointContrast (Xie et al. 2020) and CrossPoint (Afham et al. 2022) learn representations by pulling together augmented views of the same instance while pushing apart different instances via contrastive loss. Reconstruction-based methods (Yu et al. 2022; Pang et al. 2022; Xiong et al. 2023; Zha et al. 2024c; Chen et al. 2024) typically learn general-purpose 3D representations by predicting the masked regions of a point cloud from the unmasked ones, such as point-MAE (Pang et al. 2022) and Point-FEMAE (Zha et al. 2024b) utilize autoencoders, whereas PointGPT (Chen et al. 2024) employs autoregression. Some cross-modal work (Dong et al. 2023; Qi et al. 2023; Zhang et al. 2023) enhance 3D representations by incorporating vision and language priors. PointMoDE (Zha et al. 2024a) is the first to propose complementary learning between the scene and object domains in 3D space. These methods have achieved some progress across classical 3D tasks. However, these methods commonly suffer from geometric shortcut issues, leading to representational collapse in fine-grained anomaly detection.

3D Anomaly Detection

3D anomaly detection methods can be broadly classified into feature-matching-based methods and reconstruction-based methods. Feature-matching-based methods, such as Reg3D-AD (Liu et al. 2023) and Group3AD (Zhu et al. 2024), construct memory banks using features extracted from pre-trained models, comparing test samples to detect deviations. Some methods, including CPMF (Cao, Xu, and Shen 2024) and PointAD (Zhou et al. 2024a), utilize 2D projections and vision-language models for memory construction or zero-shot detection. Reconstruction-based methods, on the other hand, aim to reconstruct normal point clouds and identify anomalies through reconstruction error. IMRNet (Li et al. 2024) improves Point-MAE (Pang et al. 2022) with geometric-preserving operations, while R3D-AD (Zhou et al. 2024b) employs diffusion-based reconstruction for better localization. Shape-Guided (Chu et al. 2023) methods leverage dual memory banks from RGB and 3D features. Despite strong results, these methods are explicitly designed for anomaly detection. We explore a new perspective by adapting self-supervised models to perform anomaly detection under a unified "pre-training and fine-tuning" paradigm, offering a more general and flexible solution to 3D anomaly detection.

Methodology

Problem Definition for 3D Anomaly Detection

Given a training set $T = \{p_i\}_{i=1}^{M}$, where each $p_i \in \mathbb{R}^{N\times 3}$ is a 3D point cloud sampled from a normal object class $c \in \mathcal{C}$, and C denotes the set of all training classes. Here, M denotes the number of point clouds in the training set, and N is the number of points in each point cloud. Each training class c is associated with a small number of clean samples, and no anomalous instances are available during training. At test time, given a query point cloud x from a target class c, the objective is twofold: (1) Anomaly classification, which aims to assign a binary label $y \in \{0, 1\}$, where $y = 1$ indicates that the point cloud x is anomalous, and (2) Anomaly localization, which involves predicting a point-level anomaly score function $\phi : \mathbb{R}^3 \to \mathbb{R}^+$, where $\phi(x_i)$denotes the likelihood of each point $x_i \in x$ being anomalous.

Curvature Computation for Point Clouds

Curvature is a fundamental geometric property in 3D space. Following previous practices (Zhu et al. 2022; Zhang et al. 2024), we can directly compute the curvature of each point from the spatial coordinates of the point cloud. Specifically, we compute point-wise curvature by analyzing the local geometric structure using the covariance matrix of neighboring points. For each point $x_i \in \mathbb{R}^{1\times 3}$, we first retrieve its k-nearest neighbors $\{x_j\}_{j=1}^{k}$ and compute the centroid $\overline{x_i}$ of the neighborhood. The neighborhood is then centered by subtracting the centroid from each neighbor, and a 3×3 covariance matrix is formed as:

$$C_i = \frac{1}{k-1}\sum_{j=1}^{k}(\mathbf{x}_j - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_i)^\top. \tag{1}$$

(a) Curvature Heatmap (b) Normal vs Abnormal Points

Figure 2: (a) Curvature heatmap of an anomalous point cloud, where warmer colors indicate higher curvature. (b) Spatial distribution of normal (green) and abnormal (red) points.

We perform eigenvalue decomposition on C_i to obtain the ordered eigenvalues $\lambda_1^i \le \lambda_2^i \le \lambda_3^i$, and define the curvature at x_i as:

$$\text{Curv}(x_i) = \frac{\lambda_1^i + \lambda_2^i + \lambda_3^i}{\lambda_1^i}, \tag{2}$$

this normalized curvature metric captures the degree of variation in the local surface geometry and has been widely adopted as a property for point cloud analysis tasks.

Curvature for 3D Anomaly Detection

Curvature reflects the variation of surfaces in 3D space, such as concavities and convexities. We observe that anomalous points on 3D surfaces are highly correlated with the magnitude of curvature. As shown in Figure 2, we present a comparison between the curvature heatmap (a) and the distribution of normal and abnormal points (b) within an anomalous point cloud. It is clearly observable that the curvature at the edges of abnormal regions is significantly higher than that of the surrounding normal regions, indicating a certain relationship between curvature and the anomaly status of the point.

Furthermore, we make a interesting observation: by using the curvature of each point as its anomaly score, we achieve significantly better anomaly detection performance compared to several classical methods, despite this being a non-learning-based approach. As shown in Figure 1, the curvature-based detection method achieves significantly higher object-level AUROC and point-level AUROC on the Real3D-AD dataset compared to self-supervised approaches such as Point-BERT, Point-MAE, and PointGPT, as well as task-specific anomaly detection methods including M3DM, PatchCore, and Reg3D-AD. More quantitative results can be found in Table 1 and Table 2. This finding highlights the pivotal role of curvature as an intrinsic geometric property for identifying anomalies in 3D data. It suggests that curvature itself can serve as a highly informative cue for detecting structural irregularities and should be explicitly incorporated into 3D representation learning.

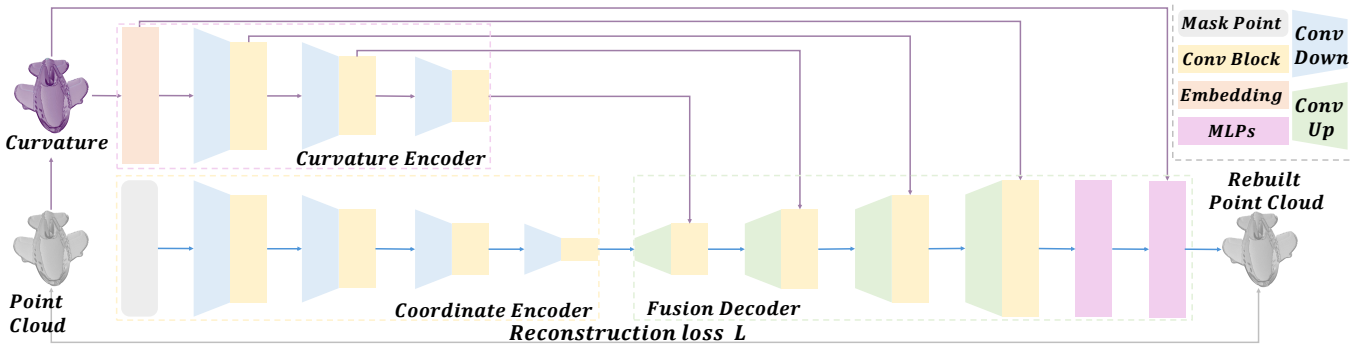

Figure 3: The pipeline of our **C**urvature-**A**ugmented **S**elf-supervised **L**earning (**CASL**) framework. Our framework, built upon the U-Net architecture, leverages multi-scale curvature prompts to guide the recovery of spatial coordinates for all masked points from random initialization.

Curvature-Augmented Self-supervised Learning

The overall pipeline of our curvature-augmented self-supervised learning framework is shown in Figure 3.

Motivation for Multi-Scale Curvature Prompts. In point cloud reconstruction, the decoder is responsible for rebuilding spatial coordinates from high-level latent features. Providing explicit geometric priors to the decoder is crucial. Traditional approaches typically utilize the coordinates of unmasked points as geometric semantic priors to reconstruct the masked ones. While this strategy has achieved some success, it inevitably suffers from the "***geometric shortcut***" issue due to the overlap between the semantic and positional domains, leading to fine-grained representational collapse in anomaly detection.

In this work, we propose addressing this issue from a curvature perspective. Curvature, as an intrinsic geometric property of 3D surfaces, plays a vital role in capturing structural irregularities. At small scales, curvature is highly sensitive to local surface variations, where subtle edges or protrusions often result in sharp curvature changes. At larger scales, curvature captures the overall shape and contour of the surface, enabling a more comprehensive understanding of global geometry. Thus, curvature serves as a meaningful geometric prior that not only provides rich structural prompts for the decoder but also avoids the geometric shortcut in traditional coordinate-to-coordinate reconstruction. To this end, we propose injecting multi-scale curvature prompts into the decoder to provide implicit geometric guidance for reconstruction.

Network Architecture and Data Pipeline. Given the information-intensive nature of 3D anomaly detection, where individual 3D object often comprise hundreds of thousands of points and require fine-grained, point-level anomaly localization, the classical Transformer (Vaswani et al. 2017) architecture is computationally inefficient. This inefficiency stems from its coarse-grained patch-based operations and quadratic time complexity of $O(n^2)$, which limit its suitability for high-resolution 3D anomaly detection. To enhance efficiency, our framework adopts the classical U-Net (Ronneberger, Fischer, and Brox 2015) architecture built upon Minkowski convolution (Choy, Gwak, and Savarese 2019; Gwak, Choy, and Savarese 2020), a well-established design for high-resolution point cloud processing that offers substantially greater efficiency than Transformer-based approaches.

As shown in Figure 3, our network architecture consists of three main components: a curvature encoder, a coordinate encoder, and a fusion decoder.

Our curvature encoder begins with an MLP-based embedding layer to extract point-wise curvature features. This is followed by three encoder blocks to capture multi-scale curvature representations. Each encoder block first applies a Minkowski Convolution (Conv Down) with a kernel size and stride of 2 to downsample the resolution, and then employs a convolutional block (Conv Block) consisting of four stride=1 convolution layers with residual connections to extract features. The detailed architecture is provided in the supplementary material.

The coordinate encoder takes as input the randomly initialized features of the N masked points and passes them through four consecutive encoder blocks, progressively mapping the masked features to high-dimensional features. The final output is a tensor of shape $N_4 \times 256$, where 256 denotes the feature dimension.

In our fusion decoder, the features from the previous resolution, with shape $N_{i-1} \times C_{i-1}$, is first upsampled to the current resolution $N_i \times C_{i-1}$ using a Minkowski Transpose Convolution (Conv Up) with both kernel size and stride set to 2. The upsampled features are then concatenated with the curvature prompts at the current resolution, resulting in a tensor of shape $N_i \times (C_{i-1} + C_i)$. This tensor is subsequently passed through a convolutional block (Conv Block) to produce an output of shape $N_i \times C_i$. This process is repeated across multiple scales until the original resolution is fully recovered.

The output of the fusion decoder is a tensor of shape $N \times 96$. We first apply MLPs to reduce its dimensionality to $N \times 32$. This representation is then concatenated with the original curvature features and passed through the second MLPs, which map it into the 3D coordinate space. In this way, the final reconstructed point cloud is obtained.

Loss Function. Due to the large number of points (typically over 100k) in the reconstructed object point clouds, classical point cloud reconstruction losses such as the Chamfer Distance (CD) (Fan, Su, and Guibas 2017) and Earth Mover's Distance (EMD) (Rubner, Tomasi, and Guibas 2000) become computationally infeasible. To ensure efficiency, we instead adopt a combination of ℓ_1 and ℓ_2 losses to constrain the distance between the reconstructed and ground truth point clouds. The loss is computed as follows:

$$\mathcal{L}_{recon} = \mathcal{L}_1(p, p_{rec}) + \mathcal{L}_2(p, p_{rec}), \tag{3}$$

Fine-tuning via Pseudo-Anomaly Classification

The prevalent strategy in transferring pretrained models for 3D understanding tasks follows a unified "pre-training and fine-tuning" paradigm, where a pretrained backbone is adapted to downstream tasks via a lightweight task-specific head. In contrast, prior anomaly detection methods based on pretrained models often rely on reconstruction or feature matching, which are specifically tailored for anomaly detection and exhibit limited generalization to other tasks. We propose a unified fine-tuning strategy based on pseudo-anomaly classification. By synthesizing pseudo-anomalies from normal samples, our method only requires appending a binary classification head after the first MLPs in Figure 3, thus enabling anomaly detection in a manner consistent with standard fine-tuning pipelines. Moreover, since reconstruction is not required during fine-tuning, our coordinate encoder leverages embeddings extracted directly from the coordinates rather than the randomly initialized masked points used during pre-training, enabling a more comprehensive 3D understanding.

Pseudo-anomaly Generation. The pseudo-anomaly generation strategy simulates abnormal regions by applying controlled perturbations to normal point clouds without requiring real anomaly data. We follow the generation method proposed by PO3AD (Ye et al. 2025), specifically, it divides the point cloud into multiple patches and randomly selects one or more of them. All points within the selected patch are displaced along either the normal direction or its opposite by a certain distance, controlled by a distance scaling factor. This process simulates realistic pseudo-anomalous regions such as protrusions or indentations.

Anomaly Score for Inference

The anomaly level of each point is evaluated based on its predicted classification logits under our point-wise classification fine-tuning. Specifically, for each point, the model outputs a probability distribution over two categories (normal and anomalous) through a softmax activation. Let x_i^0 and x_i^1 denote the predicted probabilities of point x_i belonging to the normal and anomalous classes, respectively. The point-level anomaly score $\phi(x_i)$ is defined as:

$$\phi(x_i) = -\log\left(\frac{x_i^0}{x_i^1 + \varepsilon}\right), \tag{4}$$

where ε is a small constant added for numerical stability. This formulation computes the logarithmic ratio between the normal and abnormal class probabilities, effectively capturing the degree to which a point deviates from normal patterns while enhancing the discriminability and stability of the anomaly score—larger values indicate a higher likelihood of anomaly.

To obtain the sample-level anomaly score $\phi(x)$, we adopt a top-k aggregation strategy, where only the top-k of point-level scores (with the highest anomaly likelihoods) are averaged:

$$\phi(x) = \frac{1}{k}\sum_{i=1}^{k}\phi(x_i), \quad \text{where } k = \lceil r \cdot N \rceil, \tag{5}$$

where r denotes the aggregation rate and N is the total number of points in the sample. This approach emphasizes the most anomalous regions of a point cloud, improving robustness against false positives from scattered noise.

Experiments

Pre-training

Since our primary focus is on evaluating the model's transferability in anomaly detection, we construct our pre-training dataset by collecting the training sets from two classical anomaly detection datasets: Real3D-AD (Liu et al. 2023) and Anomaly-ShapeNet (Li et al. 2024). Since both training sets consist exclusively of normal samples, our method performs self-supervised learning solely on normal data. Real3D-AD is a high-resolution point cloud dataset collected from real-world objects, covering 12 categories. Each category similarly includes 4 normal samples for training and 100 test instances. Anomaly-ShapeNet is a synthetic dataset derived from ShapeNet (Chang et al. 2015), consisting of 1,600 point cloud samples across 40 categories, with each category providing 4 normal training instances. As a result, we obtain a total of 208 distinct normal samples. To enhance diversity, we further apply data augmentation to generate 4 different variants for each sample, resulting in a pre-training dataset comprising 832 normal samples in total.

We train our model for 300 epochs using the Adam optimizer with an initial learning rate of 0.001. The learning rate is decayed by a factor of 0.5 every 10 epochs.

Fine-tuning on Downstream Tasks

We assess the efficacy of our approach by fine-tuning our pre-trained models on downstream tasks, including anomaly detection, classification, and part segmentation.

Anomaly Detection on Real3D-AD. Our anomaly detection experiments are conducted following the protocols established in previous works (Liu et al. 2023). We adopt the Area Under the Receiver Operating Characteristic Curve (AUROC) as our evaluation metric, as it provides an objective measure of both detection (object-level) and localization (point-level) performance without relying on any predefined decision threshold.

The results in the Table 1 show the performance of different anomaly detection methods on Real3D-AD. Among them, Curvature is our proposed non-learning-based method

(a) O-AUROC(↑)

Method	Reference	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Average
BTF	CVPRW'23	0.730	0.647	0.539	0.789	0.707	0.691	0.602	0.686	0.596	0.396	0.530	0.703	0.635
M3DM	CVPR'23	0.434	0.541	0.552	0.683	0.602	0.433	0.540	0.644	0.495	0.694	0.551	0.450	0.552
PatchCore	CVPR'22	**0.848**	0.777	0.570	**0.853**	0.784	0.628	0.837	0.359	0.767	0.663	0.471	0.626	0.682
CPMF	PR'24	0.701	0.551	0.552	0.504	0.523	0.582	0.558	0.589	0.729	0.653	0.700	0.390	0.586
IMRNet	CVPR'24	0.762	0.711	0.755	0.780	0.905	0.517	0.880	0.674	0.604	0.665	0.674	0.774	0.725
Reg3D-AD	NeurIPS'23	0.716	0.697	0.827	0.852	0.900	0.584	0.915	0.417	0.762	0.583	0.506	0.685	0.704
Group3AD	ACM MM'24	0.744	0.728	0.847	0.786	0.932	0.679	0.976	0.539	**0.841**	0.585	0.562	0.796	0.751
R3D-AD	ECCV'24	0.772	0.696	0.713	0.714	0.685	**0.909**	0.692	0.665	0.720	**0.840**	0.701	0.703	0.734
PO3AD	CVPR'25	0.804	0.654	0.785	0.686	0.801	0.820	0.859	0.693	0.756	0.800	0.758	0.771	0.765
ISMP	AAAI'25	0.858	0.731	0.852	0.714	0.948	0.712	**0.945**	0.468	0.729	0.623	0.660	0.842	0.767
Curvature	**Ours**	0.345	0.686	**0.889**	0.496	0.957	0.768	0.854	0.554	0.790	0.594	0.872	0.872	0.723
CASL	**Ours**	0.808	**0.799**	0.848	0.657	**0.976**	0.836	0.935	**0.769**	0.643	0.791	**0.893**	**0.924**	**0.823**

(b) P-AUROC(↑)

Method	Reference	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Average
BTF	CVPRW'23	0.738	0.708	0.864	0.693	0.882	0.875	0.709	0.891	0.512	0.571	0.501	0.815	0.722
M3DM	CVPR'23	0.530	0.607	0.683	0.735	0.618	0.678	0.600	0.654	0.561	0.748	0.555	0.679	0.637
PatchCore	CVPR'22	0.556	0.740	0.749	0.558	0.854	0.658	0.781	0.539	0.808	0.753	0.613	0.549	0.692
Reg3D-AD	NeurIPS'23	0.631	0.718	0.724	0.676	0.835	0.503	0.826	0.545	0.817	0.811	0.617	0.759	0.700
Group3AD	ACM MM'24	0.636	0.745	0.738	0.759	0.862	0.631	0.836	0.564	**0.827**	0.798	0.625	0.803	0.735
ISMP	AAAI'25	0.753	0.836	0.907	**0.798**	0.926	0.876	0.886	0.857	0.813	0.839	0.641	0.895	0.836
Curvature	**Ours**	0.715	0.699	0.810	0.705	0.860	0.860	0.727	0.904	0.545	0.614	0.519	0.797	0.729
CASL	**Ours**	**0.842**	**0.905**	**0.932**	0.713	**0.988**	**0.895**	**0.935**	**0.916**	0.814	**0.873**	**0.839**	**0.937**	**0.882**

Table 1: The experimental results for anomaly detection across 12 categories of Real3D-AD. Table (a) reports the object-level AUROC (O-AUROC) scores, while Table (b) presents the results of the point-level AUROC (P-AUROC). The results of the baselines are from their papers.

that relies purely on curvature to identify anomalies, while CASL is our learning-based method. As shown in the results, our Curvature method outperforms several classical learning-based methods specifically designed for anomaly detection, such as M3DM, CPMF, and Reg3D-AD, demonstrating the importance of curvature in identifying anomalies. Furthermore, our CASL method, which integrates multi-scale curvature prompts, achieves new state-of-the-art (SOTA) performance, surpassing the second-best approach by **5.6%** in object-level AUROC and **4.6%** in point-level AUROC, highlighting the superiority of our approach.

Anomaly Detection on Anomaly-ShapeNet. Table 2 reports the experimental results of our anomaly detection approach on Anomaly-ShapeNet, which consists of 40 distinct categories. We present both the O-AUROC and P-AUROC scores for each category, as well as the overall average performance across all categories. Due to space limitations, Table 2 only displays the average results; detailed per-category results can be found in the supplementary material. As shown in the table, our Curvature-based method also outperforms several learning-based approaches specifically designed for anomaly detection. Furthermore, our CASL method achieves a new state-of-the-art, surpassing the second-best PO3AD by **4.8%** in terms of average O-AUROC, demonstrating the superiority of our approach.

Object Classification. We further validate the effectiveness of our model on the classical object classification task. Following prior work (Yu et al. 2022; Pang et al. 2022; Chen et al. 2024), we report the classification fine-tuning results of our CASL model on three variants of the real-world ScanObjectNN (Uy et al. 2019) dataset. ScanObjectNN is a widely used benchmark consisting of approximately 15,000 point cloud samples across 15 categories. These objects represent indoor scenes and are often characterized by cluttered backgrounds and occlusions caused by other objects.

Methods	Reference	Avg. O-AUROC	Avg. P-AUROC
BTF	CVPRW'23	0.528	0.628
M3DM	CVPR'23	0.552	0.616
PatchCore	CVPR'22	0.568	0.580
CPMF	PR'24	0.559	0.573
Reg3D-AD	NeurIPS'23	0.572	0.668
IMRNet	CVPR'24	0.661	0.650
R3D-AD	ECCV'24	0.749	-
ISMP	AAAI'25	-	0.691
PO3AD	CVPR'25	0.839	0.898
Curvature	**Ours**	0.559	0.693
CASL	**Ours**	**0.887**	**0.899**

Table 2: Anomaly detection results on Anomaly-ShapeNet. We report the average O-AUROC and P-AUROC scores across its 40 categories. More detailed results are provided in the supplementary material.

Table 3 presents our experimental results. As shown, our method achieves leading performance on the OBJ-BG and OBJ-ONLY variants. On the more challenging PB-T50-RS variant, our model performs below the best-performing SFR(Zha et al. 2023a). Notably, unlike prior approaches that pretrain on the full ShapeNet (Chang et al. 2015) dataset (around 50k samples), our model is pretrained on only 832 samples, which is a significantly smaller dataset. Despite this, CASL still matches or exceeds some leading methods, demonstrating the effectiveness of the proposed multi-scale curvature prompts. By leveraging curvature guidance, our method effectively mitigates the geometric shortcut prob-

Method	Reference	OBJ-BG(↑)	OBJ-ONLY(↑)	PB-T50-RS(↑)
Supervised Learning Only				
PointNet	CVPR'17	73.3	79.2	68.0
PointNet++	NeurIPS'17	82.3	84.3	77.9
DGCNN	TOG'19	82.8	86.2	78.1
PointMLP	ICLR'22	-	-	85.7
SFR	ICASSP'23	-	-	**87.8**
with Self-supervised Representation Learning				
Point-BERT	CVPR'22	87.43	88.12	83.07
Point-MAE	ECCV'22	90.02	88.29	85.18
Point-M2AE	NeurIPS'22	91.22	88.81	86.43
PointGPT-S	NeurIPS'23	91.60	90.00	86.90
CASL	**Ours**	**92.08**	**91.05**	86.81

Table 3: Classification results on three variants of the ScanObjectNN dataset, and we report the classification accuracy(%).

Methods	Reference	mIoU_c	mIoU_I
Supervised Learning Only			
PointNet++	NeurIPS'17	81.9	85.1
DGCNN	TOG'19	82.3	85.2
with Self-supervised Representation Learning			
PointContrast	ECCV'20	-	85.1
IDPT	ICCV'23	83.8	85.9
Point-BERT	CVPR'22	84.1	85.6
Point-MAE	ECCV'22	**84.2**	86.1
PointGPT-S	NeurIPS'23	84.1	86.2
CASL	**Ours**	84.1	**86.2**

Table 4: Part segmentation results on the ShapeNetPart. The mean IoU across all categories, i.e., mIoU_c (%), and the mean IoU across all instances, i.e., mIoU_I (%) are reported.

lem and achieves strong representational ability with minimal data.

Part Segmentation. We further evaluate the performance of CASL on the part segmentation task using the ShapeNetPart dataset, which contains 16,881 samples spanning 16 object categories. For a fair comparison, we adopt the same segmentation setup following previous works (Pang et al. 2022; Chen et al. 2024). Table 4 reports our experimental results. Similarly, despite pretraining on a significantly smaller dataset, our method achieves equally good performance in the part segmentation task.

Ablation Study

The Effect of Masked Points and Curvature Prompts. We further analyze the two important components of our method in Table 5, namely the masked points and the curvature prompts. Experiment A performs reconstruction from full original coordinates back to themselves using only the coordinate encoder, without applying any masking. Experiment B also relies solely on the coordinate encoder to reconstruct coordinates, but with a 60% random masking ratio applied during encoding. Experiment C uses only the curvature encoder without any masking, aiming to reconstruct coordinates from the curvature space. Experiment D represents our proposed method, which reconstructs coordinates using curvature prompts and the coordinate encoder with fully masked points (100% masking ratio).

	Masked Points	Curvature Prompts	O-AUROC	P-AUROC	Rec. Loss
A	✘	✘	0.751	0.649	0.008
B	✔	✘	0.776	0.711	0.012
C	✘	✔	0.804	0.878	0.019
D	✔	✔	0.823	0.882	0.016

Table 5: The effect of masked points and curvature prompts.

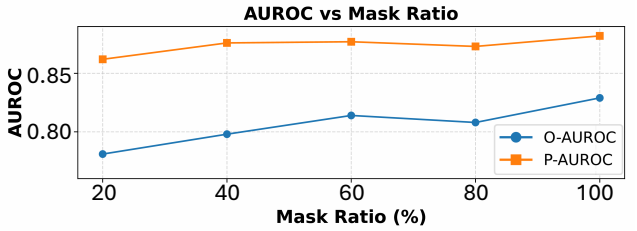

Figure 4: Detection performance with varying mask ratios.

As shown in Table 5, A achieves the lowest reconstruction loss, but detection results are very poor. Because it performs self-reconstruction from the coordinate to itself, which introduces significant geometric shortcuts, ultimately leading to representation collapse. B mitigates this issue to some extent by randomly masking a large portion of the original coordinate space and reconstructing from the masked points to the original domain. However, the learned representations remain suboptimal. Experiment C learns to reconstruct coordinates from the curvature space, resulting in substantial improvements over A and B. Our method D further incorporates curvature prompts into the decoder and operates in a fully masked coordinate space, substantially reducing geometric shortcuts and thereby enabling superior representational learning.

The Effect of Masked Ratios. We investigate the effect of different masking ratios in our coordinate encoder on anomaly detection. As shown in Figure 4, the best performance is achieved at 100% masking ratio, indicating that any additional geometric information in the coordinate space during curvature-guided reconstruction introduces geometric shortcuts, thereby limiting the representation learning.

Conclusions

In this paper, we demonstrate that classical pre-trained point cloud models perform suboptimally in anomaly detection under a unified fine-tuning paradigm due to geometric shortcuts. To address this, we propose CASL, a curvature-augmented reconstruction framework that leverages multi-scale curvature prompts and full-coordinate masking to guide learning. CASL effectively mitigates geometric shortcuts and enhances 3D structural understanding. Experiments on Real3D-AD and Anomaly-ShapeNet show that CASL achieves leading anomaly detection performance and generalizes well to other 3D understanding tasks.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China, under Grant (62302309,62571298), Shenzhen Science and Technology Program (JCYJ20220818101014030).

References

Afham, M.; Dissanayake, I.; Dissanayake, D.; Dharmasiri, A.; Thilakarathna, K.; and Rodrigo, R. 2022. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. In *CVPR*, 9902–9912.

Bergmann, P.; Fauser, M.; Sattlegger, D.; and Steger, C. 2019. MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9592–9600.

Bergmann, P.; Fauser, M.; Sattlegger, D.; and Steger, C. 2020. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4183–4192.

Bergmann, P.; Löwe, S.; Fauser, M.; Sattlegger, D.; and Steger, C. 2018. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*.

Cao, Y.; Xu, X.; and Shen, W. 2024. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition*, 156: 110761.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

Chen, G.; Wang, M.; Yang, Y.; Yu, K.; Yuan, L.; and Yue, Y. 2024. Pointgpt: Auto-regressively generative pre-training from point clouds. volume 36.

Choy, C.; Gwak, J.; and Savarese, S. 2019. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3075–3084.

Chu, Y.-M.; Liu, C.; Hsieh, T.-I.; Chen, H.-T.; and Liu, T.-L. 2023. Shape-Guided Dual-Memory Learning for 3D Anomaly Detection. In *Proceedings of the 40th International Conference on Machine Learning*, 6185–6194.

Dong, R.; Qi, Z.; Zhang, L.; Zhang, J.; Sun, J.; Ge, Z.; Yi, L.; and Ma, K. 2023. Autoencoders as Cross-Modal Teachers: Can Pretrained 2D Image Transformers Help 3D Representation Learning? Kigali, Rwanda.

Fan, H.; Su, H.; and Guibas, L. J. 2017. A point set generation network for 3d object reconstruction from a single image. In *CVPR*, 605–613. Honolulu, Hawaii, USA.

Gwak, J.; Choy, C. B.; and Savarese, S. 2020. Generative Sparse Detection Networks for 3D Single-shot Object Detection. In *European conference on computer vision*.

Li, W.; Xu, X.; Gu, Y.; Zheng, B.; Gao, S.; and Wu, Y. 2024. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 22207–22216.

Liu, J.; Xie, G.; Chen, R.; Li, X.; Wang, J.; Liu, Y.; Wang, C.; and Zheng, F. 2023. Real3d-ad: A dataset of point cloud anomaly detection. *Advances in Neural Information Processing Systems*, 36: 30402–30415.

Pang, G.; Shen, C.; Cao, L.; and Hengel, A. V. D. 2021. Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)*, 54(2): 1–38.

Pang, Y.; Wang, W.; Tay, F. E.; Liu, W.; Tian, Y.; and Yuan, L. 2022. Masked autoencoders for point cloud self-supervised learning. In *ECCV*. Tel Aviv, Israel.

Qi, Z.; Dong, R.; Fan, G.; Ge, Z.; Zhang, X.; Ma, K.; and Yi, L. 2023. Contrast with Reconstruct: Contrastive 3D Representation Learning Guided by Generative Pretraining. In *ICML*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. Munich, Germany.

Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; and Gehler, P. 2022. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14318–14328.

Rubner, Y.; Tomasi, C.; and Guibas, L. J. 2000. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2): 99–121.

Uy, M. A.; Pham, Q.-H.; Hua, B.-S.; Nguyen, T.; and Yeung, S.-K. 2019. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *ICCV*, 1588–1597. Seoul, Korea.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need. In *NeurIPS*, 30. Long Beach, CA, USA.

Xie, G.; Wang, J.; Liu, J.; Lyu, J.; Liu, Y.; Wang, C.; Zheng, F.; and Jin, Y. 2024. Im-iad: Industrial image anomaly detection benchmark in manufacturing. *IEEE Transactions on Cybernetics*, 54(5): 2720–2733.

Xie, S.; Gu, J.; Guo, D.; Qi, C. R.; Guibas, L.; and Litany, O. 2020. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *ECCV*, 574–591. Springer.

Xiong, J.; Dai, T.; Zha, Y.; Wang, X.; and Xia, S.-T. 2023. Semantic preserving learning for task-oriented point cloud downsampling. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Ye, J.; Zhao, W.; Yang, X.; Cheng, G.; and Huang, K. 2025. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 1353–1362.

Yu, X.; Tang, L.; Rao, Y.; Huang, T.; Zhou, J.; and Lu, J. 2022. Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In *CVPR*, 19313–19322. New Orleans, Louisiana, USA.

Zamanzadeh Darban, Z.; Webb, G. I.; Pan, S.; Aggarwal, C.; and Salehi, M. 2024. Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, 57(1): 1–42.

Zha, Y.; Dai, T.; Guo, H.; Wang, Y.; Chen, B.; Chen, K.; and Xia, S.-T. 2024a. Point Cloud Mixture-of-Domain-Experts Model for 3D Self-supervised Learning. *arXiv preprint arXiv:2410.09886*.

Zha, Y.; Ji, H.; Li, J.; Li, R.; Dai, T.; Chen, B.; Wang, Z.; and Xia, S.-T. 2024b. Towards compact 3d representations via point feature enhancement masked autoencoders. In *AAAI*. VANCOUVER, CANADA.

Zha, Y.; Li, N.; Wang, Y.; Dai, T.; Guo, H.; Chen, B.; Wang, Z.; Ouyang, Z.; and Xia, S.-T. 2024c. LCM: Locally Constrained Compact Point Cloud Model for Masked Point Modeling. *arXiv preprint arXiv:2405.17149*.

Zha, Y.; Li, R.; Dai, T.; Xiong, J.; Wang, X.; and Xia, S.-T. 2023a. SFR: Semantic-Aware Feature Rendering of Point Cloud. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Zha, Y.; Wang, J.; Dai, T.; Chen, B.; Wang, Z.; and Xia, S.-T. 2023b. Instance-aware dynamic prompt tuning for pre-trained point cloud models. In *ICCV*, 14161–14170. Paris, France.

Zha, Y.; Wang, Y.; Dai, T.; and Xia, S.-T. 2024d. Pre-training Point Cloud Compact Model with Partial-aware Reconstruction. *arXiv preprint arXiv:2407.09344*.

Zha, Y.; Wang, Y.; Guo, H.; Wang, J.; Dai, T.; Chen, B.; Ouyang, Z.; Yuerong, X.; Chen, K.; and Xia, S.-T. 2025. PMA: Towards Parameter-Efficient Point Cloud Understanding via Point Mamba Adapter. *arXiv preprint arXiv:2505.20941*.

Zhang, J.; Gu, W.; Huang, Y.; Jiang, Z.; Wu, W.; and Lyu, M. R. 2024. Curvature-invariant adversarial attacks for 3d point clouds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7142–7150.

Zhang, R.; Wang, L.; Qiao, Y.; Gao, P.; and Li, H. 2023. Learning 3d representations from 2d pre-trained models via image-to-point masked autoencoders. In *CVPR*, 21769–21780. Vancouver, Canada.

Zhang, X.; Zhang, S.; and Yan, J. 2024. Pcp-mae: Learning to predict centers for point masked autoencoders. *Advances in Neural Information Processing Systems*, 37: 80303–80327.

Zhou, Q.; Yan, J.; He, S.; Meng, W.; and Chen, J. 2024a. Pointad: Comprehending 3d anomalies from points and pixels for zero-shot 3d anomaly detection. *Advances in Neural Information Processing Systems*, 37: 84866–84896.

Zhou, Z.; Wang, L.; Fang, N.; Wang, Z.; Qiu, L.; and Zhang, S. 2024b. R3d-ad: Reconstruction via diffusion for 3d anomaly detection. In *European conference on computer vision*, 91–107. Springer.

Zhu, H.; Xie, G.; Hou, C.; Dai, T.; Gao, C.; Wang, J.; and Shen, L. 2024. Towards high-resolution 3d anomaly detection via group-level feature contrastive learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 4680–4689.

Zhu, L.; Chen, W.; Lin, X.; He, L.; and Guan, Y. 2022. Curvature-variation-inspired sampling for point cloud classification and segmentation. *IEEE Signal Processing Letters*, 29: 1868–1872.